Security Isolation of RISC-V — A Case Study of Penglai

Yubin Xia

Professor, Shanghai Jiao Tong University

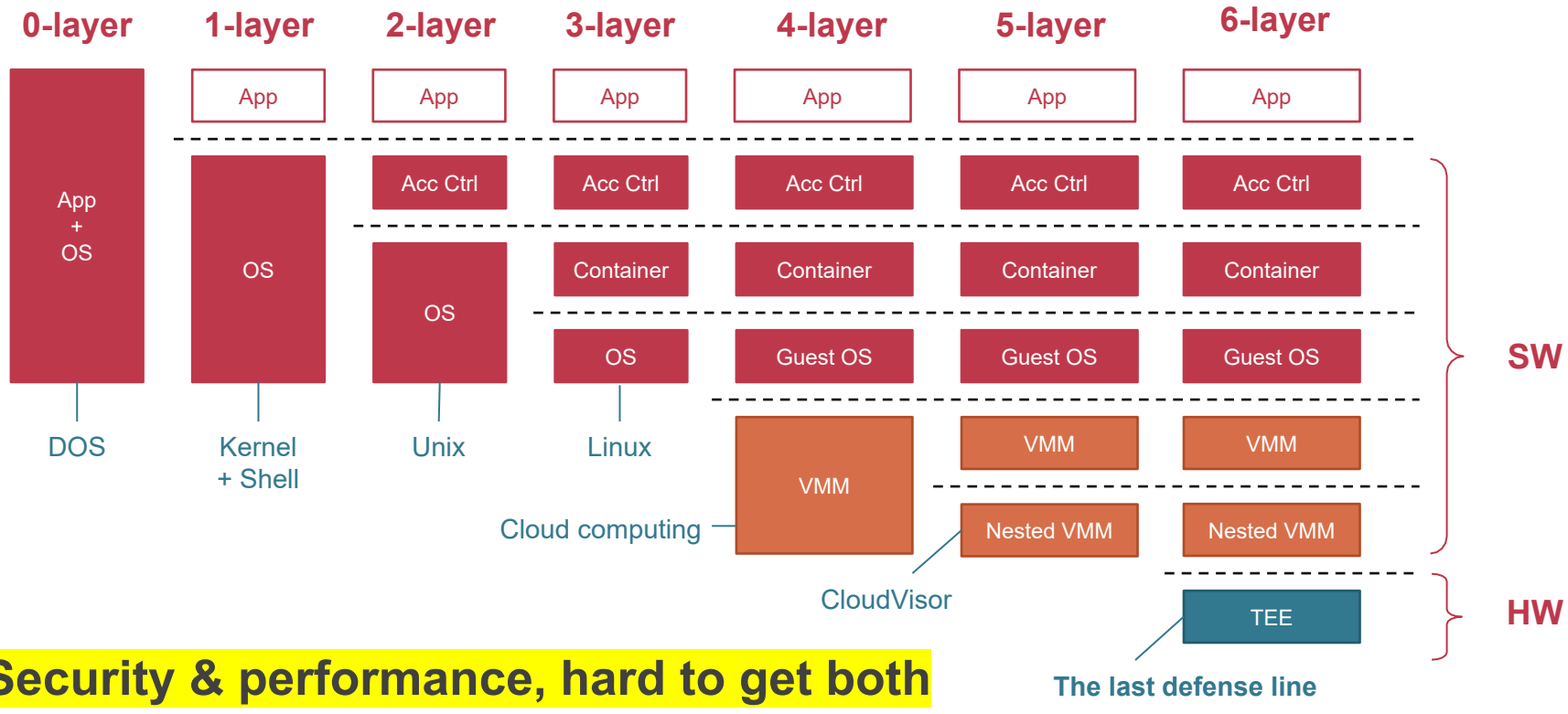
2025.7

饮 水 思 源 • 爱 国 荣 校

The need for isolation

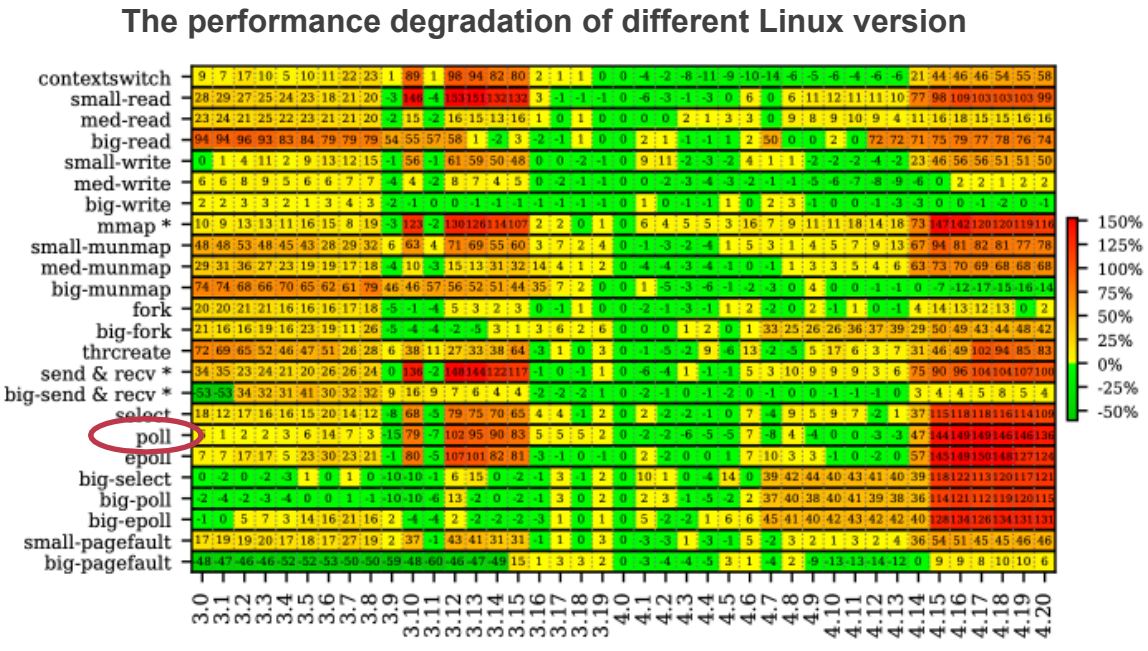
A problem with a clogged toilet cannot affect the missile-launching subsystem. Operating systems do not have this kind of isolation between components. (2006) —— Andrew Tanenbaum

Isolation is a Major Source of Security Tax



Security & performance, hard to get both

Linux: Performance is Degrading



Source: [SOSP'19] An Analysis of Performance Evolution of Linux's Core Operations

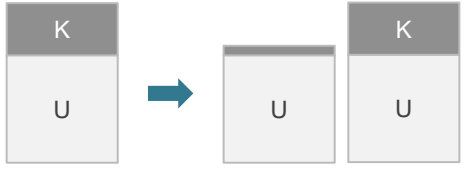
Caused by "Security Tax"

1. KPTI
2. No speculation for indirect branch
3. SLAB free list randomization
4. Strengthen user-space memcpy

Latency of `poll()` up to 146%

Example: KPTI

1. Defend Meltdown attack in 2018
2. Performance drop up to 30%



Trend: New HW Isolation Features

CPU	Security Features after 2015
Intel	SGX, TME, MPK, MPX, TDX, …
AMD	SME, SEV, SEV-ES, SNP, …
ARM	TME, PA, Bowmore, CCA, …
RISC-V	PMP, sPMP, ePMP, IOPMP, …

- **Major CPU vendors have new features**
  - Intel's new features in last 8 years are more than the sum of previous 49 years

1 New Requirements

- Cloud, big data, mobile computing
- Traditional software solution is hard to fit new requirements

2 Better Know-how

- A lot of CVEs and quantitative analysis
- Know what to do in HW

3 More Software-defined

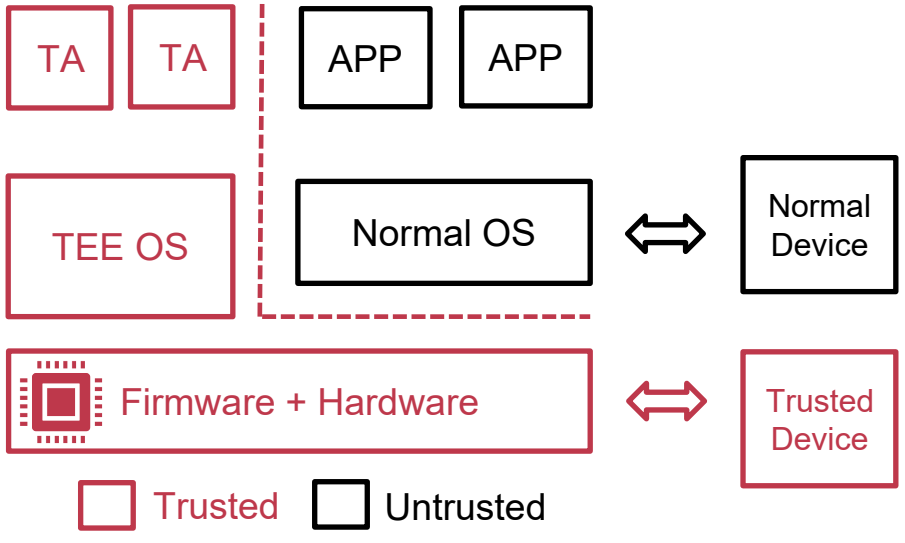
- Many security features are using software (firmware or microcode)
- CPU can be patched as software

HW/SW Co-design will bring more oppotunities

1: How can OS fit HW better?

2: How can HW fit OS better?

Trusted Execution Environment for Isolation



1. **TEE protects trusted app from untrusted software**
   - Hypervisor / OS
   - Other applications
2. **TEE contains secure hardware resources**
   - Secure CPU
   - Protected memory
   - Trusted Devices

TEE has been widely used in Clouds

- Intel SGX/TDX
- AMD SEV
- ARM TrustZone
- Keystone、Penglai

- **Major cloud vendors proposed confidential computing based on TEE**
  - 2018, Microsoft Azure proposed Confidential Computing based on Intel SGX
  - 2019, Amazon (AWS) proposed Nitro Enclave for user data protection
  - 2020, Google Cloud proposed Secure VM (Virtual Machine) based on AMD SEV

TEE is one key-enabling technical for confidential computing

TEE is also widely used in mobile systems

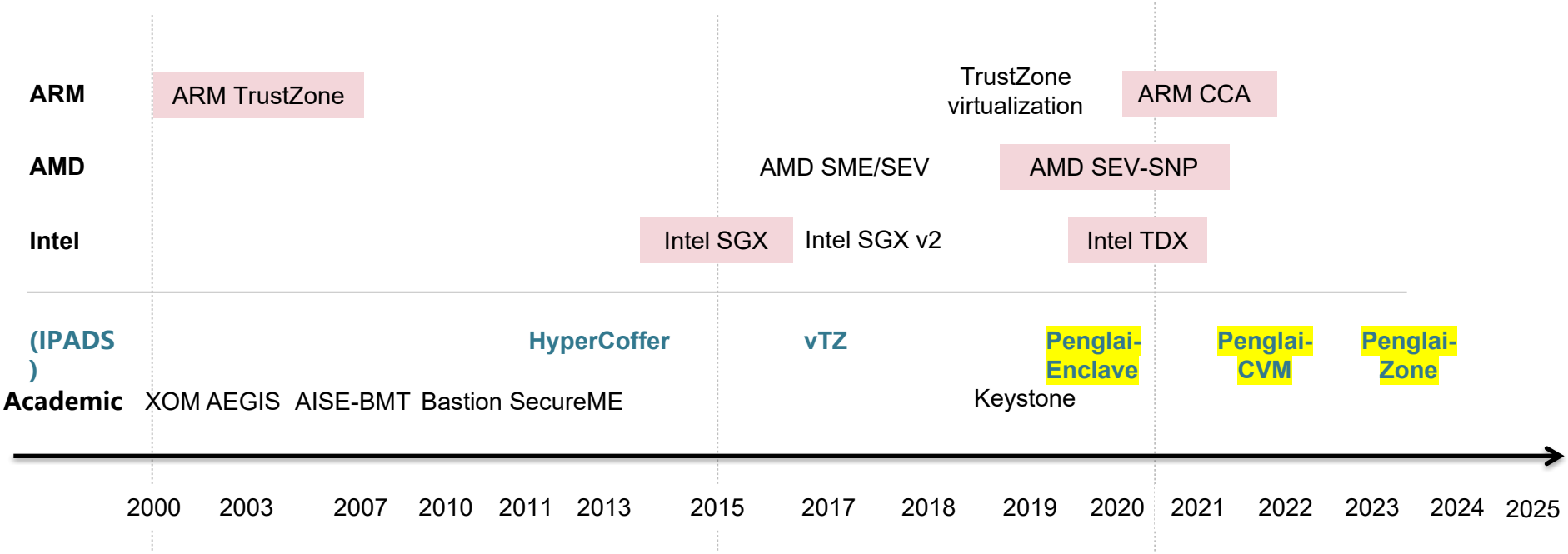
- **TEE protects the sensitive data and code for both users and developers**

Digital payment

Face recognition

Digital Right Management

A Brief History of TEE



Phase-1: ARM TrustZone

Abstraction: Zone

Phase-2: Intel SGX

Abstraction: Enclave

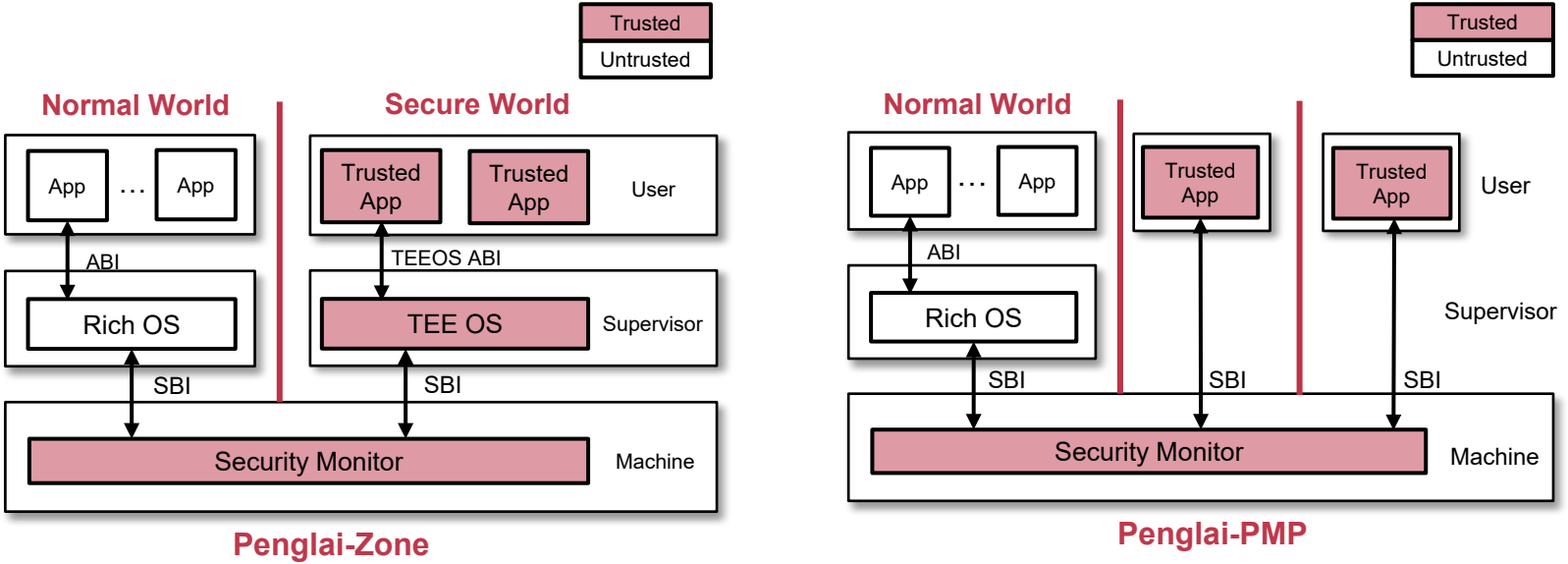
Phase-3: Secure VM

Abstraction: CVM

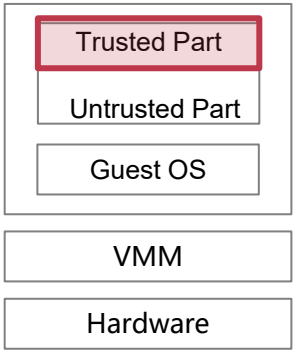
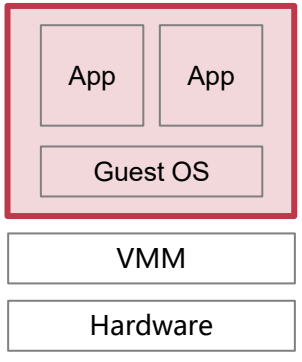
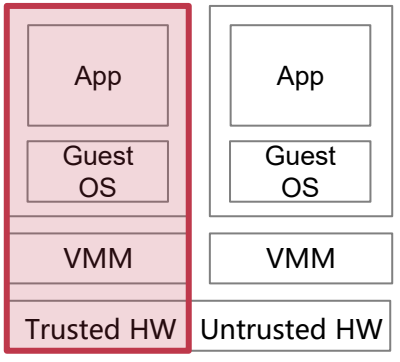
RISC-V HW & FW: PENGLAI

Penglai: An Open-Source TEE system

- **An open-sourced TEE system (hw & fw) on RISC-V platform**
  - Github repo：https://github.com/Penglai-Enclave/
  - Two modes: Penglai-Zone (Supervisor mode); Penglai-PMP (User mode)



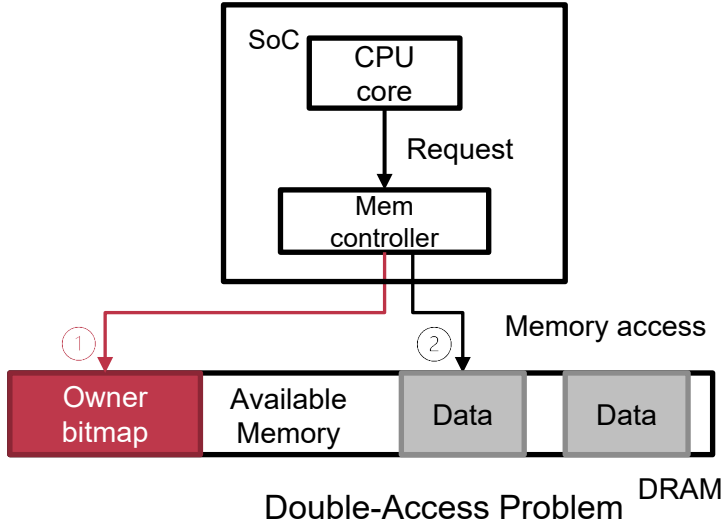
Feature of Three Different Types of TEE

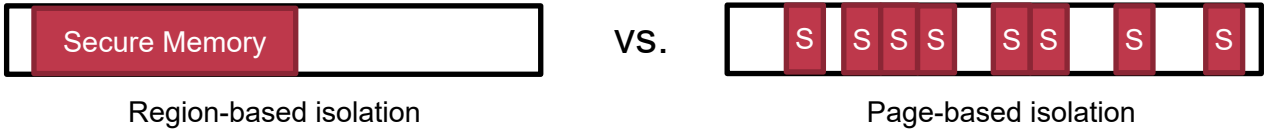
	Type-1：Enclave（Intel SGX）	Type-2：CVM（AMD/Intel/ARM）	Type-3：Zone（ARM TrustZone）
			
Scenario	Server + Mobile	Server (mostly)	Mobile (mostly)
TCB	CPU	CPU+I/O（optional）	SoC（CPU+I/O）
Good	Mem encryption	Mem encryption, flexible granularity	Whole system isolation
Bad	Side channel, weak I/O	Depend on VM abstraction	Usually no mem encryption
Penglai	**Penglai-Enclave**	**Penglai-CVM**	**Penglai-Zone**

https://github.com/Penglai-Enclave/

Problem-1: Limited Secure Memory Regions

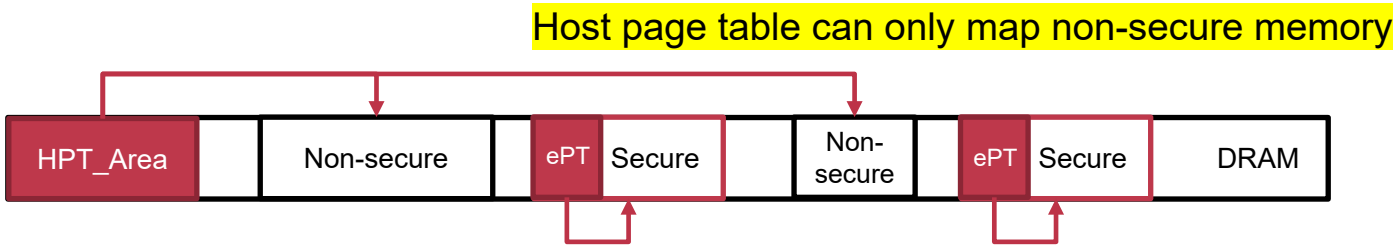
- **Coarse-grained memory isolation**
  - HW Tax: memory must be physically continuous
    - SGX has a fixed size preserved secure memory
    - TrustZone has limited secure memory regions (<16)
    - RISC-V Keystone uses 16 PMP regs for isolation
- **Page-level memory isolation**
  - Check Tax: One more access for each read/write
  - Performance overhead: 25% by average [TIMBER-V]





Research-1: Page-Table-Area for Secure Memory

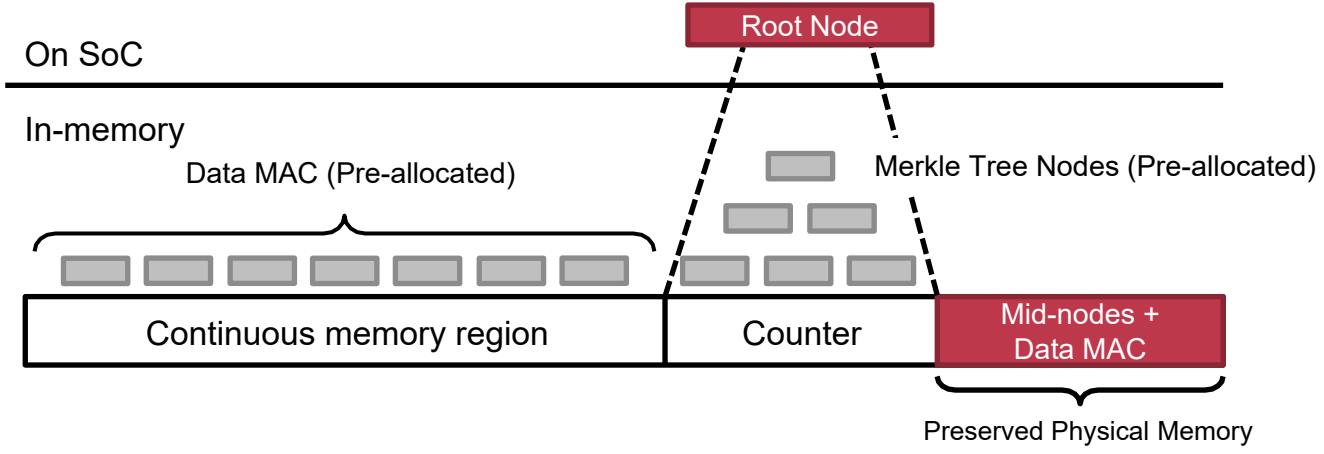
- **Penglai[1] restricts host page table (HPT) within a special area: HPT_Area**
  - An Enclave has its own page table (ePT), not accessed by the OS
  - Support **4KB** fine-grained secure memory mapping, can cover all memory
- **1. Extend MMU to do the security check for all host page table**
  - All host page table should be within the HPT_Area; only Penglai Monitor can change the host page table
- **2. Penglai Monitor will check the ownership of pages during page mapping**
  - Ensure that no Enclave's page will be mapped to host page table



1. [OSDI'21] Scalable Memory Protection in the PENGLAI Enclave.

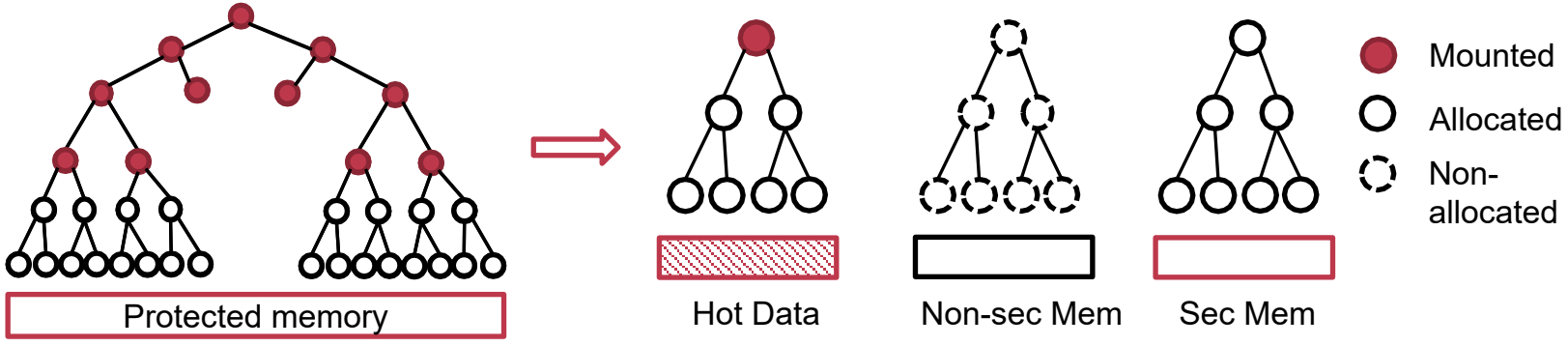
Problem-2: Scalability of Encrypted Memory

- **Hardware Tax: Scalability is Limited by the Merkle Tree**
  - Root nodes in SoC ⟷ Limited capacity in SoC
  - Pre-allocate middle nodes ⟷ Fixed memory overhead
  - Only continuous mem region ⟷ Coarse-grained protection



Research-2: Scalability of Memory Encryption

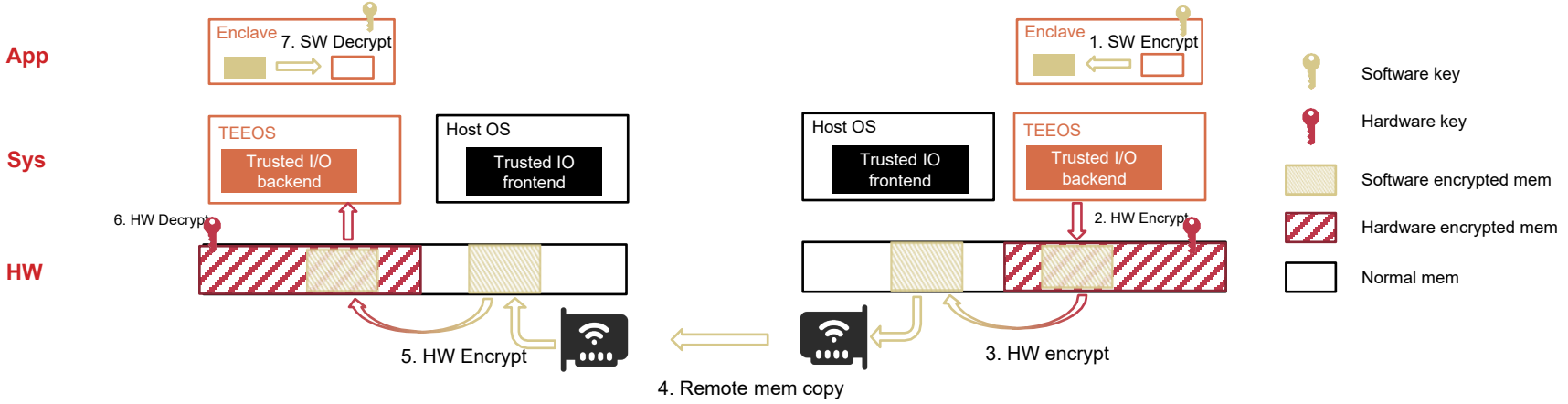
- **Mountable Merkle Tree[1]: Split the Tree to Sub-trees**
  - Three states: mounted (In-SoC), allocated (In-memory), not allocated
  - Protection range: from 256MB to 512GB



1. [OSDI'21] Scalable Memory Protection in the PENGLAI Enclave.

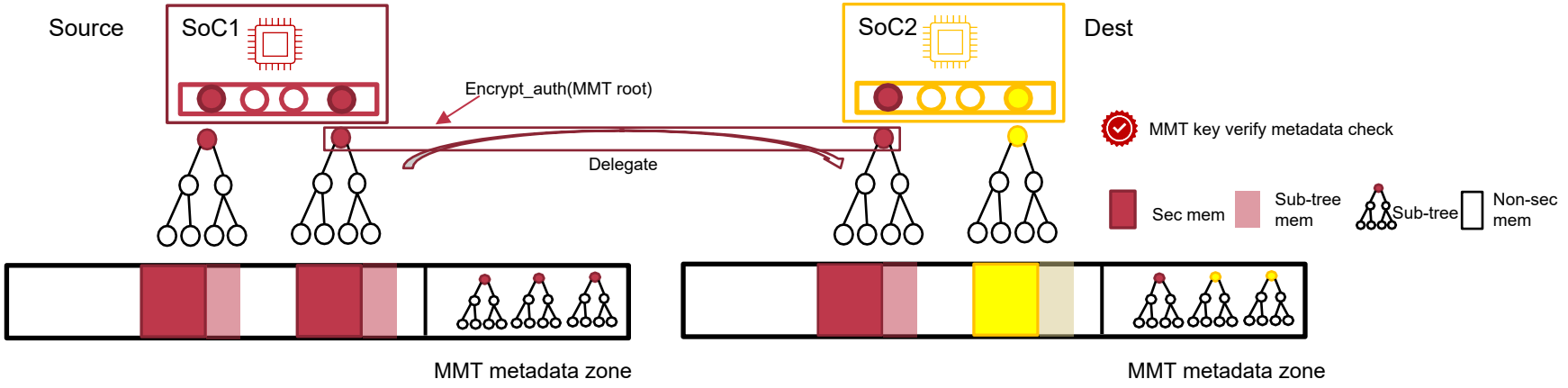
Problem-3: Encryption Tax for Network Data

- **Encryption Tax: Double Encryption from HW and SW**
  - Software encryption: TLS/SSL for secure channel within Enclaves
  - Hardware encryption: Data encryption between cache to memory
- **Encryption throughput (40Gb/s) v.s. RDMA bandwidth (400Gb/s)**



Research-3: Encryption Tax for Network Data

- **Observation: HW encryption is enough**
- **Send data and meta-data of HW encryption (Merkle Tree) together**
  - Negotiate encryption key to protect and verify Merkle Tree Root node



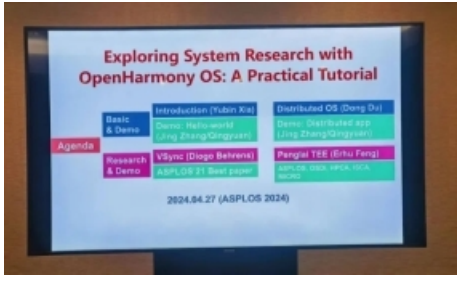
1. [HPCA'23] Efficient Distributed Secure Memory with Migratable Merkle Tree.

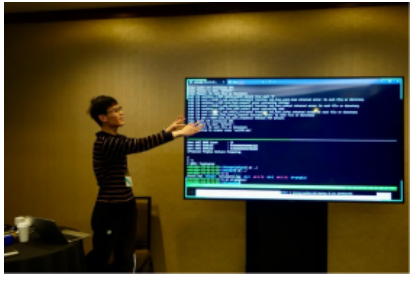
Other Research Projects

Key metrics	Current	Tech	Effort
Startup latency	Sec ~ 10 Secs	**Plugin-Enclave [ISCA'21]**	10 ms（10^2~10^3X）
		Init-less Booting [OSDI'21]	1 ~ 10 ms（10^2~10^3X）
Cross-dom latency	10^3 ~ 10^4 Cycle	**Async comm. mechanism [ISCA'19]**	10^2 Cycle（10^2~10^3X）
Secure mem size	256 MB	**Mountable Merkle Tree [OSDI'21]**	512GB（10^3X）
跨网络传输性能	<理论带宽的10%	**Migratable Merkle Tree [HPCA'23]**	87% of max thrpt（10~10^2X）
隔离实例并发数	10 ~ 10^2级	**Page-Table Area [OSDI'21]**	10^3（10~10^2X）

Academic: Technical Tutorial

- **ASPLOS 2024 tutorial: Penglai (on OpenHarmony)**
  - Report: Establish distributed TEE with the Penglai monitor
  - Tutorial: How to enable Penglai in the OpenHarmony





- **Academic Papers**
  - OSDI'21; HPCA'23; MICRO'23; ASPLOS'24; ISCA'24

Scalable Memory Protection in the PENGLAI Enclave

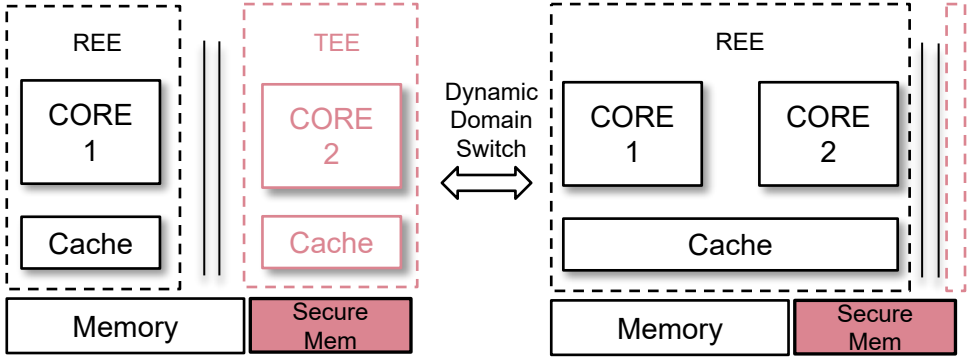
sIOPMP: Scalable and Efficient I/O Protection for TEEs

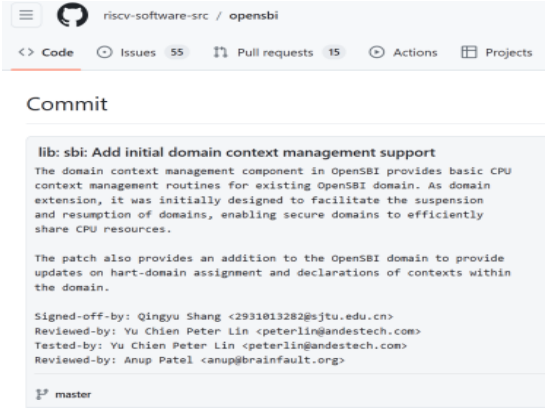
sNPU: Trusted Execution Environments on Integrated NPUs

Efficient Distributed Secure Memory with Migratable Merkle Tree

Open-Source: Penglai & OpenSBI

- **RISC-V firmware supports Penglai secure monitor**
  - OpenSBI: RISC-V Open-Source Supervisor Binary Interface
  - Underlying mechanism in Penglai: dynamic secure domain
    - Strong isolation for CPU, memory and I/O device
    - Domain switch between secure and non-secure world

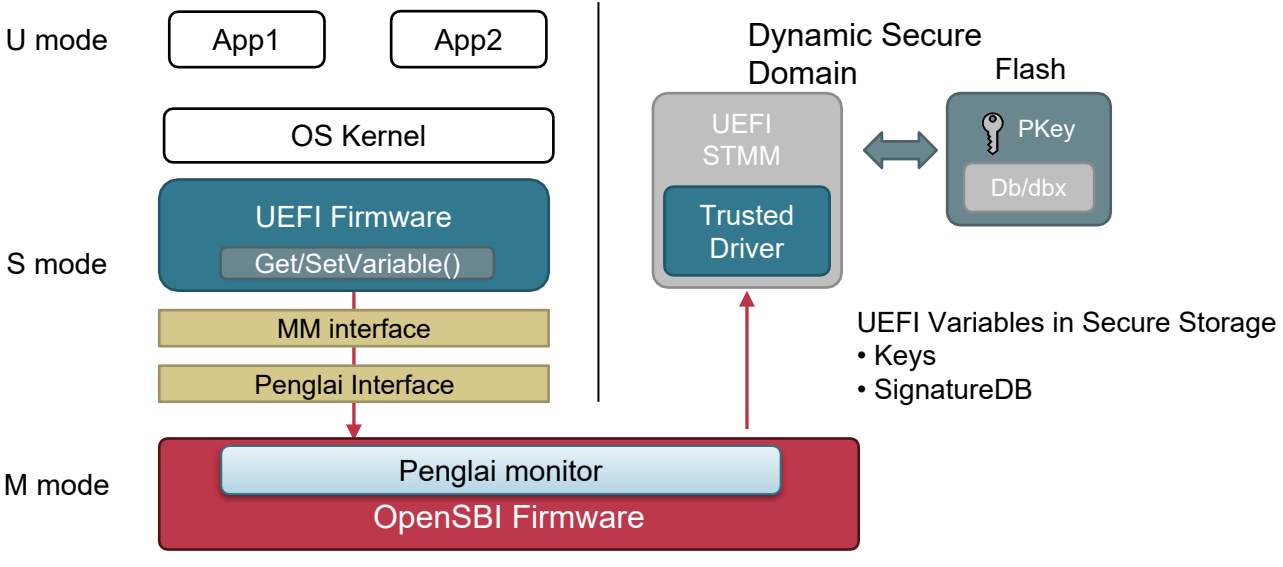




Merge in OpenSBI v1.5.1

Open-Source: Penglai & UEFI

- **RISC-V UEFI Secure Boot: using Penglai solution**
  - UEFI: UEFI Firmware (non-secure) + Standalone MM (Secure)
  - STMM runs in the dynamic secure domain (Penglai-Zone)
  - An open-sourced project supported by RISE (RISC-V Software Ecosystem)



EDK2_00_02_06 UEFI SecureBoot

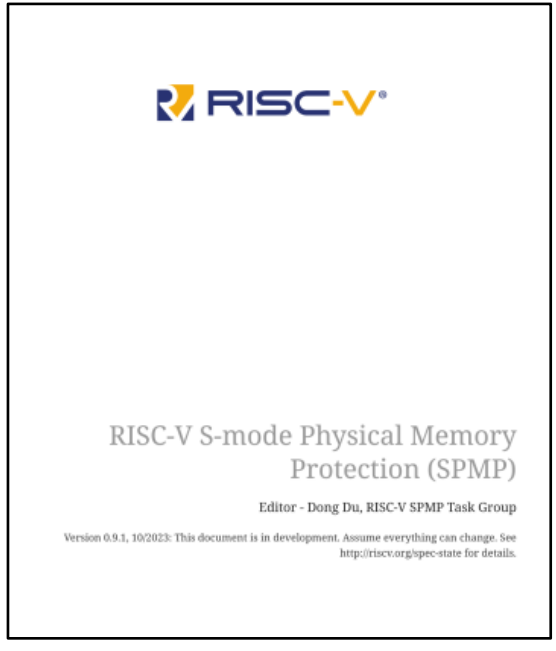
Stakeholders and Partners

- RISE
  - Ventana: Tuan Pan <tphan@ventanamicro.com>
  - Intel: Yong Li <yong.li@intel.com>
  - Rivos: Dhaval Sharma <dhaval@rivosinc.com>, Samuel Ortiz
  - Nivida: Jeff Brasen <jbrasen@nvidia.com>, Girish Mahadevan
- External
  - IPADS: Erhu Feng, Qingyu Shang
  - StarFive: Cheehong Ang <cheehong.ang@starfivetech.com>,J

Collaborate with Intel in the RISE group

RISC-V Specification: SPMP

- **RISC-V ISA Extension: S-mode Physical Memory Protection**
  - Establish the SPMP technique group (TG) from 2021 in RISC-V community
  - RISC-V SPMP spec is in *frozen* state in 2025
  - Enhance the memory security for IoT devices



SPMP extension is introduced in 2023 Linux Open Foundation Annual Report

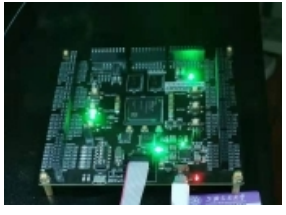
Industry: RISC-V Chip Vendors

- **Penglai supports various RISC-V chips and SoCs**

Sifive unmatched

StarFive VF-1/2

T-HEAD c910



Nucler N200

The first "Trusted Digital Space" on RISC-V with Penglai

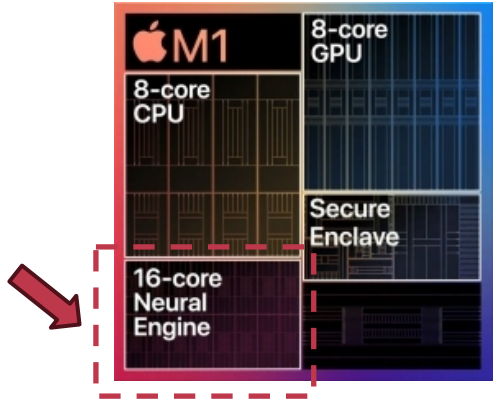
- **Penglai has been widely used in the industry**
  - Penglai is the default security firmware in Huawei cloud system for RISC-V
  - StarFive deploys Penglai in RISC-V IoT devices: water-meter, camera, router, etc.
  - Nuclei and Andes adopt Penglai extension to support OP-TEE applications

⇦ RISC-V blog: How does Penglai help industry build TEE ecosystem

NEXT: PENGLAI FOR AI

AI on more terminal devices

- **Variant AI applications:**
  - AI-agent, Self-driving, Edge LLMs, etc..
- **AI accelerators: Neural Processing Unit (NPU)**
  - Apple NPU, Intel NPU, Arm DSP/NPU



Self-Driving

AI-agent: Apple Intelligence

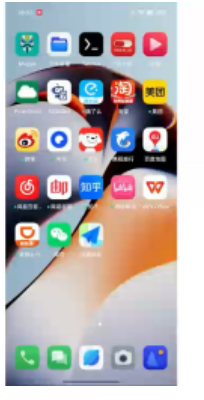
It’s coming! Mixtral (8x7B) on a Phone

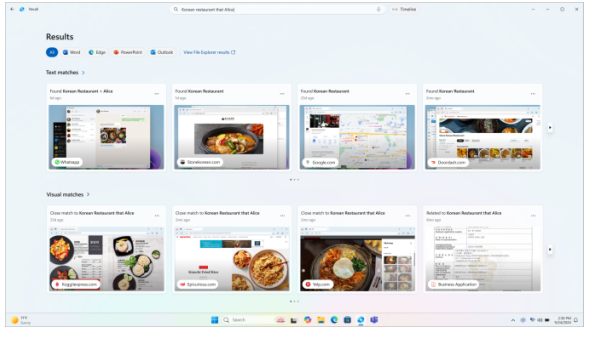
One Plus 12a

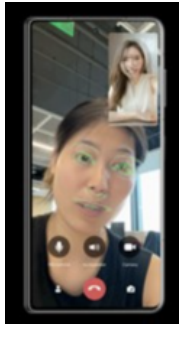
PowerInfer-2: Fast Large Language Model Inference on a Smartphone

Mobile agent based on LLM

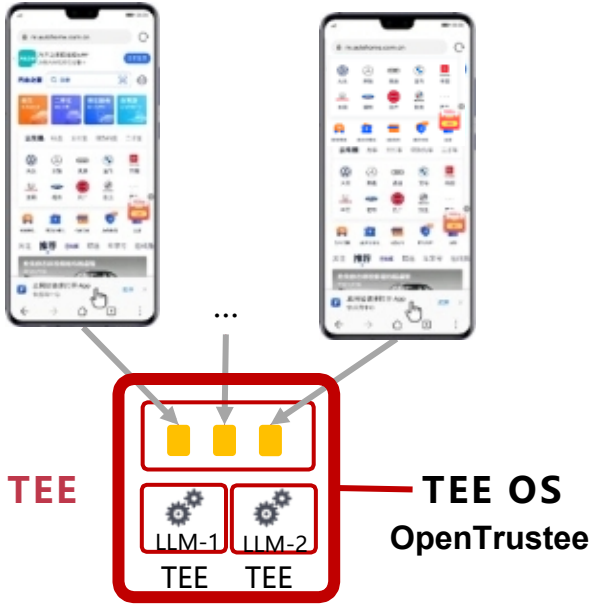
- Auto operations with high-level semantic
- Efficient information retrieval (e.g., Microsoft Recall)
- Real-time on-device deepfake detection







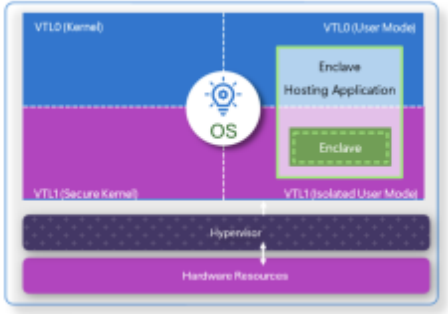
How to protect AI data? TEE is a solution



September 27, 2024

Update on Recall security and privacy architecture

By David Weston, Vice President Enterprise and OS Security at Microsoft



- Virtualization Based Security (VBS) – the hypervisor provides the secure enclave environment, which loads integrity-verified code into a confidential and isolated TEE.

Security Tax will become more critical in this scenario


上海交通大學
SHANGHAI JIAO TONG UNIVERSITY
Thanks
饮 水 思 源 爱 国 荣 校