BLUEHAT
SHANGHAI 2019

MesaTEE SGX：借助 Intel SGX 重新定义人工智能和大数据分析

Yu Ding

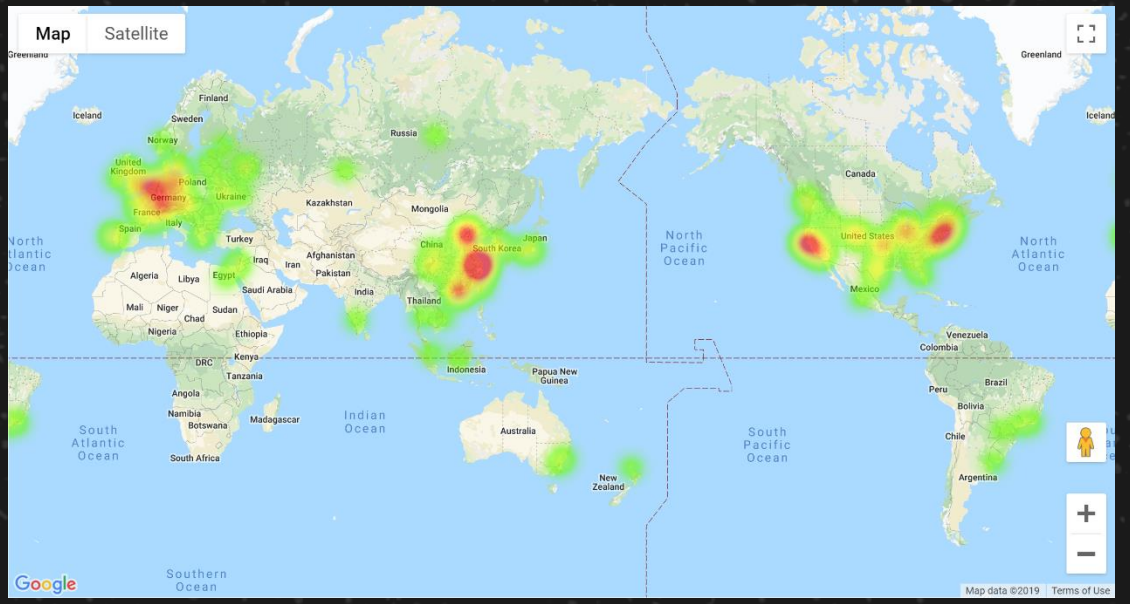
百度 X-Lab 安全研究员

May-29-2019

自我介绍

- 在百度 X-Lab 担任安全研究员
- Rust 爱好者
- 漏洞利用/缓解领域博士
- 从事 Rust-SGX 项目

- https://dingelish.com
- https://github.com/dingelish
- https://github.com/baidu/rust-sgx-sdk



MesaTEE SGX

借助 Intel SGX 重新定义人工智能和大数据分析

适用于 隐私保护 计算的 Intel SGX

- Intel SGX 背景
- 基于 Intel SGX 构建隐私保护计算软件栈所面临的挑战

Hybrid Memory Safety

- 经验法则
- Intel SGX 实践

塑造 安全 并且 可信 的人工智能/大数据分析框架

- 可信到底指什么?
- 使用 Intel SGX 实现可信赖的人工智能和大数据分析

MesaTEE SGX

借助 Intel SGX 重新定义人工智能和大数据分析

适用于 隐私保护 计算的 Intel SGX

- Intel SGX 背景
- 基于 Intel SGX 构建隐私保护计算软件栈所面临的挑战

Hybrid Memory Safety

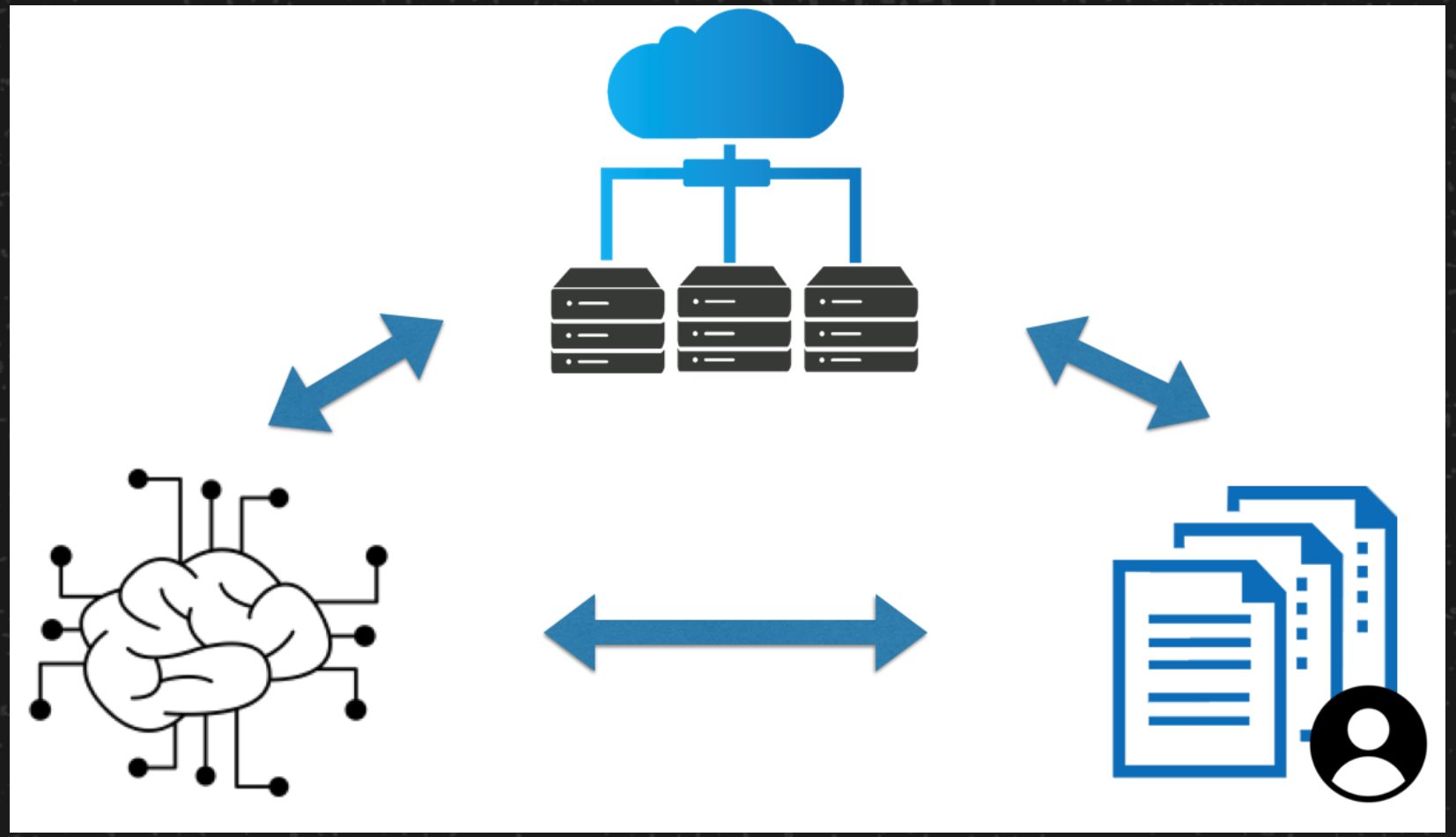
- 经验法则
- Intel SGX 实践

塑造 安全 并且 可信 的人工智能/大数据分析框架

- 可信 (Trustworthy) 到底指什么?
- 使用 Intel SGX 实现可信赖的人工智能和大数据分析

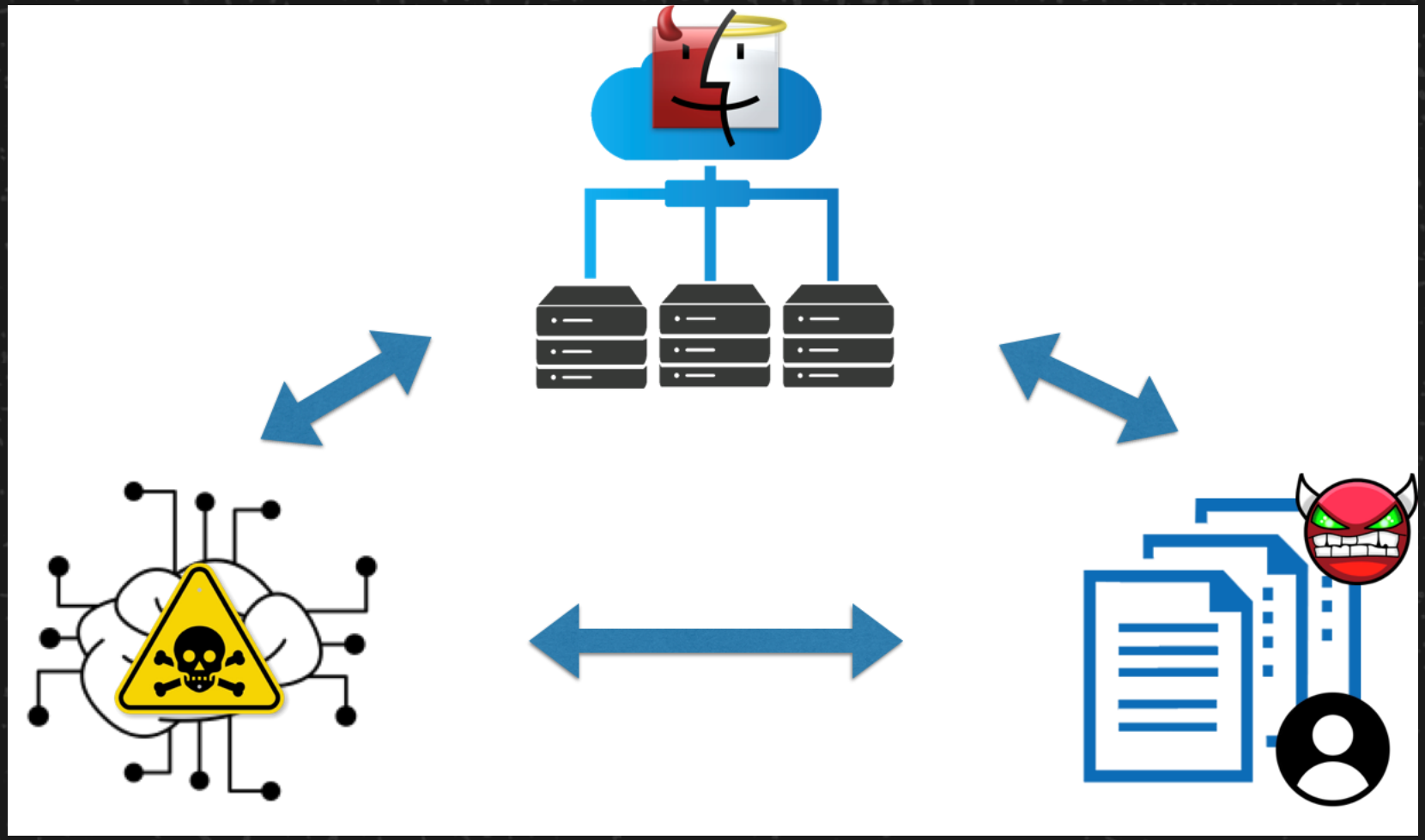
MesaTEE SGX

借助 Intel SGX 重新定义人工智能和大数据分析



MesaTEE SGX

借助 Intel SGX 重新定义人工智能和大数据分析



MesaTEE SGX

借助 Intel SGX 重新定义人工智能和大数据分析

- 云供应商
- 数据所有者
- 算法提供商（也可以是数据所有者）

- 相互之间无法信任
- 数据离开所有者后依然可以 保证 能够 受到控制

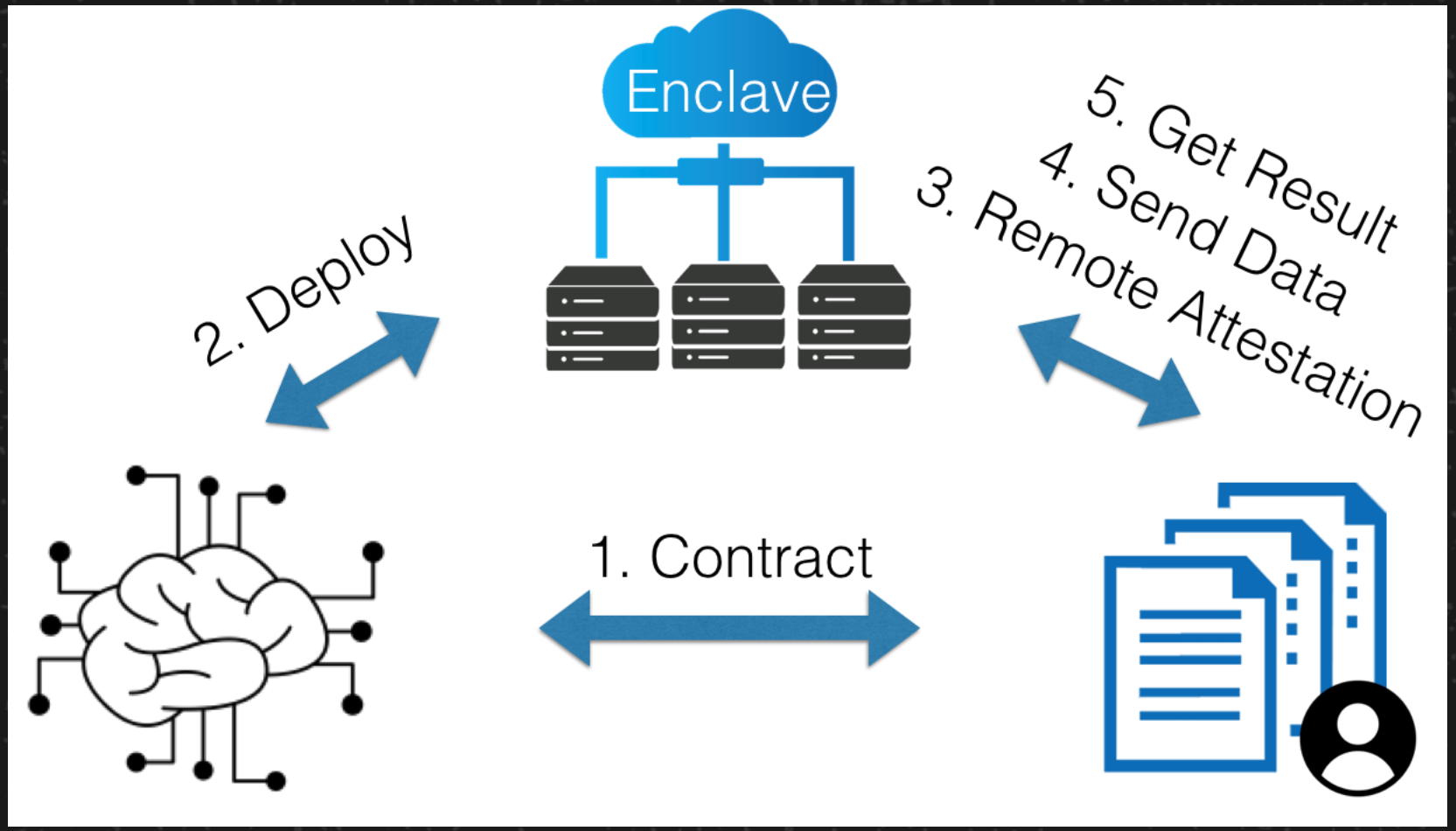
MesaTEE SGX

借助 Intel SGX 重新定义人工智能和大数据分析

- **解决方案概述**
- **使用 Intel SGX 建立信任和 TEE**
  - 安全可信的身份验证/授权
  - 安全可信的渠道
  - 安全可信的执行环境
- **使用 hybrid memory safety 构建系统**
- **可信赖的人工智能和大数据分析**

MesaTEE SGX

借助 Intel SGX 重新定义人工智能和大数据分析



MesaTEE SGX

借助 Intel SGX 重新定义人工智能和大数据分析

适用于 隐私保护 计算的 Intel SGX

- Intel SGX 背景
- 基于 Intel SGX 构建隐私保护计算软件栈所面临的挑战

Hybrid Memory Safety

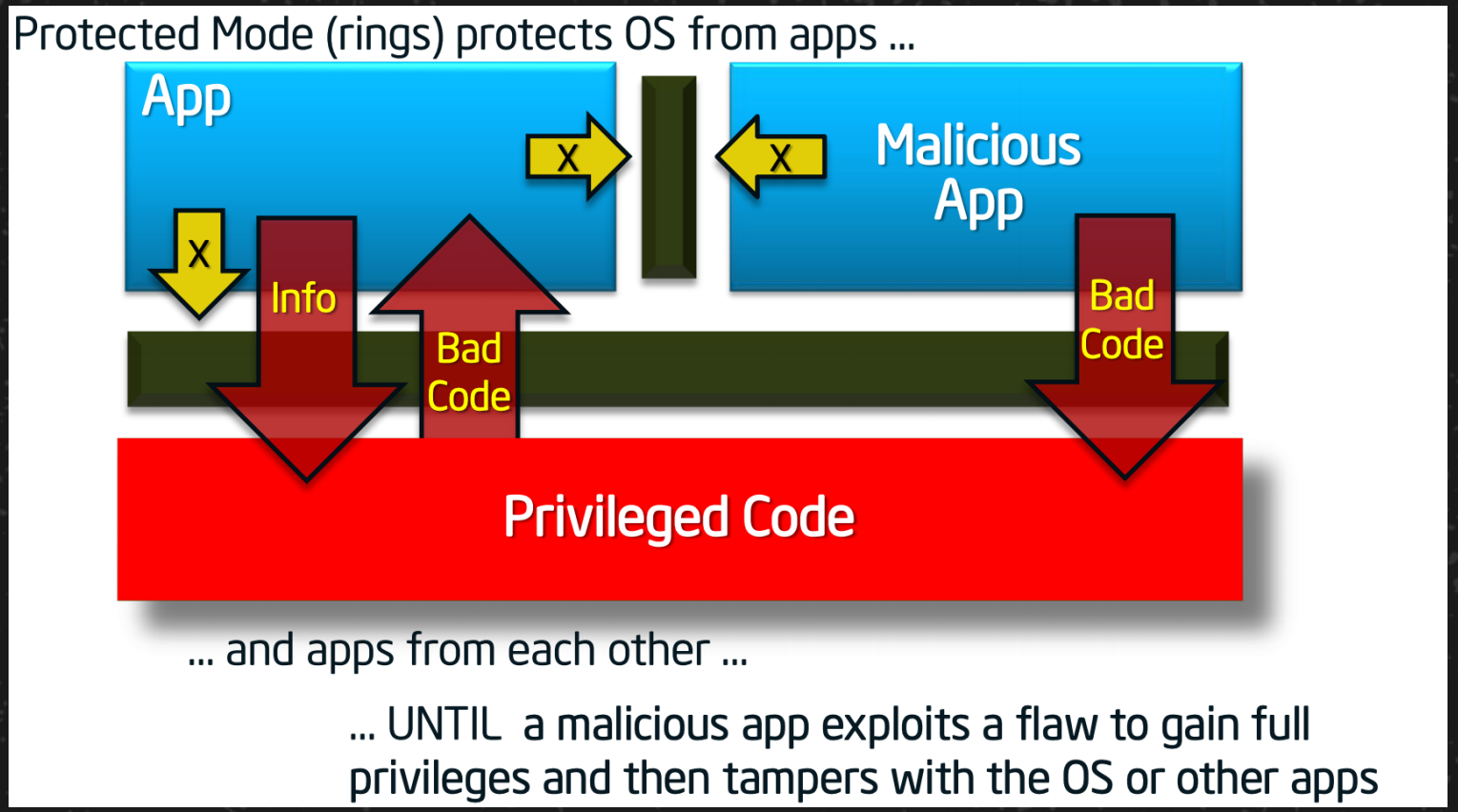
- 经验法则
- Intel SGX 实践

塑造 安全 并且 可信 的人工智能和大数据分析框架

- 可信 (Trustworthy) 到底指什么?
- 使用 Intel SGX 实现可信赖的人工智能和大数据分析

Intel SGX 背景

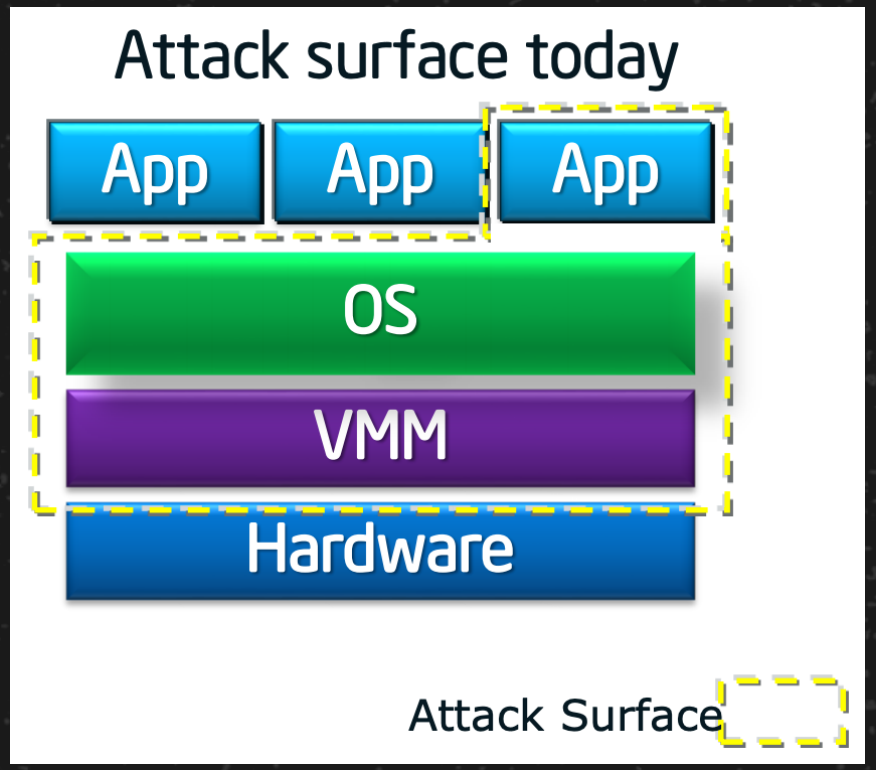
面对高特权代码攻击，应用无法受到保护

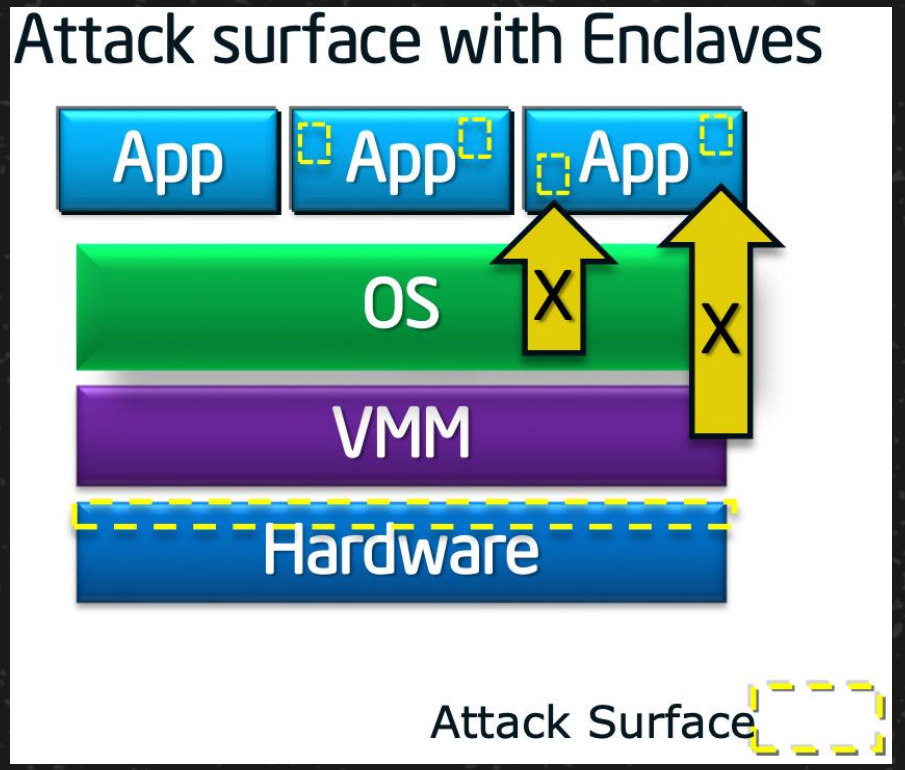


Intel® Software Guard Extensions(Intel® SGX)
Frank McKeen, Intel Labs, April 15, 2015

Intel SGX 背景

使用/不使用 Intel SGX Enclaves 时的攻击面

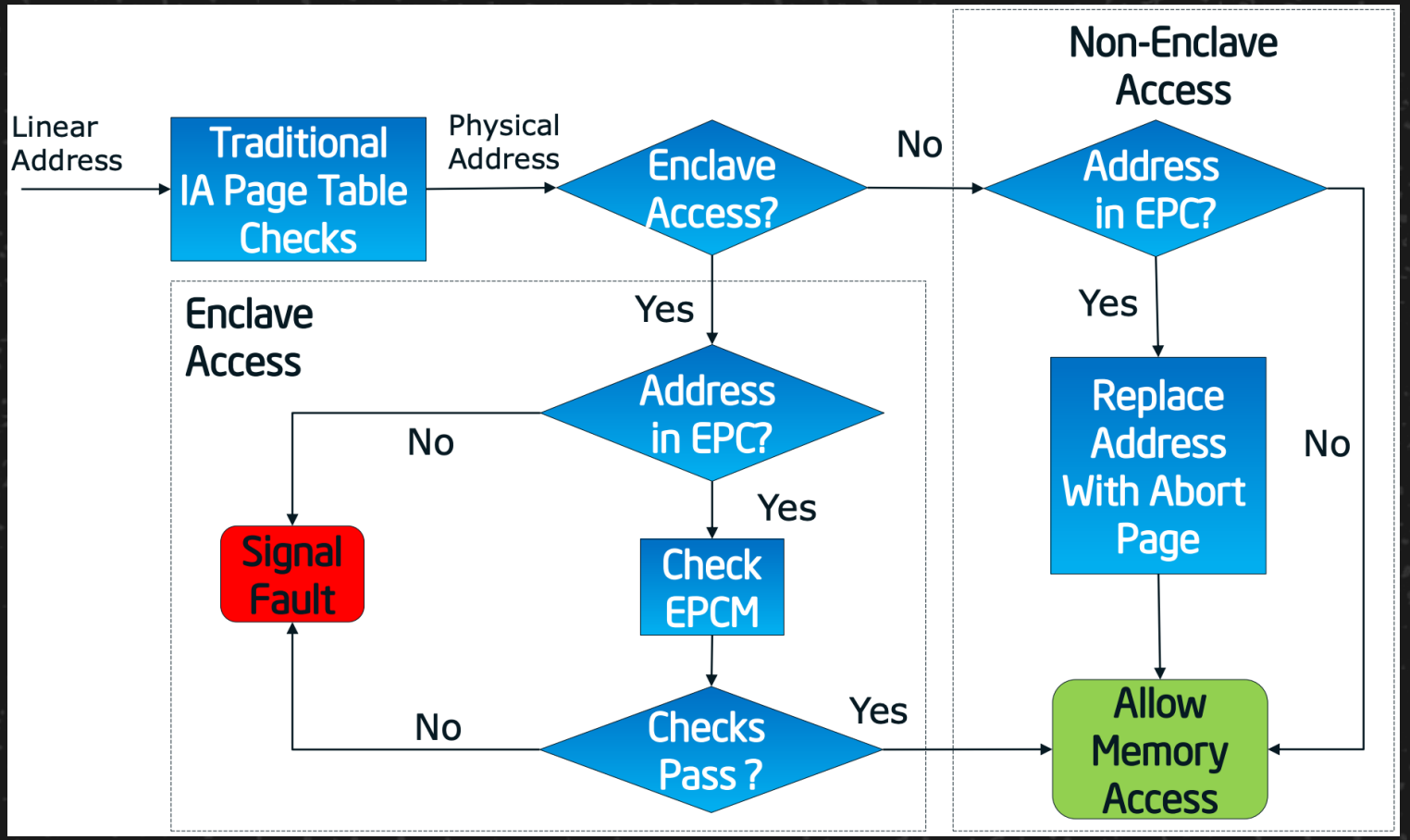




Intel® Software Guard Extensions(Intel® SGX)
Frank McKeen, Intel Labs, April 15, 2015

Intel SGX 背景

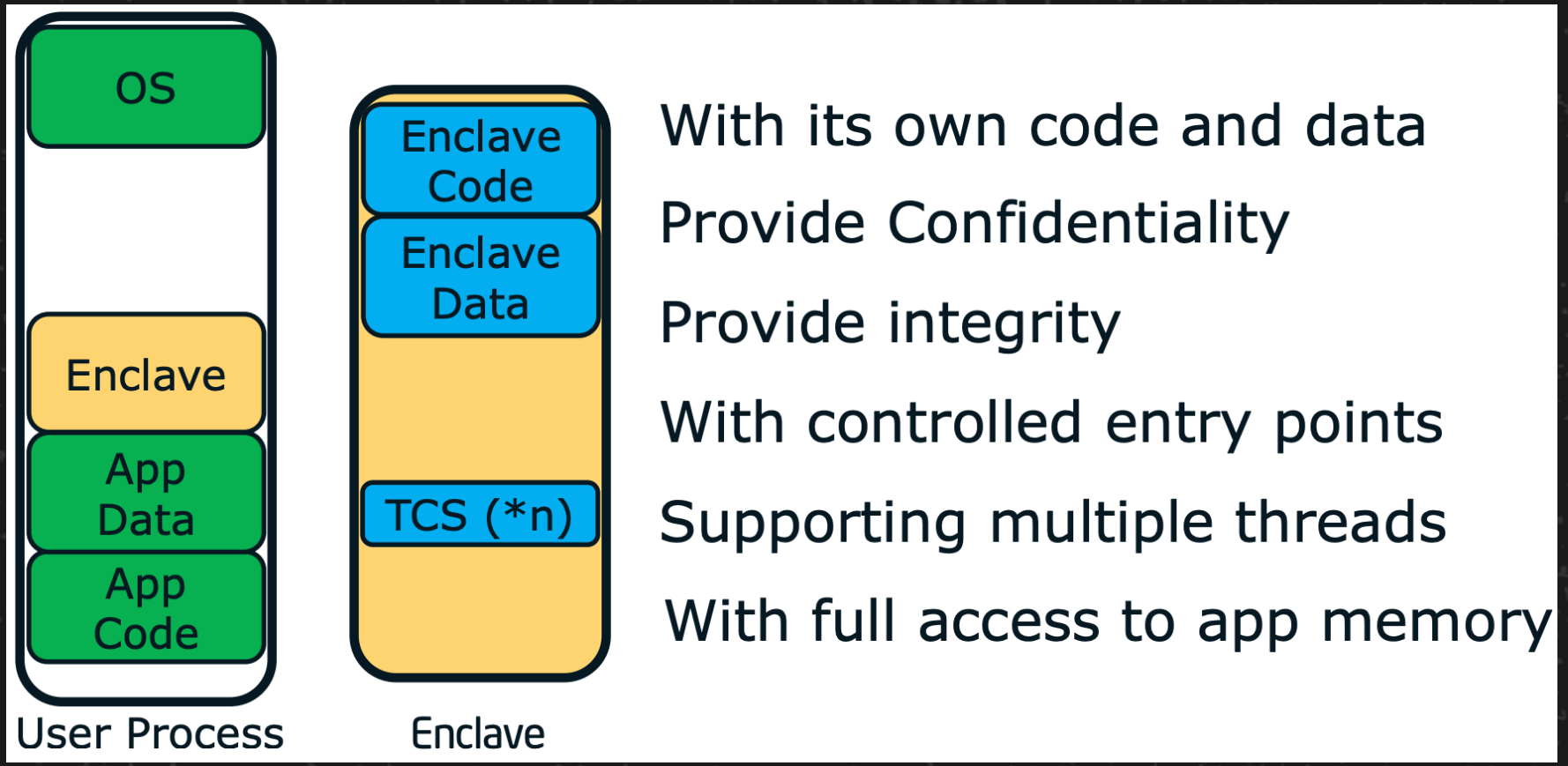
地址转换过程中的内存访问控制



Intel® Software Guard Extensions(Intel® SGX)
Frank McKeen, Intel Labs, April 15, 2015

Intel SGX 背景

机密性和完整性保证



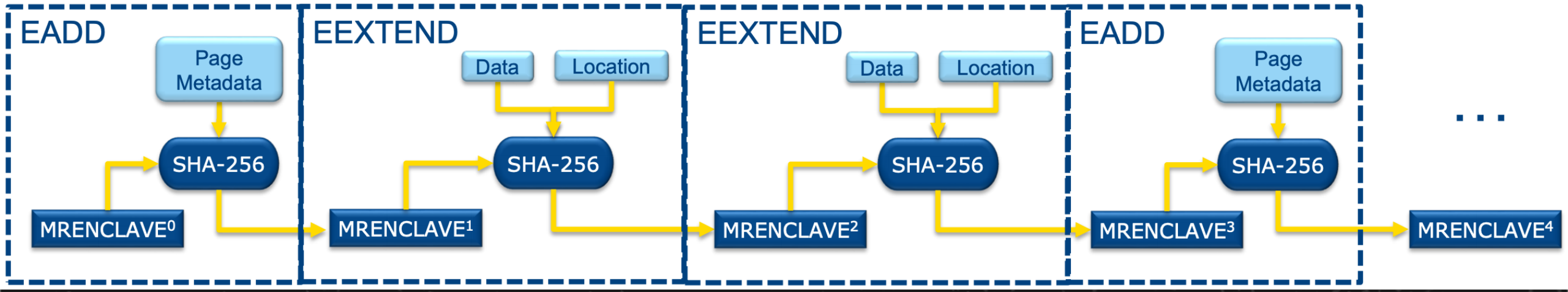
Intel® Software Guard Extensions(Intel® SGX)
Frank McKeen, Intel Labs, April 15, 2015

Intel SGX 背景

测量和证实

验证测量/签名方

通过远程证实（Remote Attestation）建立信任



Sealing and Attestation in Intel® Software Guard Extensions (SGX)
Rebekah Leslie-Hurd, Intel® Corporation, January 8th, 2016

Intel SGX 背景

远程证实

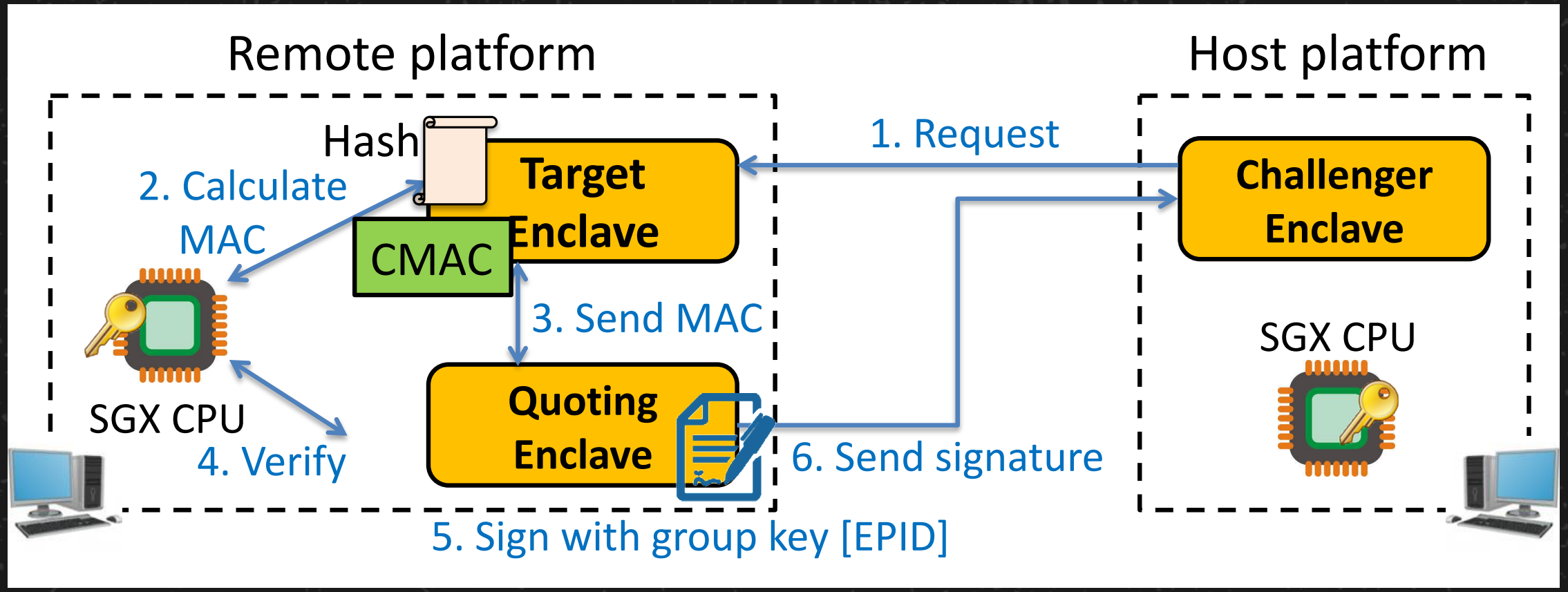


Figure is from "A First Step Towards Leveraging Commodity Trusted Execution Environments for Network Applications", Seongmin Kim et al.

Intel SGX 背景

Intel SGX 的简单总结

- **为任何应用程序提供保密能力**
- 使用全新处理器指令提供该能力
- 应用程序可支持多个飞地（Enclave）
- **提供完整性和机密性**
- 抵御硬件攻击
- 防止软件访问，包括高特权软件和 SMM
- **应用程序在操作系统环境内部运行**
- 应用程序开发者的学习曲线更低
- 面向所有开发者开放

Intel SGX 背景

基于 Intel SGX 构建隐私保护计算软件栈所面临的挑战

- **Intel SGX 的硬件局限**
- 无 syscall
- 无 RDTSC
- 无 CPUID
- 128 Mbyte 的 EPC 内存。页面错误驱动的内存交换速度缓慢
- 无 mprotect

Intel SGX 背景

基于 Intel SGX 构建隐私保护计算软件栈所面临的挑战

- **Intel SGX 的硬件局限 => 挑战**
- 无 syscall
  - 无 fs/net/env/proc/thread/...
- 无 RDTSC
  - 无可信任的时间，如何验证 TLS 证书?
- 无 CPUID
  - 为了改善性能，某些 Crypto 库需要 CPUID
- 128 Mbyte 的 EPC 内存。页面错误驱动的内存交换速度缓慢
  - 人工智能？大数据分析?
- 无 mprotect：JIT？AOT?

Intel SGX 背景

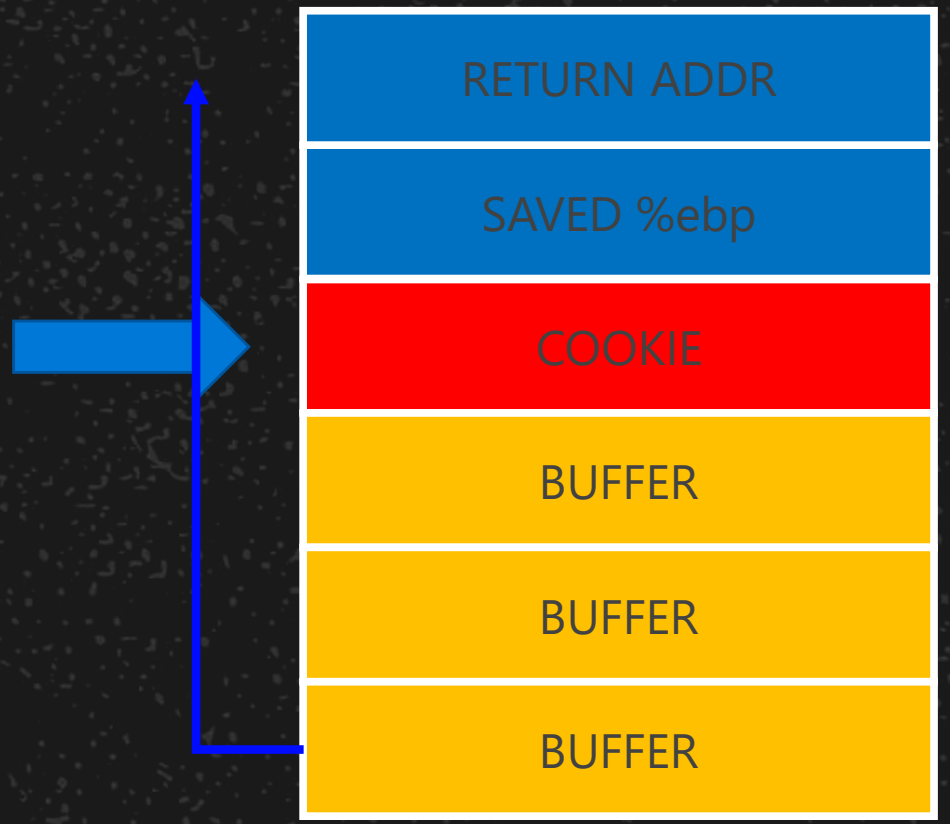
基于 Intel SGX 构建隐私保护计算软件栈所面临的挑战

- Int
- 无 sy
  - 无
- 无 R
  - 无
- 无 C
  - 为
- 128
  - 人
- 无 mprotect：JIT？AOT？

Intel SGX 背景

基于 Intel SGX 构建隐私保护计算软件栈所面临的挑战

- **Intel SGX 的软件局限**
- 存在内存 Bug
- **内存安全?**
  - 溢出?
  - UAF?
  - 数据争用?
  - ROP?



Intel SGX 背景

基于 Intel SGX 构建隐私保护计算软件栈所面临的挑战

- **Intel SGX 的软件局限**
- 存在内存 Bug
- **内存安全?**
  - 溢出?
  - UAF?
  - 数据争用?
  - ROP?

Intel SGX 背景

基于 Intel SGX 构建隐私保护计算软件栈所面临的挑战

- **简要总结**
- 挑战
  - 在 **有限的基础** 前提下，在 Intel SGX 环境中重新实现一套软件栈
  - **需要** 保证内存安全性

MesaTEE SGX

借助 Intel SGX 重新定义人工智能和大数据分析

适用于 隐私保护 计算的 Intel SGX

- Intel SGX 背景
- 基于 Intel SGX 构建隐私保护计算软件栈所面临的挑战

Hybrid Memory Safety

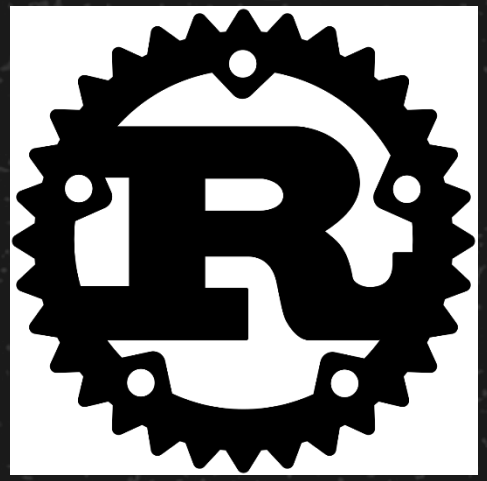
- 经验法则
- Intel SGX 实践

塑造 安全 并且 可信 的人工智能和大数据分析框架

- 可信 (Trustworthy) 到底指什么?
- 使用 Intel SGX 实现可信赖的人工智能和大数据分析

混合内存安全性 Hybrid Memory Safety

由编程语言保证内存安全性



混合内存安全性 Hybrid Memory Safety

软件栈

- 内核
- 系统调用
- Libc库、系统库
- 运行时库
- 应用程序

混合内存安全性 Hybrid Memory Safety

软件栈

- 内核
- 系统调用
- Libc库、系统库
- 运行时库
- 应用程序

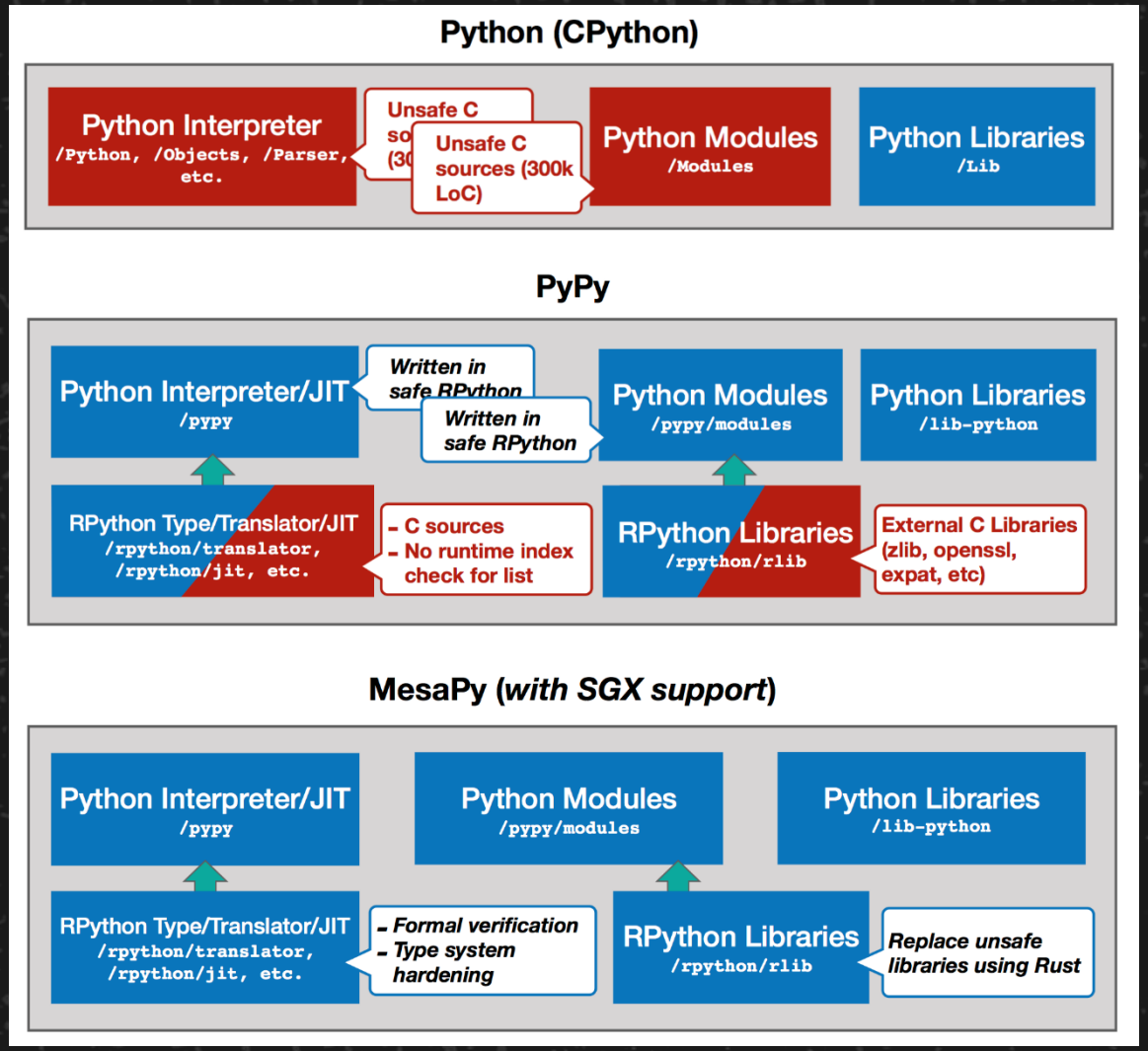
混合内存安全性 Hybrid Memory Safety

混合内存安全新——经验法则

- 不安全的组件绝对不允许污染安全的组件，对公开的 API 和数据结构，这一点尤为重要。
- 不安全的组件应当尽可能少，并与安全的组件解耦。
- 部署过程中，不安全的组件应明确标记出来并准备对其升级。

混合内存安全性 Hybrid Memory Safety

混合内存安全性——以 MesaPy 为例



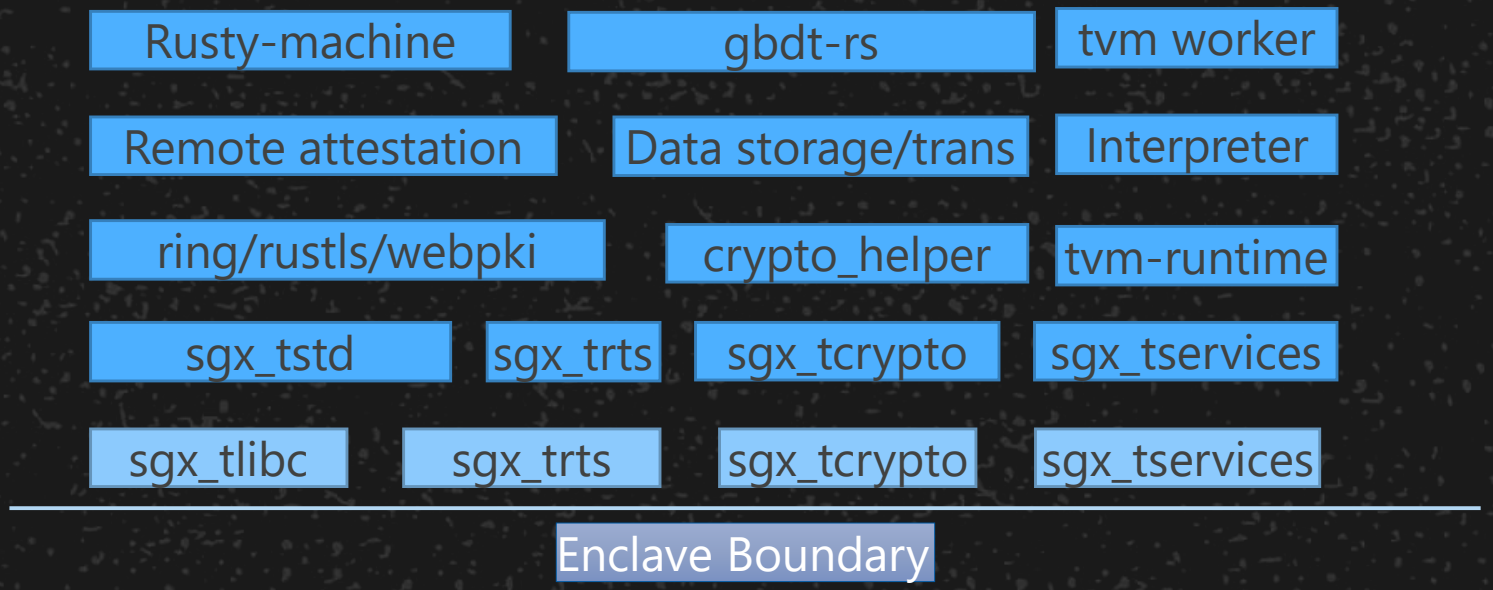
混合内存安全性 Hybrid Memory Safety

混合内存安全性——SGX 中的实践

Linux	Rust-SGX
内核	不适用
系统调用	OCALL（静态控制）
Libc	Intel – SGX tlibc
运行时	Rust-SGX sgx_tstd/…

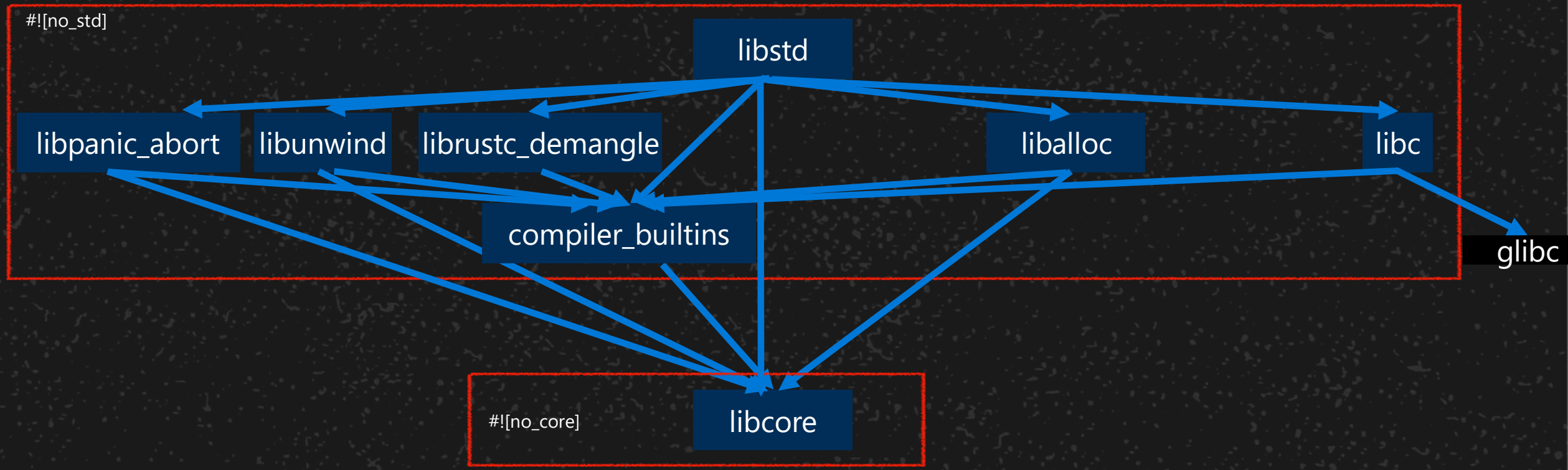
混合内存安全性 Hybrid Memory Safety

混合内存安全性——SGX 中的实践



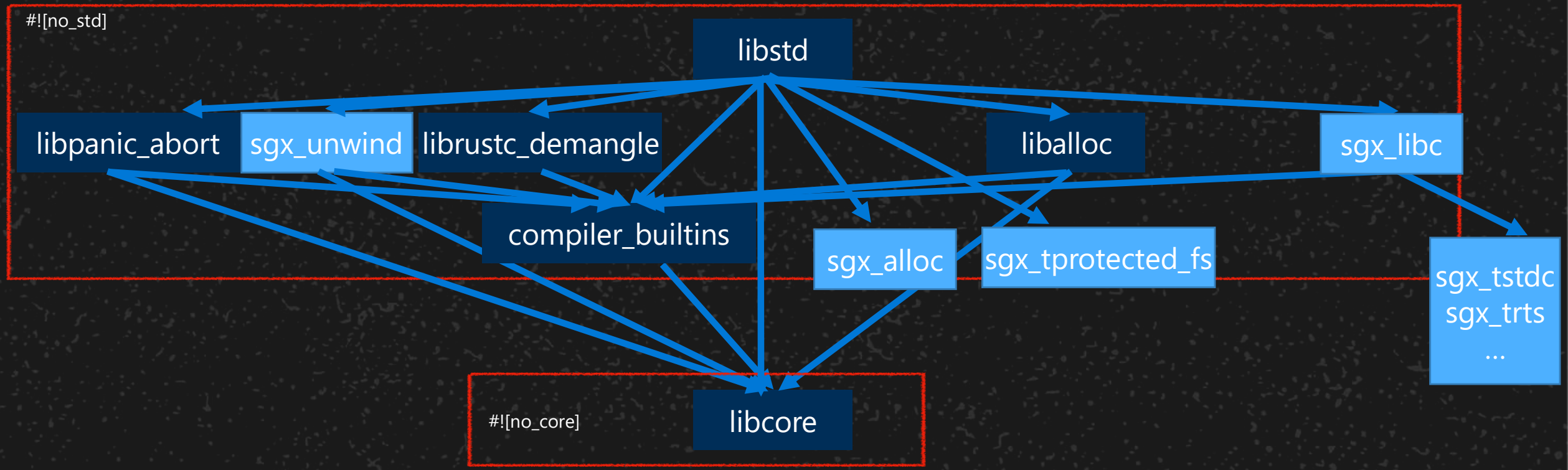
混合内存安全性 Hybrid Memory Safety

混合内存安全性——SGX 中的实践



混合内存安全性 Hybrid Memory Safety

混合内存安全性——SGX 中的实践



MesaTEE SGX

借助 Intel SGX 重新定义人工智能和大数据分析

适用于 隐私保护 计算的 Intel SGX

- Intel SGX 背景
- 基于 Intel SGX 构建隐私保护计算软件栈所面临的挑战

Hybrid Memory Safety

- 经验法则
- Intel SGX 实践

塑造 安全 并且 可信 的人工智能和大数据分析框架

- 可信 (Trustworthy) 到底指什么?
- 使用 Intel SGX 实现可信赖的人工智能和大数据分析

塑造安全并且可信的人工智能和大数据分析框架

可信 (Trustworthy) 到底指什么?

可信 (Trustworthy) 到底指什么?

塑造安全并且可信的人工智能和大数据分析框架

可信 (Trustworthy) 到底指什么?

可信赖计算（Trustworthy Computing）一词代表具备固有安全性、可用性以及可靠性的计算系统。这一概念尤其与 **微软** 曾在 2002 年发起的一项同名举措密切相关。

塑造安全并且可信的人工智能和大数据分析框架

可信到底指什么?

可信任计算 (Trusted Computing)

该术语源自可信任系统这一领域，但有着特殊含义。对于可信任计算，计算机将始终如一地按照 预期 方式运作，而具体的运作行为则可由计算机硬件和软件加以控制。

塑造安全并且可信的人工智能和大数据分析框架

使用 Intel SGX 实现可信的人工智能和大数据分析

Gradient-Boosting 决策树

如何实现可信?

- 所运行的实例是通过我想要运行的静态库启动的
- 该静态库是通过我想要使用的代码生成的
- 我所用的代码“诚实地”实现了算法
- 编译器没有作恶
- 数据以安全的方式传输

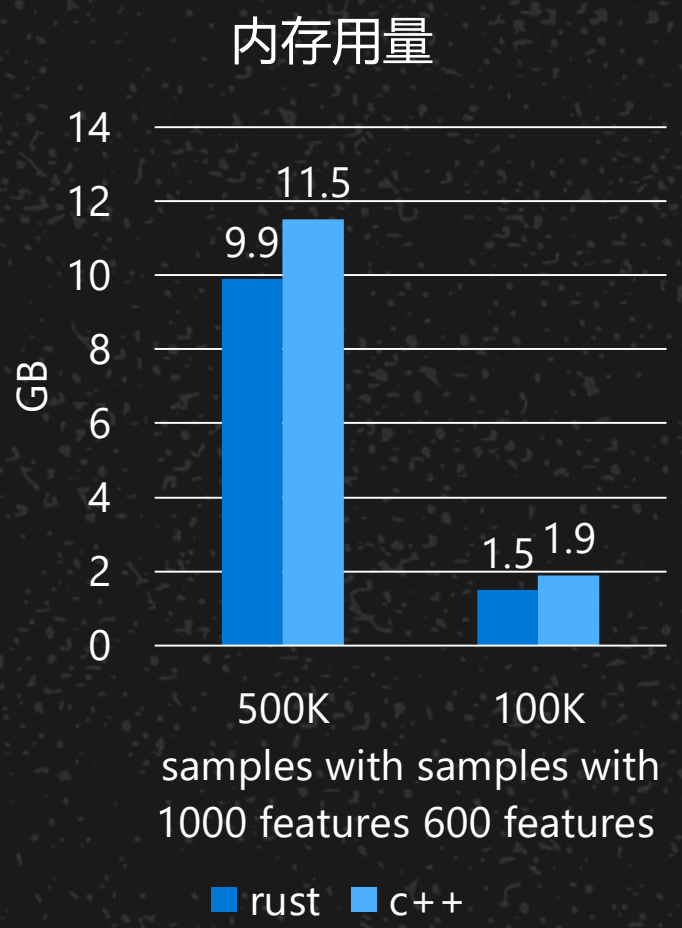
塑造安全并且可信的人工智能和大数据分析框架

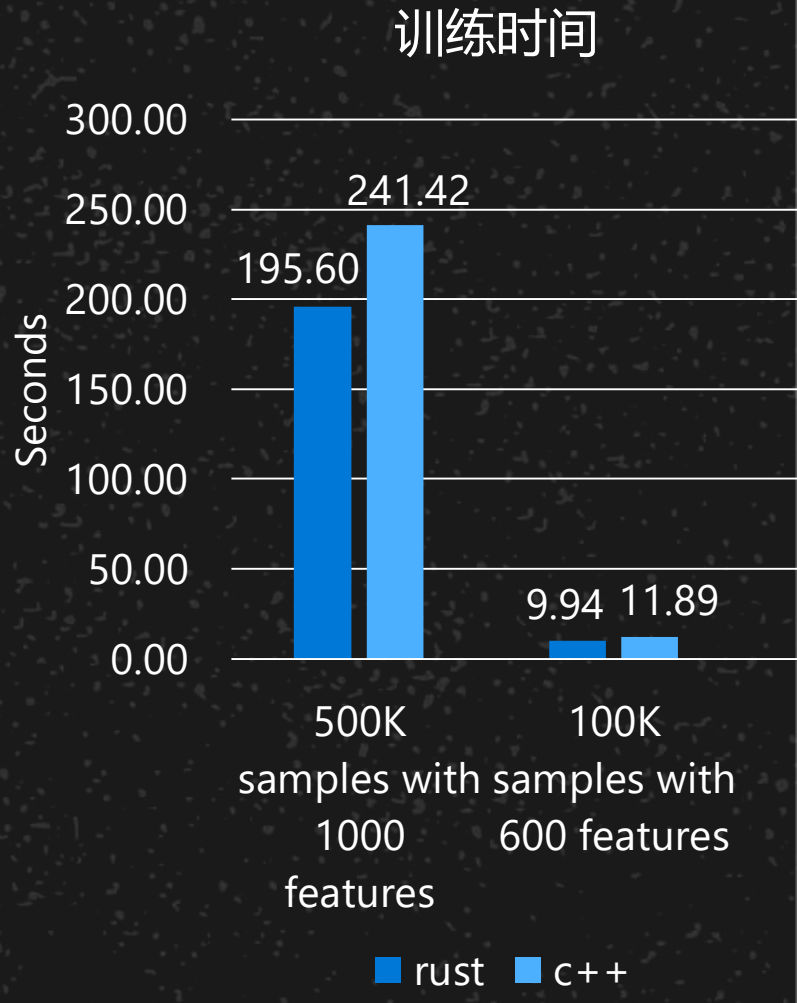
使用 Intel SGX 实现可信的人工智能和大数据分析

Gradient-Boosting 决策树

gbdt-rs

- ~2000 sloc of Rust – Self explain
- 良好的备注/文档
- 相比 XGBoost on 1thread 速度快 7 倍
- 与 SGX 无缝配合
- 简洁干净的 软件栈!



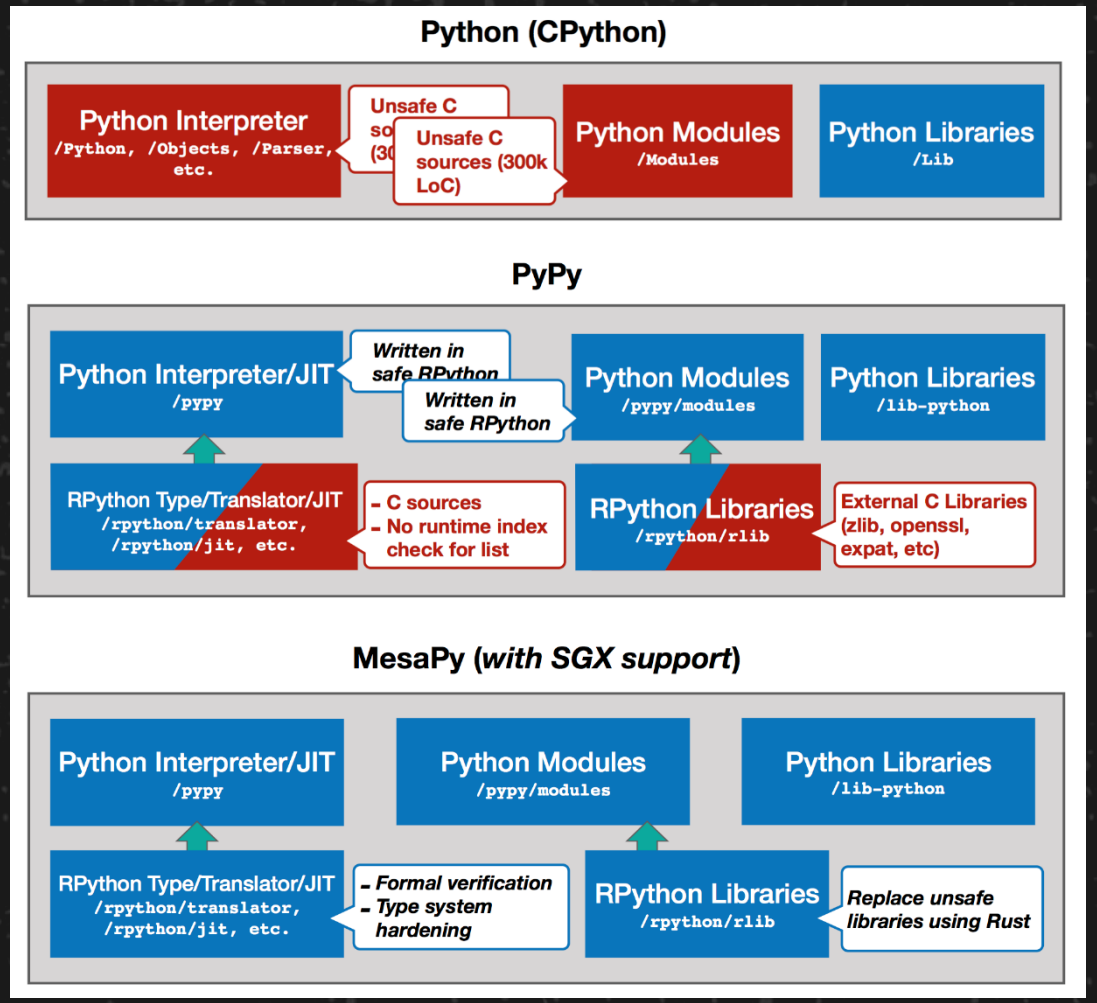


塑造安全并且可信的人工智能和大数据分析框架

使用 Intel SGX 实现可信的人工智能和大数据分析

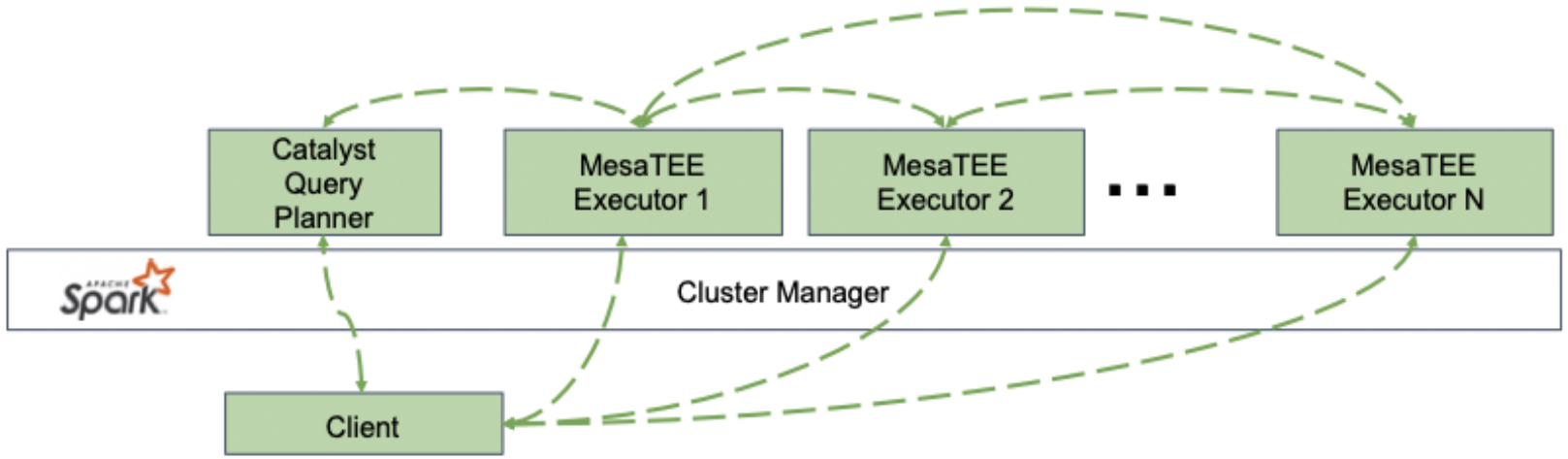
MesaPy SGX

- 移植具备强边界检查的 PyPy
- 禁用所有系统调用
- 可定制的运行时 – 有限的 ocall
- 消除非决定性
- 形式化验证
- 使用 Rust crate 替代不安全的库



塑造安全并且可信的人工智能和大数据分析框架

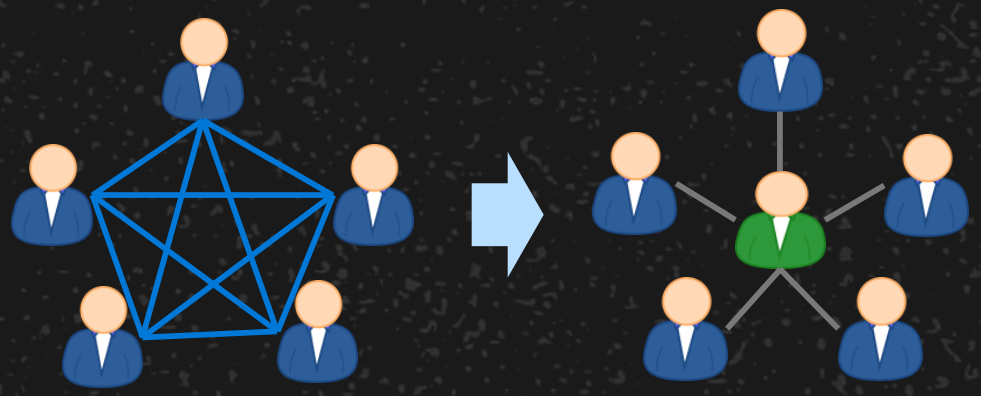
使用 Intel SGX 实现可信的人工智能和大数据分析

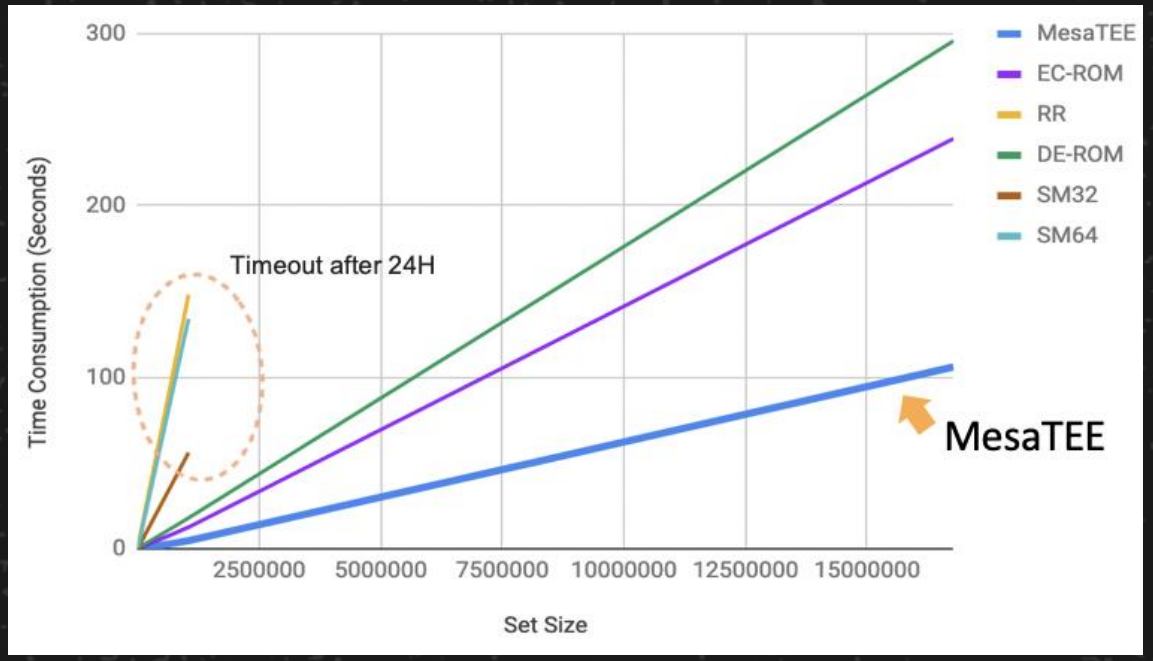


Solutions	Spark	MesaTEE Spark	GraphSC	ObliVM	Homomorphic Encryption
Data Encryption	x	√	x	x	√
Oblivious	x	√	√	√	x
Turnaround	1 sec	4-20 sec	2-6 days	>100 days	∞

塑造安全并且可信的人工智能和大数据分析框架

使用 Intel SGX 实现可信的人工智能和大数据分析

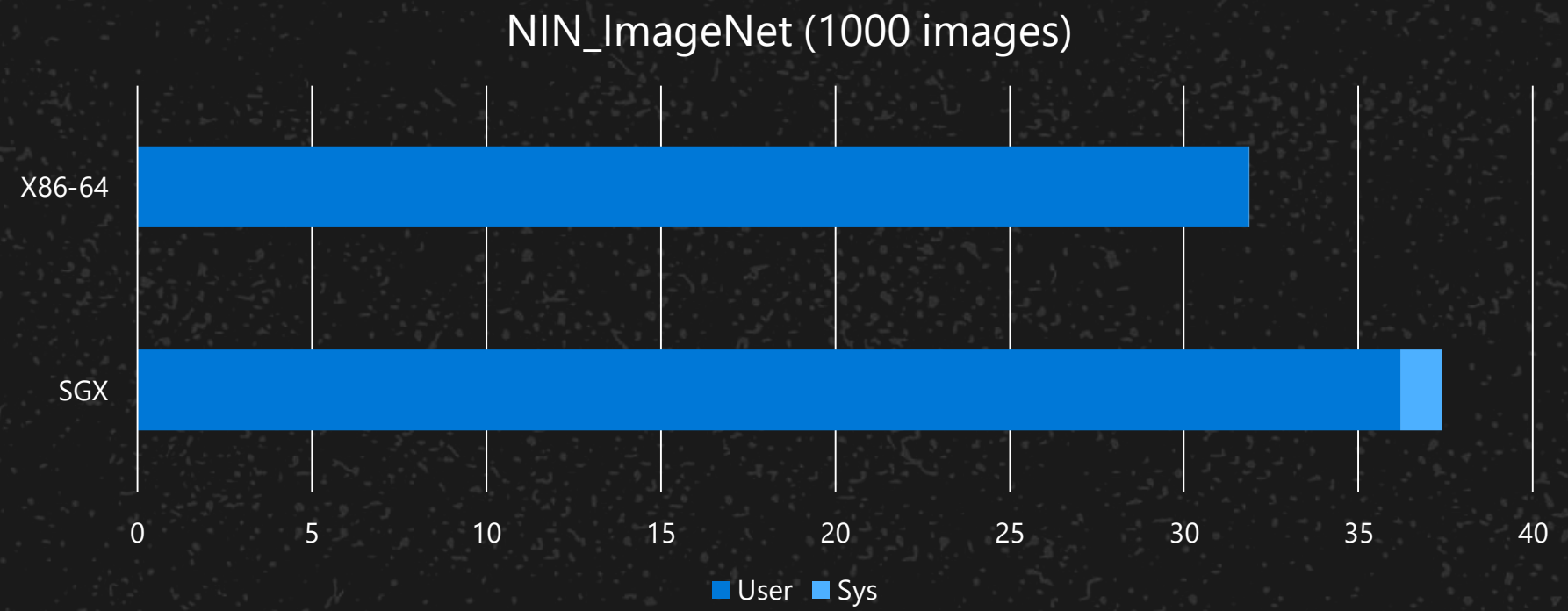




我们正与 百度 XuperData 在应用程序方面进行合作

塑造安全并且可信的人工智能和大数据分析框架

Anakin-SGX



MesaTEE SGX: 借助 Intel SGX 重新定义人工智能和大数据分析

问答

Yu Ding

百度 X-Lab 安全研究员