CCF-蚂蚁科研基金 软硬协同专项

2023年申报课题介绍

目录

课题题目

1. 基于 RISC-V 的超标量乱序 FHE 处理器设计

项目背景描述

全同态加密（FHE）可以直接计算加密数据，从而确保用户在使用高质量网络服务的同时保有私有数据的机密性。当前主要的 FHE 加速器（如 F1、CraterLake、BTS、ARK、BASALISC、RPU、Medha 等）多关注于通过提升数据级并行能力（如 SIMD）加速计算性能，对指令级并行能力、系统可编程能力缺乏深入探索。对于仍处于快速发展过程中的 FHE 领域，FHE 加速器具备可编程、易编程能力和加速计算性能同等重要，通过 DSA 架构支持 ISA 可以灵活地将新算法映射到以前设计的硬件，同时通过指令级并行设计可以进一步加速计算单元的并行能力。RISC-V 作为当下唯一的开源 ISA 可提供海量的生态资源，通过设计基于 RISC-V 的超标量乱序 FHE 处理器可以获得多种收益：1. FHE 加速器达到 CPU 级别的灵活、易用、高性能编程能力；2. 通过指令级并行设计进一步加速计算单元的并发能力；3. 借助 RISC-V 海量生态资源降低 FHE 处理器开发门槛。

预计的产出

1）仿真数据：完成 FHE 处理器架构设计、RTL 开发、处理器性能、功耗、面积指标评估
2）论文：发表 CCF-A 类或领域内顶级会议、期刊论文一篇
3）技术指标：FHE 处理器具有 RISC-V CPU 核级别的编程能力，在指令并行能力上可对标当前主流 RISC-V 超标量乱序核，在 FHE 加速效率上领先当前主要的 FHE 加速器

返回目录

2. 高性能 FHE 加速芯片计算架构设计

项目背景描述

全同态加密（FHE）支持将计算卸载到不受信任的服务器。尽管 FHE 具有较强的安全性，但由于其巨大的计算开销和访存带宽需求，比未加密计算高出约 10,000 倍，因此尚未被广泛采用。最近的 FHE 加速器研究在计算性能方面取得了长足的进步，比如 F1，Craterlake，BASALISC，ARK，BTS， TREBUCHET 等。这些 FHE 加速器，为了灵活高效的支持深度神经网络计算，比如 Resnet、LoLa 等，使用了大量的计算单元、内部缓存以及复杂的互联网络，导致芯片存在较大的功耗和面积。我们需要研究创新的计算架构，让 FHE 加速芯片有更高性能的同时，功耗和面积也具有显著优势。

预计的产出

1）仿真数据：完成芯片前端性能分析，完成芯片后端面积，功耗评估
2）论文：发表 CCF-A 类或领域内顶级会议、期刊论文一篇
3）技术指标：跟业内方案比较，芯片有更高的性能，且功耗和面积方面也有显著优势

返回目录

3. 高性能 NTT 设计与实现

项目背景描述

数论变换(NTT)是全同态加密(FHE)的实现过程中最重要的计算单元之一。尽管 NTT 将多项式乘法的复杂度降低至 O(nlogn)，其实现依然占据着极大地时间和面积开销。NTT 的设计具有较大的灵活度。一方面，设计者可以通过高效的内存管理合理调度多个高度优化的基二蝶形计算单元，完成整体 NTT 计算。这种方式需求的面积较小，但随着蝶形计算单元数量的增加，调度难度急剧增加。另一方面，基于 2D-NTT 算法，设计者可以用高基蝶形计算单元(如 128 点)简化调度逻辑。但是高基蝶形计算单元通常是基二蝶形计算单元的直接展开，其性能亟待优化。我们需要研究高性能的 NTT，提升单个蝶形计算单元的性能，简化多个蝶形计算单元的调度，使其在时间或面积上具有显著优势。从而进一步降低硬件加速的成本或者支持更多的操作。

预计的产出

1）仿真数据：证明调度的高效，实现蝶形计算单元优化，完成 RTL 设计，获得时间和面积评估
2）论文：发表 CCF-A 类或领域内顶级会议或期刊论文一篇
3）技术指标：跟业内方案比较，NTT 有更高的性能，在时间或面积上具有显著优势

返回目录

4. 面积优先的模乘设计

项目背景描述

全同态加密的硬件加速需要例化大量原语操作单元才能实现高并行计算。模乘作为最复杂的原语操作，其流水线实现的面积影响着整个设计的面积。常用的 Barrett 模乘或者 Montgomery 模乘通常可以由三个乘法和若干加减法实现。乘法也可以通过 Karatsuba 算法减少面积。在此基础上，我们需要进一步优化 Barrett 模乘或者 Montgomery 模乘、或者提出新的模乘算法，探索、提出、论证一些现有算法不曾使用的方法和策略，使新的模乘面积上具有显著优势，从而降低硬件加速的整体面积、例化更多的模乘来提高并行度。

预计的产出

1）仿真数据：提出面积优先的模乘算法，完成 RTL 设计，获得时间和面积评估
2）论文：发表 CCF-A 类或领域内顶级会议或期刊论文一篇
3）技术指标：跟业内方案比较，新的模乘在面积上具有显著优势

返回目录

5. AI 加速器和 FHE 加速器的融合设计

项目背景描述

AI 应用和同态加密应用各自都需要借助硬件加速来提升性能。自然地，如果一个 AI 应用同时需要基于同态加密来进行隐私保护，比如云端 AI 推理，那么我们就需要考虑将 AI 加速器和 FHE 加速器进行融合。如何将二者有机的融合，来达成整体上的性能、功耗、面积的最优设计目标，仍是一个开放问题。希望本课题能够对这一问题进行系统性地深入梳理，找到有机融合的设计思路，并进行原型验证。

预计的产出

1）原型系统：提出 AI 和 FHE 加速器融合设计思路，完成原型验证。
2）论文：发表 CCF-A 类或领域内顶级会议或期刊论文一篇

返回目录

6. GPU 上多项式运算的内存层次结构优化

项目背景描述

异构架构的系统可能包括 CPU、GPU、FPGA 和 ASIC。GPU 具有相比 CPU 更复杂和可编程的多级内存层次结构，综合考虑不同层次内存的容量、带宽和延迟并加以合理利用能有效提高程序性能。常见的多项式操作，如多项式系数模加（ModAdd）模乘（ModMul）和置换（Permutation）、模升（ModUp）和模降（ModDown）、数论变换（NTT）和数论逆变换（Inverse NTT）、RNS 分解（RNS-decomposition）具有不同的访存模式。优化跨 CPU，GPU 内存访问，并针对 GPU 内存层次结构去优化不同访存模式的多项式运算，对提升全同态加密（FHE）执行速度非常关键。

预计的产出

1）算法原理、实现及优化方法的源代码
2）性能指标：优化后多项式运算性能应明显高于业内已有方案
3）论文：发表 CCF-A 类或领域内顶会或期刊论文一篇

返回目录

7. 面向异构计算的 TVM 数据布局扩展和优化

项目背景描述

TVM 是一种流行的端到端 DL 程序的编译框架。TVM 能从现有 AI 框架中获取 DL 程序的高层级表示并为多种不同后端硬件平台产生底层级优化代码。在异构计算平台上，不同异构计算单元有不同的内存层级结构和参数，只有充分利用计算单元的内存层级结构和参数特性，才能充分发挥硬件效能。数据布局对 DL 算子的设计和实现非常关键，在自动布局无法充分发掘硬件性能的情况下，扩展 TVM 允许程序员指定数据布局方式成为一种非常有效的提升性能的手段。

预计的产出

1）TVM 数据布局扩展 API 设计和实现，后端实现基于 CPU 代码生成。
2）性能指标：通过合理指定数据布局扩展在不同硬件平台上实现比 TVM 默认方式更高的性能
3）论文：发表 CCF-A 类或领域内顶会或期刊论文一篇

返回目录


8. 基于 TVM 的隐私机器学习（PPML）编译器

项目背景描述

随着隐私问题越来越受到关注，具有数据隐私保护的机器学习模型推理成为了近期的研究热点。TVM 是一种流行的端到端 DL 程序的编译框架，并为多种不同后端硬件平台产生底层级优化代码。如果将全同态加密看作是一个虚拟的计算设备并为其设计虚拟指令集，利用 TVM 将高层级 DL 表示编译为相应的全同态加密计算程序，可以显著降低全同态加密程序开发难度。同时复用 TVM 中针对 DL 高层表示的各种优化可显著提高全同态加密推理程序的执行速度。

预计的产出

1）基于 TVM 的隐私机器学习（PPML）编译器设计和实现，后端基于 Nvidia GPU。
2）功能指标：能编译常见机器学习模型（LeNet 和 ResNet）推理为全同态加密程序
3）论文：发表 CCF-A 类或领域内顶会或期刊论文一篇

返回目录

9. 面向加速器的可信执行环境研究

项目背景描述

可信执行环境可以在不可信的计算平台上保护运行于其中的代码与数据的机密性和完整性。目前主流的硬件厂商，例如 Intel、AMD 等都推出了基于 CPU 的可信执行环境架构。然而，随着计算需求与应用场景的多样化，CPU 可信执行环境难以支持深度学习、大语言模型等智能计算任务。利用硬件加速器（例如 GPU 和 NPU） 对可信执行环境的计算任务进行加速成为一种重要的技术路线。然而，CPU 与加速器结合的异构计算模式对可信执行环境的安全性提出了挑战。如何扩展 CPU 可信执行环境的安全边界，有效地利用硬件加速器实现运算加速，成为了该研究方向的重要挑战。

预计的产出

1）系统实现：实现面向加速器的可信执行环境软硬件系统，完成系统的性能测试和安全性分析。
2）论文：发表 CCF-A 类或领域内顶级期刊、论文一篇。
3）技术指标：在不降低系统整体安全性的前提下，相比于 CPU 可信执行环境，实现 1～2 个数量级的性能提升。

返回目录

10. 面向机密计算的应用安全技术研究

项目背景描述

基于可信执行环境的机密计算，通过技术信任而不是运维信任手段，可方便地赋予应用机密性和完整性的保护。但是，并非把普通应用运行在 TEE 里，就自动使其成为了机密计算应用——基于安全性的考虑，仍需对应用进行精心的安全加固改造。比如，一个运行在 TEE 中的数据库应用，如果不对 SQL 语句进行安全限制，攻击者仍可能通过特殊设计或组合的查询窃取机密信息。由于存在五花八门的应用类型，我们希望设计一个系统化的统一的应用安全增强机制，从而让普通应用开发者能够极致简便地集成该机制，而无需对每个应用以一事一议的方式进行安全改造。为降低问题复杂度，一个分而治之思路是，将应用划分为若干类别，然后探索并设计类别内统一的应用安全增强机制。

预计的产出

1）系统原型：设计并实现统一的应用安全增强机制，或面向典型应用类别（如数据库）设计并实现安全增强机制。
2）论文：发表 CCF-A 类或领域内顶级期刊、论文一篇。
3）技术指标：安全增强机制对应用的性能影响控制在 10%以内。

返回目录

11. 安全增强的可信执行环境操作系统技术

项目背景描述

操作系统（OS）是计算机系统的核心软件，在可信执行环境（TEE）中亦是如此。各大 CPU 厂商已经为各自 VM TEE（如 AMD SEV、Intel TDX 等）技术适配了 Linux，并对 Linux 做了初步的安全加固（比如禁用某些与 TEE 不兼容的 OS 特性或者禁用某些影响 TEE 安全的驱动）。然而，面向 VM TEE 的 OS，包括 Linux，仍然面临诸多挑战，比如 OS 内核自身的内存安全问题、驱动程序与不可信设备之间的 I/O 接口的攻击面（如 Iago attack），TEE 系统固件（如 BIOS、UEFI、TDVF、TDSHIM 等）的攻击面、加密文件系统或存储的安全与性能的平衡、侧信道攻击等等。因此，如何增强 TEE OS 的安全性就成了一个重要的研究问题。

预计的产出

1）系统实现：实现 TEE OS 的安全加固方案的原型系统。
2）论文：发表 CCF-A 类或领域内顶级期刊、论文一篇。
3）技术指标：安全加固措施对 TEE OS 的性能影响控制在 10%以内。

返回目录