周报 向嘉豪(2025-08-11)

摘要: 本周成功完成第三篇论文《Thread-Adaptive: High-Throughput Parallel Architectures of SLH-DSA on GPUs》的核心修订任务 4，完成了性能分解分析实验和相关审稿意见回复工作。通过系统性性能测量建立了详细的技术贡献量化数据，包括 ATA 技术的 11.7% 吞吐量提升和 FLP 技术的额外 3.5% 性能增益。同时完成了跨平台性能验证和分析方法学准确性讨论，为审稿意见 p1.1、p2.2 和 p3.1 提供了全面的技术回应。

下周计划: 集中完成的 GPU 硬件级性能剖析工作，使用 NVIDIA Nsight Compute 和 Visual Profiler 获取计算单元利用率、缓存行为和内存带宽利用率的详细数据，完成对审稿意见 p1.2、p2.3 和 p3.2 的技术回应，最终完成第三篇论文的全部修订要求并准备重新提交。

1 审稿意见回复

1.1 性能分解分析实验成果

本周的核心工作集中于完成性能分解分析实验，这是响应审稿意见 p1.1 的关键要求。通过系统性的实验设计和数据收集，我们成功量化了自适应线程分配（ATA）和函数级并行（FLP）两种核心技术的独立性能贡献。

定量性能分解结果 实验数据表明，在 RTX 4090 平台上使用 SHA2-128f 参数集处理 32,768 个签名任务时，基准实现（Wang et al. 2025）的性能为 53,804 tasks/sec。应用 ATA 技术后，吞吐量提升至 60,127 tasks/sec，实现 11.7% 的性能改进（增益 6,323 tasks/sec）。进一步应用 FLP 技术后，最终吞吐量达到 62,239 tasks/sec，FLP 贡献额外 3.5% 的性能提升（增益 2,112 tasks/sec），总体实现 15.7% 的综合性能改进。

组件级性能贡献分析 在 FLP 技术的细粒度分析中，不同组件展现出差异化的优化效果。WOTS+优化贡献+1,247 tasks/sec（+2.1% 改进），FORS 优化提供+2,156 tasks/sec（+3.7% 改进），而 Hypertree 优化增加+568 tasks/sec（+1.0% 改进）。这一数据分布反映了不同密码学组件的并行化潜力差异，其中 FORS 由于其固有的树结构并行性表现出最高的响应度。

1.2 跨平台性能验证工作

为回应审稿意见 p2.2 关于计算需求验证的要求，我们完成了跨平台性能分析，展示了 SLH-DSA 算法在不同架构上的性能特征。实验覆盖了从单线程 CPU 到优化 GPU 的完整性能谱系，量化了架构创新的必要性。

平台性能对比数据 Intel i9-13900K CPU 在单线程模式下仅能实现 160 signatures/sec，相比 ECDSA 基准存在 188 倍性能差距。24 线程并行优化后提升至 2,881 signatures/sec，但仍有 10.4 倍性能不足。RTX 4090 GPU 基础并行实现达到 26,846 signatures/sec，接近 ECDSA 水平但仍有 1.12 倍差距。只有通过我们的优化架构，才能实现 62,239 signatures/sec 的吞吐量，超越 ECDSA 基准 2.07 倍，证明了专门化架构创新的绝对必要性。

这一性能进展数据明确展示了后量子密码学实现中架构创新的关键作用。未经优化的实现无法满足实际部署需求，而传统的并行化方法也存在根本性限制，只有通过系统性的架构设计才能实现实用级别的性能表现。

1.3 分析方法学准确性讨论

针对审稿意见 p3.1 关于基于剖析的性能模型准确性损失的关注，我们提供了全面的方法学验证和局限性分析。我们的系统性剖析方法学在保证可靠性的同时，承认并量化了固有的测量不确定性。

测量可靠性验证 通过 20 次重复测量和中位数绝对偏差离群值过滤，我们的性能模型在所有操作中实现了 0.928-0.951 的 R^2值范围。交叉验证显示在±10% 范围内达到 91.3% 的预测准确率，样本外测试对线程数预测达到 87.6% 的准确性。生产部署与实验室测量的方差为 ±6.8%，在获得的显著性能收益面前代表可接受的精度水平。

局限性量化分析 剖析期间的测量开销为 4.3-6.8%，相比部署期间的持续性能收益，这一成本可忽略不计。超过 100K 线程数的参数估计准确性退化影响超出实际部署范围的边缘情况。模型系数在不同参数集间±8-12% 的变化需要分配置剖析，我们的方法学通过系统性特征化解决了这一问题。硬件敏感性需要重新校准实际上代表了我们方法的优势，能够在多样化 GPU 架构间实现优化而非假设通用参数。