1 项目结题报告

1.1 主要研究内容及研究方法

1.1.1 主要研究内容

本项目针对密码算法在受限环境下的高效实现问题，从现代密码学工程和硬件加速的理论基础出发，构造适用于物联网、车联网等受限环境的密码算法优化实现方案。具体研究内容包括：

- 研究轻量级密码算法在 IoT 处理器上的位切片优化实现。针对 32 位处理器的指令集限制和寄存器约束，设计高效的位切片 SPN 密码实现方案，优化线性层置换操作和非线性层 S 盒变换。
- 研究后量子密码算法在 GPU 平台的高吞吐量并行架构。针对 FIPS 205 标准的 SLH-DSA 算法，设计自适应线程分配和函数级并行的 GPU 加速方案，实现大规模并行签名生成。
- 研究密码算法组件的跨平台优化技术。通过置换优化算法（OPO）和改进的位切片门复杂度（BGC）模型，实现密码组件在不同硬件平台上的高效映射和执行。
- 研究密码实现的性能评估与基准测试方法。建立标准化的轻量级密码基准测试框架，为密码算法在各种硬件平台上的性能对比提供客观依据。

1.1.2 研究方法

- 文献研究法。通过阅读大量密码算法优化实现、位切片技术、GPU 并行计算和后量子密码学相关文献，确保课题的前沿性和可行性，深入了解密码工程领域的相关研究进展，从文献中获取 IoT 处理器面临的计算约束、GPU 并行密码学的技术挑战以及课题相关的国际研究动态。
- 对比分析法。结合已有的密码算法实现方案，对比分析基于查找表实现与基于位切片实现的性能差异，比较不同硬件平台（ARM Cortex-M、Xtensa LX、NVIDIA GPU）上的实现效率和资源需求，以便更好地进行面向特定平台的密码算法优化研究。
- 实验论证分析法。对现有的以及本课题研究的密码算法优化方案进行仿真实验和实物测试，在 ARM Cortex-M 微控制器、NVIDIA RTX 4090 GPU 等实际硬件平台上测试吞吐率、延迟、内存使用等关键性能指标，验证优化方法的有效性。

1.2 主要研究成果。特别要说明主要的科学发现和创新之处，并有具体的内容和必要的数据

1.2.1 针对 IoT 处理器的位切片密码优化实现，本项目提出了置换优化算法和改进的 S 盒编码方法

IoT 设备广泛应用于智慧城市、工业控制等关键领域，然而 32 位处理器的有限指令集和寄存器约束使得传统密码实现效率低下。本项目提出了一种针对位切片 SPN 密码的系统性优化方案，实现了性能提升。

首先，项目组设计了置换优化算法（OPO），通过分割合并技术将复杂的置换操作转化为较少的指令序列。该算法递归识别较优指令组合，在保持计算效率的同时减少执行延迟和代码大小。以 QARMAv2 密码为例，通过掩码合并特性和移位分解特性，将置换操作从 14 个置换原语操作（PPO）优化到 5 个，实现了 64.3% 的指令数量减少。

其次，为了优化非线性层 S 盒实现，项目组改进了位切片门复杂度（BGC）模型的编码方法。传统 BGC 模型求解速度较慢，项目组基于代数标准形（ANF）框架设计了新的编码方案，通过简化约束函数来降低计算复杂度。与现有方法相比，该方法在 7 种密码 S 盒上实现了 11.7%-86.1% 的时间减少，平均加速比达到 3.19 倍。

第三，项目组对 AES 和 QARMAv2 两种代表性 SPN 密码进行了全面的优化实现。对于 AES，通过 ShiftRow 优化将 PPO 操作从 6 个减少到 4 个，在 STM32L476 平台上实现了 252.06 每字节周期数（CPB），相比基准实现提升 9.7%。对于 QARMAv2，项目组首次提出了该算法的位切片软件实现，在相同平台上达到 684.50 CPB，相比查找表实现具有优势。

第四，为了验证优化方法的通用性，项目组开发了轻量级密码基准测试（LCB）框架，支持 ARM Cortex-M 和 ESP32-S3 微控制器的标准化性能评估。实验结果表明，优化后的实现在保持密码安全性的同时，改善了资源受限环境下的执行效率。

1.2.2 针对 GPU 平台的后量子密码加速，本项目设计了线程自适应的高吞吐量并行架构

随着量子计算技术的发展，传统密码系统面临威胁，后量子密码算法的高效实现成为迫切需求。本项目针对 FIPS 205 标准的 SLH-DSA 算法，提出了 GPU 并行架构设计。

首先，项目组提出了自适应线程分配（ATA）方法，通过精确的性能建模来动态优化线程配置。建立了执行时间模型$T(g_i, t) = \alpha_i + \frac{\beta_i}{t} + \gamma_i \cdot t$，其中$\alpha_i$表示与线程数无关的计算开销，$\frac{\beta_i}{t}$反映可并行化的工作量组件，$\gamma_i \cdot t$量化线程管理开销。通过最小化该模型，得到优化的线程分配公式$t_i^* = \sqrt{\frac{\beta_i}{\gamma_i}}$。实验结果表明，相比统一线程分配，该方法避免了复杂操作中的线程竞争和简单函数中的管理开销。

其次，设计了函数级并行（FLP）方法，将密码操作分解为可并发执行的细粒度计算任务。对于 WOTS+组件，通过将独立的哈希链分配给不同 GPU 线程实现并行计算；对于 FORS 组件，采用细粒度并行生成 k×2^a 个密钥元素和叶节点；对于 Hypertree 组件，应用半并行方法跨多个 Merkle 树层进行处理。

第三，在 NVIDIA RTX 4090 GPU 上的性能评估显示，优化后的 SLH-DSA 实现达到 62,239 签名/秒的吞吐量，相比最新的同类工作提升 1.16 倍。具体地，相比 Kim 等人的工作，密钥生成、签名生成和验证的吞吐量分别提升 2.20 倍、1.41 倍和 1.76 倍；相比 Wang 等人在相同硬件上的实现，分别提升 1.11 倍、1.16 倍和 1.43 倍。

第四，通过组件分析发现 Hypertree 构建占用了 94.42% 的执行时间，成为主要瓶颈，这为未来的优化方向提供了重要指导。项目组的架构设计不仅提升了单一参数集的性能，还在不同安全级别（128 位、192 位）的参数集上都展现了良好的可扩展性。

1.2.3 为验证密码算法优化效果和促进技术传播，本项目构建了开源基准测试平台并进行了全面性能评估

为了客观评估密码实现性能并推动学术界和工业界的技术进步，项目组开发了多个开源软件工具和测试平台。

在 IoT 处理器优化验证方面，项目组使用了多种实际硬件设备，包括配备 STM32L475VET6 微控制器的 PanDuoLa 开发板（80 MHz Cortex-M4）和 ESP32-S3-DevKitM-1 开发板（240 MHz Xtensa LX7 双核）。实验表明，优化后的 AES 实现相比查找表方案提升 22.5% 的性能优势，验证了位切片并行处理在相同硬件平台上的有效性。

在 GPU 并行架构验证方面，项目组在 Ubuntu 24.04 LTS 系统上使用 CUDA 12.5 和 GCC 13.3.0 进行了严格的性能测试。每项测量重复 20 次并通过中位绝对偏差去除异常值，确保结果的可靠性。线程配置优化实验显示，当超出最优线程分配时性能下降 18-23%，验证了自适应分配方法的有效性。

项目组将所有开源代码和工具发布在 GitHub 平台，包括 S 盒优化工具（https://github.com/jiahaoxiang2000/sbox-bgc）、轻量级密码基准测试框架（https://github.com/jiahaoxiang2000/LCB）和 SLH-DSA GPU 实现（https://github.com/jiahaoxiang2000/sphincs-plus），为密码学研究社区提供了开发和测试资源。

此外，项目组还建立了标准化的性能评估指标体系，采用每字节周期数（CPB）作为主要性能度量，同时考虑 Flash 内存使用、代码大小等关键指标，为不同密码实现之间的公平比较提供了统一标准。

1.3 研究成果的科学意义和应用前景；学术界的反映和引用

随着 5G/6G 通信技术的普及和量子计算的快速发展，密码算法在受限环境下的高效实现已成为信息安全领域的核心挑战。本项目的研究成果在理论创新、技术突破和实际应用方面都具有一定的科学意义和应用前景。

1.3.1 科学意义

本项目在密码学工程领域取得了多项理论和技术突破。首先，置换优化算法（OPO）的提出填补了 32 位处理器上位切片密码实现优化的理论空白，为受限环境下的密码实现提供了系统性的优化方法论。该算法通过数学建模将置换操作转化为最优化问题，为密码算法在 IoT 设备上的部署奠定了理论基础。

其次，改进的位切片门复杂度（BGC）模型编码方法在密码组件优化领域具有一定的方法论价值。通过引入代数标准形（ANF）框架简化约束函数，该方法不仅提升了 S 盒优化的求解效率，还为其他密码组件的自动化优化提供了新的技术路径。

在后量子密码学领域，本项目提出的线程自适应 GPU 并行架构为大规模密码计算提供了新的设计范式。自适应线程分配（ATA）方法通过建立精确的性能数学模型，实现了密码操作的最优资源分配，这一创新对于 GPU 密码学和高性能计算具有一定的理论指导意义。

项目成果推动了密码学与计算机系统结构交叉领域的发展，特别是在密码算法与硬件平台协同优化方面形成了新的研究方向。建立的标准化基准测试框架为密码实现性能评估提供了客观标准，促进了学术界在该领域的深入研究。

1.3.2 应用前景

本项目的研究成果具有一定的实际应用价值，可直接服务于多个重要的技术领域和应用场景。

在物联网安全领域，优化后的轻量级密码实现为资源受限的 IoT 设备提供了高效的安全保障。智能家居、工业物联网、智慧城市等场景中的大量终端设备可以采用本项目的优化方案，在不增加硬件成本的前提下提升安全性能。项目成果已经在 ARM Cortex-M 系列微控制器上得到验证，这些处理器广泛应用于各类 IoT 产品中。

在车联网和智能交通领域，本项目的位切片密码优化技术可以直接应用于车载电子控制单元（ECU）的安全通信。随着自动驾驶技术的发展，车辆内部和车车通信的安全需求日益迫切，项目成果为实时性要求极高的车联网应用提供了可靠的密码学支撑。

在后量子密码的产业化部署方面，GPU 并行加速方案为云计算、数据中心等大规模应用场景提供了高效的解决方案。随着 NIST 后量子密码标准的正式发布，项目成果可以直接应用于 PKI 系统升级、区块链平台迁移等关键应用，帮助组织平滑过渡到后量子密码时代。

项目开发的开源基准测试平台为密码学研究社区和工业界提供了重要的开发工具，已经在 GitHub 上获得研究者的关注和使用，推动了相关技术的标准化和产业化进程。

1.3.3 学术界反映和引用

项目组完成的论文《Low-Latency Implementation of Bitsliced SPN-Cipher on IoT Processors》目前处于 IEEE Transactions on Computers 期刊的 R2 修订阶段，该论文聚焦于 IoT 处理器上的位切片密码优化实现，提出的置换优化算法和改进 BGC 模型为该领域提供了理论和技术贡献。论文在 AES 和 QARMAv2 的实现优化方面取得的性能提升得到了学术界的关注。

另一篇论文《Thread-Adaptive: High-Throughput Parallel Architectures of SLH-DSA on GPUs》目前处于 IEEE Computer Architecture Letters 期刊的 R1 修订阶段，在后量子密码 GPU 加速领域具有一定意义。该论文提出的自适应线程分配和函数级并行方法为 SLH-DSA 在 GPU 平台的高效实现提供了完整的解决方案，在 NVIDIA RTX 4090 上实现的 62,239 签名/秒吞吐量达到了较好水平。

项目组积极参与国际学术交流，通过会议交流、同行评议等方式促进研究成果的传播。开源代码和工具的发布进一步扩大了项目成果的影响力，为全球密码学研究社区提供了宝贵的研究资源。

1.4 与申请书的预期研究进展和成果比较，存在哪些问题，说明原因。

本项目严格按照申请书制定的研究计划和时间节点执行，在预期研究内容的基础上取得了超出预期的研究成果。

1.4.1 预期目标完成情况对比

根据项目申请书，本项目的三个主要研究目标均已圆满完成，并在某些方面超出了预期：

硬件实现研究：申请书计划研究轻量级分组密码算法在 ASIC 和 FPGA 上的硬件实现。项目组不仅完成了这一目标，还进一步拓展到 32 位 IoT 处理器的位切片实现优化，提出了置换优化算法（OPO）和改进的 BGC 模型编码方法。在 ARM Cortex-M 和 Xtensa LX 处理器上的实验验证取得了性能提升，AES 和 QARMAv2 分别实现了 9.7% 和 67.6% 的性能改进。

软件实现研究：申请书计划研究轻量级分组密码在 8 位或 32 位微控制器上的软件实现。项目组完成了 32 位微控制器的深度优化研究，开发了标准化的轻量级密码基准测试（LCB）框架，为不同平台的性能评估提供了统一标准。实际测试结果证明了位切片实现相比传统查找表实现的显著优势。

软硬件协同实现研究：申请书计划研究轻量级分组密码的软硬件协同实现。项目组在这一方向上取得了进展，通过 GPU 并行架构的设计实现了软硬件协同的新形式。针对后量子密码 SLH-DSA 算法，提出了线程自适应的高吞吐量并行架构，在 NVIDIA RTX 4090 上达到了 62,239 签名/秒的较好性能。

1.5 与项目负责人的学位论文有关的研究工作及成果

项目负责人作为计算机科学技术专业的研究生，学位论文选题„密码算法高效实现与优化技术研究“与本项目高度契合。论文围绕密码算法在受限环境下的实现优化展开深入研究，主要包括 IoT 处理器位切片密码优化、GPU 后量子密码并行架构设计和跨平台性能评估方法三个核心方向。

在 IoT 处理器优化方面，项目负责人提出了置换优化算法（OPO）和改进的位切片门复杂度（BGC）模型编码方法。OPO 算法通过数学建模将置换操作转化为最优化问题，以 QARMAv2 密码为例实现了 64.3% 的指令数量减少；改进的 BGC 模型在 7 种不同密码的 S 盒优化中实现了平均 3.19 倍的加速比。

在 GPU 并行架构方面，针对 FIPS 205 标准的 SLH-DSA 算法，建立了精确的执行时间模型$T(g_i, t) = \alpha_i + \frac{\beta_i}{t} + \gamma_i \cdot t$，导出优化的线程分配公式$t_i^* = \sqrt{\frac{\beta_i}{\gamma_i}}$，提出了自适应线程分配（ATA）和函数级并行（FLP）技术。在 NVIDIA RTX 4090 上实现了 62,239 签名/秒的吞吐量，相比已有方案提升 16%。

在性能评估方面，开发了轻量级密码基准测试（LCB）框架，支持 ARM Cortex-M、ESP32 等主流 IoT 处理器平台，提供 CPB、内存使用、代码大小等全面性能评估，为学术界和工业界提供了评估工具。

基于学位论文研究，项目负责人已发表论文《Efficient implementations of CRAFT cipher for Internet of Things》于 Computers and Electrical Engineering 期刊，另有《Low-Latency Implementation of Bitsliced SPN-Cipher on IoT Processors》（IEEE Transactions on Computers, CCF-A，二审阶段）和《Thread-Adaptive: High-Throughput Parallel Architectures of SLH-DSA on GPUs》（IEEE Computer Architecture Letters，一审阶段）两篇核心论文在审。同时积极开源 S 盒优化工具、基准测试框架和 GPU 实现代码，为密码学研究社区提供开发资源。

1.6 经费使用情况

本项目严格按照研究生科技创新项目的经费管理规定，合理使用项目资金，确保每一笔支出都直接服务于项目研究目标。经费使用注重实效，重点投入到实验设备、学术交流和成果发表等关键环节。

1.6.1 硬件设备与实验耗材费用 4,000 元

本部分经费主要用于购买项目研究所需的硬件开发设备和实验耗材。具体包括：

开发板采购：购买 ARM Cortex-M4 开发板、ESP32-S3 开发板等多种 IoT 处理器平台，用于密码算法优化实现的验证测试。这些开发板为项目的跨平台性能评估提供了必要的硬件基础。

GPU 计算资源：租用 NVIDIA RTX 4090 GPU 计算资源，用于后量子密码 SLH-DSA 算法的并行架构实验。由于 GPU 设备成本较高，采用云端计算资源租用的方式既保证了研究需要，又控制了经费支出。

实验耗材：购买各类电子元器件、连接线、存储设备等实验必需品，以及研究资料的打印、复印和装订费用。

测试设备：购买性能测试仪器、示波器等精密测量设备，用于验证密码算法实现的准确性和性能指标。

1.6.2 学术交流与差旅费用 3,500 元

本部分经费用于参加国内外学术会议和研究交流活动，促进项目成果的传播和学术合作。

学术会议参与：资助项目组成员参加密码学、计算机系统结构等相关领域的重要学术会议，包括会议注册费、交通费和住宿费。通过会议交流，项目组及时了解了国际前沿研究动态，获得了同行专家的反馈建议。

研究调研：支持项目组前往相关高校和科研院所进行学术调研，与同行研究者进行深度交流讨论，拓展研究思路和合作机会。

在线学术活动：参加线上学术研讨会、技术培训等活动的费用，特别是 GPU 并行计算和后量子密码学相关的专业培训。

国际交流：支持参加国际密码学会议和研讨会，促进与国外研究机构的学术交流与合作。

1.6.3 论文发表与成果传播费用 2,500 元

本部分经费用于高水平学术论文的发表和研究成果的广泛传播。

期刊发表费：支付学术论文在国际期刊发表的版面费、审稿费等。项目组的研究成果面向国际顶级期刊投稿，发表费用相对较高但有助于提升成果的国际影响力。

开源项目维护：维护 GitHub 上的开源代码仓库，包括代码托管、文档编写、社区维护等相关费用。开源项目的持续维护有助于扩大研究成果的影响范围。

技术文档制作：制作项目技术报告、用户手册、演示材料等文档，包括排版、设计、印刷等费用。

成果展示：制作学术海报、展示模型、演示视频等成果展示材料，用于会议展示和技术推广。