

论文阅读总结报告

概述与理解

近年来，操作系统设计正经历深刻变革，传统 monolithic 架构面临资源利用率低、弹性差和故障隔离不足等挑战。三篇论文——LegoOS（OSDI 2018）、DBOS（VLDB 2022）和 FlexOS（ASPLOS 2023）——分别从硬件资源解聚、数据库导向和灵活隔离角度提出了创新方案，共同指向操作系统未来向分布式、专用化和软硬件协同方向演进。通过深入分析这些工作，我们可以窥见操作系统如何更好地适应现代数据中心和异构计算需求。DBOS 论文提出了一种以数据库管理系统为核心的操作系统堆栈，将传统 OS 状态（如调度、文件系统）建模为关系表，利用分布式 DBMS 提供事务性、高可用性和可查询性。DBOS 的核心思想是“一切即表”，通过 SQL 实现 OS 服务的简洁表达和高效管理。例如，其四层堆栈中，底层是微内核服务，上层是 DBMS 和 OS 功能，用户空间支持无服务器计算模型。

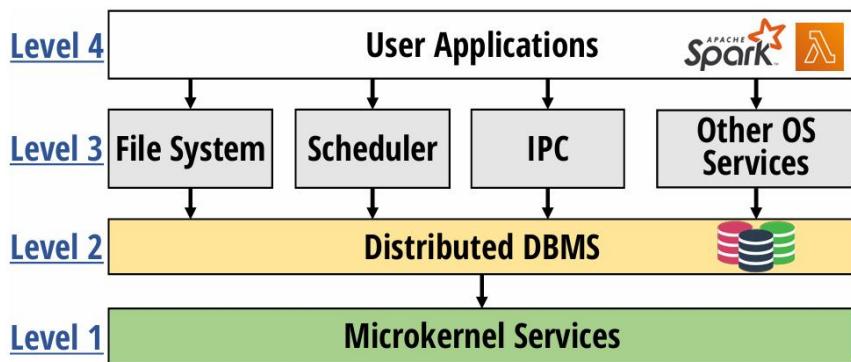


Figure 1: Proposed DBOS stack. Level 1 is the bottom layer.

这种设计显著降低了代码复杂性，并支持跨集群的监控和分析，但性能挑战在于网络延迟和 DBMS 开销。DBOS 的实验显示，在调度、IPC 和文件系统方面，性能可与现有系统竞争，尤其在分析任务中优势明显。FlexOS 则专注于操作系统的隔离灵活性，允许用户在编译或部署时定制安全策略，而非在设计时锁定。它基于模块化 LibOS（如 Unikraft），将 OS 组件分解为细粒度模块，每个模块可独立隔离（如使用 Intel MPK 或 VM/EPT），并支持软件硬化机制（如 CFI）。

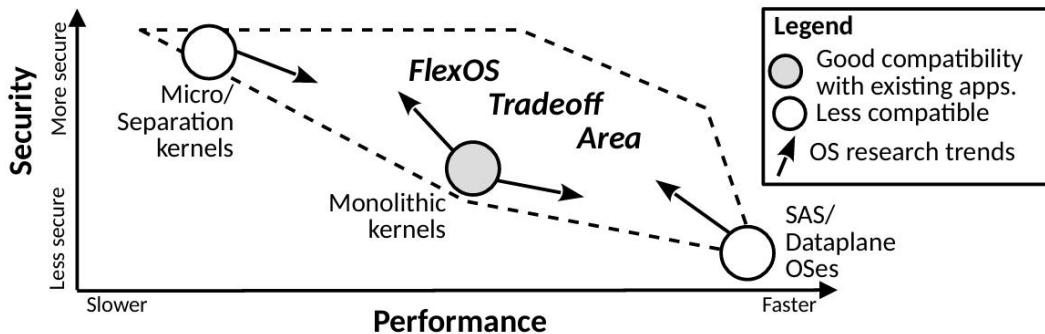
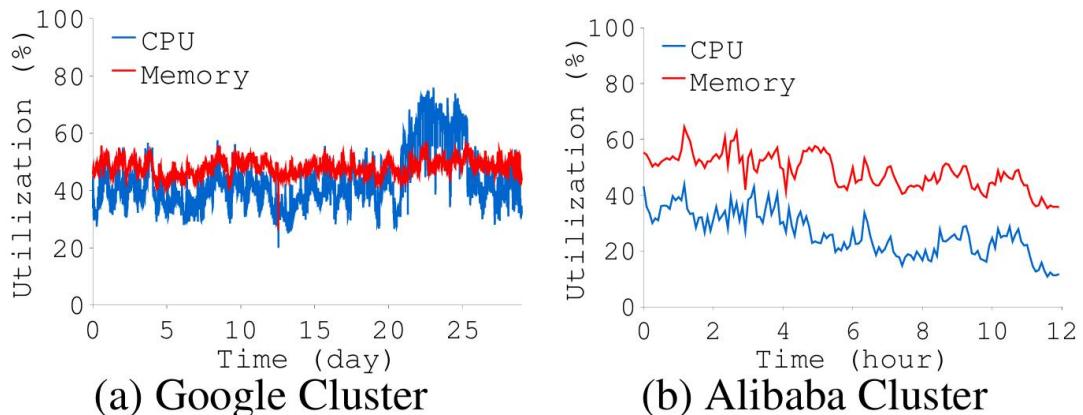


Figure 1: Design space of OS kernels.

FlexOS 通过部分安全排序技术帮助用户导航庞大的设计空间，实现安全与性能的权衡。例如，在 Redis 和 Nginx 测试中，FlexOS 展示了从无隔离到多隔离配置的广泛性能区间，突出了按需 specialization 的潜力。这种灵活性使 OS 能快速响应漏洞或硬件变化，但代价是增加了配置复杂性。LegoOS 针对硬件资源解聚架构，提出了 splitkernel 模型，将 OS 功能分解到独立的处理器、内存和存储组件上，每个组件由本地监控器管理，通过消息传递协作。



LegoOS 通过虚拟缓存 (ExCache) 和分布式虚拟内存管理，在非一致性硬件上实现高性能，同时提升资源利用率和故障隔离。评估显示，LegoOS 在 TensorFlow 等应用中性能接近 monolithic Linux，但资源打包更高效，故障率更低。这体现了硬件解聚下 OS 需重新设计以处理网络延迟和分布式资源管理。这些工作共同揭示了操作系统未来发展方向：一是从 monolithic 向分布式、模块化演进，以应对规模化和异构硬件；二是强化软硬件协同，通过硬件抽象和专用加速优化性能；三是提升自适应能力，支持动态配置和故障恢复。结合我的研究方向——软硬件协同，未来操作系统可能更深度集成硬件特性，例如利用定制加速器（如 DPU、FPGA）处理 OS 功能，或通过内存语义网络减少软件开销。软硬件协同将

推动 OS 成为智能资源中介，实现能效和性能的平衡，同时需解决验证、安全性和编程模型挑战。总之，操作系统正走向更灵活、专用和协同的范式，为云原生和边缘计算奠定基础。

展望

在软硬件协同研究领域，操作系统的未来发展方向将深刻依赖于硬件特性的暴露与软件层的智能适配，推动系统从静态资源管理向动态、感知驱动的范式转变。当前，如 LegoOS 所展示的资源解聚架构，已凸显出硬件组件（如专用处理单元、内存类存储）的独立管理需求，未来操作系统需进一步抽象硬件异构性，提供统一编程接口，同时允许应用直接调度加速器资源，例如通过编译器与运行时协同，将计算任务映射到 FPGA 或 AI 芯片，实现低延迟与高能效。软硬件协同也将强化系统的自适应性，利用机器学习模型预测负载波动，动态调整隔离策略（如 FlexOS 的灵活配置）或资源分配（如 DBOS 的查询优化），从而在保证安全的前提下最大化性能。此外，随着新兴硬件（如 CXL 互连、存算一体设备）的普及，操作系统将演变为“协同中介”，负责协调跨设备的数据流和一致性，减少软件开销，最终向构建高能效、弹性和可验证的云边端一体化系统演进。