Nishant Sharma

CS-GY 6233 - Introduction to Operating Systems

Week 11: Related Research Section

November 9, 2025

Related Research

Prout et al. (2018) investigated the performance impact of Spectre and Meltdown mitigations on high-performance computing workloads, addressing the critical need to quantify real-world overhead across diverse application types. Their work represents one of the first comprehensive empirical studies measuring mitigation-induced performance degradation in production-like environments. The authors conducted extensive benchmarking on multiple HPC systems, testing both synthetic benchmarks and realistic scientific computing workloads including linear algebra operations, graph processing, and I/O-intensive data analytics. Their experimental results demonstrated overhead ranging from 5% to 30% depending on workload characteristics, with I/O-intensive applications and database operations suffering disproportionately higher impact compared to CPU-bound mathematical computations. While Prout et al. provide crucial empirical evidence that mitigation overhead varies significantly by workload type, their work stops at measurement without providing system operators practical tools for workload classification or performance prediction. Our proposed eBPF-based profiling approach directly addresses this gap by transforming their observational findings into actionable operational guidance, enabling administrators to proactively identify which of their specific applications will experience high mitigation sensitivity based on system call frequency patterns.

Building on these performance measurements, Canella et al. (2019) presented a systematic evaluation of transient execution attacks and their defenses, uncovering critical gaps in mitigation completeness. Their work addressed the proliferation of ad-hoc Spectre and Meltdown variants by developing a comprehensive classification framework for transient execution attacks based on root causes and exploitation techniques. The authors implemented and tested attack variants across

three major CPU vendors (Intel, AMD, ARM) and evaluated the effectiveness of deployed software and hardware mitigations including KPTI, retpoline, and microcode updates. Through rigorous testing, they discovered that most defenses, including those already deployed in production systems, cannot fully mitigate all attack variants, with several newly identified Meltdown and Spectre exploitation strategies remaining unaddressed by existing patches. While Canella et al. systematically demonstrate that deployed mitigations remain incomplete and performance-costly, they do not address the operational challenge facing system administrators: how to balance security and performance when perfect mitigation is unavailable. Our workload classification tool provides operators with the data needed to make informed security-performance trade-offs, enabling granular deployment strategies where high-sensitivity workloads receive dedicated isolation while lower-sensitivity applications run with full mitigations at minimal cost.

Addressing these mitigation limitations, Behrens et al. (2020) proposed the Unmapped Speculation Contract (USC), a novel approach to efficiently mitigating transient execution attacks with significantly reduced performance overhead. Their work tackled the fundamental problem that existing software-only mitigations like KPTI impose excessive performance penalties by isolating kernel memory on every context switch, even when speculative execution poses no actual threat. The authors developed USC by establishing a contract between software and hardware that permits safe speculative access to unmapped memory regions, validated through extensive modifications to the Linux kernel and RISC-V processor implementations. Their evaluation demonstrated that USC reduces mitigation overhead from approximately 30% to just 8% for system-call-intensive workloads, while maintaining equivalent security guarantees against Meltdown-type attacks. However, Behrens et al. acknowledge that USC's benefits are most pronounced for specific workload classes, yet they provide no methodology for system operators to identify which of their applications would benefit most from USC deployment versus traditional KPTI. Our eBPF-based profiler directly complements their work by providing the missing classification layer that identifies USC candidates through real-time system call frequency analysis. This research gap—the absence of practical operator-facing tools bridging academic performance measurements and production de-

ployment decisions—persists because prior work focused on either attack discovery or mitigation design, while the operational challenge of deploying mitigations with minimal performance impact remained understudied, representing an intersection between systems security and operations research that neither community fully addressed.

References

[1] A. Prout, W. Arcand, D. Bestor, B. Bergeron, C. Byun, V. Gadepally, M. Houle, M. Hubbell, M. Jones, A. Klein, P. Michaleas, L. Milechin, J. Mullen, A. Rosa, S. Samsi, C. Yee, A. Reuther, and J. Kepner, "Measuring the Impact of Spectre and Meltdown," in *Proc. IEEE High Performance Extreme Computing Conference (HPEC)*, Waltham, MA, USA, 2018, pp. 1–5.

[2] C. Canella, J. Van Bulck, M. Schwarz, M. Lipp, B. von Berg, P. Ortner, F. Piessens, D. Evtyushkin, and D. Gruss, "A Systematic Evaluation of Transient Execution Attacks and Defenses," in *Proc. 28th USENIX Security Symposium (USENIX Security 19)*, Santa Clara, CA, USA, 2019, pp. 249–266.

[3] J. Behrens, A. Cao, T. Zhai, S. Radhakrishnan, D. Wentzlaff, and A. Morrison, "Efficiently Mitigating Transient Execution Attacks using the Unmapped Speculation Contract," in *Proc. 14th USENIX Symposium on Operating Systems Design and Implementation (OSDI 20)*, 2020, pp. 1321–1337.