Cache Coherence Protocol Overhead in Non-Uniform Memory Access Architectures: An Empirical Analysis of MESIF State Transitions Under Heterogeneous Workload Contention

Rowan Turner
Professor
Technische Universität München
Arcisstraße 21, 80333 Munich, Bavaria, Germany

Riley Garcia
Associate Professor
Korea Advanced Institute of Science and Technology (KAIST)
291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea

Quinn Green
PhD
University of Toronto
27 King's College Circle, Toronto, Ontario M5S 1A1, Canada

Taylor Green
D.Sc
Delft University of Technology
Mekelweg 5, 2628 CD Delft, South Holland, Netherlands

Abstract. Non-Uniform Memory Access (NUMA) architectures present persistent coherence overhead challenges when MESIF-protocol state machines experience high-frequency invalidation cascades under heterogeneous thread workloads. This study empirically characterizes cache coherence traffic patterns across a 4-socket Intel Xeon Scalable (Sapphire Rapids) system running mixed read-write workloads derived from real-world HPC and database benchmarks. Using hardware performance counters and cycle-accurate simulation via gem5, we quantify inter-node snoop filter saturation thresholds, false-sharing penalties, and directory-entry eviction rates. Our findings demonstrate that workload-induced state-transition storms increase inter-socket memory latency by 34–61% beyond theoretical NUMA penalties. We further identify a novel mitigation strategy involving adaptive snoop-filter partitioning, reducing coherence traffic by up to 28% without architectural modification. These results carry immediate implications for OS-level NUMA-aware thread scheduling and microarchitectural design in next-generation many-core processors.

Keywords: NUMA architecture, MESIF cache coherence protocol, snoop filter saturation, false sharing penalty, hardware performance counters, inter-socket memory latency, heterogeneous workload contention, directory-based coherence, many-core microarchitecture

Introduction

Modern high-performance computing (HPC) systems, enterprise database servers, and cloud-native infrastructure increasingly rely on Non-Uniform Memory Access (NUMA) architectures to scale computational throughput beyond the physical limits of single-socket memory bandwidth. As of 2025, multi-socket server deployments—anchored by Intel Xeon Scalable, AMD EPYC Genoa, and ARM Neoverse V2 platforms—constitute the dominant substrate for latency-sensitive workloads including in-memory analytics, real-time transaction processing, and large-scale machine learning inference. These systems distribute physical memory across discrete NUMA nodes interconnected via high-speed coherence fabrics such as Intel Ultra Path Interconnect (UPI) or AMD Infinity Fabric, each carrying a non-trivial coherence tax that compounds unpredictably under realistic, heterogeneous workload conditions. The MESIF protocol—an extension of the classical MESI coherence model introducing a Forward state to optimize peer-to-peer cache-to-cache transfers—governs data consistency across private L1/L2 and shared LLC hierarchies in Intel's multi-socket implementations. While its theoretical operation is well-characterized in formal models, empirical behavior under concurrent mixed-access workloads remains inadequately quantified in the open literature. Specifically, the interaction between high-frequency write invalidations, snoop-filter capacity limits, and directory-entry eviction pressure produces emergent latency phenomena that neither analytical NUMA models nor synthetic microbenchmarks faithfully capture. This gap represents a critical obstacle to accurate performance prediction and OS-level scheduling optimization in production deployments. Prior investigations have addressed cache coherence overhead in isolation—either through simulation-only studies lacking hardware validation, or through narrowly scoped benchmarks targeting synthetic access patterns. Notably, the compound effect of simultaneous MESIF state-transition storms across multiple memory controllers interacting with OS-level page migration daemons (e.g., AutoNUMA) remains underexplored. Furthermore, existing snoop-filter management policies embedded in current microarchitectures treat filter capacity as a static resource, an assumption that breaks down under workloads exhibiting high temporal locality variance—a hallmark of collocated HPC and OLTP tasks in converged infrastructure. This paper addresses these deficiencies through a rigorous empirical methodology combining cycle-accurate gem5 simulation, hardware performance counter instrumentation (via Linux perf and Intel VTune Profiler), and controlled workload injection derived from SPEC CPU 2017, TPC-C, and Graph500 benchmarks. The remainder of this article is structured as follows: Section 2 reviews related work on MESIF-protocol behavior and NUMA-aware scheduling. Section 3 describes the experimental platform, workload configuration, and performance counter selection methodology. Section 4 presents quantitative results on coherence traffic, snoop-filter saturation thresholds, and inter-socket latency amplification. Section 5 introduces and evaluates the proposed adaptive snoop-filter partitioning heuristic. Section 6 discusses implications for OS scheduler design and future microarchitectural directions. Section 7 concludes with a summary of key findings and directions for future work.

This is a preliminary version. To read the full version of the article, please purchase a subscription.

References

1. Kumar, N., & Kataria, V. Enhanced Sentiment Classification using a Multi-layered Stacked Ensemble Architecture.
2. Semeniuk, V. (2024). Аналіз та оптимізація продуктивності баз даних у розподілених системах. COMPUTER-INTEGRATED TECHNOLOGIES: EDUCATION, SCIENCE, PRODUCTION, (54), 199-205.
3. D Satyanarayana, Perspective Chapter: Postquantum Security', Coding Theory - Advances in Communications Engineering and Information Security. IntechOpen, Nov. 19, 2024. doi: 10.5772/intechopen.1007829.
4. D. Satyanarayana and J. M. H. Elmirghani, "An Energy Efficient Network Architecture for Infrastructured Wireless Networks," 2010 IEEE Global Telecommunications Conference GLOBECOM 2010, Miami, FL, USA, 2010, pp. 1-6, doi: 10.1109/GLOCOM.2010.5683234.
5. D. Satyanarayana and J. M. H. Elmirghani, "A Voronoi Based Energy Efficient Architecture for Wireless Networks," 200S Third International Conference on Next Generation Mobile Applications, Services and Technologies, Cardiff, UK, 2009, pp. 377-382, doi: 10.1109/NGMAST.2009.79.
6. D. Satyanarayana and S. V. Rao, "Local Delaunay Triangulation for Mobile Nodes," 2008 First International Conference on Emerging Trends in Engineering and Technology, Nagpur, India, 2008, pp. 282-287, doi: 10.1109/ICETET.2008.253.
7. D. Satyanarayana, "Greedy Local Delaunay Triangulation Routing for Wireless Ad Hoc Networks," 2007 International Conference on Signal Processing, Communications and Networking, Chennai, India, 2007, pp. 49-53, doi: 10.1109/ICSCN.2007.350694
8. D. Satyanarayana, S. Chattopadhyay and J. Sasidhar, "Low power combinational circuit synthesis targeting multiplexer-based FPGAs," 17th International Conference on VLSI Design. Proceedings., Mumbai, India, 2004, pp. 79-84, doi: 10.1109/ICVD.2004.1260906.
9. Alkalbani, A.S., Mantoro, T., Degala, S. (2016). Energy-Distance Aware Clustering Scheme (E-DACS) for Wireless Sensor Networks. In: Meesad, P., Boonkrong, S., Unger, H. (eds) Recent Advances in Information and Communication Technology 2016. Advances in Intelligent Systems and Computing, vol 463. Springer, Cham. https://doi.org/10.1007/978-3-319-40415-8_27
10. Satyanarayana, D., Sathyashree, A. A. K., & Alkalbani, A. (2015, November). Voronoi LEACH for Energy Efficient Communication in Wireless Sensor Networks. In Proceedings for the International Conference on Electronics Computer networks and Information Technology, Dubai.

11. Hashimov, A. M., Guliyev, H. B., & Babayeva, A. R. (2017). Managing shunt reactors in accordance to fuzzy controllers for stabilization of voltage in high tension tire-cover. International Journal on Technical and Physical Problems of Engineering (IJTPE), 9(1), 18.
12. Yusifbayli, N. A., & Guliyev, H. B. (2011). Intelligent control system of voltage regimes in the electrical networks. IEEE PES innovative smart grid Europe, 5-7.
13. Rahmanov, N., Guliyev, H., İbtahimov, F., & Mammadov, Z. (2024). Determination of optimal dimensions of hybrid AC/DC distributed generation system with renewable sources for autonomous power supply of remote locations. In E3S Web of Conferences (Vol. 584, p. 01020). EDP Sciences.
14. Guliyev, H. B. (2024). Management modes of reactive power compensation facilities in networks with renewable energy sources with distorting loads. IJTPE Journal International Journal on Technical and physical problems of engineering (IJTPE), 16(1), 14-20.
15. Hashimov, A. M., Babayeva, A. R., & Guliyev, H. B. (2019). Shunt reactors control algorithm using fuzzy sets theory. International Journal on Technical and Physical Problems of Engineering, 11(1), 10-15.