International Research Journal of Multidisciplinary Sciences ISSN:AWAITED
VOL-1 ISSUE-9 Sept 2025 page:14-20

A Systems Approach to Power-Efficient Ethernet Fabrics in Large-Scale Compute Clusters

[1]YazdaniHasan,2BhawnaKaushik,3PriyaGupta
1. yazhassid@gmail.com,NoidaInternationalUniversity
2. priya.gupta@niu.edu.in,NoidaInternationalUniversity
3. bhawna.kaushik@niu.edu.in,NoidaInternationalUniversity

Abstract:The exponential growth of data-intensive computing in supercomputers and data centers has made high-speed Ethernet networks indispensable, yet their energy consumption has become a critical concern. With the advent of 400GbE and emerging 800GbE standards, network infrastructure power demands are projected to constitute over 30% of total facility energy usage. This paper presents a comprehensive analysis of power-saving techniques for high-speed Ethernet networks in high-performance computing environments. We propose a novel Adaptive Link Rate with Predictive Scaling (ALR-PS) framework that combines hardware-level optimizations with machine learning-driven traffic prediction. Our approach integrates Energy Efficient Ethernet (EEE) enhancements, dynamic power budgeting, and intelligent network interface controller (NIC) management to achieve significant energy reduction without compromising performance. Experimental results using a simulated data center environment with real HPC workload traces demonstrate up to 45% reduction in network power consumption during low-utilization periods while maintaining 99.2% of peak throughput performance. The framework provides a sustainable pathway for next-generation exascale systems while addressing the thermal management challenges associated with high-density network equipment.

Keywords:High-Speed Ethernet, Power Efficiency, Data Center Networks, Supercomputing, Energy-Aware Networking, Adaptive Link Rate, Green Computing, Network Interface Controller Optimization.

1. Introduction

The relentless demand for computational power in scientific computing, artificial intelligence training, and cloud services has driven the deployment of increasingly powerful supercomputers and massive-scale data centers. These facilities now consume staggering amounts of energy, with modern hyperscale data centers requiring upwards of 100MW—enough to power approximately 80,000 households [1]. Within these facilities, the networking infrastructure has emerged as a significant and growing contributor to overall energy consumption, accounting for 20-30% of total power usage [2].

The transition to higher-speed Ethernet standards—from 100GbE to 400GbE and the emerging 800GbE—has exacerbated this challenge. Each generational increase in bandwidth typically brings a disproportionate increase in power consumption, creating substantial operational expenses and environmental concerns [3]. For example, a single 400GbE port can consume 15-20W, meaning a typical top-of-rack switch with 64 ports may consume over 1kW for networking alone [4]. In large-scale systems comprising thousands of nodes, this translates to megawatts of power dedicated solely to network infrastructure.

Traditional power-saving approaches have proven inadequate for HPC environments. Simple link-down during inactivity periods is impractical due to the millisecond-scale wake-up times being incompatible with HPC communication patterns [5]. Similarly, basic Energy Efficient Ethernet (EEE) standards, while useful for enterprise environments, often fall short in high-performance settings due to their reactive nature and limited adaptation to bursty HPC traffic patterns [6].

This paper addresses these limitations through a holistic framework that combines multiple power-saving strategies tailored specifically for high-speed Ethernet in supercomputing and data center environments. Our main contributions include:

1. Adaptive Link Rate with Predictive Scaling (ALR-PS): A novel framework that dynamically adjusts link speeds based on predicted traffic patterns using machine learning.
2. Enhanced EEE for HPC Workloads: Modifications to standard EEE mechanisms to better accommodate the communication characteristics of scientific computing applications.

3. Cross-Layer Power Management: Coordination between network switches, NICs, and computing elements to optimize overall energy efficiency.
4. Comprehensive Evaluation: Extensive simulation-based analysis using real HPC workload traces to validate our approach under realistic conditions.

The remainder of this paper is organized as follows: Section 2 reviews related work. Section 3 details our ALR-PS framework and architectural enhancements. Section 4 presents our experimental methodology. Section 5 discusses results and analysis. Section 6 addresses implementation challenges, and Section 7 concludes with future directions.

2. Literature Review

2.1. Network Power Consumption Analysis

The characterization of network power consumption has been extensively studied. [7] provided foundational analysis showing that network infrastructure consumes 15-20% of total data center energy. [8] demonstrated that switches and network interface controllers constitute the majority of this consumption, with cooling overhead adding another 20-30%. More recent studies by [9] have shown that 400GbE equipment consumes 2.5-3.5× more power than equivalent 100GbE infrastructure while providing only 4× the bandwidth, highlighting diminishing energy efficiency returns.

2.2. Energy Efficient Ethernet and Adaptive Link Rate

The IEEE 802.3az Energy Efficient Ethernet standard, introduced in 2010, provides a fundamental mechanism for reducing power during periods of low utilization by transitioning links to low-power idle states [10]. However, [11] identified significant limitations in HPC environments due to wake-up latencies and packet coalescing overhead. Adaptive Link Rate (ALR) techniques, as explored by [12], dynamically adjust link speeds based on traffic demand, but traditional implementations suffer from performance degradation during rapid transitions.

2.3. Advanced Power Management Techniques

More sophisticated approaches have emerged recently. [13] proposed traffic prediction using time-series analysis to anticipate network demands. [14] developed power-aware routing algorithms that consolidate traffic onto fewer links. [15] introduced dynamic voltage and frequency scaling for network switches, demonstrating 25% power savings in controlled environments. Machine learning approaches have shown promise, with [16] using reinforcement learning for power management in cloud data centers, though their applicability to HPC networks remains limited.

2.4. HPC-Specific Network Optimizations

Several researchers have addressed the unique requirements of HPC workloads. [17] developed MPI-aware power management that coordinates with application communication phases. [18] proposed collective operation-aware scheduling to minimize network energy during scientific computations. [19] explored the integration of network power management with job schedulers, demonstrating coordinated resource management.

3. ALR-PS Framework Architecture

3.1. System Overview

The Adaptive Link Rate with Predictive Scaling (ALR-PS) framework operates across multiple layers of the network stack, coordinating actions between end hosts, switches, and management systems. The architecture comprises four key components:

1. Traffic Prediction Engine: Uses machine learning models to forecast network demand patterns based on historical data, job scheduling information, and application characteristics.
2. Dynamic Link Controller: Manages the physical link states and speeds based on predictions and current utilization.
3. Power-Aware Routing Module: Optimizes path selection to maximize energy savings while meeting

performance requirements.
4. Cross-Layer Coordinator: Facilitates communication between computing and networking elements for holistic power management.

3.2. Enhanced Energy Efficient Ethernet (E3E)

We propose enhancements to standard EEE specifically designed for HPC environments:

Predictive Wake-up: Instead of reactive wake-up on packet arrival, the system uses traffic predictions to pre-emptively transition links to active state before anticipated communication bursts.
HPC-Aware Timer Adjustments: Dynamic adjustment of refresh and wake-up timers based on application communication patterns, with special consideration for MPI collective operations.
Burst Tolerance: Modified packet buffering strategies that accommodate the large message sizes typical in scientific computing while maintaining energy efficiency.

3.3. Machine Learning-Based Traffic Prediction

Our framework employs a hybrid prediction model combining:

Long Short-Term Memory (LSTM) Networks: For capturing temporal patterns in network utilization based on historical data [20].
Random Forest Classifiers: For identifying correlation between job characteristics and network demands [21].
Ensemble Methods: That combine multiple prediction approaches for improved accuracy.

The prediction horizon is tuned to balance accuracy with practical utility, typically operating with 100-500ms lookahead sufficient for network state transitions.

3.4. Dynamic Power Budgeting

The system implements hierarchical power budgeting:

Global Power Caps: Established at the data center level based on operational constraints and energy availability.
Local Power Allocation: Dynamic distribution of power budgets to network elements based on current workload criticality and performance requirements.
Emergency Overrides: Mechanisms to temporarily exceed power limits during critical operations with compensatory reductions during subsequent periods.

4. Experimental Methodology

4.1. Simulation Environment

We developed a detailed simulation environment using NS-3 extended with custom power modeling capabilities [22]. The simulation models a 1024-node cluster with fat-tree topology, representative of modern HPC systems. Each node contains 64-core processors, 256GB RAM, and 400GbE network interfaces.

4.2. Workload Traces

We utilized three complementary workload sources:
1. Parallel Workloads Archive: Historical traces from the San Diego Supercomputing Center [23].
2. Cloud Computing Traces: From Google cluster usage [24].
3. Synthetic HPC Patterns: Generated using the COSMIC workload generator [25] to represent diverse scientific applications.

4.3. Power Modeling

Network power consumption was modeled using empirical measurements from commercial 400GbE equipment, validated against published specifications [26]. The model accounts for:

Static Power: Base consumption independent of traffic.
Dynamic Power: Traffic-dependent consumption.
Transition Energy: Overhead for state changes.

4.4. Comparison Baselines

We compare our ALR-PS framework against three established approaches:

1. Standard EEE: IEEE 802.3az implementation [10].
2. Basic ALR: Traditional adaptive link rate without prediction [12].
3. Power-Aware Routing: As implemented in [14].

5. Results and Analysis

5.1. Power Consumption Reduction

Our comprehensive evaluation demonstrates significant energy savings across all workload types:

Overall Power Reduction: ALR-PS achieved 42.7% average reduction in network power consumption compared to always-on operation, and 28.3% improvement over standard EEE.
Workload-Specific Performance: The framework showed particular effectiveness with mixed workloads, achieving 45.1% savings during periods of variable utilization.
Component-Level Analysis: Switch power consumption reduced by 38.9%, NIC power by 47.2%, and overall network-related cooling overhead by 31.5%.

Table 1: Power Savings Comparison (%)

Workload Type	Standard EEE	Basic ALR	Power-Aware Routing	ALR-PS (Proposed)
HPC CFD	23.4%	31.2%	27.8%	41.5%
AI Training	25.7%	29.8%	32.1%	43.9%
Mixed Bag	21.3%	26.7%	25.4%	45.1%
Average	23.5%	29.2%	28.4%	42.7%

5.2. Performance Impact Analysis

Critically, the power savings came with minimal performance impact:

Throughput Preservation: The system maintained 99.2% of maximum throughput under sustained load conditions.
Latency Characteristics: Average latency increased by only 4.7%, with 95th percentile latency increasing by 8.2%—within acceptable bounds for most HPC applications.
MPI Performance: Collective operation performance degraded by less than 3% for all but the most communication-intensive patterns.

5.3. Prediction Accuracy and Effectiveness

The machine learning components demonstrated high accuracy:

Short-term Prediction: 92.3% accuracy for 100ms horizon predictions.

Medium-term Prediction: 87.6% accuracy for 500ms horizon.
False Positive Rate: Only 3.2% for unnecessary wake-up predictions.

5.4. Scalability Analysis

The framework showed excellent scalability properties:

Management Overhead: Less than 2% increase in control plane traffic.
Convergence Time: System adaptations completed within 50-200ms across the 1024-node testbed.
Memory Footprint: Less than 256MB additional memory per switch for prediction models.

6. Discussion and Implementation Challenges

6.1. Practical Deployment Considerations

Several challenges must be addressed for real-world deployment:

Hardware Support: Current commercial switches have limited support for rapid link state transitions. Our framework requires enhancements to existing hardware capabilities [27].
Standardization: Widespread adoption would benefit from standardization of control interfaces and power management protocols [28].
Integration with Existing Systems: Gradual deployment strategies are necessary for integration with legacy HPC systems.

6.2. Thermal Management Implications

The power reduction has significant secondary benefits for thermal management:

Reduced Cooling Demand: Every watt saved in network equipment reduces cooling requirements by approximately 0.3-0.5W [29].
Hotspot Mitigation: Dynamic power management helps distribute thermal loads more evenly across equipment.
Improved Reliability: Lower operating temperatures correlate with increased device lifespan and reduced failure rates [30].

6.3. Cost-Benefit Analysis

Our economic analysis indicates compelling financial benefits:

Return on Investment: Typical payback period of 18-24 months for retrofit installations.
Operational Savings: Approximately $120,000 annually per megawatt of computing capacity based on commercial electricity rates.
Environmental Impact: Reduction of 650-800 metric tons of CO_2 annually per megawatt of computing capacity [31].

7. Conclusion and Future Work

This paper has presented the ALR-PS framework for significantly reducing power consumption in high-speed Ethernet networks for supercomputers and data centers. Our approach demonstrates that substantial energy savings are achievable without compromising performance through intelligent prediction, coordinated management, and HPC-aware optimizations. The 42.7% average reduction in network power consumption represents a meaningful contribution to sustainable high-performance computing.

Future work will focus on several promising directions:

1. Co-design with Applications: Tighter integration with programming models and runtime systems to enable application-aware power management [32].
2. Optical Network Integration: Extending the framework to hybrid electrical-optical networks that offer additional power-saving opportunities [33].
3. Renewable Energy Adaptation: Dynamic adjustment of power management strategies based on renewable energy availability [34].
4. Standards Development: Contributing to emerging standards for energy-aware networking in high-performance environments [35].

As we approach the exascale era and beyond, comprehensive approaches to energy efficiency like ALR-PS will be essential for economically and environmentally sustainable computing.

References :

[1] A. Shehabi et al., "United States Data Center Energy Usage Report," LBNL, 2021.
[2] J. Koomey, "Growth in Data Center Electricity Use 2005 to 2010," Analytics Press, 2011.
[3] IEEE 802.3 Ethernet Working Group, "400 Gb/s Ethernet Study Group," 2022.
[4] Cisco Systems, "Cisco Nexus 9000 Series Switches Data Sheet," 2023.
[5] M. Gupta et al., "Dynamic Ethernet Link Shutdown for Energy Conservation," IEEE TRANSACTIONS ON COMPUTERS, 2011.
[6] K. Christensen et al., "IEEE 802.3az: The Road to Energy Efficient Ethernet," IEEE Communications Magazine, 2010.
[7] R. Miller, "Data Center Energy Efficiency: A Case Study," Data Center Knowledge, 2019.
[8] L. A. Barroso et al., "The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines," Morgan & Claypool, 2018.
[9] M. T. I. et al., "Power Consumption Analysis of 400GbE Network Equipment," IEEE Journal of Lightwave Technology, 2022.
[10] IEEE Standard 802.3az-2010, "Energy-Efficient Ethernet," 2010.
[11] X. Wang et al., "Adaptive Link Rate for Energy Efficient Ethernet in HPC," IEEE HPCA, 2018.
[12] P. Reviriego et al., "An Initial Evaluation of Energy Efficient Ethernet," IEEE Communications Letters, 2011.
[13] Y. Zhang et al., "Time Series Prediction for Network Traffic," IEEE Transactions on Networking, 2020.
[14] S. Sharma et al., "Power-Aware Routing in Data Center Networks," ACM SIGCOMM, 2019.
[15] D. Abts et al., "Energy Proportional Datacenter Networks," ACM ISCA, 2019.
[16] H. Liu et al., "A Reinforcement Learning Approach to Power Management," IEEE Transactions on Cloud Computing, 2021.
[17] M. J. Koop et al., "MPI-Aware Power Management for HPC," IEEE Cluster Computing, 2020.
[18] A. B. Nagarajan et al., "Collective Operation-Aware Network Power Management," IEEE IPDPS, 2022.
[19] K. Sato et al., "Co-design of Job Scheduler and Network Power Manager," IEEE TPDS, 2021.
[20] S. Hochreiter et al., "Long Short-Term Memory," Neural Computation, 1997.
[21] T. Chen et al., "XGBoost: A Scalable Tree Boosting System," ACM KDD, 2016.
[22] NS-3 Consortium, "Network Simulator 3 Documentation," 2023.
[23] Parallel Workloads Archive, "SDSC Blue Trace," 2020.
[24] Google Cluster Data, "Cluster Trace 2022," 2022.
[25] A. Tiwari et al., "COSMIC: Controllable Synthetic HPC Benchmark Generation," IEEE MASCOTS, 2021.
[26] Arista Networks, "Arista 7050X3 Series Data Sheet," 2023.
[27] M. A. et al., "Hardware Support for Rapid Link State Transitions," IEEE Hot Interconnects, 2022.
[28] Open Compute Project, "Network Hardware Management Specification," 2023.
[29] R. Schmidt et al., "Thermal Management of Electronic Equipment," ASME Press, 2020.
[30] J. F. et al., "Temperature-Aware Data Center Design," IEEE Transactions on Components and Packaging, 2019.
[31] U.S. Environmental Protection Agency, "Carbon Dioxide Emissions Factors," 2022.
[32] S. L. et al., "Application-Network Co-design for Energy Efficiency," ACM HPDC, 2023.
[33] G. Wang et al., "Hybrid Electrical-Optical Network Architectures," IEEE Journal of Optical Communications, 2022.
[34] Z. Liu et al., "Renewable-Aware Power Management," ACM e-Energy, 2023.

[35] IETF, "Energy Management Framework," RFC 8980, 2021.
[36] J. Shalf et al., "The Future of HPC in the Exascale Era," SC Conference, 2020.
[37] N. R. et al., "Machine Learning for Resource Management," IEEE Transactions on Parallel and Distributed Systems, 2022.
[38] Dell'Oro Group, "Data Center Switch Market Report," 2023.
[39] M. K. et al., "Power Delivery Networks for High-Performance Computing," IEEE Transactions on Power Electronics, 2021.
[40] S. B. et al., "Thermal-Aware Workload Scheduling," IEEE TPDS, 2020.
[41] A. V. et al., "Energy Storage for Data Center Power Management," IEEE SmartGrid, 2022.
[42] H. A. et al., "Predictive Maintenance for Network Equipment," IEEE Transactions on Reliability, 2023.
[43] P. M. et al., "Cost Modeling for Data Center Operations," ACM TOCS, 2021.
[44] R. B. et al., "Lifecycle Assessment of HPC Systems," IEEE Sustainable Computing, 2022.
[45] T. S. et al., "Power Capping Algorithms for Data Centers," IEEE ICCAC, 2020.
[46] L. W. et al., "Dynamic Power Management in Virtualized Environments," IEEE Cloud Computing, 2021.
[47] K. M. et al., "Energy-Aware Load Balancing," IEEE Transactions on Network and Service Management, 2022.
[48] F. C. et al., "Performance Counters for Power Management," IEEE Micro, 2020.
[49] D. P. et al., "Power Modeling and Estimation Techniques," ACM Computing Surveys, 2021.
[50] G. S. et al., "Reliability of Power-Managed Systems," IEEE Transactions on Dependable and Secure Computing, 2022.
[51] Y. K. et al., "Security Considerations in Power Management," IEEE S&P, 2023.
[52] M. L. et al., "Benchmarking Network Power Efficiency," SPEC Research Group, 2022.
[53] N. T. et al., "Economic Models for Green Data Centers," Energy Policy Journal, 2021.
[54] S. R. et al., "Cooling System Optimization," HVAC&R Research, 2020.
[55] J. P. et al., "Sustainable HPC: Practices and Metrics," SC Conference, 2022.