Efficient Reinforcement Learning Implementations for Sustainable Operation of Liquid Cooled HPC Data Centers

Avisek Naug, Antonio Guillen-Perez, Vineet Gundecha, Ashwin Ramesh Babu, Sahand Ghorbanpour, Ricardo Luna Gutierrez, Soumyendu Sarkar

HPE Labs, Hewlett Packard Enterprise

Introduction

The demand for high-performance computing (HPC) and AI has led to a significant increase in data center energy consumption. Traditional air cooling is often insufficient for modern high-density servers. Liquid cooling presents a more efficient alternative, capable of reducing cooling energy usage by up to 63%.

However, optimizing these complex liquid cooling systems is a major challenge. Current methods often rely on static strategies, failing to unlock the full potential of liquid cooling. While reinforcement learning (RL) has shown promise, existing approaches face challenges in scalability and real-time application.

System Overview

We use a high-fidelity digital twin of the Oak Ridge National Laboratory's Frontier supercomputer cooling system. This environment models an end-to-end liquid cooling system, from cooling towers to individual server blade groups. The system is modular, allowing for customizable data center configurations.

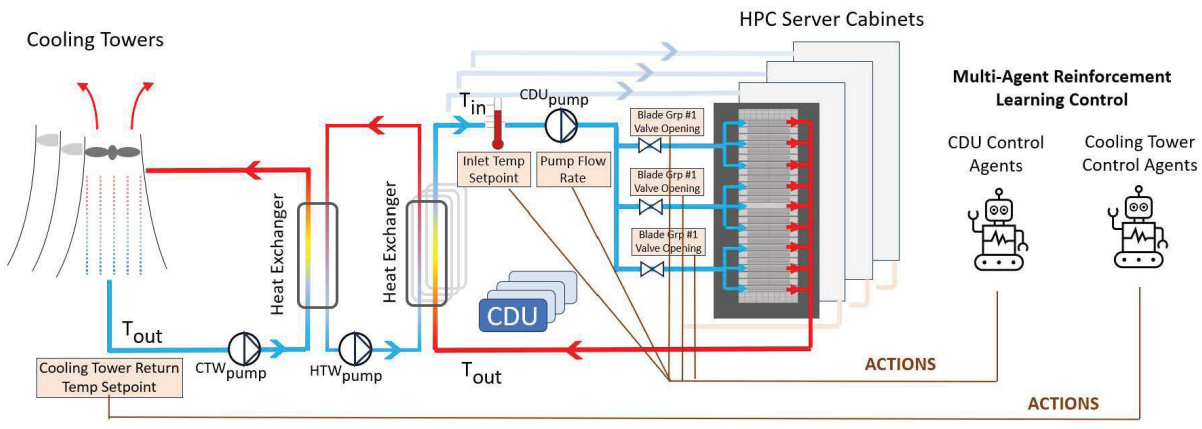

Figure: Overview of the liquid-cooled data center system. RL agents control the Cooling Towers and the Cooling Distribution Units (CDUs).

Methods: RL for Efficient Control

We developed a multi-agent reinforcement learning (MARL) approach to manage the cooling system.

Centralized Action Execution: To enhance efficiency, observations from similar components (e.g., all blade groups in a cabinet) are batched together. A single forward pass through the policy network is then performed, which significantly accelerates the inference process.

Centralization is grounded using the theory Markov D-Separation

Enables massively scalable and parallel deployment

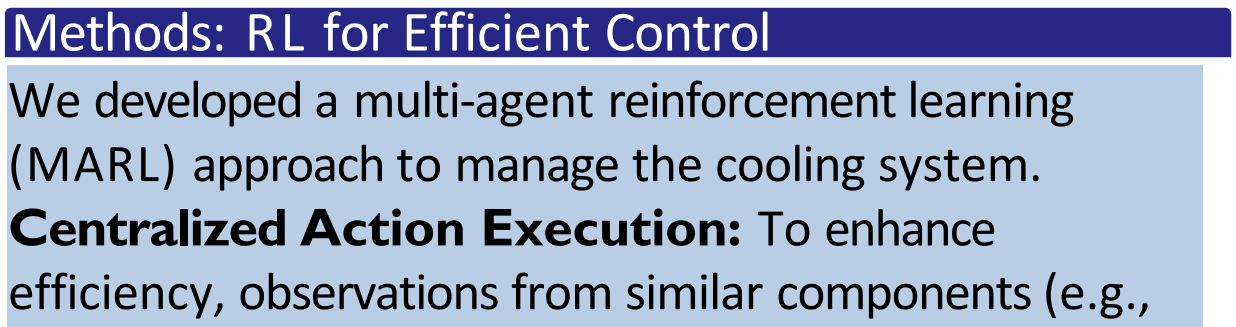

Figure: Centralized action execution for scalable inference.

Multi-Head Policy: The policy network for the blade group agents has two heads:

- **Head 1:** Determines coolant supply temperature and pump flow rate.
- **Head 2:** Controls valve openings for each blade group using a Dirichlet distribution to ensure proportional coolant distribution.

This decomposition provides more direct reward feedback and leads to more effective control strategies.

Experiments and Results

We evaluated our RL strategies against the industry standard ASHRAE G36 baseline in a simulated data center with 2 cooling towers, 5 cabinets, and 3 blade groups per cabinet.

Ablation Study: Our multi-agent RL with centralized action and a multi-head policy (Case 7) achieved the best performance. It reached 95.63% temperature compliance while also having the lowest cooling tower power consumption.

Table 1: **Ablation of RL Agent Design.** We incrementally replace the static baseline (Case 1) with RL controllers for: Cooling Tower (Case 2), CDU coolant setpoint/flow (Case 3), and Blade Group valves (Case 4). Case 5 introduces a single multi-agent RL controller, Case 6 adds batching for state space reduction, and Case 7 uses a multi-head policy. Experiments use N=2 towers, m=2 cells, C=5 cabinets with B=3 blade groups each, and are evaluated on an unseen exogenous trace. Blade-group temperature compliance $D_{blade,avg}$ is computed with $\mathcal{U}_T$=40°C and $\mathcal{L}_T$=20°C.

Metric → / Agent/Control Type ↓	Control Details	$D_{blade,avg}$% (% of time Temp within ideal range)	$\sum P_{ij}(kW)$ (Cooling Tower Avg Power)	$\sum Q_i$ (IT Level Avg Cooling Power)	Avg Episode Reward per Cabinet	Avg Episode Reward per Cooling Tower
1. Baseline Control	ASHRAE G36	76.92	237.31	235.28	1697.08	360.17
2. CT RL + BG Baseline	Only CT RL control	79.21	246.46	235.03	1702.16	352.28
3. CT Baseline + BG RL	No Valve Control	64.91	217.6	203.96	1638.48	372.97
4. CT Baseline + BG RL	With Valve Control	77.13	217.37	211.83	1698.36	373.52
5. Multiagent RL	Decentralized Action	78.24	218.11	212.94	1697.49	370.51
6. Multiagent RL	Centralized Action (CA)	90.46	207.37	208.69	1714.65	395.88
7. Multiagent RL	CA & Multihead policy	**95.63**	**206.52**	**197.18**	**1726.31**	**396.24**

Figure: Ablation study of different RL agent designs.

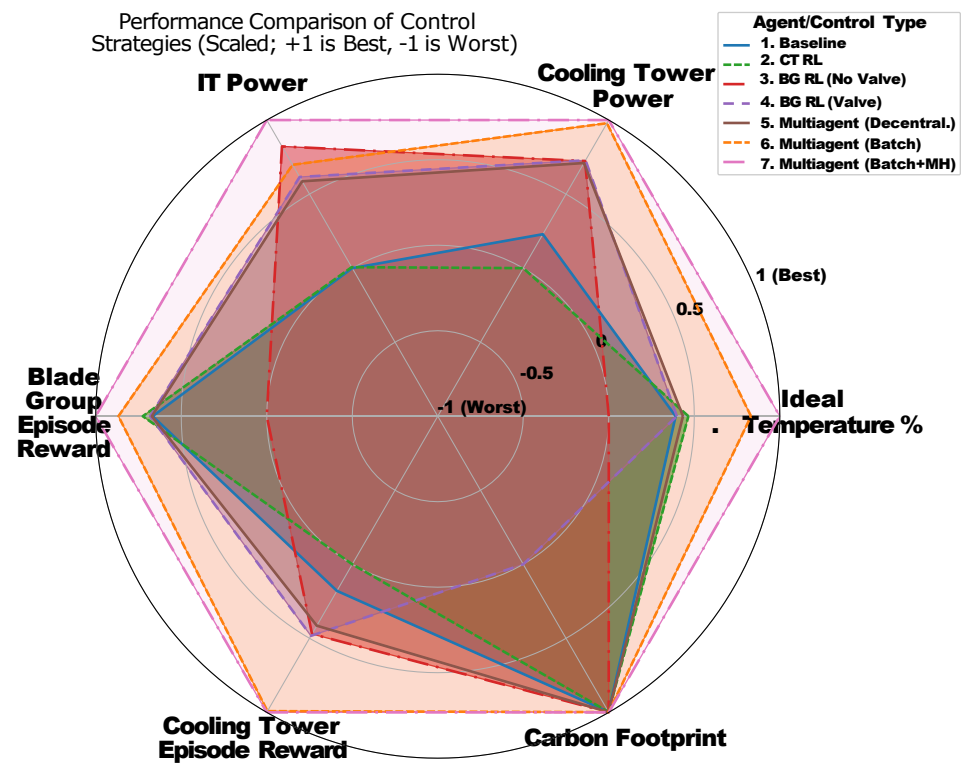

Figure: Relative performance of different RL approaches.

Sustainability on a Larger Scale: The multi-head centralized action RL policy consistently outperformed the baseline at larger data center scales, achieving better temperature compliance and significant carbon footprint savings.

Metric↓	Ablation→	$N=2, m=2, C=10, B=3$	$N=2, m=2, C=15, B=3$	$N=3, m=2, C=20, B=3$	$N=4, m=2, C=15, B=3$
$D_{\{blade,avg\}}$%	ASHRAE G36	71.92	68.24	75.31	83.08
	CA Policy	96.28	86.19	94.07	92.61
$CFP(kgCO_2)$	ASHRAE G36	4448.96	6254.15	10518.78	15753.9
	CA Policy	4432.78	6392.21	9932.36	12652.1

Table: Performance at Scale: Evaluation of Rul-Based Control vs Multhi-head Centralized Action Policy for Scalable Blade-Group Agent with increasing Data Center sizes.

Conclusion and Climate Impact

Our work introduces an efficient and scalable suite of RL agents for controlling liquid-cooled data centers. The experiments demonstrate significant improvements in energy efficiency, thermal management, and carbon footprint reduction compared to traditional methods.

This research provides a practical pathway for implementing RL in sustainable data centers, contributing to the reduction of the environmental impact of our digital infrastructure.