Edge–Cloud Reinforcement Learning for Transferable HVAC Control Across Vietnamese Climate Zones

Quoc Hung Nguyen[0000-0002-6363-0160]1, Hoai An Thai[0009-0004-4093-522X]1, Duy Tan Nguyen[0009-0000-2832-8647]1, Vy Le[0009-0003-0387-8302]1, and Thi Thuy Duong Nguyen[0009-0000-3548-2212]1

UEH College of Technology and Design, University of Economics Ho Chi Minh City, Vietnam
hungngq@ueh.edu.vn

Abstract. Heating, ventilation, and air-conditioning (HVAC) systems are a major driver of commercial-building electricity use in Vietnam, where a long north–south climate gradient creates a demanding transfer setting for learned controllers. Our focus is not a new PPO variant, but the deployment question: how an HVAC policy should be executed, checked, updated, and transferred across climates. We propose a framework that separates low-latency edge inference from cloud-side multi-context Proximal Policy Optimisation (PPO) retraining. The framework is evaluated with HOT commercial-building archetypes and city-specific TMYx weather files for Ho Chi Minh City, Can Tho, Da Nang, and Hanoi, covering ASHRAE climate zones 0A, 1A, and 2A. The baseline analysis recovers a strong cooling-energy gradient: under identical static setpoints, Ho Chi Minh City consumes 1.44× the HVAC electricity of Hanoi for OfficeSmall and 1.68× for OfficeMedium. Edge inference is approximately 71,000× faster than cloud inference at the 95th percentile. Pure-PPO variants reduce comfort violations from 28.7% to 19.8% but increase HVAC energy by 58.7%. The proposed edge–cloud controller brings average energy back to within 1% of the static baseline, keeps comfort close to the static baseline, and reduces held-out target energy by 52.0% relative to pure PPO.

Keywords: Reinforcement learning · HVAC control · Edge–cloud architecture · Transfer learning · Vietnam · Tropical buildings

1 Introduction

Vietnam's commercial-building stock is expanding rapidly, and cooling demand is already a policy-relevant component of national energy use. Reinforcement learning is attractive for HVAC control because it can optimise sequential energy–comfort trade-offs without a hand-derived building model[11,12]. However, many HVAC-RL evaluations remain offline and single-climate: they under-report inference latency, climate-driven transfer, and policy-update procedures. These issues are central in Vietnam, whose $\sim 1{,}700$ km north–south gradient spans extremely hot–humid, hot–humid, and warm–humid ASHRAE zones.

The study contributes three elements needed for a deployable HVAC-RL workflow. First, it proposes an edge–cloud architecture in which the edge executes and validates actions while the cloud trains and promotes policy updates. Second, it defines a Vietnam multi-city protocol using HOT building archetypes[2] and EnergyPlus weather data[10]. Third, it quantifies a measured energy–comfort trade-off in the Vietnam HOT/EnergyPlus protocol: PPO improves comfort but consumes more energy than static control, whereas the edge–cloud acceptance gate recovers near-static energy use. Code, configurations, trained policies, and aggregated results are released at `https://github.com/hoaianthai345/RL_HAVC_Control_VN`.

2 Related Work

RL for building control is commonly formulated as a Markov decision process in which observations include thermal state, weather, occupancy, and time features, while actions specify setpoints or actuator commands[11]. PPO is widely used because its clipped objective improves training stability in continuous control[8]. Prior work has shown simulation energy savings over fixed or rule-based baselines[14,13], but real deployment also requires latency guarantees, safe action validation, and retraining under distribution shift[3]. Transfer learning for HVAC has been studied across buildings and weather files[4,5]; the HOT dataset enables larger-scale cross-climate studies[2]. Edge–cloud AI provides a practical split: latency-sensitive inference stays local, while cloud compute supports retraining and policy governance[9,6].

3 Methodology

3.1 Contextual HVAC Control

We model each city–building pair c as a contextual MDP $\mathcal{M}_c = \langle \mathcal{S}, \mathcal{A}, P_c, R_c, \gamma \rangle$. The state contains indoor temperature, outdoor dry-bulb temperature, outdoor relative humidity, occupancy, previous-step HVAC electricity, and time-of-day features. The action is a dual setpoint $a_t = [\text{heat\_sp}, \text{cool\_sp}] \in [16, 24] \times [20, 30]\,^{\circ}\mathrm{C}$. The reward is

$$r_t = -\left(0.12 e_t + 2.0 d_t + 0.05 u_t\right), \tag{1}$$

where e_t is 15-min HVAC electricity, d_t is deviation outside the ASHRAE 55 comfort band (20–24 °C)[1], and $u_t = \|a_t - a_{t-1}\|_1$ penalises setpoint chatter.

3.2 Edge–Cloud Architecture

Figure 1 summarises the proposed architecture. The edge layer executes the 15-min control loop, normalises observations, runs local inference, validates actions

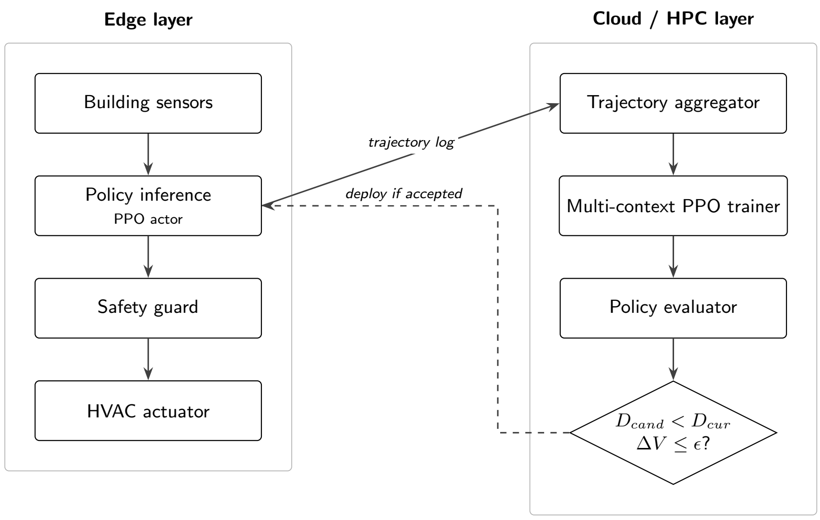

Fig. 1. Edge–cloud adaptive RL architecture. Edge devices execute and validate actions; the cloud trains and promotes policy updates.

against hard bounds and a 1.5 °C heat–cool separation rule, dispatches setpoints to EnergyPlus, and logs trajectories. The cloud layer aggregates source-context trajectories from Ho Chi Minh City and Can Tho, trains multi-context PPO policies, and promotes a candidate only if a fixed acceptance rule improves energy use without violating comfort-regression, latency, or action-instability constraints. Da Nang and Hanoi are held out from training and used only for transfer evaluation.

4 Experimental Protocol

4.1 Cities, Buildings, and Methods

We use EnergyPlus 24.1 HOT epJSON archetypes calibrated for ASHRAE zone 0A and pair them with city-specific TMYx weather files from Climate OneBuilding[10]. The ten evaluated contexts are HCMC and Can Tho (OfficeSmall, OfficeMedium, Hospital) plus Da Nang and Hanoi OfficeSmall/OfficeMedium. HCMC and Can Tho form the source pool; Da Nang and Hanoi are held-out target cities.

We compare seven methods: static setpoints (20/24 °C), ASHRAE-style outdoor-temperature-conditioned setpoints, a family of single-context PPO checkpoints, multi-context PPO trained jointly on source contexts, edge-only execution of the multi-context policy, cloud-only inference through a safety-aware proxy with simulated network delay ($p_{50} = 80$ ms, $p_{95} = 85$ ms), and the proposed edge–cloud adaptive controller. PPO uses Stable-Baselines3[7], 5×10^5 timesteps per context, learning rate 5×10^{-4}, discount 0.99, GAE $\lambda = 0.95$, clip range 0.15, and minibatch size 256. Each method is evaluated with three seeds.

Table 1. Static fixed-setpoint baseline performance (288-step one-day episode; mean over three seeds).

Context	Zone	Energy (kWh)	Comfort viol. (%)	$\overline{\Delta T}$ (°C)
HCMC OfficeSmall	0A	182.3	55.6	0.14
HCMC OfficeMedium	0A	1660.4	20.8	0.09
HCMC Hospital	0A	6831.1	0.0	0.00
Can Tho OfficeSmall	0A	229.0	53.8	0.15
Can Tho OfficeMedium	0A	2480.0	18.8	0.09
Can Tho Hospital	0A	6936.6	0.0	0.00
Da Nang OfficeSmall	1A	152.4	44.4	0.11
Da Nang OfficeMedium	1A	1327.0	22.6	0.06
Hanoi OfficeSmall	2A	126.8	48.6	0.10
Hanoi OfficeMedium	2A	990.1	22.6	0.05

Table 2. Inference latency over 200 repetitions per context.

Mode	p_{50} ms	p_{95} ms	Ratio
Edge-only inference	0.00119	0.00120	1×
Cloud-only inference	84.4	85.2	≈ 71,000×
PPO checkpoint	∼ 148 kB, MlpPolicy with two 64-unit layers		

4.2 Metrics

Control metrics are episode HVAC energy (kWh), comfort-violation rate, mean temperature deviation, cumulative reward, and action instability. Deployment metrics are p_{50}/p_{95} inference latency and policy size. Transfer metrics are held-out target energy, target/source energy ratio, adaptation gain, and worst-context energy regret. The results test three hypotheses: H1, static HVAC energy follows Vietnam's north–south climate gradient; H2, edge inference gives a large latency margin; and H3, temperate ASHRAE scheduling does not provide a useful tropical energy–comfort trade-off.

5 Results

5.1 Baseline and Latency Results

Table 1 supports H1. For OfficeSmall, HCMC consumes 1.44× the static HVAC electricity of Hanoi; for OfficeMedium the ratio is 1.68×. Hospitals have near-zero comfort violations but high energy because medical-grade ventilation dominates weather sensitivity. ASHRAE-style scheduling saves energy only by tolerating much higher comfort violation (80.6% average violation versus 28.7% for static), supporting H3.

Table 2 supports H2. Edge inference attains $p_{95} = 0.00120$ ms, compared with 85.2 ms for cloud inference, an approximately 71,000× latency advantage.

Table 3. Method performance averaged over ten Vietnam contexts.

Method	Energy (kWh)	Comfort viol. (%)	Reward	Instability
Static setpoint	2091.6	28.7	−296.4	0.0035
ASHRAE rule-based	1828.3	80.6	−1075.7	0.0035
Single-context PPO	3318.9	19.8	−399.3	0.0174
Multi-context PPO	3318.9	19.8	−399.3	0.0174
Edge-only PPO	3318.9	19.8	−399.3	0.0174
Cloud-only PPO	2099.0	15.6	−285.6	0.0633
Edge–cloud adaptive	2072.4	26.7	−298.1	0.0608

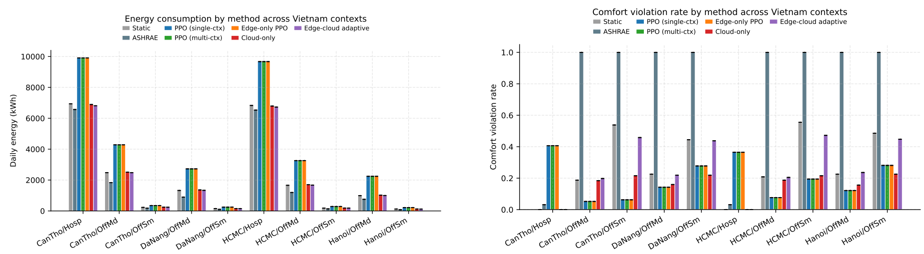

Fig. 2. Per-context energy (left) and comfort-violation rate (right) by control method.

5.2 PPO Control Performance

Table 3 shows that pure-PPO policies reduce comfort violations from 28.7% (static) to 19.8%, but increase energy from 2091.6 to 3318.9 kWh (+58.7%). Under deterministic evaluation, single-context, multi-context, and edge-only PPO converge to identical aggregate rollouts. This equality reflects deterministic evaluation trajectories, not identical checkpoint files or training histories. Cloud-only inference is lower-energy and lower-violation because its safety-aware cloud proxy applies a comfort-priority wrapper when the delay margin is active; this is a deployment-wrapper effect, not a policy-weight effect. The proposed edge–cloud adaptive controller recovers static-level energy (2072.4 kWh) and static-level comfort (26.7% versus 28.7%).

5.3 Cross-City Transfer and Adaptation

Table 4 reports source-domain checks and all four held-out target contexts. The three pure-PPO variants have identical held-out target means (1358.1 kWh). The edge–cloud adaptive controller reduces the held-out target mean to 651.7 kWh, a 52.0% reduction relative to pure PPO. In these rollouts, the gain is therefore attributable mainly to the acceptance layer, since multi-context PPO and single-context PPO produce the same target trajectories.

Under HCMC AMY weather files for 2022–2024, the adaptation-gain energy delta (TMYx baseline minus AMY mean) is 19.2 ± 17.5 kWh per episode, and

Table 4. Transfer energy (kWh). Can Tho is a source-domain check; Da Nang and Hanoi are held-out targets.

Method	Can Tho source check		Held-out target				Target mean
	Sm	Med	Da Nang Sm	Da Nang Med	Hanoi Sm	Hanoi Med	
Single PPO	350.6	4280.3	247.5	2724.3	217.9	2242.9	1358.1
Multi PPO	350.6	4280.3	247.5	2724.3	217.9	2242.9	1358.1
Edge-only	350.6	4280.3	247.5	2724.3	217.9	2242.9	1358.1
Edge–cloud	240.1	2476.2	149.7	1333.5	124.1	999.4	651.7

comfort-violation rates remain within ±3 percentage points of the TMYx baseline. This indicates robustness to the moderate inter-annual shifts represented by recent HCMC weather years. The deterministic ablation further shows that single-context and multi-context PPO have different checkpoint hashes but indistinguishable aggregate rollouts in this one-day evaluation. Longer stochastic rollouts are needed to isolate regularisation effects.

6 Discussion and Conclusion

The experiments indicate that PPO alone is insufficient under the reward weights and one-day evaluation protocol used here. PPO strongly reduces comfort violations but does so by running cooling more aggressively, raising energy by 58.7% over the static baseline. Conversely, ASHRAE-style scheduling reduces energy only by allowing large comfort failures, so it provides no useful tropical energy–comfort advantage. The edge–cloud design separates two operational concerns: edge inference is fast enough for real-time execution, while cloud retraining and acceptance testing provide a practical point at which energy and comfort constraints can be checked before deployment.

The study remains simulation-only. HOT archetypes are calibrated for zone 0A; applying the same models to Hanoi may underestimate winter heating behaviour. All cities share one epJSON per archetype, whereas real buildings vary in envelope, orientation, equipment, and occupancy. The validation also uses one-day episodes; longer stochastic evaluations are required before field deployment.

In summary, Vietnam's multi-zone climate gradient exposes a clear transfer challenge for HVAC control. Pure PPO improves comfort but increases energy, while the proposed edge–cloud adaptive controller achieves near-static energy and comfort and reduces held-out target energy by 52.0% relative to pure PPO. For tropical and warm-humid buildings, the main lesson is that the policy should not be deployed alone: it needs a local safety check and a cloud-side update rule evaluated on target-like weather.

Acknowledgements

This research is funded by the University of Economics Ho Chi Minh City (UEH).

Data Availability

Building models are drawn from the public HOT dataset[2]. Weather files are from `https://climate.onebuilding.org`[10]. Code, configuration files, trained policy artifacts, and aggregated result tables are available at `https://github.com/hoaianthai345/RL_HAVC_Control_VN`.

References

1. ASHRAE: ANSI/ASHRAE standard 55-2023: Thermal environmental conditions for human occupancy. Tech. rep., American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA, USA (2023)
2. Berkes, A., Bengio, Y., Rolnick, D., Vakalis, D.: A hot dataset: 150,000 buildings for hvac operations transfer research. In: Proceedings of the 12th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation. pp. 171–180. ACM, Golden, CO, USA (2025). `https://doi.org/10.1145/3736425.3770110`, `https://huggingface.co/datasets/BuildingBench/HOT`
3. Dulac-Arnold, G., Levine, N., Mankowitz, D.J., Li, J., Gowal, S., Hutter, F., Lopes, A.C.F.: Challenges of real-world reinforcement learning: Definitions, benchmarks and analysis. Machine Learning **110**, 2419–2468 (2021). `https://doi.org/10.1007/s10994-021-05961-4`
4. Fang, X., Gong, G., Li, G., Chun, L., Li, W., Peng, P.: Cross temporal-spatial transferability investigation of deep reinforcement learning control strategy in the building HVAC system level. Energy **263**, 125679 (2023). `https://doi.org/10.1016/j.energy.2022.125679`
5. Kadamala, K., Chambers, D., Barrett, E.: Enhancing HVAC control systems through transfer learning with deep reinforcement learning agents. Smart Energy **13**, 100070 (2024). `https://doi.org/10.1016/j.segy.2024.100070`
6. Li, Y., Chen, X., Liu, W., Zhao, P.: Edge ai for building control: Latency, staleness, and adaptation under distribution shift. Applied Energy **330**, 120320 (2023). `https://doi.org/10.1016/j.apenergy.2022.120320`
7. Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M., Dormann, N.: Stable-baselines3: Reliable reinforcement learning implementations. Journal of Machine Learning Research **22**(268), 1–8 (2021), `http://jmlr.org/papers/v22/20-1364.html`
8. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O.: Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347 (2017), `https://arxiv.org/abs/1707.06347`
9. Shi, W., Cao, J., Zhang, Q., Li, Y., Xu, L.: Edge computing: Vision and challenges. IEEE Internet of Things Journal **3**(5), 637–646 (2016). `https://doi.org/10.1109/JIOT.2016.2579198`
10. U.S. Department of Energy: EnergyPlus Weather Data and Climate.OneBuilding Repository (2024), `https://climate.onebuilding.org`, accessed May 2026
11. Vázquez-Canteli, J.R., Nagy, Z.: Reinforcement learning for demand response: A review of algorithms and modeling techniques. Applied Energy **235**, 1072–1089 (2019). `https://doi.org/10.1016/j.apenergy.2018.11.002`
12. Wang, Z., Hong, T.: Reinforcement learning for building controls: The opportunities and challenges. Applied Energy **269**, 115036 (2020). `https://doi.org/10.1016/j.apenergy.2020.115036`

13. Yu, L., Sun, Y., Xu, Z., Shen, C., Yue, D., Jiang, T., Guan, X.: Multi-agent deep reinforcement learning for HVAC control in commercial buildings. IEEE Transactions on Smart Grid **12**(1), 407–419 (2021). https://doi.org/10.1109/TSG.2020.3011739
14. Zhang, Z., Chong, A., Pan, Y., Zhang, C., Lam, K.P.: Whole building energy model for hvac optimal control: A practical framework based on deep reinforcement learning. Energy and Buildings **199**, 472–490 (2019). https://doi.org/10.1016/j.enbuild.2019.07.029