

Risk-Aware Multi-Level Energy-Saving Control for O-RAN-Based Next-Generation Mobile Networks

Jae-Bin Kwon, Kyung-Seop Lim, Hyun-Min Yoo, Een-Kee Hong
Dept. of Electronics and Information Convergence Engineering
Kyung Hee University, Yongin, Republic of Korea
Emails: {kckwon, daniel9456, yhm1620, ekhong}@khu.ac.kr

Abstract—Radio Access Networks (RANs) constitute a major portion of mobile network energy consumption, making RAN energy saving a critical concern for 5G and 6G systems. Open RAN (O-RAN) enables policy-driven and data-driven control that can overcome conventional static, hardware-specific load-threshold schemes. In this paper, we propose a risk-aware multi-level energy-saving framework for O-RAN. The framework follows the O-RAN Software Community (O-RAN SC) reference design, with its control commands mapped to standardized management-plane configuration operations. The framework anticipates rollback risk and blocks unsafe actions before they are applied. In a 3,600-second simulation, the proposed framework attains a 50.92% energy-saving ratio, comparable to a risk-aware baseline (51.01%), while reducing the rollback rate from 19.48% to 0.06%, achieving a substantially better energy-reliability balance.

Index Terms—O-RAN, network energy saving, Non-RT RIC, SMO, O1 interface, M-Plane.

I. INTRODUCTION

The rapid growth of mobile traffic and base-station densification have made RAN energy consumption one of the most pressing operational concerns for 5G and emerging 6G systems. 3GPP consolidates Network Energy Saving (NES) techniques for NR—cell Discontinuous Transmission (DTX), Discontinuous Reception (DRX), and various sleep modes—and analyzes their interaction [1]. These techniques offer significant energy savings, but only if applied at the right time and in a manner suited to network conditions [2].

Conventional load-threshold-based control selects energy-saving commands from an instantaneous load value, and does not adequately account for operational constraints such as active sessions, fault conditions, and state-transition costs. An aggressive command issued without regard to these constraints may have to be revoked shortly afterwards. Such a rollback cancels the energy gain, incurs additional transition cost and signaling, and degrades quality of service (QoS) during the transition. Energy saving must improve power efficiency without degrading QoS, so a control approach that ignores these operational constraints is inadequate for practical deployment.

Open RAN (O-RAN) can address these limitations by enabling programmable, data-driven control through open interfaces and RAN Intelligent Controllers (RICs) [3], [4]. The Service Management and Orchestration (SMO) framework and the Non-Real-Time RIC (Non-RT RIC) apply network configuration to the O-DU and O-RU through the management-plane

interfaces, using NETCONF on standardized YANG models. These configuration actions are split across two interfaces depending on the target node: O-DU configuration is carried over the O1 interface, whereas O-RU configuration is carried over the Open Fronthaul (FH) M-Plane. This split is reflected in the O-RAN Software Community (O-RAN SC) reference implementation, where the O-DU and O-RU expose O1 and O-FH M-Plane as distinct NETCONF endpoints [5]. This programmable management-plane structure makes it possible to design a closed-loop energy-saving framework that observes network state and applies configuration commands in a coordinated manner.

In this paper, we propose a risk-aware multi-level energy-saving control scheme that explicitly models early-rollback risk and operational constraints together with energy gain and transition overhead. The contributions of this paper are summarized as follows:

- We define three energy-saving control commands (A1 DRX_profile_update over O1, A2 carrier_sleep and A3 RU_energy_saving_enable over the O-FH M-Plane) plus a no-action baseline (A0), each mapped to a YANG branch and an edit-config operation on the management plane.
- We construct a normalized power model in which the power consumption after applying a command is determined by the resulting DRX/carrier/RU state rather than by manually assigned weights.
- We formulate a risk-aware objective combining energy gain, transition overhead, and a history-based rollback probability, paired with an operational-constraint filter that excludes aggressive commands under recent high session counts, active alarms, or minimum-dwell violations.
- An evaluation over a 3,600-second scenario, in which control decisions rely only on past and current observations, shows that the proposed scheme sustains an energy-saving ratio comparable to a risk-aware baseline while substantially reducing the rollback rate and eliminating operational rollbacks.

II. RELATED WORK

Network energy saving for cellular RANs has been studied extensively. 3GPP TR 38.864 catalogs NES techniques and their impact on signaling, delay, and QoS [1]. Early Cell DTX work suppresses transmission during idle intervals [6],

[7], and Advanced Sleep Modes (ASM) manage the energy-delay trade-off via optimal-policy and reinforcement-learning approaches [8], [9]; energy consumption of ML-enhanced Open RAN is reviewed in [10]. These works operate at the cycle or slot level and assume sleep transitions are directly actuatable, which is not the case for many current O-RAN deployments. On the O-RAN side, the Alliance has published use cases for RIC-driven energy saving over A1/O1/E2 [3], and implementation studies have built rApps/xApps adjusting sleep configurations via aggregated indicators or AI/ML inference [11], [12], including multi-vendor interoperability tests [13], DRL-based xApps [14], mobility-aware power optimisation [15], and energy- and QoS-aware load balancing [16]. These efforts confirm feasibility but largely select energy-saving actions from instantaneous indicators, leaving open how to design control when only a small set of YANG-backed configuration changes is exposed and early rollbacks are costly.

A second strand concerns the safety of automated configuration changes: SMO/Non-RT RIC operations are constrained by service-level commitments and operational events such as alarms and hardware-imposed minimum-dwell requirements, under which greedy or threshold-driven controllers tend to oscillate and roll back their decisions. To the best of our knowledge, such risk modeling has not been combined with a management-plane-grounded multi-level command abstraction in the open-RAN energy-saving context. Our work bridges these lines: we model the controllable surface at the YANG level across both O1 and the O-FH M-Plane, then combine a risk-aware selection objective with a hard operational-constraint filter so that energy savings and rollback rate are managed together.

III. METHODOLOGY

A. System Model

1) *Architecture*: We consider a single-cell, single-carrier deployment disaggregated into an O-CU, O-DU, and O-RU. Energy-saving control is hosted as a Non-RT RIC rApp in the SMO. The rApp observes management-plane indicators, decides a command each decision interval, and the SMO is modeled to apply the configuration update through the appropriate interface: DU-side changes over O1 and O-RU-side changes over the O-FH M-Plane, both represented as NETCONF operations on standardized YANG models, following the O-RAN SC reference implementation [5]. The Near-RT RIC and E2 are not used, because the energy-saving actions considered here are persistent configuration changes on managed YANG nodes rather than near-real-time, per-message control. Three observable quantities are exposed each second t : a normalized load $L(t) \in [0, 1]$ from cell-level measurements (PRB utilization, throughput counters), an active-session indicator $C(t)$ modeled after `RRC.ConnMean` and capped by the UE pool size, and a binary fault indicator $F(t)$ set to 1 while an alarm is active. The decision interval is $\Delta = 1$ s and the risk-model horizon is $H = 30$ s; all policies see identical synthetic observations.

TABLE I
CONTROL COMMANDS: INTERFACE/YANG MAPPING, RESULTING STATE, AND OPERATIONAL PARAMETERS

Cmd	Iface	YANG target	DRX	Carr.	RU	Dwell	$\rho_{\max}$
A0	—	no change	NORM	ACT	AWAKE	0 s	1.00
A1	O1	DRX DU profile	LOW	ACT	AWAKE	20 s	0.70
A2	M-Pl.	uplane: SLEEP	NORM	SLEEP	AWAKE	40 s	0.40
A3	M-Pl.	hw: ES=TRUE	NORM	SLEEP	SLEEP	60 s	0.20

Iface: O1 = O1 interface (O-DU); M-Pl. = O-FH M-Plane (O-RU).

2) *Management-Plane-Based Control Commands*: Four commands are considered, summarized in Table I. A0 issues no change. A1 (`DRX_profile_update`) switches the DU to a lower-power DRX profile via the O1-exposed `o-ran_3gpp-nr-nrm-gnbdfunction`. A2 (`carrier_sleep`) sets the `tx-/rx-array-carrier` active-state on the O-FH M-Plane model `o-ran-uplane-conf` to SLEEP, leaving the RU powered. A3 (`RU_energy_saving_enable`) sets `energy-saving-enabled` to TRUE on the O-FH M-Plane model `o-ran-hardware`. Thus A1 is an O1 (DU) action while A2 and A3 are O-FH M-Plane (RU) actions. The state resulting from each command—(`drx_profile`, `carrier_state`, `ru_power_state`)—determines the power used in the energy model, not a per-action weight.

3) *Normalized Power Consumption Model*: The active-state power is load-dependent:

$$P_{\text{active}}(L) = P_{\text{static}} + P_{\text{dyn}} \cdot L, \quad (1)$$

with $P_{\text{static}} = 0.60$ and $P_{\text{dyn}} = 0.40$ in the normalized unit used throughout. Because A1 keeps the cell active and only changes the DRX duty cycle, its power is derived physically from a conservative DRX activity-reduction factor:

$$R_{\text{drx}} = T_{\text{on}}/T_{\text{cycle}}, \quad (2)$$

$$P_{A1}(L) = P_{\text{active}}(L) \cdot (1 - \eta_{\text{drx}} \cdot (1 - R_{\text{drx}})), \quad (3)$$

with $T_{\text{on}} = 4$ ms, $T_{\text{cycle}} = 40$ ms, and $\eta_{\text{drx}} = 0.10$ reflecting the limited reduction when the RU stays powered. In contrast, A2 and A3 place the carrier or the RU into a vendor-defined low-power state whose residual consumption is not derivable from a single duty-cycle parameter; we therefore model them with fixed residual-power ratios:

$$P_{A2}(L) = \rho_{\text{cs}} \cdot P_{\text{active}}(L), \quad P_{A3}(L) = \rho_{\text{ru}} \cdot P_{\text{active}}(L), \quad (4)$$

with $\rho_{\text{cs}} = 0.35$ and $\rho_{\text{ru}} = 0.15$. The energy-saving ratio (ESR) over the horizon is

$$\text{ESR} = \frac{E_{\text{base}} - E_{\text{pol}}}{E_{\text{base}}} \times 100 \quad [\%]. \quad (5)$$

B. Risk-Aware Multi-Level Control

1) *Energy Gain and Transition Overhead*: At each second t , the instantaneous energy gain of command a is the relative power reduction against the active baseline:

$$\Delta E_t(a) = \frac{P_{\text{active}}(L(t)) - P_a(L(t))}{P_{\text{active}}(L(t))}. \quad (6)$$

When the applied command changes, the rApp pays a normalized transition cost $\Omega(s_t, a)$: sleep-transition costs

0.0/0.1/0.3/0.7 for entering A0–A3, wake-up costs 0.0/0.1/0.3/1.0 for returning to A0, and the maximum of the two sleep costs between two non-A0 commands.

2) *Early-Rollback Risk Estimation*: We define early-rollback risk as the probability that the load over the prediction horizon H exceeds the sustainability threshold $\rho_{\max}(a)$, i.e., that the command could not be sustained for its required dwell time. Crucially, the rApp may not look at the actual future load: it forecasts from past observable history only. Using a 60-s trailing window, we compute the recent linear load trend $\hat{\delta}$ over a 10-s window and the recent dispersion $\hat{\sigma}$ (0.03 floor for numerical stability), and form a conservative peak estimate

$$\hat{L}_{\text{peak}} = L(t) + \max(0, \hat{\delta} \cdot H) + k \cdot \hat{\sigma}, \quad (7)$$

$$\text{Risk}_t(a) = \sigma\left(\frac{2 \cdot (\hat{L}_{\text{peak}} - \rho_{\max}(a))}{\hat{\sigma}}\right), \quad (8)$$

where $\sigma(\cdot)$ denotes the logistic sigmoid and k is a fixed forecast-margin coefficient. The estimator approaches 1 when the forecasted peak comfortably exceeds the threshold and 0 when comfortably below; since no element of $\hat{L}_{\text{peak}}$ depends on values after t , the risk term relies only on past and current observations.

3) *Risk-Aware Objective and Selection*: The rApp minimizes a weighted sum:

$$J_t(a) = -\alpha \cdot \Delta E_t(a) + \beta \cdot \Omega(s_t, a) + \gamma \cdot \text{Risk}_t(a), \quad (9)$$

$$a_t^* = \arg \min_{a \in \mathcal{A}_{\text{feasible}}(t)} J_t(a), \quad (10)$$

with $\alpha + \beta + \gamma = 1$; we use $\alpha = 0.40$, $\beta = 0.30$, $\gamma = 0.30$, giving slight priority to energy gain.

4) *Operational-Constraint-Based Action Filtering*: Before the objective is evaluated, a filter prunes $\mathcal{A}_{\text{feasible}}(t)$ using only current or recent observable state: (i) if the maximum session count over a recent 10-s window exceeds θ_{conn} , A3 is blocked, since RU sleep under non-negligible sessions may cause an immediate forced wakeup; (ii) if $F(t) = 1$, A2 and A3 are blocked, since aggressive sleep during an alarm risks mis-handling fault recovery; (iii) if the elapsed dwell of the current command is below its minimum, transitions are blocked. A0 is always retained as a safe fallback. The filter removes unsafe aggressive commands before the optimizer sees them, decoupling hard feasibility from the soft trade-off in J_t .

5) *Overall Control Procedure*: Algorithm 1 summarizes the per-second loop: the filter runs first, J_t is evaluated over feasible commands using the history-based forecast, and the result is committed. A separate event-driven rollback check reverts a committed command to A0 only when an actual violation of its sustainability or operational constraint is observed; the controller never inspects future load, sessions, or alarms.

IV. EXPERIMENT

A. Simulation Setup

We conduct numerical simulations based on a single-cell, single-carrier topology comprising one logical O-CU, one

Algorithm 1 Risk-Aware Multi-Level Energy-Saving Control

Require: recent observations $L_{\text{hist}}(t)$, $C_{\text{hist}}(t)$, $F(t)$; state s_t

Ensure: applied command $a_{\text{applied}}(t)$

- 1: $\mathcal{A}_{\text{feasible}} \leftarrow \text{constraint_filter}(\{A0, \dots, A3\}, C_{\text{hist}}, F(t), s_t)$
- 2: $\hat{L}_{\text{peak}}, \hat{\sigma} \leftarrow \text{forecast_from_history}(L_{\text{hist}})$
- 3: **for** each a in $\mathcal{A}_{\text{feasible}}$ **do**
- 4: compute $\Delta E_t(a)$ by (6), $\Omega(s_t, a)$ from the transition-cost model, and $\text{Risk}_t(a)$ by (7)–(8)
- 5: $J_t(a) \leftarrow -\alpha \cdot \Delta E_t(a) + \beta \cdot \Omega(s_t, a) + \gamma \cdot \text{Risk}_t(a)$
- 6: **end for**
- 7: commit $a_{\text{applied}}(t) \leftarrow \arg \min_{a \in \mathcal{A}_{\text{feasible}}} J_t(a)$ via edit-config (O1 for A1, O-FH M-Plane for A2/A3)
- 8: **if** violation_observed($a_{\text{applied}}, L(t), C(t), F(t)$) **then** revert $a_{\text{applied}} \leftarrow A0$ {observed-violation rollback}

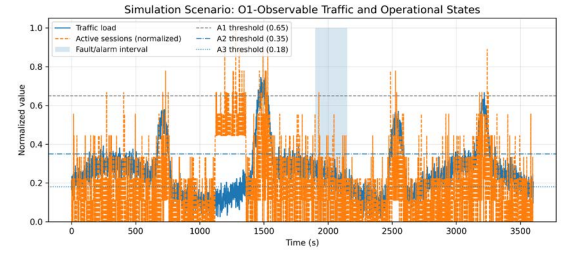


Fig. 1. Simulation scenario: observable load, normalized active sessions, and the fault interval, with the A1/A2/A3 thresholds of the load-threshold baseline.

logical O-DU, one logical O-RU, and a UE pool of 12 over 3,600 s, with a simulation interval of $\Delta = 1$ s and a prediction horizon of $H = 30$ s. Note that the O1-based management procedure is modeled in the simulation according to the O-RAN specifications to emulate the corresponding configuration and control operations.¹ The rApp observes $L(t)$, $C(t)$, and $F(t)$ each second and either keeps the current command or selects a configuration action mapped to the YANG attribute in Table I over the corresponding interface (O1 for the DU action and the O-FH M-Plane for the RU actions). The NETCONF exchange itself is not implemented, and the selected command is applied to the simplified DU/RU state model. Rollback is event-driven—a command is reverted to A0 only when the realized load or operational state violates its constraint—so all policies rely only on past and current observations, without access to future values.

The synthetic load $L(t)$ is a slow sinusoidal baseline $0.20 + 0.10 \cdot \sin(2\pi t/1400)$ plus Gaussian noise ($\sigma = 0.035$) and four sinusoidal bursts centered on (715, 1490, 2535, 3210) s with peak amplitudes (0.35, 0.45, 0.40, 0.30) and 120–180 s durations, clipped to $[0.01, 0.95]$. The session indicator $C(t)$ is Poisson with mean $0.10 + 4.5 \cdot L(t)$, clipped at the pool size, with +4 sessions injected during $[1, 120, 1, 360]$ s; $F(t)$ is 1 during $[1, 900, 2, 150]$ s. The scenario is plotted in Fig. 1 and the model parameters are listed in Table II; the qualitative results are robust to small perturbations of the objective weights.

The power and transition parameters are normalized mod-

¹Evaluation using real O-RAN hardware and executable O-RAN SC software components is left for future work.

TABLE II
SIMULATION AND MODEL PARAMETERS

Parameter	Value	Role
Duration / Δ / H	3600 s / 1 s / 30 s	Horizon, cadence, forecast
UE pool size	12	Cap on active sessions
$P_{\text{static}} / P_{\text{dyn}}$	0.60 / 0.40	Active-state power model
$T_{\text{on}} / T_{\text{cycle}} / \eta_{\text{drx}}$	4 ms / 40 ms / 0.10	DRX model (A1)
$\rho_{\text{cs}} / \rho_{\text{ru}}$	0.35 / 0.15	Sleep power ratios (A2/A3)
Dwell (A1/A2/A3)	20 / 40 / 60 s	Minimum dwell times
ρ_{max} (A1/A2/A3)	0.70 / 0.40 / 0.20	Sustainability thresholds
$\alpha / \beta / \gamma$	0.40 / 0.30 / 0.30	Objective weights
θ_{conn}	1	Active-session limit for A3
Alarm interval	[1900, 2150] s	Fault model
$k / W_{\text{hist}} / \text{trend} / W_c$	1.0 / 60 s / 10 s / 10 s	Forecast and filter windows

eling assumptions selected to represent progressively deeper energy-saving states from DRX-profile adaptation to carrier sleep and RU-level energy saving. The parameter values preserve the corresponding qualitative ordering of energy reduction, transition overhead, and minimum dwell requirements rather than representing measurements from a specific vendor platform.

Four policies are compared on the same scenario: B0 (Always-active) always issues A0 and serves as the energy baseline; B1 (Load-threshold) selects A1/A2/A3 purely on the current $L(t)$ using descending load thresholds 0.65/0.35/0.18, with no risk or constraint reasoning; B2 (Risk-aware only) minimizes $J_t(a)$ without the constraint filter; P (Proposed) minimizes $J_t(a)$ over the filtered set $\mathcal{A}_{\text{feasible}}(t)$. All share the same event-driven rollback stage, and none inspects future observations.

B. Results

1) *Energy-Saving Ratio and Rollback Rate*: Fig. 2(a) shows the ESR over the full 3,600 s: B0 yields 0 by construction, B1 attains 48.34%, B2 51.01%, and the Proposed scheme 50.92%—essentially the same as B2, so the constraint filter costs no meaningful savings (within 0.1 percentage points). Fig. 2(b) reports the rollback rate—the fraction of energy-saving (ES) action attempts revoked after commit—decomposed into traffic-triggered and operational-triggered components (which may co-occur and need not sum to the total). B1 rolls back 19.48% of its attempts, dominated by operational rollbacks (18.06%) from active sessions or alarms, with 2.36% traffic-triggered. B2 reduces the total to 7.56%, of which 7.44% are operational and only a small fraction traffic-triggered, because the load forecast cannot anticipate alarms and session bursts. The Proposed scheme reduces total rollbacks to 0.06% and operational rollbacks to 0%; the remainder is traffic-triggered forecast mis-prediction.

2) *Energy-Reliability Trade-off and Command Composition*: Fig. 3(a) plots the policies in the (rollback rate, ESR) plane, where better policies sit in the upper-left corner. B1 lies far right for a moderate ESR, B2 moves toward the upper-left, and P is closest to the corner: almost the same ESR as B2 (50.92% vs. 51.01%) with a rollback rate lower by more than two orders of magnitude (0.06% vs. 7.56%)—the practical advantage of pairing the risk-aware objective with hard constraints. Fig. 3(b) breaks down the time spent in

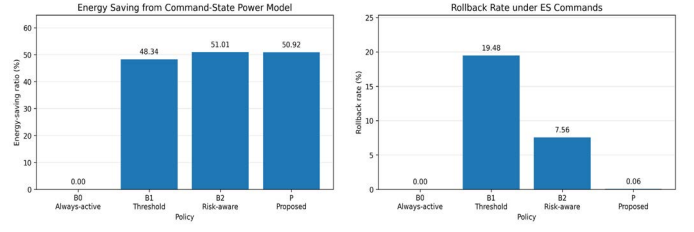


Fig. 2. (a) Energy-saving ratio across policies. (b) Rollback rate with traffic- and operational-triggered components; the Proposed scheme eliminates operational rollbacks, leaving only a 0.06% traffic-triggered rate.

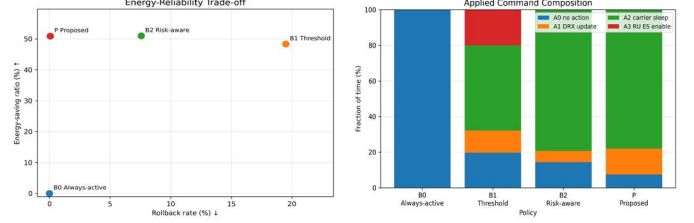


Fig. 3. (a) Energy-reliability trade-off: the Proposed scheme shifts B2 substantially leftward at negligible ESR cost. (b) Distribution of applied commands over the horizon.

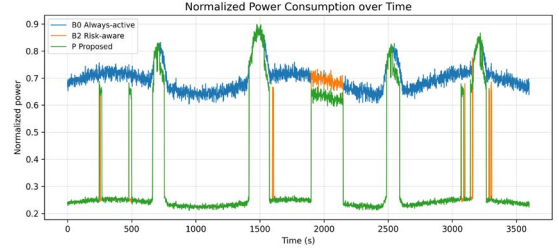


Fig. 4. Normalized power over time for B0, B2, and the Proposed scheme. The spikes in B2 correspond to forced rollbacks at burst and alarm boundaries.

each applied command after the rollback stage. B1 frequently selects A2/A3 but later reverts much of it to A0; B2 increases A0 time as rolled-back attempts degrade to A0; the Proposed scheme is the most balanced, spending substantial time in A1 and A2 while issuing fewer A3 attempts that would have violated the session or fault constraint.

3) *Effect of Filtering and Power Over Time*: To isolate the filter's effect, we measure the unsafe-aggressive-command rate—the fraction of attempted A2/A3 commands violating an operational constraint at the attempt time. B2 exhibits 7.99%, while the Proposed scheme drives it to 0% by construction, with the filter blocking 53.13% of aggressive candidates over the alarm and high-session intervals; this is a supporting view of Fig. 2(b) rather than a standalone claim. Fig. 4 plots the normalized power of three representative policies: B0 closely tracks $P_{\text{active}}(I(t))$ and is highest throughout; B2 reduces power but shows spikes at burst and alarm boundaries, where rolled-back commands bounce between aggressive sleep and A0; the Proposed scheme stays in a low-power band and behaves markedly cleaner across the alarm interval [1,900, 2,150] s, where the filter pre-emptively excluded A2/A3 for the entire window.

V. DISCUSSION

The results reveal a clear pattern across the three energy-saving policies. The load-threshold policy achieves a non-trivial ESR but suffers a high rollback rate, most of which is operational, so the nominal ESR alone overstates its practical quality. The risk-aware objective reduces rollbacks substantially, yet residual operational rollbacks remain because alarms and session bursts are not captured by the load-history forecast. The largest improvement comes from pairing the risk-aware objective with the hard constraint filter, which provides exactly the operational information that the load-only estimator lacks: the proposed scheme retains essentially the same ESR while reducing rollbacks by more than two orders of magnitude and eliminating operational rollbacks. This supports a two-stage design—hard constraints first, soft costs second—for energy-saving control that integrates readily into an O-RAN SC-style deployment.

These results also point to several directions for future work. First, the current scheme uses a simple history-based predictor; integrating an AI/ML model such as an LSTM trained on historical PM data could forecast load more accurately around burst boundaries and further improve energy saving. Second, while the present evaluation considers a single cell and carrier, the framework can be extended to multi-cell deployments by incorporating neighbor-cell handover load and mobility prediction. Finally, the normalized power model can be refined with vendor-calibrated power profiles to report absolute energy figures, building toward validation on the O-RAN SC testbed with real O1 and O-FH M-Plane configuration exchange.

VI. CONCLUSION

This paper proposed a risk-aware multi-level energy-saving control scheme for O-RAN. The scheme defines three configuration commands—DRX profile update over O1, and carrier sleep and RU energy-saving enable over the O-FH M-Plane—each mapped to an edit-config operation, and combines a risk-aware selection objective with an operational-constraint filter that excludes unsafe commands before they are applied. In a 3,600-second simulation, the scheme reduced the rollback rate from 19.48% to 0.06% and eliminated operational rollbacks, while maintaining an energy-saving ratio of 50.92%, comparable to that of the risk-aware baseline. By validating the framework on the O-RAN SC testbed with real NETCONF exchange and extending it with a learned load predictor and multi-cell support, we expect this work to contribute toward practical energy-saving frameworks for O-RAN.

VII. ACKNOWLEDGMENT

This work was supported by the IITP (Institute of Information & Communications Technology Planning & Evaluation) - ITRC (Information Technology Research Center) (IITP-2026-RS-2021-II212046, 50%) grant funded by the Korea government (Ministry of Science and ICT) and in part by IITP grant funded by the Korea government(MSIT) (No.RS-2024-00395824, Development of Cloud virtualized RAN (vRAN) system supporting upper-midband, 50%)

REFERENCES

- [1] 3GPP, "Study on network energy savings for NR," 3rd Generation Partnership Project (3GPP), Tech. Rep. TR 38.864, Mar. 2023, v18.1.0.
- [2] S. Buzzi, C.-L. I, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone, "A survey of energy-efficient techniques for 5G networks and challenges ahead," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 697–709, Apr. 2016.
- [3] O-RAN Working Group 1, "Network energy saving use cases technical report," O-RAN Alliance, Tech. Rep. O-RAN.WG1.Network-Energy-Savings-Technical-Report-R003-v02.00, Jun. 2023.
- [4] M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "Understanding O-RAN: Architecture, interfaces, algorithms, security, and research challenges," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 1376–1411, 2023, 2nd Quart.
- [5] O-RAN Software Community, "sim/o1-ofhmp-interfaces: O1 and Open Fronthaul M-Plane interface simulator (PyNTS)," <https://github.com/o-ran-sc/sim-o1-ofhmp-interfaces>, 2026, accessed: 2026-06-14.
- [6] S. Tombaz, S.-W. Han, K. W. Sung, and J. Zander, "Energy efficient network deployment with cell DTX," *IEEE Communications Letters*, vol. 18, no. 6, pp. 977–980, Jun. 2014.
- [7] Q. Wang, F. Zhao, and T. Chen, "A base station DTX scheme for OFDMA cellular networks powered by the smart grid," *IEEE Access*, vol. 6, pp. 63 442–63 451, 2018.
- [8] F. E. Salem, T. Chahed, E. Altman, A. Gati, and Z. Altman, "Optimal policies of advanced sleep modes for energy-efficient 5G networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2019.
- [9] D. Renga, Z. Umar, and M. Meo, "Trading off delay and energy saving through advanced sleep modes in 5G RANs," *IEEE Transactions on Wireless Communications*, vol. 22, no. 11, pp. 7172–7185, Nov. 2023.
- [10] X. Liang, Q. Wang, A. Al-Tahmeesschi, S. B. Chetty, D. Grace, and H. Ahmadi, "Energy consumption of machine learning enhanced Open RAN: A comprehensive review," *IEEE Access*, vol. 12, pp. 81 889–81 909, 2024.
- [11] M. Kim, K. S. Lee, S. Jung, J.-H. Na, S. D'Oro, L. Bonati, and T. Melodia, "An open RAN development framework with network energy saving rApp implementation," in *Proc. EuCNC/6G Summit*, 2025.
- [12] D. Roh and J.-H. Na, "Performance analysis of different cell DRX/DTX configurations for energy saving in O-RAN system," in *Proc. 15th Int. Conf. Inf. Commun. Technol. Convergence (ICTC)*, Jeju Island, South Korea, Oct. 2024.
- [13] A. Akman, P. Oliver, M. Jones, P. Tehrani, M. Hoffmann, and J. Li, "Energy saving and traffic steering use case and testing by O-RAN RIC xApp/rApp multi-vendor interoperability," in *Proc. IEEE Future Netw. World Forum (FNWF)*, 2023.
- [14] Q. Wang, S. Chetty, A. Al-Tahmeesschi, X. Liang, Y. Chu, and H. Ahmadi, "Energy saving in 6G O-RAN using DQN-based xApp," in *Proc. IEEE 29th Int. Workshop Comput. Aided Model. Des. Commun. Links Netw. (CAMAD)*, 2024.
- [15] M. M. H. Qazzaz, A. Salama, M. Hafeez, and S. A. R. Zaidi, "ORAN-MAP: A hybrid approach to mobility-aware power optimisation in Open Radio Access Networks (ORAN)," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Wkshps)*, 2024.
- [16] G. Z. Bruno, G. M. Almeida, A. Da Silva, L. A. DaSilva, J. F. Santos, A. Huff, K. V. Cardoso, and C. B. Both, "Towards energy- and QoS-aware load balancing for 6G: Leveraging O-RAN to achieve sustainable and energy-efficient 6G," *IEEE Transactions on Network and Service Management*, 2024, to be published.