Mobile Network Control with a World Model

Maxime Bouton, Ioanna Mitsioni, Jaeseong Jeong, Alessandro Previti
Ericsson Research, Stockholm

***Abstract*—The increasing complexity of mobile networks necessitates intelligent and dynamic control strategies for efficient and energy-conserving management. We propose a world model-based approach for network control, trained on pre-collected network data. This method enables adaptive configuration of crucial parameters without manual tuning. We demonstrate its effectiveness by evaluating predictions on real data and simulating closed-loop control of a 5G network energy-saving feature. Our results show improved performance in balancing energy savings with quality of service compared to traditional methods and reinforcement learning approaches.**

I. Introduction

Rapid growth in the usage, density, and complexity of mobile networks poses significant challenges to operators in terms of management and energy efficiency. With rising energy costs and sustainability goals, optimizing network configuration has become crucial. Automated network configuration solutions have the potential to enhance quality of service, reduce manual operations and operational costs, and save energy by intelligently managing resources based on varying demands. However, optimal control of the configuration parameters is challenging due to hard-to-model factors such as network usage and radio propagation conditions, and the scale of networks that involve thousands of base stations (BSs). This work focuses on the example of controlling the sleep state of capacity carriers, a key aspect of energy management in mobile networks that requires balancing hardware deactivation for energy savings with maintaining sufficient quality of service.

Previous works have addressed this problem through hand-engineered strategies, optimization algorithms, and reinforcement learning (RL) for specific parameters [1]. Hand-engineered strategies consist of rule-based methods based on expert knowledge of the system and often lack adaptability to site-specific information. Optimization methods with mathematical models of the network and user behavior often resort to simplifying assumptions to reduce computational costs. RL has been shown to be promising in antenna tilt configuration [2] and energy efficiency optimization [3], [4]. However, it requires extensive training using network simulators or offline data and often needs to be retrained when changing objectives.

Our approach employs model-based planning by learning a world model (WM) from available 5G mobile network data. Data-driven world WMs have been extensively studied in model-based RL to increase sample efficiency and long-term planning in robotics and game environments [5], [6]. Their direct application to existing network data however, can lead to models insensitive to actions since exploratory, diverse actions might be unsafe for the live network. Here, we propose a model architecture that mitigates data coverage issues by better feature modeling. The resulting action-sensitive model can be used with an online planner for energy-saving applications. In contrast to RL, online planning with a WM allows a network operator to change the cost function without retraining. We evaluate the prediction capabilities of our model on real data from a 5G network and then demonstrate its closed loop performance when used with a cross entropy planner in a simulation environment. We show that the planning method outperforms RL alternatives trained in simulation and a hand-engineered baseline. These initial results establish a first step towards expanding the WM-based control paradigm in network automation, improving how modern networks are managed for optimal service quality and energy efficiency.

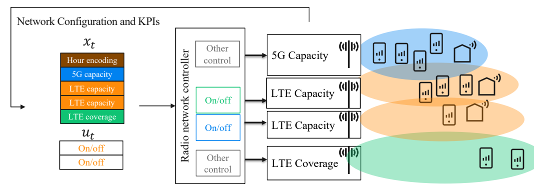

Fig. 1. Illustration of our capacity cell control problem. Mobile networks is composed of different frequency band covering the same area, we control the on/off state of one or more LTE frequency band.

II. Proposed Approach

Mobile networks comprise multiple BSs, each equipped with antennas operating on various carrier frequencies to serve the same sectors, as seen in Fig. 1. This redundancy ensures sufficient bandwidth during high-demand periods. During low-demand periods, disabling redundant frequencies and their hardware reduces energy consumption. We focus on optimizing the on/off control of these frequencies, which can be managed by a centralized system at up to 15-minute intervals, with faster local systems handling unforeseen traffic spikes.

We represent the network as a dynamical system and formulate the problem of saving energy while maintaining service requirements as a model predictive controller (MPC):

$$\min_{u_t,\dots,u_{t+H-1}} J = \sum_{i=0}^{H-1} u_{t+i} \tag{1}$$

$$\text{s.t. } x_{t+i+1} = f(x_{t+i}, u_{t+i}, \mathcal{H}_{t+i}), i = 0 \dots H-1 \tag{2}$$

$$g(x_{t+i}, u_{t+i}) \leq 0, \quad i = 0 \dots H-1, \tag{3}$$

where $u_t \in \{0,1\}^n$ represents the on/off state of n carriers, x_t is the predicted state of the network, g represents a constraint function, f is the dynamics model, and $\mathcal{H}_t$ is a history of the past state and actions up to t.

The network state x_t comprises key performance indicators for each frequency band, including resource utilization percentage, aggregated user throughput, and traffic volume. We implement a simple service requirement constraint $g(x_t, u_t) = 1_{[u_t=1]} 1_{[x_t[\text{throughput LTE Coverage}] \leq 50\,\text{Mbps}]}$, ensuring that the average user throughput in coverage frequencies remains above a threshold.

Training the WM: The WM is trained to predict the future states of the network given different network configuration parameters. It is trained from an existing dataset using teacher forcing. Standard WM training uses regression with a mean squared error loss [5], however, due to the discontinuous nature of the on/off action we developed a multi-task sequence-to-sequence architecture handling continuous, binary, and semi-continuous features explicitly.

Given input sequence $x_1, ..., x_T$ and action sequence $u_1, ..., u_{T+H-1}$, where T is the input history length and H is the prediction horizon, the model predicts $\hat{\mathbf{Y}} = \hat{x}_2, ..., \hat{x}_{T+H}$, with auto-regressive predictions for steps $T+2$ to $T+H$. States and actions are embedded into a common latent space and processed by a sequence-to-sequence backbone (e.g., GPT2 transformer or LSTM). Task-specific output heads produce predictions for each feature type, with semi-continuous predictions conditioned on binary activations.

The model is trained end-to-end using teacher forcing to optimize a weighted multi-task loss function:

$$\mathcal{L} = \sum_{t=1}^{T+H} \left[\lambda_c \mathcal{L}_{\text{c}}(x_t, \hat{x}_t) + \lambda_b \mathcal{L}_{\text{b}}(x_t, \hat{x}_t) + \lambda_s \mathcal{L}_{\text{sc}}(x_t, \hat{x}_t)\right] \quad (4)$$

where $\mathcal{L}_{\text{c}}$, $\mathcal{L}_{\text{b}}$, and $\mathcal{L}_{\text{sc}}$ represent the losses for continuous, binary, and semi-continuous features, respectively, and λ_c, λ_b, and λ_s are the corresponding weight factors. We found that this loss function and separate action embeddings are key for ensuring control input sensitivity and correct predictions of the discontinuous features. The WM can then be used for planning. We solve Eq. (3) using the trained model to represent f with a cross entropy planner [5] and a scalar cost function that rewards sleep and penalizes constraints violations.

III. Results

We evaluate the WM's capabilities to predict over extended horizons, and show in Section III that the multi-task loss approach provides better accuracy compared to standard regression models. To assess the model's ability to predict counterfactual actions, we replay network usage and KPI data with modified actions proposed by the cross-entropy planner. The multi-task model correctly predicts the action effect, predicting zero utilization when the carrier is turned off, while regression models fail to account for these unseen actions.

Closed-loop performance is evaluated in simulation through 4-day episodes for 5 networks with varying traffic patterns. We compare our approach with a rule-based baseline (turning off during the night according to fixed rules) and a soft actor-critic (SAC) agent in Table I. The return represents a scalar combination of J and g. The cross-entropy method (CEM) with our WM outperforms the baseline, achieving more sleep time and fewer constraint violations. While the RL agent shows competitive performance, it requires more extensive reward engineering and simulation data for training.

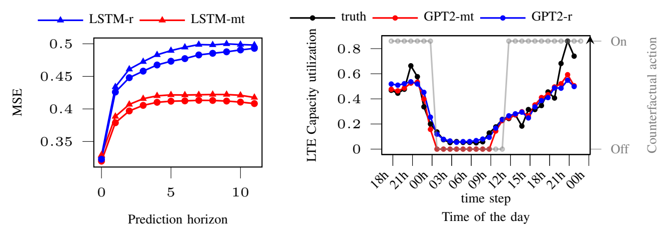

Fig. 2. **Left:** Difference in MSE on a test dataset between the regression models and multi-task models for different sequence to sequence back-bones as the prediction horizon increases. **Tight:** Predictions of the world model when providing counterfactual actions.

TABLE I
Closed loop performance in simulation

policy	sleep per day (h)	constraint violation (%)	return
baseline	3.12	2.19	45.7
CEM	7.25	0.33	57.4
SAC	5.90	1.70	51.1

IV. Conclusion

In this abstract, we present initial steps towards controlling mobile network parameters by planning with a world model. We show that we can train controllable models from real data even with limited action coverage and demonstrate that they can be used with planning algorithms such as CEM. Future work will consist of evaluating more complex cost functions that require long-term planning. We will iterate on the model architecture to extend the control actions and observability of the existing model, as well as add uncertainty modeling.

References

[1] F. Salahdine, J. Opadere, Q. Liu, T. Han, N. Zhang, and S. Wu, "A survey on sleep mode techniques for ultra-dense networks in 5g and beyond," *Computer Networks*, vol. 201, p. 108 567, 2021, ISSN: 1389-1286. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389128621004801.

[2] A. Mendo, J. Outes-Carnero, Y. Ng-Molina, and J. Ramiro-Moreno, "Multi-agent reinforcement learning with common policy for antenna tilt optimization," *IAENG International Journal of Computer Science*, 2023.

[3] J. Ye and Y.-J. A. Zhang, "Drag: Deep reinforcement learning based base station activation in heterogeneous networks," *IEEE Transactions on Mobile Computing*, vol. 19, no. 9, pp. 2076–2087, 2019.

[4] "Telenor: Autonomous AI agent with Ericsson," *GSMA use case library*, 2025. [Online]. Available: https://www.gsma.com/get-involved/gsma-foundry/gsma_resources/telenor-autonomous-ai-agent-with-ericsson/.

[5] K. Chua, R. Calandra, R. McAllister, and S. Levine, "Deep reinforcement learning in a handful of trials using probabilistic dynamics models," *Advances in neural information processing systems*, vol. 31, 2018.

[6] D. Hafner, J. Pasukonis, J. Ba, and T. Lillicrap, "Mastering diverse domains through world models," *arXiv preprint arXiv:2301.04104*, 2023.