Digital-Twin-Enabled Channel Access and Power Control for Smart Grids in Communication Networks

Qihao Li*, Qiang Ye†, Fengye Hu*
*College of Communication Engineering, Jilin Univeristy, Changchun 130021, China
†Department of Electrical and Software Engineering, University of Calgary, Calgary, T2N 1N4, Canada
Email: {qihaol, hufy}@jlu.edu.cn, qiang.ye@ucalgary.ca;

2025 IEEE 101st Vehicular Technology Conference (VTC2025-Spring) | 979-8-3315-3147-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/VTC2025-SPRING65109.2025.11174650

***Abstract*—In this paper, we propose a novel channel access and power control scheme for smart grids in communication networks. The scheme is named digital twin-based memory recall optimization (DMRO), which aims to extract meaningful patterns from noisy network traffic measurements and support more sophisticated decision-making processes for optimizing channel access and power control for smart grids. Specifically, we design a pattern extraction method that minimizes the Frobnius norm between the collected measurements and the expected k-rank approximation of the measurements in order to extract useful information. Then, considering the interference and signal-to-interference-plus-noise ratio (SINR) constraints in the wireless environment, we develop a digital twin-based distributed channel access and power control scheme to improve the latency taming and energy utilization efficiency of the phasor measurements units (PMU). We consider both the real-time traffic prediction and the paired optimization scheme on the digital twin side, and utilize memory recall to enhance local model robustness to optimize from a more diverse set of situations by replaying underrepresented experiences. Simulation results demonstrate that the proposed DMRO scheme can achieve high traffic prediction accuracy and improve the latency taming and energy utilization efficiency even increasing industrial channel interference or the number of PMUs.**

I. Introduction

The smart grid, offering advanced electricity services, has evolved significantly due to intelligent monitoring devices and real-time communication [1], [2]. Central to this system are phasor measurement units (PMUs), which provide precise, real-time data critical for monitoring distributed energy resources (DERs) like solar and wind power [2]. However, the heterogeneous nature of smart grids, involving diverse communication technologies and network conditions, presents challenges in efficient channel access and power control, directly impacting overall grid performance. Efficiently managing channel access and power control for PMUs in smart grids requires balancing the delay in queuing and the conservation of energy. Improving energy efficiency is essential for the long-term sustainability of grid operations, while reducing latency is essential to guarantee quick decision-making. At this point, the advantages of utilising a digital twin (DT) become apparent [3]. As a virtual replica of the physical PMU, the DT enables precise simulations and predictions, guiding the optimization process and facilitating informed decision-making [4]. The DT can be utilized to serve as a advanced tool to explore and resolve the tradeoff between queuing latency and energy efficiency, enhancing overall system performance.

However, there are several challenges in the way of achieving optimum data transmission latency and energy efficiency performance. One of the primary challenges lies in the difficulty of predicting data traffic patterns, which is essential for informed decision-making in control processes. The variable nature of distributed energy resource outputs, the stochastic behavior of consumer energy demand, and the intricate interplay of network components contribute to this unpredictability [5], [6]. Data collected from PMUs often contain a degree of noise and ambiguity that can obscure critical patterns necessary for forecasting [7]. Conventional predictive methods may become less effective when faced with high-dimensional and non-linear data characteristics typical of smart grid environments [8]. More advanced approaches are required to navigate through the data's intricacies and to ensure that the digital twin's predictive model remains robust and dependable.

Second, another major challenge is the integration of the DT within the optimization process, considering its potential to simulate the tradeoffs between queuing latency and energy efficiency against the backdrop of channel interference and data traffic predictions in a multi queue competition system [2], [9]. Therefore, this calls for advanced Markov decision process (MDP) solutions that can integrate the different aspects of smart grid operations, particularly the queuing latency boundaries in multi-queue systems, into the decision-making capabilities of the DT.

Third, the problem of overestimation presents a significant challenge in the pursuit of optimal control strategies. The tendency to select actions based on their maximum estimated reward values, even when they may not be the most advantageous, accentuates the difficulty in determining the "best" control actions [3], [4], [10]. Overcoming this issue necessitates the adoption of advanced technologies that utilize parallel learning between the DT and PMU.

979-8-3315-3147-8/25/$31.00 ©2025 IEEE

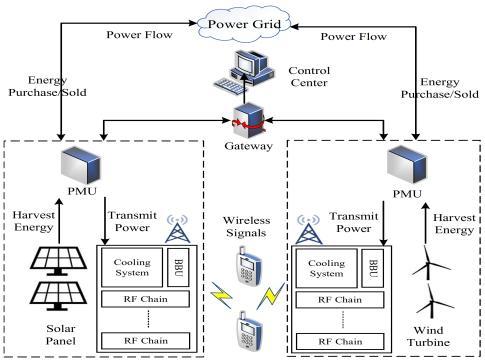


Fig. 1: Network scenario.

In this paper, we primarily address how to predict network traffic while considering the noise and interference and improve decision-making processes for optimizing channel access and power control for smart grids network. We propose a digital twin-based memory recall optimization (DMRO) scheme for taming transmission latency and increasing energy utilization efficiency. The following are the paper's primary contributions.

- We minimize the Frobenius norm between the collected PMU data and its expected k-rank approximation to improve the clarity of information extracted from the original data. The proposed prediction method can be used to refine the useful information extracted from noisy network traffic measurements.
- Considering the multi-queue single-server framework, we develop an advanced MDP model that leverages predictive traffic analysis, channel interference, and dynamically channel access adjustment, to optimize the trade-off between maintaining low-latency data transmission and enhancing energy efficiency. The investigated dual-objective optimization can be employed to ensure PMU data gathering efficiency and energy conservation with the consideration of network conditions and power distribution.
- We develop a DT model that synergies with the physical PMU to mitigate overestimation, employing memory recall to strengthen the robustness of the local model. By revisiting and reevaluating experiences that are typically underrepresented in the MDP model, the digital twin model can achieve a level of optimization that encompasses a broader spectrum of potential scenarios.

II. System Model

A. Network Model

As shown in Fig. 1, we consider a smart grid network configured for smart grid networks, featuring two singular macrocell base stations (MBS). Their primary function is to facilitate traffic offloading, thereby optimizing service rates within the network. They are situated near two types of energy-generating installations, like Photovoltaic (PV) plants or wind turbines, each connected to a PMU. The PMUs play a pivotal role in harvesting data pertinent to the operational status of these energy resources. The PMUs are critical for real-time monitoring and control of the power grid, ensuring stability and efficiency by measuring the electrical waves on an electricity grid. Some mobile users (MUEs) are connected to the MBSs as well.

A DT of these PMUs is employed to optimize network performance and make proper control decision process. The DT is a virtual replica of the PMUs in the network and is connected to all the MBSs. It serves the purpose of predicting network traffic and schedules the optimal channel access and energy control decision for the PMUs in the heterogeneous network. The smart metering packets are transmitted from the PMUs, through the digital-twin, and then toward the controller channelled through the gateway within the core network. Leveraging observed traffic measurements and other voltage and current phasor measurements, the DT employs a novel scheme known as DT-based memory recall optimization (DMRO) scheme. Time-division access is used in the smart grid networks, where each time slot, K sub-channels are allocated by the $\mathcal{M} = \{1, \ldots, M\}$ MBS to $\mathcal{U} = \{1, \ldots, U\}$ PMUs, with one sub-channel per PMU. The same K sub-channels are shared among these PMUs. The $\mathcal{E} = \{1, \ldots, E\}$ MUEs independently access the sub-channels, monitored by the MBS. Each PMU experiences interference from MUE-MBS communications on the same sub-channel and additive white Gaussian noise (AWGN).

B. Traffic Prediction Model

Denote by spatiotemporal sequence $\mathbf{a}_t^u$ the data rate of PMUs over a series of time intervals $t = 1, 2, \ldots, T$. Each traffic vector $\mathbf{a}_t^u$ encapsulates data for a specific time interval within a defined geographic region and is comprised of elements $[a_t^{(1,u)}, \ldots, a_t^{(M,u)}]$. To elucidate patterns from the traffic sequence $\mathbf{a}_t^u$, we apply matrix factorization, expressing $\mathbf{a}_t^u$ as the product $\mathbf{B}_t \mathbf{\Sigma}_t (\mathbf{C}_t)^\top$, where $\mathbf{B}_t$ and $\mathbf{C}_t$ are orthogonal matrices of dimensions $\mathbb{R}^{T \times T}$ and $\mathbb{R}^{(L \times W) \times (L \times W)}$, respectively, and $\mathbf{\Sigma}_t$ is a diagonal matrix encompassing the singular values of $\mathbf{a}_t^u$. This decomposition helps isolate the key components of variability within the data. We focus on the most significant patterns by selecting the leading k singular values and their associated vectors, obtaining a reduced approximation $\mathbf{a}_{t,k}^u = \mathbf{B}_{t,k} \mathbf{\Sigma}_{t,k} (\mathbf{C}_{t,k})^\top$. Here, $\mathbf{B}_{t,k}$ and $\mathbf{C}_{t,k}$ contain the primary k columns of $\mathbf{B}_t$ and $\mathbf{C}_t$, and $\mathbf{\Sigma}_{t,k}$ includes the principal k singular values. This approximation preserves the essence of the cellular traffic data in a lower-dimensional form.

We introduce kernel matrix $\mathbf{K}_t(\sigma_t)$ to map the data into a higher-dimensional feature space where linear patterns can better capture the underlying dynamics of the data traffic. The kernel matrix $\mathbf{K}_t(\sigma_t)$ can be denoted as a function of time-dependent parameters and the network environment:

$$\mathbf{K}_t(\sigma_t) = \exp\left(-\frac{\|\mathbf{a}_t^u - \mathbf{a}_{t'}^u\|^2 + \lambda_1 \|\mathbf{I}_t - \mathbf{I}_{t'}\|^2}{2\sigma_t^2}\right) \quad (1)$$

where $\mathbf{I}_t$ is the data interference matrix, σ_t is the parameter controlling the smoothness of the kernel, and λ_1 is scaling factor for the influence of the interference, such that $\mathbf{K}_t(\sigma_t)$ can capture the spatial and temporal dependencies in the traffic patterns while adapting to changes in network conditions. Randomly select m representative samples ($m \ll M$) from the dataset. Approximate $\mathbf{K}_t(\sigma_t)$ as $\mathbf{K}_t \approx \mathbf{K}_t^{M,m}(\mathbf{K}_t^{m,m})^{-1}(\mathbf{K}_t^{M,M})^\top$, where Denote by $\mathbf{K}_t^{m,m} \in \mathbb{R}^{m\times m}$ the matrix for the selected m samples, $\mathbf{K}_t^{M,m} \in \mathbb{R}^{M\times m}$ the matrix between all data points and the selected m samples, and $\mathbf{K}_t^{M,M} \in \mathbb{R}^{M\times M}$ the full kernel matrix. Perform SVD on $\mathbf{K}_t^{m,m}$ and select the top k eigenvalues and their corresponding eigenvectors to obtain $\mathbf{K}_{t,k} \approx \mathbf{K}_t^{M,m}\mathbf{U}_{t,k}^{\mathbf{K}}(\mathbf{\Lambda}_{t,k}^{\mathbf{K}})^{-1}(\mathbf{U}_{t,k}^{\mathbf{K}})^\top(\mathbf{K}_t^{M,M})^\top$, where $\mathbf{U}_{t,k}^{\mathbf{K}}$ is the matrix of eigenvectors and $\mathbf{\Lambda}_{t,k}^{\mathbf{K}}$ is the diagonal matrix of eigenvalues. Then, an attention mechanism is incorporated to selectively focus on the current state of interest matrix $\mathbf{Q}_t$, the historical information matrix $\mathbf{H}_t$ and the information content $\mathbf{V}_t$ associated with $\mathbf{H}_t$. Suppose $\mathbf{Q}_t = \mathbf{W}_q \cdot [\mathbf{a}_t^u; \mathbf{I}_t]$, $\mathbf{H}_t = \mathbf{W}_h \cdot [\ddot{\mathbf{a}}_t^u; \mathbf{I}_t]$, $\mathbf{V}_t = \mathbf{W}_v \cdot [\ddot{\mathbf{a}}_t^u; \mathbf{I}_t]$, where $\mathbf{W}_q$, $\mathbf{W}_h$, and $\mathbf{W}_v$ are the related learned weight matrix of each matrix, $\ddot{\mathbf{a}}_t^u$ is the historical traffic data. Thus, the attention mechanism can be computed as $\mathrm{Am}(\mathbf{Q}_t, \mathbf{H}_t, \mathbf{V}_t) = \mathrm{soft}(\frac{\mathbf{Q}_t\mathbf{H}_t^\top}{\sqrt{d_k}}) \cdot \mathbf{V}_t$, where $\mathbf{Q}_t\mathbf{H}_t^\top$ computes the similarity, d_k is the dimensionality of $\mathbf{H}_t$, used for scaling to stabilize gradients, soft ensures the similarity scores sum to 1 , converting them into attention weights. Approximate $\mathbf{Q}_t\mathbf{H}_t^\top$ using k-rank SVD method, which leaves $\mathbf{Q}_t\mathbf{H}_t^\top \approx \mathbf{U}_{t,k}^{\mathrm{Am}}\mathbf{\Sigma}_{t,k}^{\mathrm{Am}}(\mathbf{R}_{t,k}^{\mathrm{Am}})^\top$, which reduces the computational complexity of $\mathbf{Q}_t\mathbf{K}_t^\top$ from $O\left(n^2 d_k\right)$ to $O\left(nkd_k\right)$. Thus, the k-ranked approximation $\mathbf{a}_{t,k}^u$ can be further given as

$$\mathbf{a}_{t,k}^u \approx \left[\mathrm{soft}\left(\frac{\mathbf{U}_{t,k}^{\mathrm{Am}}\mathbf{\Sigma}_{t,k}^{\mathrm{Am}}(\mathbf{R}_{t,k}^{\mathrm{Am}})^\top}{\sqrt{d_k}}\right)\mathbf{V}_t\mathbf{K}_{t,k}\right]\mathbf{B}_{t,k}\mathbf{\Sigma}_{t,k}(\mathbf{C}_{t,k})^\top \quad (2)$$

Optimizing the matrix approximation involves minimizing the Frobenius norm of the discrepancy $\mathbf{E}_t$, which quantifies the error between $\mathbf{a}_t^u$ and its k rank approximation. The objective is to approximate $\mathbf{a}_t^u$ by $\mathbf{a}_{t,k}^u$, articulated mathematically as: $\min_{\mathbf{B}_{t,k},\mathbf{\Sigma}_{t,k},\mathbf{C}_{t,k}} \|\mathbf{E}_t\|^2 = \min_{\mathbf{B}_{t,k},\Sigma_{t,k},\mathbf{C}_{t,k}} \|\mathbf{a}_t^u - \mathbf{a}_{t,k}^u\|^2$.

Then, we employ adaptive regression with Long Short-Term Memory (LSTM) networks, renowned for their efficacy in capturing the complexities of sequential data. We utilize a widely adopted LSTM architecture featuring hyperbolic tangent activation functions without peep-hole connections. The model comprises a single hidden layer based on this LSTM framework, paired with an output layer that also employs a hyperbolic tangent function, establishing a streamlined yet robust modeling approach. Please refer to [10] for more details.

C. Multi Queuing Model

In each time slot, every PMU generates data segmented into equal-sized packets. The number of packets produced by PMU u for MBS $\{1,\ldots,M\}$ in time slot t is denoted as $\mathbf{b}_t^u$, occurring at a rate λ. At each slot, MBS m is assigned to a connected PMU queue based on connectivity information and the queue lengths, represented by the vector $\mathbf{e}_{m,t}^u = \left[(\delta_{m,j})_{j\in\mathcal{M}}\right]_{m\in\mathcal{M}}$, where $\delta_{m,j} = 1$ if $m = j$, and $\delta_{m,j} = 0$ otherwise. Initially, this data is placed in a queue and is scheduled to be transmitted in the subsequent time slot, adhering to a first-in-first-out (FIFO) policy. It is assumed that the buffer capacity is sufficiently large to prevent data loss due to overflow. The queue length for PMU i at time slot $k+1$ is given by $q_{t+1}^u = q_t^u - \mathbf{a}_{t,k}^u\mathbf{e}_{m,t}^u + \mathbf{b}_t^u\mathbf{e}_{m,t}^u$, where q_t^u signifies the queue length of PMU u at time slot t.

Suppose that for this queueing system characterized by a Discrete Time Markov Chain (DTMC) with a state vector y_t^u within a countable state space H, which itself is a subset of the Cartesian product of a countably infinite state space H_X and a finite state space H_K, stability can be asserted under certain conditions. Specifically, if one can identify a Lyapunov function $L(q_t^u)$, mapping from $\mathbf{N}^N$ to $\mathbb{R}$, that is lower bounded and satisfies two crucial conditions, then the system exhibits a desired stability behavior. For a queueing system modeled by a discrete time Markov chain (DTMC) with state vector y_t^u in a countable state space H, stability can be ensured under specific conditions. A Lyapunov function $L(q_t^u)$, mapping from $\mathbf{N}^N$ to $\mathbb{R}$, is used to establish stability, provided it meets the following criteria: 1). the expected value of the Lyapunov function for the next state remains finite across all states, $E\left[L\left(q_{t+1}^u\right) \mid y_t^u\right] < \infty$, for all y_t^u; 2). for states where $\|q_t^u\| > B$, the expected change in the Lyapunov function is strictly negative $E\left[L\left(q_{t+1}^u\right) - L\left(q_t^u\right) \mid y_t^u\right] < -\epsilon$, where $\epsilon > 0$. When these conditions are satisfied, the system ensures positive recurrence of all states in the DTMC.

Let H_B denote the set of states y_t^u where the norm of q_t^u does not exceed B, a scenario where the third condition remains inapplicable. Demonstrating that H_B forms a compact set is straightforward. For states not encompassed by this compact set, the third condition is applicable, namely,

$$E\left[L\left(q_{t+1}^u\right) - L\left(q_t^u\right) \mid y_t^u\right] < -\epsilon\left\|q_t^u\right\|. \quad (3)$$

For all instances of y_t^u outside H_B, the averaged form of the aforementioned inequality yields:

$$\begin{aligned} &E\left[L\left(q_{t+1}^u\right) \mid y_t^u \notin H_B\right] - E\left[L\left(q_t^u\right) \mid y_t^u \notin H_B\right] \\ &< -\epsilon E\left[\left\|q_t^u\right\| \mid y_t^u \notin H_B\right]. \end{aligned} \quad (4)$$

Conversely, for y_t^u within H_B, a compact set, it follows that:

$$E\left[L\left(q_{t+1}^u\right) \mid y_t^u \in H_B\right] \le M < \infty, \quad (5)$$

where M represents the supremum of $E\left[L\left(q_{t+1}^u\right) \mid y_t^u\right]$ for y_t^u within H_B. Integrating both preceding expressions, we deduce:

$$\begin{aligned} &E\left[L\left(q_{t+1}^u\right)\right] \\ =&E\left[L\left(q_{t+1}^u\right) \mid y_t^u \in H_B\right] P\left\{y_t^u \in H_B\right\} \\ &+ E\left[L\left(q_{t+1}^u\right) \mid y_t^u \notin H_B\right] P\left\{y_t^u \notin H_B\right\} \\ <&MP\{y_t^u \in H_B\} + P\{y_t^u \notin H_B\}\{E\left[L\left(q_t^u\right) \mid y_t^u \notin H_B\right] \\ &- \epsilon E\left[\left\|q_t^u\right\| \mid y_t^u \notin H_B\right]\} \\ <&M + E\left[L\left(q_t^u\right)\right] - \epsilon E\left[\left\|q_t^u\right\|\right] + M_0 \end{aligned}$$

with M_0 being a constant satisfying:

$$M_0 > \{-E\left[L\left(q_t^u\right) \mid y_t^u \in H_B\right] + \epsilon E\left[\|q_t^u\| \mid y_t^u \in H_B\right]\} P\left\{y_t^u \in H_B\right\} \quad (6)$$

Thus, for any N_0, it follows that:

$$\frac{\epsilon}{N_0} \sum_{t=0}^{N_0-1} E\left[\|q_t^u\|\right] < M + \frac{1}{N_0} E\left[L\left(q_0^u\right) - L\left(q_{N_0}^u\right)\right] + M_0.$$

Assuming $E\left[L\left(q_{N_0}^u\right)\right] > K_0$ allows for the assertion that:

$$\frac{\epsilon}{N_0} \sum_{t=0}^{N_0-1} E\left[\|q_t^u\|\right] < M + \frac{1}{N_0} E\left[L\left(q_0^u\right)\right] - \frac{K_0}{N_0} + M_0. \quad (7)$$

As N_0 approaches infinity, with both $E\left[L\left(q_0^u\right)\right]$ and K_0 finite, we have:

$$\frac{\epsilon}{N_0} \sum_{t=0}^{N_0-1} E\left[\|q_t^u\|\right] < M + M_0 \quad (8)$$

and get the the limit behavior of the average queuing length $\lim_{t\to\infty} E\left[\|q_t^u\|\right] \leq \frac{1}{\epsilon}\left(M + M_0\right)$.

D. Latency and Energy Efficiency

Given the constraint that the average queue length is less than $\frac{1}{\epsilon}(M + M_0)$, the average throughput $\bar{H}_u$ measures queue efficiency. According to Little's Law, the average queue length equals the product of throughput and latency. Thus, the average latency Delay_u for PMU u is $\mathrm{Delay}_u = \frac{M+M_0}{\epsilon \bar{H}_u}$. Here, $\frac{1}{T}\bar{H}_u = \sum_{i=1}^{T} \mathbf{a}_{t,k}^u \mathbf{e}_{m,t}^u$ defines throughput, limited by queue length or rate $\mathbf{a}_{t,k}^u \mathbf{e}_{m,t}^u$ based on channel conditions. The total power consumed by PMU u at time t, denoted as $\mathrm{PT}_{u,t}$, is the sum of circuit power P_c and static power P_s. The circuit power is given by $\mathrm{P}_c = \mathrm{P}_s + \beta_1 \mathbf{a}_{t,k}^u \mathbf{e}_{m,t}^u$, where β_1 represents the dynamic power consumption per unit of data rate, reflecting the energy usage for data processing and communication tasks. Static power, $\mathrm{P}_s = V I_{\text{leak}} + A_s C f V^2$, accounts for power consumed regardless of activity, with V as the supply voltage, I_{leak} as the leakage current, A_s as the fraction of actively switching gates, C as the circuit capacitance, and f as the clock frequency. Energy efficiency, $\mathrm{EE}_u = \frac{\mathrm{a}_{t,k}^u \mathrm{e}_{m,t}^u}{\mathrm{PT}_{u,t}}$, quantifies the effective power usage, expressed as the ratio of data rate to total power consumption.

The tradeoff between minimizing Delay_u and maximizing EE_u can be modeled as a weighted sum optimization problem, where we assign weights to each objective based on their relative importance $\max \beta_w \cdot \mathrm{EE}_u - \alpha_w \cdot \mathrm{Delay}_u$, where α_w and β_w are weighting factors that determine the relative importance of minimizing latency versus maximizing energy efficiency.

III. The Proposed DMRO Scheme

A. Physical Entity Model

In addressing the tradeoff between minimizing latency and optimizing energy efficiency within smart grid network for PMUs, we propose a formulation rooted in MDP on the physical PMU side, which allows for dynamic decision-making in channel access and power control, crucial for operational efficiency in smart grids.

State and Action Space: The MDP framework consists of the state space $\mathcal{S}$, action space $\mathcal{A}$, and transition probabilities, along with a reward function that encapsulates the delay-energy tradeoff. For each timeslot t, the state observed by PMU u is denoted as $s_t^u = \{(\xi_t^u, \zeta_t^u, q_t^u)\}$, which includes local observations crucial for decision-making. Here, $\xi_t^u = 1$ if the SINR of PMU u exceeds a threshold ${\mathrm{Ths}_u}^{\min}$, and 0 otherwise. ζ_t^u is a binary variable indicating whether the interference caused by PMU u on other nodes is within acceptable limits; $\zeta_t^u = 1$ if the product of the channel gain and transmission power is below the interference threshold $\mathrm{Thi}_u^{\min}$, ensuring no negative impact on the network's performance. The action $a_t^u = \{(c_t^u, p_t^u) \mid a_t^u \in \mathcal{A}\}$ corresponds to adjustments in channel access c_t^u and transmission power p_t^u, where the power level is chosen from a predefined set up to a maximum value $P^{\max}$.

Transition Probability: Given a state $s_t^u = (\xi_t^u, \zeta_t^u, q_t^u)$ observed by PMU u at time t, the transition probability to a new state $s_{t+1}^u = \left(\xi_{t+1}^u, \zeta_{t+1}^u, q_{t+1}^u\right)$ can be modeled as $\Pr\left(s_{t+1}^u \mid s_t^u, a_t^u\right) = \Pr\left(\xi_{t+1}^u, \zeta_{t+1}^u, q_{t+1}^u \mid \xi_t^u, \zeta_t^u, q_t^u, c_t^u, p_t^u\right)$. To encapsulate system dynamics with dependencies between SINR threshold exceeding, interference threshold exceeding, and queuing state changes, we consider $\Pr\left(s_{t+1}^u \mid s_t^u, a_t^u\right) = \Psi\left(\xi_{t+1}^u \mid c_t^u, p_t^u\right) \Phi\left(\zeta_{t+1}^u \mid c_t^u, p_t^u\right) \Omega\left(q_{t+1}^u \mid q_t^u, \xi_{t+1}^u\right)$, where $\Psi\left(\xi_{t+1}^u \mid c_t^u, p_t^u\right)$ represents the probability of achieving the required SINR given the selected channel and power level; $\Phi\left(\zeta_{t+1}^u \mid c_t^u, p_t^u\right)$ denotes the probability that the interference remains below the threshold given the selected channel and power level; $\Omega\left(q_{t+1}^u \mid q_t^u, \xi_{t+1}^u\right)$ captures the probability of transitioning to a new queuing state, which may depend on the achieved SINR (as it influences the service rate).

Reward Function: The reward function aims to maximize energy efficiency while minimizing delay, and can be formulated as: $R\left(s_t^u, a_t^u\right) = \beta_w \cdot \mathrm{EE}_u - \alpha_w \cdot \mathrm{Delay}_u$, where $R\left(s_t^u, a_t^u\right)$ represents the reward received by PMU u when taking action a_t^u in state s_t^u. The state-action value function $Q(s, a)$ represents the expected return of taking action a in state s and following a certain policy π thereafter. It captures the long-term expected reward and can be expressed as: $Q^\pi\left(s_t^u, a_t^u\right) = \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k R\left(s_{t+k}^u, a_{t+k}^u\right) \mid s_t^u, a_t^u\right]$, where γ is the discount factor that values future rewards. The Bellman equation for $Q(s, a)$ is given by:

$$Q^\pi\left(s_t^u, a_t^u\right) = \mathbb{E}_{s_{t+1}^u}\left[R\left(s_t^u, a_t^u\right) + \gamma \sum_{a_{t+1}^u \in \mathcal{A}} \pi\left(a_{t+1}^u \mid s_{t+1}^u\right) \cdot Q^\pi\left(s_{t+1}^u, a_{t+1}^u\right) \mid s_t^u, a_t^u\right], \quad (9)$$

where s_{t+1}^u denotes the state at the next timeslot resulting from taking action a_t^u in state s_t^u, and $\pi\left(a_{t+1}^u \mid s_{t+1}^u\right)$ is the policy specifying the probability of taking action a_{t+1}^u when in state s_{t+1}^u.

B. Digital twin Model

In the scheme, the physical PMU, denoted as $\mathbb{PMU}$, and its digital twin counterpart, denoted as $\mathbb{DT}$, function independently, each governed by its own principles. Knowledge acts as a model that integrates data from one system and provides parameters to the other, thereby establishing a connection between the two. Data, viewed as a record of transitions from actions to states, extracts patterns from the real system, which are then distilled into experience. This experience, aligned with specific objectives, is used to update the artificial system. The inverse model of this experience, embedded in the knowledge framework, uses the updated policy from the digital twin to guide the actors, collecting feedback from the physical environment. Specifically, the relationship between the state and the action in the digital twin can be expressed as $a' = \Pi_{\mathbb{DT}}(s'|w)$, where $\Pi_{\mathbb{DT}}$ represents the digital twin's policy function, which prescribes an action based on its state, in contrast, $s' = f_{\mathbb{DT}}(a'|\theta)$, delineates the causation of system states by actions. To optimize for reward and steer the system toward a designated state, we define f and Π as:

$$f_{\mathbb{DT}}(a) \triangleq \arg\max_{s \in S} L_{\mathbb{DT}}(a, s \mid \theta), \tag{10}$$

$$\Pi_{\mathbb{PMU}}(s) \triangleq \arg\max_{a \in A} Q_{\mathbb{PMU}}(s, a \mid w), \tag{11}$$

where $L_{\mathbb{DT}}(a, s|\theta)$ indicates the empirical likelihood of a sequential occurrence of action a_i and state s_j, and $Q_{\mathbb{PMU}}(s, a|w)$ encapsulates the long-term reward yielded by state s_i and action a_j. In essence, the inductive function $f_{\mathbb{DT}}$ is reliant on the probability of observing certain states following specified actions, while the policy function $\Pi_{\mathbb{PMU}}$ is contingent on the reward of actions taken in given circumstances.

Fig. 2 shows an interactive process between a PMU, its digital twin, and a control center, integrating real-time data acquisition, predictive analytics, and decision-making. Data from the PMU is uploaded to the DT through a gateway. The PMU detects and uploads state-action-reward tuples $(s, a, s'_{[\text{Real}]}, r)$ to the related DT memory. The $\mathbb{DT}$ framework maintains a function, $\Pi_{\mathbb{DT}}$, which predicts the optimal action given a prediction state $s'_{[\text{Predict}]}$. In the DT memory, it stores tuples of state based on prediction, action, and reward $(s, a, s'_{[\text{Predict}]}, r)$ in its memory. When updating the $\mathbb{DT}$ and $\mathbb{PMU}$ framework, it calculates the temporal difference:

$$\delta = \Big| r + \gamma \Pi_{\mathbf{DT}} \left(s'_{[\text{Predict}]}, a^{\max}_{t,u}; w^{\text{now}}_{\mathbb{DT}} \right) - Q_{\mathbb{PMU}}(s, a_{t,u}; w^{\text{now}}_{\mathbb{PMU}}) \Big| \tag{12}$$

in order to incorporate new information from the real world to refine the digital twin's predictions. The $\mathbb{PMU}$ framework is updated to $w^{\text{new}}_{\mathbb{PMU}}$ by combining the current weights $w^{\text{now}}_{\mathbb{PMU}}$ with the temporal difference scaled by a learning rate $\alpha_{\mathbb{PMU}}$. Then, the $\mathbb{DT}$ framework w^{DT} are updated through a weighted combination of the current DT parameters and the PMU's new parameters, $w^{\text{new}}_{\mathbb{DT}} = \tau \cdot w^{\text{now}}_{\mathbb{PM}} + (1-\tau) \cdot w^{\text{now}}_{\mathbb{DTT}}$, where $\tau \in (0, 1)$. Common values for τ range from 0.001 to 0.1. The system continuously updates both the $\mathbb{DT}$ and the real $\mathbb{PMU}$ frameworks to keep them in synchronization and improve the predictions and actions over time.

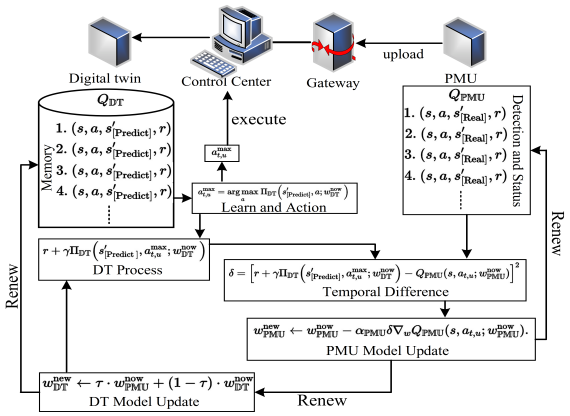


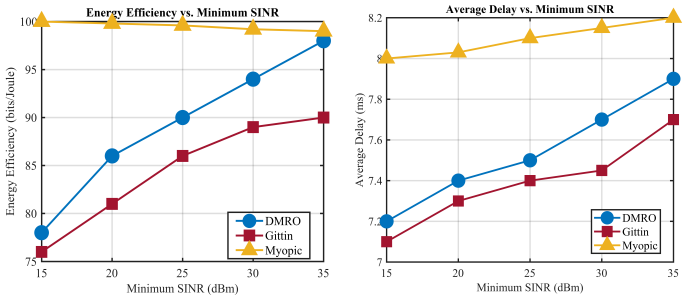
Fig. 2: The Proposed DMRO Scheme.

C. Memory Recall

The overestimation problem occurs when the $\mathbb{PMU}$ framework consistently selects suboptimal actions simply because they have the highest reward value. The framework's optimal policy assumes that the best action in a given state is the one with the highest projected reward. However, at the outset, the $\mathbb{PMU}$ framework lacks knowledge about the environment and must estimate $Q_{\mathbb{PMU}}(s, a_{t,u}; w^{\text{now}}_{\mathbb{PMU}})$, refining this estimate over iterations. Due to uncertainty in these initial reward estimates, the highest expected or estimated $Q_{\mathbb{PMU}}(s, a_{t,u}; w^{\text{now}}_{\mathbb{PMU}})$ may not correspond to the actual optimal choice.

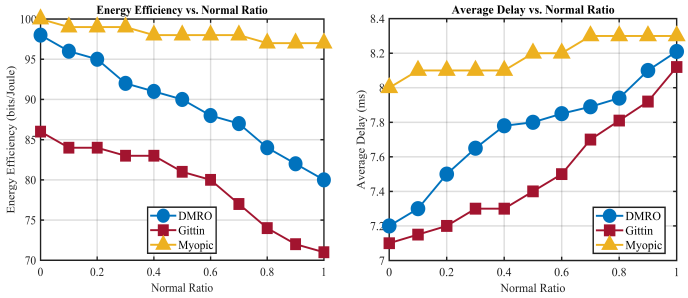
To update the scheme parameters, we estimate the gradient of the expected return $g(w)$ with respect to the parameters w. This gradient provides guidance for adjusting the parameters to maximize the expected return. Each update step computes a sample gradient based on either a single memory or a batch of memories. When a single memory (s, a, r, s') is sampled directly from the environment to compute $\nabla_w Q_{\mathbb{PMU}}(s, a_{t,u}; w^{\text{now}}_{\mathbb{PMU}})$, it may not fully represent the environment's dynamics, as consecutive experiences tend to be correlated. This correlation can introduce high variance in the gradient estimate, destabilizing the learning process. By storing memory in the $\mathbb{DT}$ framework and sampling randomly from it, we obtain a more representative and uncorrelated batch of memories. This improves the accuracy of the gradient estimate $\nabla_w g(w)$, reducing its variance compared to using a single experience.

Denote by $B = \{(s_t, a_t, r_t, s'_{t,[\text{Predict}]})\}_{t=1}^{N}$ the batch of memories sampled from the $\mathbb{DT}$ framework buffer, where N is the buffer size. The batch gradient estimate is given by $\nabla_w \hat{g}(w) = \frac{1}{N} \sum_{t=1}^{N} [\delta \cdot \nabla_w Q_{\mathbb{PMU}}(s_t, a_{t,u}; w^{\text{now}}_{\mathbb{PMU}})]$, which shows that random sampling from the replay buffer reduces the variance in updates by providing a more representative sample of the state space, which can be expressed as a reduction



(a) Energy Efficiency vs SINR (b) Average Delay vs SINR

Fig. 3: Energy Efficiency and Average Delay with varying SINR.



(a) Energy vs. Normal Ratio (b) Delay vs. Normal Ratio

Fig. 4: Energy Efficiency and Average Delay with varying Ratio.

in the variance of the gradient estimate $\mathrm{Var}\left(\nabla_w \hat{g}(w)\right) < \mathrm{Var}\left(\delta \cdot \nabla_w Q_{\mathbb{PMU}}\left(s_t, a_{t,u}; w^{\text{now}}_{\mathbb{PMU}}\right)\right)$.

IV. Performance Evaluation

In this section, we present the experiments conducted to evaluate the effectiveness of the proposed DRMO for channel access and power control optimization with traffic prediction. We conducted a comparison between the proposed scheme and two classic employed decision making policies: 1) Myopic policy [11]; 2) Gittin policy [12]; We consider a rural cell with a 400-meter radius. Within this cell, we uniformly distribute 10 PMUs and 10 Small Cell Users SUEs, uniformly. Each PMU is responsible for generating packets of 52 Bytes at a rate of 60 packets per second. The simulation advances in discrete time slots, each lasting 1 millisecond.

Fig. 3a shows that the DMRO scheme outperforms the Gittin scheme in terms of energy efficiency across all SINR levels. The DMRO curve would be higher, indicating better performance, due to its advanced prediction and control optimization capabilities. Fig. 3b shows that DMRO would typically have lower delays compared to the other schemes, especially as the SINR increases, suggesting that DMRO efficiently balances the trade-off between power control and latency. Fig.4a shows how the DMRO scheme maintains high energy efficiency across varying normal ratios, possibly due to its dynamic adaptation to changing network conditions using the digital twin's predictive insights. Fig.4b shows the DMRO scheme can provide stable delay performance even as the normal ratio increases, highlighting the scheme's effective management of traffic and channel access based on real-time data analysis.

V. Conclusion

In this paper, we have proposed a channel access and power control scheme tailored for smart grids. The scheme uses a DT-based framework to extract information from noisy network traffic data and optimize power control and channel access. Simulation results have demonstrated the DMRO scheme's effectiveness in improving traffic prediction, latency, and energy efficiency. For the future work, we will delve into exploiting these predictive insights to select the optimal radio access technology connection points for smart units at the most opportune moments, leveraging the capabilities of DT.

Acknowledgment

This work has been supported in part by the National Natural Science Foundation of China under Grant No. 62201148, in part by the Guangdong Province Basic and Applied Basic Research Foundation under Grant No. 2022KQNCX.

References

[1] M. W. Khan, G. Li, K. Wang, M. Numan, L. Xiong, and M. A. Khan, "Optimal control and communication strategies in multi-energy generation grid," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 4, pp. 2599–2653, 2023.

[2] J. Han, Q. Hong, M. H. Syed, M. A. U. Khan, G. Yang, G. Burt, and C. Booth, "Cloud-edge hosted digital twins for coordinated control of distributed energy resources," *IEEE Transactions on Cloud Computing*, vol. 11, no. 2, pp. 1242–1256, 2023.

[3] H. Xu, J. Wu, Q. Pan, X. Guan, and M. Guizani, "A survey on digital twin for industrial internet of things: Applications, technologies and tools," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 4, pp. 2569–2598, 2023.

[4] L. U. Khan, Z. Han, W. Saad, E. Hossain, M. Guizani, and C. S. Hong, "Digital twin of wireless systems: Overview, taxonomy, challenges, and opportunities," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 4, pp. 2230–2254, 2022.

[5] H. Liao, Z. Zhou, N. Liu, Y. Zhang, G. Xu, Z. Wang, and S. Mumtaz, "Cloud-edge-device collaborative reliable and communication-efficient digital twin for low-carbon electrical equipment management," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 2, pp. 1715–1724, 2023.

[6] P. Moutis and O. Alizadeh-Mousavi, "Digital twin of distribution power transformer for real-time monitoring of medium voltage from low voltage measurements," *IEEE Transactions on Power Delivery*, vol. 36, no. 4, pp. 1952–1963, 2020.

[7] Y. Ai, M. Cheffena, and Q. Li, "Radio frequency measurements and capacity analysis for industrial indoor environments," in *Proc. of EuCAP*, 2015, pp. 1–5.

[8] Q. Li, J. Chen, M. Cheffena, and X. Shen, "Channel-aware latency tail taming in industrial iot," *IEEE Transactions on Wireless Communications*, vol. 22, no. 9, pp. 6107–6123, 2023.

[9] Q. Li, N. Zhang, M. Cheffena, and X. Shen, "Channel-based optimal back-off delay control in delay-constrained industrial wsns," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 696–711, 2020.

[10] Q. Li, W. Wu, W. Zhang, and X. Sherman Shen, "Online traffic prediction in multi-rat heterogeneous network: A user-cybertwin asynchronous learning approach," in *Proc. of IEEE PIMRC*, 2023, pp. 1–7.

[11] K. Liu and Q. Zhao, "Indexability of restless bandit problems and optimality of whittle index for dynamic multichannel access," *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5547–5567, 2010.

[12] J. Ai and A. A. Abouzeid, "Opportunistic spectrum access based on a constrained multi-armed bandit formulation," *Journal of Communications and Networks*, vol. 11, no. 2, pp. 134–147, 2009.