Optimal Workload Allocation for Distributed Edge Clouds With Renewable Energy and Battery Storage

Duong Thuy Anh Nguyen[†], Jiaming Cheng[‡], Ni Trieu[†], Duong Tung Nguyen[†]
[†]Arizona State University, Tempe, AZ 85281, USA, {dtnguy52, nitrieu, duongnt}@asu.edu
[‡]University of British Columbia, Vancouver, BC V6T1Z4, Canada, jiaming@ece.ubc.ca

***Abstract*—This paper studies an optimal workload allocation problem for a network of renewable energy-powered edge clouds serving users across various geographical areas. Each edge cloud has on-site renewable energy generating units and a battery storage unit. Due to the discrepancy in electricity pricing and the diverse temporal-spatial characteristics of renewable energy generation, how to optimally allocate workload to different edge clouds to minimize the total operating cost while maximizing renewable energy utilization is a crucial and challenging problem. To this end, we introduce an optimization framework designed for Edge Service Providers (ESPs), aiming to reduce energy costs and environmental impacts, while ensuring essential quality-of-service standards. Numerical results demonstrate the effectiveness of the proposed model and solution in maintaining service quality as well as reducing operational costs and emissions. Furthermore, the impacts of renewable energy generation and battery storage on optimal system operations are rigorously analyzed.**

***Index Terms*—Cloud/edge computing, data centers, edge clouds, renewable energy, battery storage, carbon footprint.**

I. Introduction

Over the past decade, Cloud/Edge Service Providers (ESPs) have emerged as indispensable drivers of digital transformation. They enable the delivery of a wide spectrum of digital services, encompassing tasks such as data storage, processing, software applications, and beyond [1]. Each ESP oversees a portfolio of edge clouds (EC), also known as edge data centers. The ESPs typically manage extensive networks comprising varying sized and configured ECs, each housing a diverse array of computing resources. The ECs are distributed across different geographical locations to ensure proximity to end-users, reduce latency, and optimize service delivery. These ECs serve as the fundamental building blocks of their cloud infrastructure, facilitating the provisioning of a wide range of services to customers, ranging from virtual machines and storage to machine learning and content delivery.

Energy efficiency is a critical problem in cloud/edge computing. ECs consume a considerable amount of energy to operate servers, networking equipment, cooling systems, and other infrastructure components. Most ECs are connected to the electrical grid, relying on utility power as their principal source of electricity. Given their enormous energy consumption, ECs contribute to a large amount of greenhouse gas (GHG) emissions. While the cost of electricity receives growing attention, the environmental impacts of power-intensive operations are often overlooked. Indeed, inexpensive electricity can sometimes come at the expense of environmental harm. According to the 2021 data from the U.S. Energy Information Administration [2], states like Wyoming, Utah, and North Dakota have some of the lowest electricity prices but significantly higher carbon footprints in their power sectors compared to the national average.

The exponential surge in data generation and computing demands has ushered in a relentless increase in the energy consumption of ECs. This heightened energy demand poses both environmental and economic challenges. The commitment to utilizing more renewable energy sources, including solar, wind, and hydropower, has become one of the foremost strategic and operational goals for ECs, offering a solution to mitigate their carbon footprint, align with sustainability objectives, and decrease dependence on fossil fuels. A primary challenge associated with renewable energy sources is their intermittent nature. To address this, energy storage solutions like batteries are widely recognized as attractive options for promoting the sustainability and efficiency of EC operations. By storing surplus energy from renewable sources and/or low-cost grid electricity during off-peak hours to power ECs during peak periods or outages, they effectively tackle the variability and intermittency of renewable energy sources, facilitate their efficient utilization, and contribute to a resilient edge system.

In regards to green EC design, [3] introduces an architecture for real-time monitoring, live virtual machine (VM) migration, and VM placement optimization to minimize power consumption. This approach aims to enhance server utilization and optimize power management in ECs, ultimately reducing their carbon footprint. Another line of research focuses on energy-cost-aware request routing among ECs by considering geographically dependent electricity costs [4]. Recently, renewable energy resources have been integrated into ECs, thereby advancing sustainability [5]. Reference [6] explores energy-information transmission trade-offs across various optimization problems, encompassing electricity costs, request routing, data center locations, proximity to renewable energy sources, server quantities, and the implications of carbon taxes.

Reference [7] introduces the concepts of “green workload” and “green service rate” in contrast to “brown workload” and “brown service rate” to distinctly address the separation of maximizing green energy utilization and minimizing brown energy costs. Reference [8] proposes a game theory-based resource management framework that incorporates renewable energy to minimize cloud operating costs and queuing delays. In [9], a cooperative framework is considered where multiple electricity retailers work together to implement incentive-based demand response in distributed data centers, aiming to maximize profits. In [10], authors consider a workload allocation model for data

The first two authors have contributed equally to this work.

979-8-3503-7099-7/24/$31.00 ©2024 IEEE

center in electricity market.

Contributions: Motivated by the compelling considerations outlined above, this paper proposes a holistic model for optimal EC operations. Given the diverse edge environments, characterized by varying electricity costs and carbon footprints, it becomes imperative for ESPs to implement efficient workload allocation strategies to ensure high quality-of-service (QoS), minimize costs, and enhance sustainability. Furthermore, the potential co-location of renewable energy generation and battery storage at EC facilities drives our investigation into tailored optimization models for co-optimizing EC provisioning and power procurement with these technologies. Specifically, we introduce an optimization framework for ESPs to simultaneously minimize energy costs and environmental impact by integrating renewable energy sources and battery storage systems while maintaining essential QoS standards.

Unlike previous work that delves into decisions regarding EC placement, our primary objective centers on enhancing operational efficiency, guaranteeing the delivery of high-quality services while simultaneously curbing power consumption and reducing our environmental footprint. We employ a systematic and analytical framework, with an emphasis on environmental sustainability, wherein we meticulously account for emissions and carbon taxes. Our approach features the integration of renewable energy sources and battery storage as key components. Our model also facilitates agile energy trading with the grid. We aim to analyze the impacts of integrating battery storage units, onsite renewable energy, and two-way energy trading with the grid on the optimal operation of networked ECs. Our numerical results illuminate the profound interdependencies among these factors and provide pragmatic insights for ESPs to reduce renewable energy curtailment. Previous work has often overlooked these aspects, focusing primarily on technical solutions to address intricate optimization models.

II. System Model

We consider an ESP that owns and operates multiple ECs in different geographical locations. Each EC is equipped with servers, computing resources, and networking functions to provide low-latency, high-performance cloud services to users situated in various areas. The ESP aims to deliver high-quality cloud services to its diverse user base while simultaneously optimizing several critical aspects of its operations, which encompass minimizing unmet demand, reducing energy consumption, and mitigating emissions.

The ESP aggregates user requests, which are then directed to ECs for further processing. The ESP strives to ensure that users' demand is consistently met, resulting in a seamless and reliable user experience. ECs typically consume a substantial amount of energy to power servers, networking equipment, and other infrastructure components. Also, due to their enormous energy consumption, ECs exert a considerable influence on the electric grid, contributing significantly to GHG emissions and carbon footprints. Thus, energy efficiency is crucial to reduce operational costs and minimize the environmental impact of ECs' operations. This involves deploying energy-efficient infrastructure components and strategically integrating renewable sources, such as solar panels and wind turbines, along with battery storage systems. Consequently, surplus energy generated during off-peak hours can be stored and later utilized during peak demand periods. The ESP also has the opportunity to sell excess power generated from its renewable sources back to the grid, generating extra revenue to offset energy costs and reduce both GHG emissions and reliance on fossil fuels.

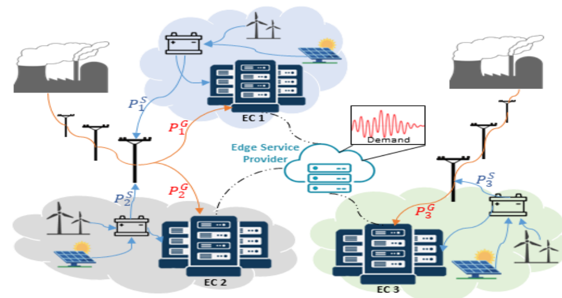


Fig. 1: System model

The ESP adheres to operational constraints while upholding stringent QoS standards. These constraints encompass various factors, such as optimizing server activation based on varying resource demand, effectively distributing workloads to align with user needs, and maintaining certain delay and server utilization requirements. The ESP also faces additional constraints related to grid capacity limits and renewable energy availability. These constraints necessitate efficient energy management. By balancing these operational and environmental constraints, the ESP can provide robust, sustainable, and environmentally responsible cloud services that meet the evolving demands. The system model is depicted in Fig. 1. The main notations are presented in Table I.

Notation	Definition
Sets and indices	
EN, AP, ESP	Edge cloud, Access point, Edge Service Provider
$i, \mathcal{M}, M$	Index, set and number of areas (APs)
$j, \mathcal{N}, N$	Index, set and number of edge clouds (ECs)
$t, \mathcal{T}$	Index for time period and time interval
Parameters	
e_j^t, a_j^t	Buy-more and sell-back electricity price at EC j at time t
$C_j^{\max}$	Available servers at EC j
α	Average computing resource to serve 1 request
λ_i^t	Expected resource demand of users in AP i at time t
ϕ_i	Unmet demand penalty
$d_{i,j}, D^{\max}$	Propagation delay between AP i and EC j, and threshold
$P_j^{R,t}$	Renewable energy generation at EC j at time t
P_j^{idle}, P_j^{peak}	Idle/Peak server power consumption
E_j^{usage}	Power Usage Effectiveness (PUE)
θ_j, δ_j	Emission factor and carbon tax at EC j
Variables	
$x_{i,j}^t$	Requests from AP i to EC j at time t
q_i^t	Unmet demand in area i at time t
c_j^t	Number of active computer servers at EC j at time t
γ_j^t	Average server utilization at EC j at time t
$P_j^{G,t}$	Power to be purchased from the grid at EC j at time t
$P_j^{C,t}, P_j^{D,t}$	Charged/Discharged battery energy at EC j
$P_j^{U,t}$	Power demand at EC j at time t
$P_j^{S,t}$	Sell-back power from EC j at time t
E_j^t	Battery energy level at EC j at time t

TABLE I: Notations

III. Problem Formulation

Consider an ESP with a total of N ECs, denoted as $\mathcal{N}$, serving users situated in various areas. These geographical user areas are represented by Access Points (APs), collectively defined as the set $\mathcal{M}$, with M individual APs identified using index i, while each EC is denoted by index j. In our model, we consider the time interval denoted as $\mathcal{T}$, where ΔT represents the duration of a single period within this interval.

A. Computing Capacity

To enhance efficiency, optimize costs, and manage resources effectively, the ESP determines the number of servers to stay active based on demand. Let c_j^t represent the number of active servers at EC $j \in \mathcal{N}$ at time $t \in \mathcal{T}$. The number of active servers should not surpass the total available servers at EC j, denoted as $C_j^{\max}$. Thus, we have:

$$0 \le c_j^t \le C_j^{\max}, \quad \forall j. \tag{1}$$

B. Workload Allocation

For each area $i \in \mathcal{M}$ and every time period $t \in \mathcal{T}$, we use λ_i^t to represent the total expected resource demand of users in area i at time t. Demand volumes can vary significantly throughout the day. We denote the number of requests from area i allocated to EC j during period t as $x_{i,j}^t$, and α represents the average computing resource required to serve one request. Service requests from each area must either be served by ECs or considered unmet, denoted as q_i^t. Thus, we impose the following constraint:

$$\alpha \left(\sum_j x_{i,j}^t + q_i^t \right) = \lambda_i^t, \quad \forall i. \tag{2}$$

Let ϕ_i denote the penalty cost of for unmet demand in area i. Therefore, the total unmet demand penalty for the ESP is:

$$C^{\mathsf{u}} = \sum_{i,t} \phi_i q_i^t. \tag{3}$$

To illustrate the impact of delay requirements on the system's performance, we denote $d_{i,j}$ as the propagation delay between each AP $i \in \mathcal{M}$ and EC $j \in \mathcal{N}$. To ensure QoS, we introduce the binary indicator parameter $b_{i,j}$, which depends solely on the propagation delay $d_{i,j}$ and the maximum delay threshold $D^{\max}$ [11]. Specifically, the round-trip propagation delay must always be kept below $D^{\max}$. In other words,

$$b_{i,j} = \begin{cases} 1, & 2d_{i,j} \le D^{\max} \\ 0, & 2d_{i,j} > D^{\max} \end{cases}, \quad \forall i, j. \tag{4}$$

C. Energy Consumption Model

1) Average Server Utilization: Let ρ represent the service rate signifying the maximum number of service requests that a single server can effectively handle within a time period. The average server utilization at EC j during time t, denoted as γ_j^t and discussed in [6], quantifies the proportion of the EC's capacity for that specific period and is determined as follows:

$$\gamma_j^t = \frac{\sum_{i \in \mathcal{M}} x_{i,j}^t}{\rho c_j^t}, \quad \forall j. \tag{5}$$

To control queuing delay, we set a limit $\gamma^{\max} \in (0, 1]$ on the average server utilization at each EC as follows:

$$\gamma_j^t \le \gamma^{\max}, \quad \forall j. \tag{6}$$

The choice of the $\gamma^{\max}$ parameter depends on the service request traffic pattern and the QoS requirements [6]. If $\gamma^{\max}$ is sufficiently small, the waiting time for a service request at an EC before server handling becomes negligible, with most of the overall latency in responding to service requests determined by the bounded propagation delay, $D^{\max}$.

2) Power Consumption: Let P_j^{idle} denote the average power consumption of an individual server in the idle state, and let P_j^{peak} denote the peak power consumption when the server is actively processing service requests. Additionally, we introduce the term Power Usage Effectiveness (PUE)[1], represented as E_j^{usage}. The total power consumption (or power demand) $P_j^{U,t}$ at each EC location j and for each period t can be computed as follows [12], [13]:

$$\begin{aligned} P_j^{U,t} &= c_j^t \left(P_j^{idle} + \left(E_j^{usage} - 1 \right) P_j^{peak} \right) \\ &\quad + c_j^t \left(P_j^{idle} - P_j^{peak} \right) \gamma_j^t, \quad \forall j, t. \end{aligned} \tag{7}$$

The ratio P_j^{peak}/P_j^{idle} serves as a metric for assessing the power elasticity of servers. A higher value of this ratio indicates greater elasticity, resulting in reduced power consumption during periods of server inactivity. When $P_j^{peak} = P_j^{idle}$, the power consumption is $P_j^{U,t} = c_j^t E_j^{usage} P_j^{peak}$. In this scenario, power consumption becomes solely dependent on the quantity of servers, without consideration for the number of routed requests or the operational period.

D. Energy Model

1) Cost of Electricity: ECs usually rely on the electrical grid as their primary power source. In North America, the electric grid operates on a regional basis. Most regions have regulated electricity markets where prices remain fixed throughout the day. In areas with deregulated markets, energy prices from the grid can fluctuate significantly throughout the day and across seasons, reflecting the dynamics of the wholesale electricity market. Let e_j^t represent the electricity price at each EC j and time t. Also, the amount of energy imported from the grid by EC j at time t is denoted by $P_j^{G,t}$. Thus, the electricity cost is $e_j^t P_j^{G,t}$.

2) Grid Capacity Limit: The amount of electricity imported from the grid cannot exceed the power limit $P_j^{G,\max}$ of the point of coupling between an EC and the grid, i.e., we have:

$$0 \le P_j^{G,t} \le P_j^{G,\max}, \quad \forall j. \tag{8}$$

3) Renewable Energy: To manage power costs effectively, reduce carbon emissions, and ensure uninterrupted operations, ESPs often employ a diverse range of energy sources and technologies. Let $P_j^{R,t}$ indicate the renewable energy generation at EC j at time t. If the renewable energy is sufficient to meet the current power demand, i.e., $P_j^{R,t} \ge P_j^{U,t}$, then no further procurement is necessary, meaning $P_j^{G,t} = 0$. Otherwise, the

[1]PUE is a metric used to assess the energy efficiency of data centers. As reported in the 2016 U.S. EC Energy Report, the average annualized PUE across various data centers typically falls within the range of 1.8 to 1.9.

ESP must determine how much additional power ($P_j^{G,t}$) to purchase from the main grid.

4) Battery Storage System: Batteries with finite capacity can be incorporated to store excess energy during periods of low demand or when renewable sources produce surplus electricity, ensuring stable supply and lower costs. When no procurement is required, the surplus energy can be used for charging the battery. Conversely, if procurement is necessary, the ESP can decide whether to opportunistically discharge energy from the battery, in addition to the procured power, to meet the current demand.

5) Energy Sell-back: We assume that the ESP is allowed to sell back the electricity to the grid at a sell-back price of a_j^t. For example, the ESP may want to sell back the surplus energy when the onsite generation exceeds the demand. Let $P_j^{S,t}$ represent the amount of electricity to be sold to the grid from EC j at time t. Then, we can calculate the electricity adjustment cost as follows:

$$C^{\mathsf{e}} = \sum_{j,t} \left(e_j^t P_j^{G,t} - a_j^t P_j^{S,t} \right). \tag{9}$$

6) Power Balance Equation: Let $P_j^{C,t}$ and $P_j^{D,t}$ denote the charging and discharging power of the battery at EC j at time t. Constraints on maximum charging and discharging rates for batteries are expressed as:

$$0 \leq P_j^{C,t} \leq P_j^{C,\max} \text{ and } 0 \leq P_j^{D,t} \leq P_j^{D,\max}, \quad \forall j. \tag{10}$$

For simplicity, we do not consider power transmission losses. The energy balance equation is expressed as follows:

$$P_j^{R,t} + P_j^{G,t} + P_j^{D,t} = P_j^{U,t} + P_j^{C,t} + P_j^{S,t}, \quad \forall j. \tag{11}$$

7) Battery Energy Dynamics: For simplicity, we assume uniform efficiency for battery charging and discharging, represented by the parameter $\eta \in [0, 1]$. The energy level of the battery at location j and time t is denoted by E_j^t. The battery energy dynamics can be expressed as follows:

$$E_j^{t+1} = E_j^t + \Delta T \left(\eta P_j^{C,t} - \frac{P_j^{D,t}}{\eta} \right), \quad \forall j, t. \tag{12}$$

8) Battery Capacity Constraint: Let $E_j^{\max}$ denote the battery capacity at EC j. The battery also needs to maintain a minimum energy level of $E_j^{\min}$ to ensure its longevity, i.e.,

$$E_j^{\min} \leq E_j^t \leq E_j^{\max}, \quad \forall j. \tag{13}$$

E. Environmental Impact

1) Carbon Emission Factor: The carbon emission factor for electricity varies significantly based on its source. In the case of electricity from the grid, the emission factor is location-dependent and closely tied to the energy composition of the region. Regions heavily reliant on fossil fuels, such as coal and natural gas, tend to exhibit higher carbon emission factors. Conversely, on-site renewable energy sources, such as solar panels and wind turbines, are renowned for their eco-friendliness. They typically boast significantly lower carbon emission factors or even approach zero emissions.

In our model, we denote θ_j as the emission factor associated with each unit of electricity purchased from the grid to power EC j. Consequently, the total carbon emissions linked to powering EC j at time t are calculated as follows:

$$EM_j^t = \theta_j P_j^{G,t}. \tag{14}$$

2) Carbon Tax: To address environmental concerns, several states in the United States and Canadian provinces have implemented carbon taxes [14]. Typically, these carbon taxes are imposed on power plants, which then transfer the cost of the carbon pricing to consumers through increased electricity prices. Thus, environmental considerations are factored into our system model through the electricity cost. However, given that carbon taxes are not yet widely adopted, we introduce them as distinct parameters in our research to better understand their impacts. To this end, we designate the carbon tax for each EC location j as δ_j, resulting in an additional cost of

$$C^{\mathsf{c}} = \sum_{j,t} \delta_j EM_j^t = \sum_{j,t} \delta_j \theta_j P_j^{G,t}. \tag{15}$$

F. ESP Optimization Model

The model for the ESP can be formulated as a Mixed-Integer Linear Programming (MILP) problem as follows:

$$\underset{\mathbf{P},\mathbf{E},\gamma,\mathbf{x},\mathbf{q},\mathbf{c}}{\text{minimize}} \quad C^{\mathsf{u}} + C^{\mathsf{e}} + C^{\mathsf{c}} \tag{16}$$

$$\begin{aligned} \text{subject to} \quad & (1), (2), (6), (8), (11) - (13) \qquad (17) \\ & P_j^{G,t}, P_j^{U,t}, P_j^{C,t}, P_j^{D,t}, P_j^{S,t}, E_j^t, \gamma_j^t \geq 0, \ \forall j, t \\ & 0 \leq x_{i,j}^t \leq b_{i,j} \lambda_i^t, \ \forall i, j, t; \quad q_i^t \geq 0, c_j^t \in \mathbb{Z}^+, \forall j, t. \end{aligned}$$

IV. Numerical Results

We consider an edge system comprising 8 ECs ($N = 8$) and 10 APs ($M = 10$). The edge network topology is based on the cities and locations of randomly selected Equinix ECs [2]. In the *default setting*, we assume that all ECs are eligible to serve demand from every area, i.e., $b_{i,j}$ is set to 1 for all i and j. We will also perform sensitivity analysis on larger networks with more than 10 areas. In our setup, we assume that all ECs are eligible to serve demand from every area. We consider P^{idle} to be randomly generated from U$[0.45, 0.55]$ kWh, while P^{peak} is randomly generated from U$[1.2, 1.5]$ kWh. As reported in the 2016 U.S. EC energy report, we adopt E^{usage} to fall within the interval of U$[1.8, 1.9]$. The electricity price is taken from the range of $[0.1, 0.35]$ \$/kWh [2]. Due to the absence of carbon tax data in certain U.S. states, we generate δ_j randomly from the range $[0.6, 0.7]$ [14].

The "sell-back" price is expected to be less than the procurement cost, as denoted by $a_j = \zeta e_j, \forall j$, with ζ equal to 0.8. By utilizing the trace data from GWA-DAS [3], we randomly generate the expected demand λ_i^t, ranging from 10 to 30 per hour. In our problem, we assume that each EC is directly connected to the grid. Thus, the grid capacity at location j is randomly generated from U$[1, 1.5]$ megawatts (MW). Regarding the emission factor for each EC j, we will focus solely on CO2 emissions. According to [14], we assume that the carbon tax for the selected location j follows U$[20, 50]$

[2]https://www.equinix.com/data-centers/americas-colocation, Access 2022.
[3]http://gwa.ewi.tudelft.nl/datasets/gwa-t-1-das2/report

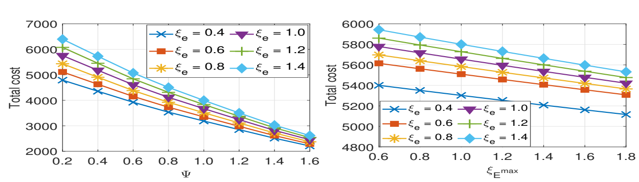
$/ton. The carbon emission factor θ_j is randomly generated from U$[0.1, 0.8]$, based on data from [15].

In the *default setting*, we consider that $P^R_{j,t}$ is randomly generated from U$[80, 100]$ kW. The maximum battery capacity $E^{\max}_j$ is generated from U$[90, 100]$ kW, while $E^{\min}_j$ follows U$[30, 50]$ kW. The charging capacity ($P^{C,\max}_j$) and discharging capacity ($P^{D,\max}_j$) are generated from U$[70, 80]$ kW and U$[70, 80]$ kW, respectively. Additionally, we consider the values $\gamma^{\max} = 0.9$, $\alpha = 0.5$, $\rho = 0.8$, $\Delta T = 1$, and $T = 12$. We will also vary these parameters during sensitivity analysis. All the experiments are conducted in MATLAB using CVX[4] and Gurobi[5] on a desktop with an Intel Core i7-11700KF CPU and 32GB of RAM. The source code and models pertinent to this study are accessible online[6].

A. Sensitivity analysis

This section presents sensitivity analyses to assess the influence of key system parameters on the optimal solution. These parameters include renewable energy (P^R_j), electricity price (e^t_j), and the "sell-back" ratio (ζ). To evaluate the impact of renewable energy on system performance, we introduce a scaling factor Ψ for P^R_j, where $\Psi = 1$ indicates the default value. Specifically, the value of P^R_j generated in the default setting is multiplied by Ψ to either scale up or down the renewable energy. Similarly, we use $\xi E^{\max}, \xi_e$, and $\xi_{D^{\max}}$ as scaling factors for the maximum battery size at EC ($E^{\max}_j$), electricity price e_j, and average utilization $D^{\max}$.

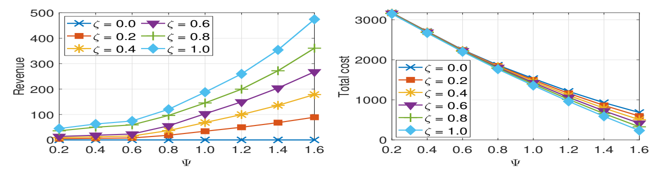
1) Benefits of battery storage: As depicted in Figure 2(a), the total cost decreases as Ψ increases, signifying an increase in the available renewable energy at each EC. This allows the operator to have greater flexibility in supplying power from renewable sources, resulting in a reduced reliance on grid energy. Furthermore, it is evident that a higher electricity price (e) can motivate the operator to maximize the utilization of renewable energy resources, reducing the need for energy procurement from the grid. Similarly, we also examine the impact of battery capacity size on EC operations. Recall that $\xi_{E^{\max}}$ is a scaling factor for battery size. When $\xi_{E^{\max}}$ is set to a higher value, each EC can potentially have a greater capacity to absorb surplus energy. This advantage becomes particularly significant when electricity prices (e) are elevated.



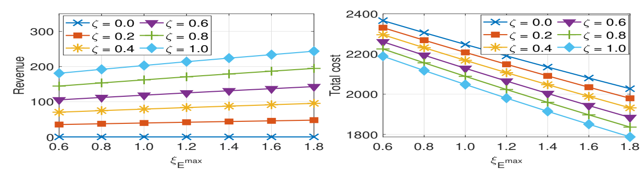
(a) varying Ψ and e (b) varying e and $E^{\max}$

Fig. 2: The benefits of renewable energy

2) Benefits of "sell-back" option: Fig.3(a) - 3(d) shows how "sell-back ratio" influences the system performance. Recall that ζ is defined as "sell-back" ratio between the electricity price and "sell-back" price (i.e., $e_j = \zeta a_j, \forall j$). Once the power demand of each EC has been met, any surplus renewable energy tends to be given higher priority for selling back to the grid, especially when ζ is set to higher values. As illustrated in Fig.3(c) and 3(d), larger values of E signify a greater battery capacity at each EC, allowing for more energy storage. In such cases, the operator may opt to store excess energy and subsequently sell it back to the grid, particularly when the "sell-back" price is favorable.



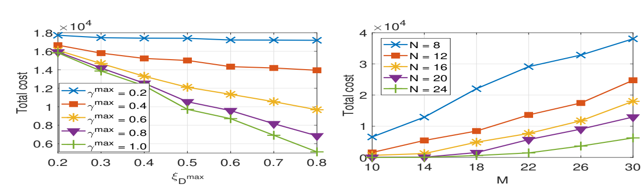
(a) Revenue: varying Ψ and ζ (b) Total cost: varying Ψ and ζ



(c) Revenue: varying $E^{\max}$ and ζ (d) Total cost: varying $E^{\max}$ and ζ

Fig. 3: The benefits of "sell-back" option

3) The impact of other system parameters: Notice that $\gamma^{\max}$ and $D^{\max}$ directly influence QoS. Specifically, $\gamma^{\max}$ imposes a constraint on the average server utilization at each EC, while $D^{\max}$ sets a limit on the propagation delay between APs and ECs. As shown in Figures 4(a), when ω_γ is decreased, it signifies a stricter restriction on the average utilization at each EC, which can result in a higher chance of unmet demand due to these more stringent requirements. As described in (4), an EC j can only serve user requests from area i when the propagation delay between them is within the threshold $D^{\max}$. In other words, a decrease in $D^{\max}$, indicating a more stringent delay requirement, leads to an increase in the number of $b_{i,j}$ values that become zero, indicating a reduction in the number of eligible ECs to serve user demand. Consequently, the total cost of the system increases as $D^{\max}$ decreases. Furthermore, Figure 4(b) shows the impact of network size on the optimal solution. As expected, with a fixed number of ECs, the total cost rises as the number of APs increases.



(a) varying $\gamma^{\max}$ and $D^{\max}$ (b) varying M and N

Fig. 4: The impact of other system parameters

B. Performance Comparisons

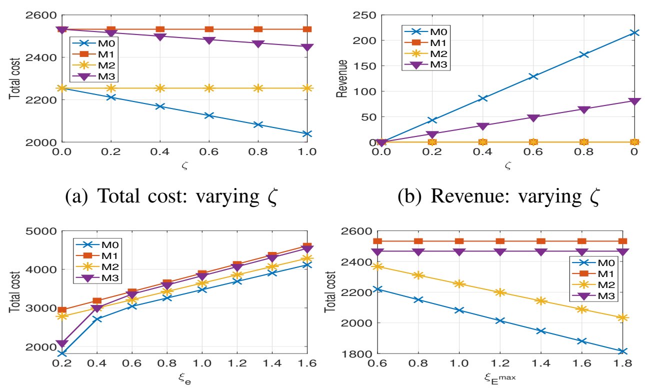
In this section, we aim to compare the performance of the proposed model with the following benchmarks:

- **M1**: This model lacks the capability for the "sell-back-to-grid" option, and ECs do not have batteries.

[4]http://cvxr.com/cvx/
[5]https://www.gurobi.com/
[6]https://github.com/JJmingcc/Renewable_EC

- **M2**: This model exclusively focuses on using batteries to store energy and does not allow the operator to sell surplus energy back to the grid.
- **M3**: This model is designed to enable the selling of excess energy back to the grid, while ECs do not feature battery installations.

For the mathematical formulations of the benchmark models **M1** - **M3**, please refer to the Appendix. A in the technical report[7]. The evaluation and comparison of the four schemes are based on their total cost in four different settings. To simplify, we refer to our proposed model as "**M0**". These four models can be straightforwardly categorized into those that incorporate or omit considerations for sell-back and battery storage options.



(a) Total cost: varying ζ (b) Revenue: varying ζ

(c) Total cost: varying ξ_e (d) Total cost: varying $\xi_{E^{\max}}$

Fig. 5: Model comparisons

As illustrated in Fig. 5(a)–5(d), our proposed model significantly outperforms the other schemes. **M1** achieves the worst performance since it lacks both the "sell-back-to-grid" option and battery storage. Fig. 5(a) shows that **M0** and **M3** do not vary regardless of changes in ζ because these two models lack considerations for "sell-back-to-grid". Thus, when $\zeta = 0$, indicating a sell-back price of 0, there is no distinction in performance between **M0** and **M2** or **M1** and **M3**. The difference between these two pairs demonstrates the advantages of battery storage, as energy can be discharged. When ζ increases, the advantages of models that consider "sell-back-to-grid" options (such as **M0** and **M3**) become increasingly pronounced due to the elevated sell-back prices, which can be verified in Fig. 5(b).

Furthermore, as depicted in Fig. 5(c), all four models exhibit an increase as electricity prices (e) rise. Notably, the advantages of the sell-back option can be emphasized, especially when electricity prices (e) are low. In such scenarios, selecting energy supply from the grid is not that expensive, offering greater operational flexibility. However, as the electricity price (e) escalates, a model with solely a sell-back option becomes insufficient, as these resources would be wasteful without storing surplus energy in a battery. Therefore, the cost of **M2** gradually surpasses that of **M3**. The reduction in the maximum battery size ($E^{\max}$) signifies decreased capacity for storing renewable energy, potentially resulting in limited availability of renewable energy and, consequently, the possibility of renewable energy curtailment. Figure 5(d) illustrates that costs decrease across models that take battery storage into account, as the operator can prioritize the utilization of renewable energy sources by discharging energy from the batteries to meet power demand. These advantages become more prominent as the maximum battery size ($E^{\max}$) increases. In summary, the risk of renewable energy curtailment is heightened when EC systems lack both battery storage and the sell-back option.

[7]Technical report: https://arxiv.org/abs/2310.00742

V. Conclusion

This paper has unveiled valuable insights and potential benefits of integrating renewable energy sources, battery systems and energy trading in edge service operations. Notably, when renewable energy sources are considered, our approach underscores the allocation of workload not only to ECs with low electricity prices but also to those with high renewable energy generation. Moreover, the introduction of batteries adds another dimension to the workload allocation problem as the battery capacity can efficiently absorb surplus energy, acting as an essential tool for energy arbitrage.

References

[1] D. T. A. Nguyen, J. Cheng, D. T. Nguyen, and A. Nedić, "CrowdCache: A decentralized game–theoretic framework for mobile edge content sharing," *Proc. of IEEE WiOpt*, 2023.
[2] "Annual electric power industry report," U.S. Energy Information Administration, 2021, https://www.eia.gov/electricity/data/state/.
[3] L. Liu, H. Wang, X. Liu, X. Jin, W. B. He, Q. B. Wang, and Y. Chen, "Greencloud: A new architecture for green data center," in *Proc. of 6th ICAC- INDST*, 2009, p. 29–38.
[4] L. Rao, X. Liu, L. Xie, and W. Liu, "Minimizing electricity cost: Optimization of distributed internet data centers in a multi-electricity-market environment," in *Proc. of IEEE INFOCOM*, 2010, pp. 1–9.
[5] M. Ghamkhari and H. Mohsenian-Rad, "Energy and performance management of green data centers: A profit maximization approach," *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 1017–1025, 2013.
[6] A.-H. Mohsenian-Rad and A. Leon-Garcia, "Energy-information transmission tradeoff in green cloud computing," in *Proc. of IEEE GLOBECOM*, 2010.
[7] A. Kiani and N. Ansari, "Toward low-cost workload distribution for integrated green data centers," *IEEE Communications Letters*, vol. 19, no. 1, pp. 26–29, 2015.
[8] N. Hogade, S. Pasricha, and H. J. Siegel, "Energy and network aware workload management for geographically distributed data centers," *IEEE Trans. Sustain. Comput.*, vol. 7, no. 2, pp. 400–413, 2021.
[9] P. Zhang, P. Yang, Z. Zhao, C. S. Lai, L. L. Lai, and M. Shahidehpour, "A framework for several electricity retailers cooperatively implement demand response to distributed data center," *IEEE Trans. Smart Grid.*, vol. 14, no. 1, pp. 277–289, 2022.
[10] S. Bahrami, V. W. S. Wong, and J. Huang, "Data center demand response in deregulated electricity markets," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2820–2832, 2019.
[11] J. Cheng, D. T. Nguyen, and V. K. Bhargava, "Resilient edge service placement under demand and node failure uncertainties," *IEEE Trans. Netw. Service Manag.*, pp. 1–1, 2023.
[12] X. Fan, W.-D. Weber, and L. A. Barroso, "Power provisioning for a warehouse-sized computer," *SIGARCH Comput. Archit. News*, vol. 35, no. 2, p. 13–23, 2007.
[13] H. Wang, J. Huang, X. Lin, and H. Mohsenian-Rad, "Proactive demand response for data centers: A win-win solution," *IEEE Trans. Smart Grid*, vol. 7, no. 3, pp. 1584–1596, 2016.
[14] "Carbon tax US," U.S. Energy Information Administration, 2020, https://ourworldindata.org/carbon-pricing.
[15] "Emissions by plant and by region," U.S. Energy Information Administration, 2021, https://www.eia.gov/electricity/data/emissions/.