Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 06 Issue 01
2023

Research Article
https://doi.org/10.56979/601/2023

Renewable and Temperature Aware Load Balancing for Energy Cost Minimization in Data Centers: A Study of BRT, Peshawar

Muhammad Imran Khan Khalil[1], Ahmar Mubeen[2], Amer Taj[1], Naveed Jan[3], and Sheeraz Ahmed[4]

[1]University of Engineering & Technology, Peshawar, 25000, Pakistan.
[2]Bus Rapid Transit (BRT), Peshawar, 25000, Pakistan.
[3]University of Technology, Nowshera, 24100, Pakistan.
[4]Iqra National University, Peshawar, 25000, Pakistan.
*Corresponding Author: Sheeraz Ahmad. Email: sheeraz.ahmad@inu.edu.pk

Received: October 09, 2023 **Accepted:** November 28, 2023 **Published:** December 05, 2023

Abstract: Power management in data centers (DSs) is a pressing contemporary concern, particularly in the context of cost reduction. DCs are the key source of energy consumption as they strive to meet customer demands, leading to negative impact on environment and elevated energy expenses. Existing research in this domain primarily focuses on task allocation techniques that leverage renewable energy sources and lower electricity rates to mitigate energy costs. In our research, we address the cost reduction problem in data centers and consider Bus Rapid Transit (BRT) service system as a case study in Peshawar, Pakistan. We introduce a novel strategy called Renewable and Temperature-aware Load Balancing (RTLB), which employs an online greedy algorithm design technique to optimize the processing of user requests within a DC. Our proposed algorithm considers various factors, including ambient and internal temperatures, on-site renewable energy availability, conventional energy consumption, active server count, and compliance with predefined constraints. The experiments and their results, conducted using real-world data, validate the higher performance of RTLB when compared to existing workload allocation strategies, ultimately reducing the overall operating expenses of the DC.

Keywords: Energy Efficiency; Optimization; Geo-distributed DC; Geograhical Load Balancing; Renewable Energy; Bus Rapid Transit.

1. Introduction

In the past decade, online applications and services have seen an exponential surge in popularity. This led to the establishment of massive Internet data centers (IDCs) as part of the cloud computing trend, aimed at enhancing reliability, manageability, and cost-efficiency [1]. However, the substantial energy consumption associated with IDC operations has emerged as a critical concern. Recent research findings reveal that many IDC operators, including major players like Microsoft and Google, incur electricity bills exceeding $30 million annually, constituting a substantial portion of their operational expenses [2].

Concurrently, the field of electrical power systems is advancing toward the concept of the smart grid. This term now characterizes the next-generation power system that incorporates more advanced information and communication technologies throughout electricity generation, distribution, and consumption. In an effort to minimize data center operating costs, several studies have explored the integration of smart grid solutions due to their notable efficiency gains [2], [3], [4].

As cloud computing services gain widespread popularity, the prevalence of large-scale data centers (DCs) is on the rise [1], [5]. These DCs play a central role in delivering various cloud services, such as web search, Internet of Things (IoT), and data analytics. To enhance the efficiency and reliability of their services, Cloud Service Providers (CSPs) establish multiple data centers across different geographical locations. Each of these globally distributed DCs is composed of cooling systems, network servers, and network equipment [6].

The substantial power consumption of each DC is required to execute workloads and meet the demands of the DC. The US Department of Energy in 2021, estimated that all data centers collectively consumed approximately 70 billion kw-h, which is 2.3% of the overall electricity consumption [8]. Projections indicate that by 2023, DCs in the US will require roughly 200 billion kw-h of energy considering the prevailing level of demand. Notably, several prominent companies, such as Microsoft, Google, and Amazon, allocate a heavy portion of their running budgets to cover energy expenses [7], [9].

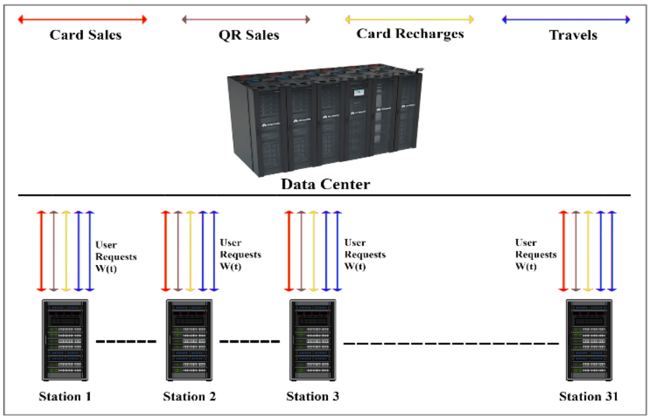

Figure 1. Matching user requests between data centers and the stations

Among the components within data centers (DCs), IT equipment and cooling systems account for the lion's share of electricity consumption. The objective is to enhance network equipment efficiency by maintaining temperature balance within the DCs, thus reducing the substantial power requirements of cooling equipment [1]. This efficiency is often quantified using a metric known as Power Usage Effectiveness (PUE), which gauges the cooling system's energy usage. The calculation of PUE involves dividing the total facility power by the collective load of IT equipment, as outlined in reference [9]. To gain a deeper understanding of the energy utilization during workload processing at BRT, Peshawar, please refer to the comprehensive breakdown illustrated in Figure 1.

In the realm of overall energy consumption, IT equipment stands out as the primary contributor and is a critical factor when assessing data center energy efficiency. Engineering optimizations have succeeded in reducing PUE through improved cooling techniques, virtualization, and refined power conversion designs for data centers [10]. These efforts are primarily directed towards reducing the overall energy consumption of data centers.

Some tech giants, such as Google, Microsoft, and Apple, are actively working to diminish their reliance on conventional energy sources, known as 'brown energy,' by embracing green power solutions like solar and wind energy. They also invest directly in the development of on-site renewable energy sources. However, complete reliance on renewable energy can be challenging due to its intermittent and variable nature. To address this challenge, data centers often employ a combination of green energy sources, such as solar panels, alongside conventional brown energy sources [11].

The paper contributes significantly in the following ways:

1. We tackle the challenge of minimizing the overall energy cost by considering workload for BRT, Peshawar, framing it as an optimization problem.
2. We present a solution to this workload distribution problem in the form of a linear programming.
3. After that, we introduce the RTLB algorithm, employing a greedy algorithm design method to address the problem.
4. By leveraging incoming user requests from BRT Peshawar and other real-world data, we conduct a comparative analysis to assess the effectiveness of RTLB against established workload allocation strategies. The outcomes demonstrate the superiority RTLB over existing techniques.

The remainder of the paper is structured as follows: Section II offers a comprehensive literature review summary. Section III outlines the problem's context and formulation. Section IV delves into the details of our linear optimization solution. Section V provides insights into our experimental setup, while Section VI presents a summary of the experimental findings that validate the effectiveness of our proposed algorithm. Lastly, Section VII concludes the paper and highlights potential areas for future research.

2. Literature Review

2.1 Energy Cost Minimization

To date, numerous studies have been conducted in the pursuit of developing techniques to reduce energy costs. The notion of minimizing energy expenses was initially introduced by [1], who explored the concept of reduction in the total usage of energy cost within the bulk market framework. They devised an analytical pricing optimizer that intelligently routed workloads to data centers with the minimum costs of energy utilization. In a different approach to addressing the same challenge, [2] proposed a model based on Integer Linear Programming (ILP) and resolved it using the simplex method. Their objective was to reduce the operating costs of distributed DCs significantly. However, both [1] and [2] did not take into account potential service delays or alternative power sources.

In pursuit of greener and more cost-effective data center operations, [4] introduced a workload allocation in data center using three-tier structure of interactive user requests for allocation. Their aim was to maximize the utilization of renewable energy sources while simultaneously reducing the total costs of energy of the data center. Nevertheless, this approach might introduce delays when workloads are moved across globally distributed data centers.

In a distinct effort, [5] developed a web based software that employs an online greedy base approach to optimize the usage of renewable energy sources while reducing the total operational cost of the data center. In cases where green energy sources are unavailable, the unallocated user request is assgined to DCs with the lowest energy consumption, while also guaranteeing that there is enough equipment capacity to manage the tasks effectively Their approach involved the utilization of actual user request data to every data center to simulate renewable energy production, and results indicated a significant reduction in brown energy usage.

To further mitigate energy expenditures and maximize the adoption of green power sources, [6] devised the algorithms based on power effectiveness. These energy efficient solutions were designed to optimize load distribution, thereby reducing energy costs and revenue losses due to delays in a linear combination. Meanwhile, [9] conducted a thorough analysis of the feasibility of data centers operating on green energy. They provided recommendations for the construction and operation of optimal green energy sources to achieve this ambitious goal.

In the realm of workload distribution across data centers, [14] outlined strategies based on linear programming techniques. Their research aimed to process user queries while staying within non-renewable energy budgets, meeting Service Level Agreements (SLAs), and curbing energy expenditures. Another approach, employed by [21], leveraged Global Load Balancing (GLB) to exploit variations in energy prices in DCs. This strategy directed more user requests to DCs with lower energy costs or a higher proportion of green energy.

However, it's worth noting that a lot of studies have primarily considered an delay intolerant workload. While [14] , [21], and [22] all delve batch workload processing.

2.2 GLB and Energy Efficiency

There is a substantial body of research focused on optimizing user request processing through the utilization of Global Load Balancing (GLB) to decrease the overall expenses incurred by data centers. The policies governing workload distribution, often referred to as geographical load balancing, form the cornerstone of various energy-saving initiatives within the realm of cloud computing.

Within the framework of GLB, front-end proxies receive user requests (incoming workloads), which are subsequently directed to geographically dispersed data centers (DCs) [19]. Consequently, Cloud Service Providers (CSPs) assume a pivotal role in the development and assessment of energy-efficient solutions designed to dynamically reduce electricity costs. Research studies have underscored the capacity of GLB to augment the utilization of renewable energy sources while curbing reliance on conventional power sources [20], [21], and [22].

Recent years have witnessed focused investigations, such as [20], addressing the formulation of problems related to GLB considering green energy sources, variable energy prices, and levels of server utilization. The above factors hold paramount importance in achieving different efficiency parameters associated with geographical load balancing. Addressing the pressing concern of reducing total energy costs, [20] introduced the concept of renewable energy based workload distribution strategy. The concept of this algorithm, aimed at mitigating data center operational expenses, has been extensively explored in important and pertinent researches [13], [16], [17], [21], [22], [31].

Furthermore, researchers like [18], [23], [27], [28], and [29] have primarily concentrated their efforts on contemporary challenges related to data center cost reduction. These scholars have delved into the intricacies of geographical load balancing under conditions of variable electricity prices, fluctuating customer demands, and the integration of on-site green energy sources.

It is important to note that one drawback associated with GLB is the potential for increased DC energy utilization due to accessibility of minimum energy price. This possibility, while economically advantageous, may have environmental implications, such as increased carbon emissions. As posited by [28], GLB can contribute to a reduction in brown energy usage if electricity costs are contingent upon the ratio of conventional electricity sources within the overall energy generation [32].

2.3 Renewable Energy

Numerous research work have studied strategies for the effective utilization of different renewable energy sources, such as Photovoltaic (PV) modules, with the aim of reducing the consumption of conventioal power sources for the execution of user requests in DCs [15], [24], [27], [29], [30], [32]. In the context of interactive workload distribution within data centers, [32] introduced an energy efficient algorithm to minimize the cost of the DC. This algorithm formulated as linear programming by taking into account factors like energy prices, availability of green energy and user requests. The proposed approach demonstrated the optimal workload distribution without considering any statistical data. However, it is worth noting that the proposed algorithm namely TTOA exhibited relatively slower convergence compared to the stochastic dual gradient approach.

In another endeavor, [31] devised a demand response technique aimed at incentivizing Internet Service Providers (ISPs) to adopt renewable energy sources, thereby reducing reliance on brown energy and decreasing carbon emissions. The role of storage units in the context of green energy utilization was emphasized by [5]. They asserted that small-scale storage, combined with geographical load balancing, plays a crucial role in the transition toward 100 percent renewable energy adoption. In small-scale data centers, Uninterruptible Power Supplies (UPSs) can be employed to store limited quantities of renewable energy. Moreover, [19] designed an energy efficient algorithm with the assumption that Internt Service Providers possess detail data regarding incoming workload and availability of renewable energy. They Identified that factors such as bandwidth costs, energy consumption, and delay are interrelated in every time slot.

3. Problem Setting

Table 1. Presents an overview of the symbols and ideas that will be employed throughout this research to enhance clarity and understanding.

Notations	Definitions
$t \in \{1, T\}$	Time slot index
$i \in \{1, N\}$	Data center index
$W(t)$	Total workload at time t
$w_i(t)$	Workload at DC i at time t
R_i^{max}	Upper bound renewable energy at DC i
$R_i(t)$	Green energy level at time t in DC i
$q_i(t)$	Electricity price at time t in DC i
S_i^{max}	Upper bound of servers at DC i
$S_i^{ac}(t)$	Total servers in active mode at t in DC i
$S_i^{in}(t)$	Total server in inactive mode at t in DC i
μ_i	Processing speed of workload at DC i
$PUE_i(t)$	Energy efficiency ratio at time t in DC i
$P_i^{IT}(t)$	Power of IT equipments at t in DC i

$P_i(t)$	Overall power utilization at t in DC i
$P_i^{ac}(t)$	Power consumption of active servers
$P_i^{in}(t)$	Power consumption of inactive servers
$C_i(t)$	Energy cost at DC i
$C(t)$	Overall energy cost
$\tau_{i^*}^{in}(t)$	Temperature within DC i^*
$\tau_{i^*}^{out}(t)$	Temperature outside DC i^*
α_{in}	Temperature threshold within DC (21℃)
α_{out}	Temperature threshold outside DC
$K_{i^*}^{Load}$	Data center cooling load
$B_{i^*}(t)$	Conventional energy at DC i^*

3.1 Problem Formulation

In this context of this research work, we make the assumption of having three separate data centers, each equipped with a specific number of uniform servers. These data centers are configured such that they rely on green energy as a secondary energy source, while brown energy functions as their primary energy supply. At each time instance denoted as 't' user tasks represented by $W(t)$ are directed to the global-LB, as illustrated in Figure 1. The primary objective of the Global Load Balancer (Global-LB) is to efficiently distribute user requests to the most suitable data center, with a focus on minimizing power utilization in DCs. After the selection of data center is chosen considering predefined parameters, the user requests is then passed on to the local-LB, as shown in Figure 2. Subsequently, the local proxy is responsible for routing the new workload to a computer system which has low utilization level.

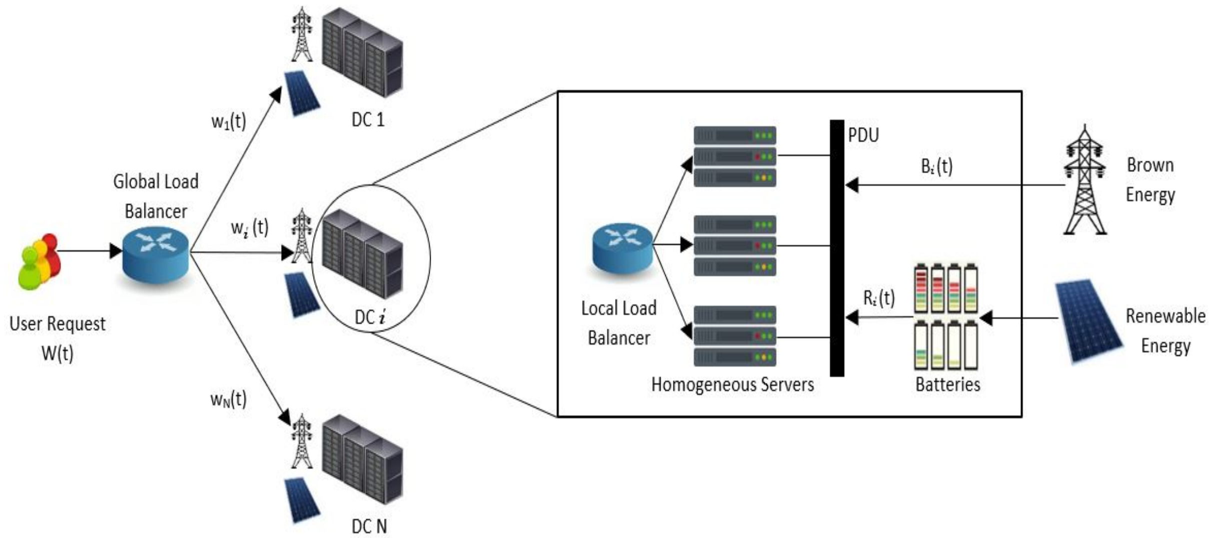

Figure 2. User Query Distribution and DC Architecture

3.2 Workload Model

Broadly speaking, incoming user requests can be categorized into two main types: batch and interactive, as noted in prior studies [26], [34]. In the context of our research, we specifically focus on characterizing the incoming workload as interactive. This designation is made based on the predominant nature of user requests within the context of BRT Peshawar, where a significant proportion of user tasks are relatively small in size and of an interactive nature. Furthermore, we make the assumption that the workload is indivisible and necessitates processing within a single data center. Consequently, we can consider that at time 't,' the workload denoted as $W(t)$ arrives at DC i.

$$\sum_{i=1}^{3} \lambda_i w_i(t) = W(t) \quad (1)$$

$$\sum_{i=1}^{3} \lambda_i = 1 \; where \; \lambda_i \in [0,1] \qquad (2)$$

The new user requests which indivisible in nature is denoted by λ_i, which takes values within the range of [0,1]. Constraints 1 and 2 are formulated to explicitly ensure the indivisibility of the incoming workload and its allocation to a single, unique data center. Each data center is equipped with a substantial number of servers to handle the processing requirements of the incoming workload. As a crucial condition, the total count of active servers, denoted as $S_i^{ac}(t)$, within a data center $'i'$ should not surpass the maximum server capacity, denoted as S_i^{max}. Consequently, we arrive at the following relationship:

$$S_i^{ac}(t) \le S_i^{max} \qquad (3)$$

3.3 Renewable Energy Generation Model

In the context of data centers, the predominant sources of electricity are typically renewable energy, such as solar panels, and conventional brown energy [21]. In our problem scenario, we specifically examined the utilization of Photovoltaic modules for renewable energy. The generation of renewable energy, denoted as $R_i(t)$, at data center $'i'$ at time $'t'$ can be quantified using the following model:

$$0 \le R_i(t) \le R_i^{max} \qquad (4)$$

The source of green power consistently maintain a positive value, and it is impossible to exceed its upper limit, as illustrated in eq. 3.4.

3.4 Power Consumption Model

Data center's power consumption is primarily attributed to two key factors: networking infrastructure and cooling mechanisms. Among these factors, power consumption of the cooling mechanisms notably surpasses that of the network equipment. This is largely due to the critical role played by maintaining an optimal temperature within data centers for the efficient operation of servers [28]. In our analysis, we choose to employ the metric of Power Usage Effectiveness (PUE) because it provides an accurate measure of the energy utilized by the cooling systems [21]. PUE represents the ratio between the total electricity consumed by the data center and the energy consumed by the network equipment. For the purposes of our research, we consider PUE values within the range of [1,2] [16], [33]. Specifically, the energy consumed by the data center's IT equipment is expressed as follows:

$$P_i^{IT}(t) = S_i^{ac}(t)\left[P_i^{in}(t) + P_i^{ac}(t)U_i(t)\right] \qquad (5a)$$

In data center $'i'$ at time $'t'$ the typical estimate of utilization level of active servers is denoted as $U_i(t)$, calculated as the ratio of the incoming workload, $w_i(t)$, to the product of the total count of active servers, $S_i^{ac}(t)$, and the service rate, μ_i. Furthermore, at time $'t'$ the total energy utilization within data center $'i'$ is represented as:

$$P_i(t) = PUE_i(t) \cdot S_i^{ac}(t)\left[P_i^{in}(t) + P_i^{ac}(t)\frac{w_i(t)}{S_i^{ac}(t)\mu_i}\right] \qquad (5b)$$

In our assumptions, we consider that DCs are auto-reliant when it comes to renewable energy. Additionally, we acknowledge that operational costs take precedence over the initial expenses associated with renewable energy installations. Consequently, we make the assumption that the generation of renewable power incurs no marginal cost, as indicated by previous studies [27], [35]. Accordingly, the calculation of the overall energy utlilization expenses across all DCs at timeslot $'t'$ can be approximated using the model given below:

$$C(t) = \sum_{i=1}^{N} q_i(t)\left[P_i(t) - R_i(t)\right] \qquad (6)$$

3.5 Optimization Problem of Load Balancing

In the context of this study, our main goal is to determine which data center $'i'$ at time $'t'$ possesses the more economical energy cost. The optimization problem presented in equation (7) revolves around the decision variable $C_i(t)$, with the consideration that λ_i ˙signifies the indivisibility of the workload.

$$min \sum_{t=1}^{T} \sum_{i=1}^{N} \lambda_i[C_i(t)] \qquad (7)$$

Subject to;

$$\sum_{i=1}^{N} \lambda_i w_i(t) = W(t) \quad \forall t \in [1, T] \tag{8}$$

$$\sum_{i=1}^{N} \lambda_i = 1,\ \lambda_i \in [0,1] \quad \forall t \in [1, T] \tag{9}$$

$$0 \leq S_i^{ac}(\mathrm{t}) \leq S_i^{max} \quad \forall i \in [1, N] \tag{10}$$

$$0 \leq R_i(t) \leq R_i^{max} \quad \forall i \in [1, N] \tag{11}$$

The incoming user requests $W(t)$ at timeslot t exclusively assigned to only one DC is explained in Constraints (8) and (9). Constraint (10) ensures that the number of active servers at data center $'i'$ does not exceed the upper limit in data center $'i'$. Constraint (11) guarantees the proper management of sources of renewable energy and cannot exceed maximum threshold.

4. Solution of the Optimization Problem

4.1 Workload Distribution using RTLB

The functionality of the proposed algorithm is elucidated in Algorithm 1

Algorithm 1: RTLB

$for\ \ t \in [1, T]\ \ do$
1. At the beginning of each time t, exaime the new user requests $W(t)\ \forall\ i\ \in [1 - N]$
- Data center $\boldsymbol{i}^*$Selection
2. Calculate $P_i(t), R_i(t), S_i^{ac}(t), S_i^{in}(t)$ to select the i^*

$$i^* = \min_{\arg i \in [1,N]} \sum_{i=1}^{N} \lambda_i [C_i(t)]$$

Considering the constraints (4.6a), (4.6b), (4.6c), and (4.6e)

3. Allocate the new user request to DC i^*
- Cooling Systems' Power Selection
4. $\boldsymbol{if}\quad \tau_{i^*}^{in}(t) > \alpha_{in}\ \&\ \tau_{i^*}^{out}(t) \geq \alpha_{out}\quad \boldsymbol{then}$
5. $K_{i^*}^{Load} \leftarrow R_{i^*}(t)$
6. $else$
7. $K_{i^*}^{Load} \leftarrow B_{i^*}(t)$
8. Update $S_i^{ac}(t), S_i^{in}(t)$ and $w_i(t)\quad \forall\ i\ \in [1, N]$

$\boldsymbol{end\ for}$

The pseudocode of our proposed algorithm for user request allocation and utilization of green power considering workload allocation strategy, all while adhering to various constraints. This algorithm is employed by Global-LB to manage user requests and minimize the overall operational expenses of DC.

In Line 1, essential information is collected at timlot t of every DC. Line 2 involves calculations aimed at solving the optimization problem. This includes assessing the status of inactive servers $S_i^{in}(t)$ and active servers $S_i^{ac}(t)$, green power level $R_i(t)$, and the overall power consumption of each DC $'i'$ at time $'t'$. Subsequently, these calculations are subjected to scrutiny against the constraints, namely, (4.6a), (4.6b), (4.6c), and (4.6e), of every DC and time slot $'t'$. Once the specified requirements are met, DC i^* is chosen.

The Data center $'i'$ already chosen in Line 3, a decision must be made regarding the power source, specifically whether DC $'i'$ should operate on conventional and green sources of power. The selected DC is contingent upon the fulfillment of certain conditions.

These conditions are sequentially examined from Lines 4 to 6. If the temperature within data center $\tau_{i^*}^{in}(t)$ of data center i^* at time $'t'$ exceeds the internal temperature threshold α_{in}, set at 21°C, and the external temperature $\tau_{i^*}^{out}(t)$ of DC i^* at time $'t'$ is at or above the threshold of the external temperature α_{out}, then load of the cooling system $K_{i^*}^{Load}$ at $'t'$ in i^* is allocated to green power. If in cases where aforementioned criteria (Lines 4 to 6) should not met then load of cooling system $K_{i^*}^{Load}$ of i^* at $'t'$ remains on brown energy $B_{i^*}(t)$.

Finally, in Line 8, the statuses of servers that are fall active mode $S_i^{ac}(t)$ and in active mode $S_i^{in}(t)$ for all DCs are combined and kept for subsequent iterations. This information is essential for processing the workload assigned to data center $'i'$ at time slot $'t'$.

5. Experimental Setup

We utilize hourly workload data from BRT Peshawar spanning thirty days, corresponding to a total of 720 time slots $T = 720$. The choice of an hourly time interval is made due to the dynamic nature of workload changes, which fluctuate as passengers flow in and out. We consider three geographically dispersed data centers, denoted as $N = 3$, with each data center housing a total of 15 servers. We assign a service rate (μ_i) of 1.00 to each individual server. This service rate value aligns with similar assumptions made in prior research works [23], [25].

For the Power Usage Effectiveness (PUE), we adopt a value of 1.20 [22] as the default setting. Within each data center 'i,' the network servers consume the specified energy in DC. A single server consumes 120 Watts in active position, whereas 60 Watss in their inactive mode, as indicated in previous studies [32-35].

5.1 Workload Description

We utilize an actual workload trace from BRT Peshawar for the month of March 2022, encompassing a total of 720 hours over a span of 30 days. The daily influx of workload related to the card recharges of BRT, the card sales of BRT, the QR sales of BRT, and total travels is visually presented in Figure 3. Each line within the graph below corresponds to the user requests data associated with a speicifed parameter. It's important to note that the workload depicted in the graph represents only 1% of the entire workload. The incoming workload encompasses the following elements:

- The Card Recharges of BRT
- The Card Sales of BRT
- The QR Sales of BRT
- Total Travels

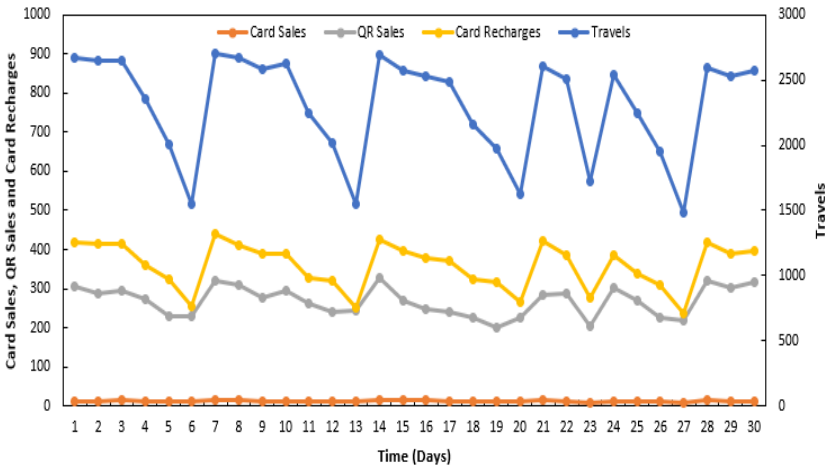

Figure 3. Daily Incoming Workload

5.2 Electricity Prices

As previously stated, BRT Peshawar has three data centers situated within the same city. Consequently, the electricity rates for all three data centers remain consistent. For our numerical assessments, we take into account the commercial electricity rates provided by the Water and Power Development Authority (WAPDA). The key differentiating factor among these data centers is the energy resource they rely upon.

5.3 Renewable Energy Description

It's worth observing that green energy sources exhibit a degree of volatility and intermittency. As a result, many data centers incorporate energy storage devices (ESD) to ensure a consistent supply of renewable energy [11]. In our assumptions, we assumed the renewable energy level at time t denoted as $R_i(t)$, constitutes approximately 26% of the overall green power utilized by the DCs.

5.4 Benchmark Algorithms

Our proposed RTLB algorithm is subject to comparison with two distinct workload distribution approaches, as detailed below:

***B*1**: Energy storage devices-Oblivious (ESD-Oblivious): In this approach, the allocation of workload is determined based on the nearest data center, factoring in brown energy availability, workload processing delays, electricity pricing, and the total count of active servers. However, this approach does not take into account on-site renewable energy sources.

***B*2**: Energy storage devices-Aware (ESD-Aware): In ESD-Aware incoming user request allocation policy, as introduced in [14], involves the distribution of user requests among data centers while considering energy prices, the presence of green energy, brown energy availability, the presence of ESD, the total servers in active mode, with the fulfillment of a predefined limitations for every time t.

6. Experimental Result

Evaluating RTLB's performance based on user requests from BRT Peshawar and the existing electricity pricing, we find the following:

The overview of the energy costs incurred by the data centers under different user requests distribution policies is depicted in Figure 4. Notably, B1, which without considering green energy sources into account for user requests allocationis and solely relies on brown energy, results in a considerably higher electricity cost. In contrast, B2 achieves relatively minimum power expenditures by leveraging sources of renewable power and energy storage devices. The differences in the outcomes of these workload distribution strategies can be ascribed to the algorithms' functioning, where workloads are directed to data centers with the lowest current workload without optimization strategies.

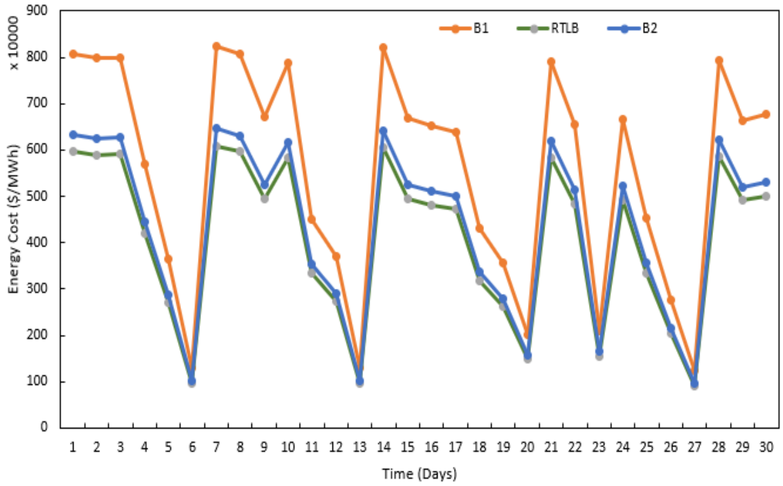

Figure 4. Power Consumption of RTLB and Benchmark Techniques

RTLB stands out as a highly effective solution among the three workload distribution techniques considered. It consistently achieves substantial reductions in the operational expenses of data centers. RTLB accomplishes this by predominantly relying on brown energy as the primary energy source, while allocating green energy exclusively for the CRAC (Computer Room Air Conditioning) system, and utilizing IT equipment in conjunction with brown energy, except those hours where renewable energy utilized.

Table 2. The Analysis of Energy Consumption

Comparison Factor	Improvement of RTLB over	
	B1	**B2**
Energy Cost	25.98%	6%

Table 2 provides a comparative summary of RTLB's performance in relation to B1 and B2. The data in the table clearly indicates that RTLB surpasses B1 by 25.98 percent and outperforms B2 by 6 percent. Consequently, RTLB leads to a substantial reduction in the electricity costs incurred by DC in BRT.

7. Conclusions

The power utilization within data centers holds paramount importance for Internet service providers (ISPs). ISPs employ various methods for task distribution and power management to effectively curtail their overall energy expenditures. In this study, we addressed the pressing challenge of reducing the overall energy costs while considering factors such as servers in active mode, green power level, and brown energy. To tackle this issue, we introduced a an online energy efficient algorithm based on greedy approach namely Renewable and Temperature-aware Load Balancing (RTLB), to distribute the incoming user requests. The implementation of our strategy resulted in a substantial minimization in the overall expenses of the data centers in BRT Peshawar.

In our future endeavors, we aim to enhance the recommended model by incorporating additional components, including considerations of bandwidth costs and the utilization of renewable energy through energy storage devices (ESDs). Furthermore, we plan to conduct extensive analyses of heterogeneous server-based data centers dispersed across various locations to estimate potential cost savings.

References

1. Qureshi, A.; Weber, R.; Balakrishnan, H.; Guttag, J.; and Maggs, B. "Cutting the electric bill for internet-scale systems," Proceedings of the ACM SIGCOMM, 2009 conference on Data communication pp. 123-134.
2. Rao, L.; Liu, X. Xie, L.; and Liu, W. "Minimizing electricity cost: optimization of distributed internet data centers in a multi-electricity-market environment," in INFOCOM, 2010 Proceedings IEEE, 2010, pp. 1-9.
3. Barroso, L. A. "The price of performance: An economic case for chip multiprocessing," Queue, vol. 3, no. 7, pp. 48-53, 2005.
4. Toosi, A. N.; Qu, C.; De Assunção, M. D.; and Buyya, R. "Renewable-aware geographical load balancing of web applications for sustainable data centers," Journal of Network and Computer Applications, 2017, vol. 83, pp. 155-168.
5. Liu, M.; Lin, A.; Wierman, S. H.; and Andrew, L. L. "Geographical load balancing with renewables," ACM SIGMETRICS Performance Evaluation Review, 2011, vol. 39, no. 3, pp. 62-66.
6. Zheng, X.; and Cai, Y. "Energy-aware load dispatching in geographically located internet data centers," Sustainable Computing: Informatics and Systems, 2011, vol. 1, no. 4, pp. 275-285.
7. Bertsekas, D. P. "Dynamic programming and optimal control 3rd edition, volume ii," Belmont, MA: Athena Scientific, 2011, vol. 2, pp 235-238.
8. Georgiadis, L.; Neely, M. J.; and Tassiulas, L. "Resource allocation and cross-layer control in wireless networks," Foundations and Trends® in Networking, 2066, vol. 1, no. 1, pp. 1-144.
9. Neely, M. J. "Stochastic network optimization with application to communication and queueing systems," Synthesis Lectures on Communication Networks, 2010, vol. 3, no. 1, pp. 1-211.
10. Xu, D. and Liu, X. ''Geographical Trough Filling for Internet Datacenters,'' in Proc. IEEE INFOCOMMini-Conf., 2012, pp. 2881-2885.
11. Lin, M.; Wierman, A.; Andrew, L. L. H.; and Theresa, E. "Dynamic right-sizing for power-proportional data centers," IEEE/ACM Transactions on Networking, 2012, vol. 21, no. 5, pp. 1378-1391.
12. Polverini, M.; Cianfrani, A.; Ren, S.; and Vasilakos, A. V. "Thermal-aware scheduling of batch jobs in geographically distributed data centers," IEEE Transactions on Cloud computing, 2013, vol. 2, no. 1, pp. 71-84.
13. Khalil, M. I. K.; Ahmad, I.; and Almazroi, A. A. "Energy efficient indivisible workload distribution in geographically distributed data centers," IEEE Access, 2019, vol. 7, pp. 82672-82680.
14. Gu, C.; Li, Z.; Liu, C.; and Huang, H. "Planning for green cloud data centers using sustainable energy," IEEE Symposium on Computers and Communication (ISCC), 2016, pp. 804-809.
15. Rao, L.; Liu, X.; Xie, L.; and Liu, W. "Minimizing electricity cost: Optimization of distributed internet data centers in a multi-electricity-market environment," Proceedings IEEE INFOCOM, 2010, pp. 1-9.
16. Wang, H.; Huang, J.; Lin, X.; and Mohsenian-Rad, H. "Proactive demand response for data centers: A win-win solution," IEEE Transactions on Smart Grid, 2015, vol. 7, no. 3, pp. 1584-1596.
17. Xu, Y.; Zhan, Y.; and Xu, D. "Building cost-efficient cloud data centers via geographical load balancing," IEEE Symposium on Computers and Communications (ISCC), 2017, pp. 826-831.
18. Khalil, M. I. K.; Ahmad, I.; Shah, S. A. A.; Jan, S.; and Khan, F. Q. "Energy cost minimization for sustainable cloud computing using option pricing," Sustainable Cities and Society, 2020, vol. 63, p. 102440.
19. Aust, B.; and Horsch, A. "Negative market prices on power exchanges: Evidence and policy implications from Germany," The Electricity Journal, 2020, vol. 33, no. 3, p. 106716.
20. Zhang, Y.; Deng, L.; Chen, M.; and Wang, P. "Joint bidding and geographical load balancing for data centers: Is uncertainty a blessing or a curse? " IEEE/ACM Transactions on Networking, 2018, vol. 26, no. 3, pp. 1049-1062.
21. Khalil, M. I. K.; and Taj, A. "Factors Affecting the Efficacy of Software Development A Study of Software Houses in Peshawar, Pakistan." Academy of IRMBR, 2021, vol. 9, no. 3, pp. 423-431.
22. Ahmad, I.; Yousaf, M.; Yousaf, S.; and Ahmad, M. O. "Fake news detection using machine learning ensemble methods," Complexity, 2020, vol. 2020.
23. Huisman, R.; Huurman, C.; and Mahieu, R. "Hourly electricity prices in day-ahead markets," Energy Economics, 2019, vol. 29, no. 2, pp. 240-248.
24. Boogert A.; and Dupont, D. "When supply meets demand: the case of hourly spot electricity prices," IEEE Transactions on Power Systems, 2018, vol. 23, no. 2, pp. 389-398.
25. Bello, A. Renesas, J.; Muñoz, A.; and Delgadillo, A. "Probabilistic forecasting of hourly electricity prices in the medium-term using spatial interpolation techniques," International Journal of Forecasting, 2016, vol. 32, no. 3, pp. 966-980.
26. Ahmad, I.; Khalil, M. I. K.; and Shah, S. A. A. "Optimization-based workload distribution in geographically distributed data centers: A survey," International Journal of Communication Systems, 2020, vol. 33, no. 12, p. e4453.
27. Guo, Y.; Gong, Y.; Fang, Y.; Khargonekar, P. P.; and Geng, X. "Energy and network-aware workload management for sustainable data centers with thermal storage," IEEE Transactions on Parallel and Distributed Systems, 2013, vol. 25, no. 8, pp. 2030-2042.
28. Zhang, H.; Kai, C.; Wei, B.; Dongsu, H.; Chen, T. et al., "Guaranteeing deadlines for inter-data center transfers," IEEE/ACM Transactions on Networking, 2016, vol. 25, no. 1, pp. 579-595.
29. Polo, J.; Yolanda, B.; David, C.; Malgorzata, S.; Ian, W. et al., "Deadline-based MapReduce workload management," IEEE Transactions on Network and Service Management, 2013, vol. 10, no. 2, pp. 231-244.

30. Paul, D.; and Zhong, W.D. "Price and renewable aware geographical load balancing technique for data centers," in Information, Communications and Signal Processing (ICICS) 2013 9th International Conference IEEE, 2013, pp. 1-5.
31. Ahmad, I.; Ahmad, M. O.; Alqarni, M. A.; Almazroi, A. A.; and Khalil, M. I. K. "Using algorithmic trading to analyze short-term profitability of Bitcoin," PeerJ Computer Science, 2021, vol. 7, p. e337.
32. Khalil, M. I. K.; Shah, S. A. A.; Khan, I. A.; Hijji, M.; Shiraz, M.; and Shaheen, Q. "Energy cost minimization using string matching algorithm in geo-distributed data centers," Computers, Materials, and Continua, 2023, vol. 75, no. 3, pp. 6305-6322.
33. Muhammad, D.; Ahmad, I.; Khalil, M. I. K.; Khalil, W.; and Ahmad, O. A. "A generalized deep learning approach to seismic activity prediction," Applied Sciences, MDPI, 2023, vol. 13, p. 1698.
34. Khalil, M. I. K.; Shah, S. A. A.; Taj, A.; Shiraz, M.; Alamri, B.; Murawat, S.; and Hafeez, G. "Renewable aware geographical load balancing using option pricing for energy cost minimization in data centers," Processes, MDPI, 2022, vol. 10, no. 10, p. 1983.