

Lyapunov-Based Energy Optimization for Deadline-Constrained Tasks across Geo-Distributed Data Centers

Muhammad Imran Khan Khalil

Department of Computer Science, University of Engineering & Technology Peshawar
25000, Pakistan

Ahmad Ali

Department of Computer Science, University of Engineering & Technology Peshawar
25000, Pakistan

Syed Adeel Ali Shah*

Department of Computer Science, University of Engineering & Technology Peshawar
25000, Pakistan Email: adeel@uetpeshawar.edu.pk

Alauddin

Department of Computer Science, University of Engineering & Technology Peshawar
25000, Pakistan

Amer Taj

Department of Computer Science, University of Engineering & Technology Peshawar
25000, Pakistan

Izaz Ahmad Khan

Department of Computer Science, Bacha Khan University, Charsadda 24420,
Pakistan

ABSTRACT

Electricity cost is one of the key contributors to the data center's overall operational cost. Many researchers worked on reducing the overall electricity cost of the data center (DC) operations. In this paper, we examine the problem of minimizing the overall electricity cost of geographically distributed data centers while fulfilling the deadline requirement of the delay-tolerant workloads under time-varying system dynamics. One way of reducing the energy cost of the DCs is to exploit the dynamic electricity prices across time and location for scheduling workloads among geographically distributed data centers. Based on the Lyapunov Optimization technique, we present an online algorithm for scheduling delay-tolerant workloads called Strict Deadline-aware online Algorithm (SDA). A delay-tolerant workload means that the workload can be delayed up to a specified amount of time. The main aim of our work is to minimize the overall electricity cost of the operator while completing the workload within its acceptable deadline. The basic idea of the algorithm is to forward the workload arriving at each time slot to a specific data center considering the currently offered prices of electricity, delay of current

workload, total unprocessed workload, and respecting other sets of constraints. Experiments were conducted considering multiple data centers using real-world workload traces and power rates. The results show that SDA outperforms existing workload distribution approaches to reduce the operator's overall power costs. Aside from lowering average power costs, the suggested algorithm completes all tasks within the deadline.

Keywords: Lyapunov Optimization technique, Geographical Load Balancing, Geographically Distributed Data Centers, Dynamic Electricity Prices, Energy Efficiency.

Introduction

Large-scale data centres are expanding day by day to meet the increasing demand for IT applications and their services, such as online gaming, web search, data analytics, etc. Many cloud service providers (CSPs) such as Google, Microsoft, and Facebook construct multiple geographically distributed data centers worldwide to deliver their services efficiently. Tens of megawatts of electricity are required to keep the data center active [1-4]. Every data center consumes significant power, resulting in high electricity bills for the CSPs. As a result, these large companies pay millions of dollars of annual charges for electricity costs. According to an estimation, the electricity cost constitutes about 30–50 percent of the operational expenses of the data centers [5-7].

Several researchers focused on the cost minimization of these geographically distributed data centers during workload processing. The research work for electricity cost minimization can be classified into two broad categories: hardware-based techniques such as chip multiprocessor [8-12], dynamic voltage and frequency scaling [13-15], and software-based strategies, i.e., active power management [16], visualization [17], proactive demand response [18], and incoming workload distribution to the appropriate data center to reduce the electricity cost of these operators.

The data center's workload can be broadly classified into two categories: a *delay-tolerant* workload that can be delayed up to a specified amount of time and a *delay-intolerant* workload that is the most time-sensitive data and needs immediate processing within fixed time constraints [19-24]. One key feature of the data center is processing the incoming workload within strict deadlines. The deadlines of the workload are either hard or soft. A hard deadline means the workload execution is useless once late. In contrast, a soft deadline means the workload processing degrades after the deadline's expiration, which affects performance [25-31]. Thus, meeting the deadline requirements during workload processing in geographically distributed data centers is essential.

In the preceding approach for energy cost minimization during workload processing, the incoming user request with a given deadline arrives at the workload aggregator at each time slot. If the user requests are processed within the set deadline, they are accepted and permitted to join the system [32-36]. The user request cannot be completed within the user-specified deadline in very few situations. In such cases, the user is offered the next nearest possible and available timeslot, and the Service Level

Agreement Violation (SLAV) occurs. Upon accepting the proposed timeslot, the system allows the workload with the new agreed deadline. The workload is executed within its deadline at a suitable time at the local (data center level) level. After that, the workload is forwarded to the data center with the minimum unfinished workload and the minimum per unit electricity cost [37-41].

This paper proposes a new delay-tolerant workload distribution strategy with specified deadlines to minimize the data center's electricity cost. This paper uses the Lyapunov Optimization technique to present an online algorithm for scheduling delay-tolerant workloads called Strict Deadline-aware online Algorithm (SDA). SDA ensures that the incoming workload must be executed entirely within specified deadlines. SDA considers deadline-based jobs having release time and job volume or length. The jobs can be postponed by a certain amount of time.

The paper's significant contributions and structure are summarized as follows. Section 2 presents a brief overview of the current literature. In Section 3, we consider the delay-tolerant workload and formulate the geographical load balancing as an optimization problem to minimize the average electricity cost within its agreed deadline. The solution to the optimization problem is provided using the Lyapunov optimization technique. This technique works without the availability of future data (electricity cost and workload). Section 4 experiments on a dataset of real-world workload traces and per-unit electricity prices to evaluate the proposed algorithm against the benchmark algorithms. The findings confirm that the proposed algorithm performs well compared to benchmark algorithms. Besides, the SDA completes all the workloads within its specified assigned deadlines. Section 5 concludes the work and provides direction for future research work.

Literature Review

Many researchers are motivated toward energy-efficient data centers to reduce the average electricity cost of the data centers. Researchers have worked on various aspects, such as using stored energy, renewable energy, thermally aware scheduling, etc., to minimize electricity prices [42-45]. The researchers used variations in electricity prices at different times and locations. Some of the work done in this direction is mentioned below briefly:

Chong et al. [13] proposed an online algorithm based on the Lyapunov Optimization technique for minimizing electricity costs by utilizing the storage device's uninterrupted power supply (UPS). The intuition behind the algorithm is to charge the battery when prices in the market are lower and use it when electricity prices are high. The main limitation is that the battery has limited capacity and can only be used for a limited time. Secondly, this technique reduces the average delay of the overall workload. Khalil et al. [8] considered the problem of minimizing the electricity cost of data centers by utilizing renewable energy. The workload is distributed among data centers located in different localities based on the availability of renewable energy. The major problem with this approach is that one can only rely partially on renewable energy due to unknown workload amounts and unpredictable weather conditions.

Lei et al. [19] considered the problem of scheduling workload among distributed data centers using renewable energy. They proposed an online algorithm to minimize the electricity cost and average delay. The proposed algorithm reduces the electricity bill

by dynamically turning servers on/off while considering the fixed electricity price over time. However, in practice, the cost varies with time. Paul et al. [28] proposed an online algorithm named GreFar for scheduling batch jobs among various data centers. The algorithm aimed to minimize the energy cost and allocate the available resources fairly among different requests while satisfying the average delay and maximum inlet temperature constraints. The algorithm was compared with an offline algorithm that knew future input. A simulation-based technique evaluated the performance of the algorithm. The main drawback, or limitation, is that the GreFar minimizes the average delay and does not guarantee the completion of each job within their deadlines.

Other works in the domain include

Moving toward greening Geographical Load Balancing [46],
Minimizing the cost in the multi-electricity market environment [47],
Proactive demand response for data centers [48], and
Cutting down on electricity prices for interactive workload [49].

Problem Setting and Proposed Solution

Because of unpredictable future workload arrivals and electricity prices, the workload distribution among multiple data centers to minimize electricity costs is an open research problem. Even if they were available ahead of time, tackling the challenge of reducing electricity costs would be hampered by the curse of dimensionality [50]. Before discussing the problem statement, the various notations and definitions used in this work are presented in Tab. 1.

Table 1: Notations

Notations	Definitions
$t \in \{1, T\}$	Index for timeslot
$i \in \{1, N\}$	Index for datacenter
$W(t)$	Total workload arrived at global load balancer at timeslot t
W_{max}	The maximum workload that can arrive at the global load balancer at timeslot t
$a_j(t)$	Arrival time of workload arrived at timeslot t
$d(t)$	Deadline of workload arrived at timeslot t
$C_i(t)$	Per unit Electricity cost at DC i at timeslot t
$C_{min,i}$	Minimum per unit electricity cost at DC i at any timeslot t
$C_{max,i}$	Maximum per unit electricity cost at DC i at any timeslot t
$P_i(t)$	Total power received from grid at DC i at timeslot t
$U_i(t)$	Unfinished workload at DC i at timeslot t
U_{max}	Maximum unfinished workload at timeslot t
V	Trade-off variable b/w Cost & Delay
V_{max}	Maximum value of V

Problem Formulation

We assume multiple data centers in various geographical areas. Each data center comprises thousands of identical servers and is exclusively fueled by brown energy. Decisions are made at two levels at each timeslot: the global and local levels. The workload arrives at the aggregator at the global level, and from there, the suggested algorithm routed it to one of the data centers. At the local (datacenter level), the proposed method determines the quantity of power acquired from the grid to reduce the operator's overall electricity cost while meeting certain limitations. The fundamental concept is to take advantage of the delay provided by each task and attempt to execute the workload at a time and location where the power price is low. Nevertheless, it must be assured that a maximum number of jobs are done within their deadline. Fig. 1 depicts the entire system model.

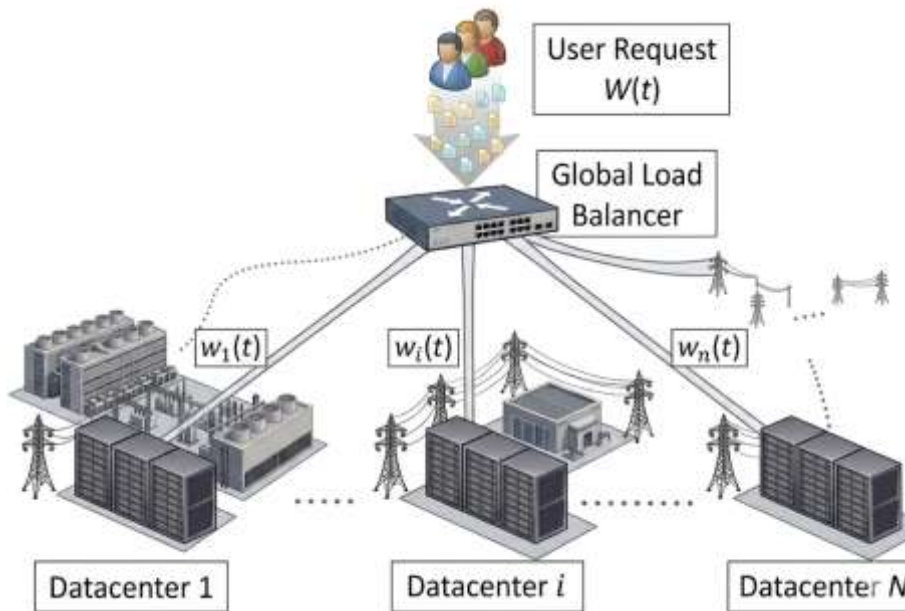


Figure 1: System Model

Problem Setting

The workload is divided into two categories: delay-tolerant and delay-intolerant. In this research, we focused on delay-tolerant or deadline-based jobs only. A delay-tolerant workload will have an arrival time ($a_j(t)$), job volume or job length, and a deadline ($d(t)$). The delay-tolerant jobs can be postponed by a certain amount of time.

Let $W(t)$ represent the amount of workload that arrived at the aggregator at time t .

$$\min_{arg\ i \in [1, N]} DC\_Cost[t, i] = W(t) \quad (1)$$

Constraint (1) ensures that at each time, the incoming workload should be assigned to that data center with the minor backlog and per-unit cost of electricity at that time. $U(t)$ is used to represent the unfinished workload at the time slot t . The unfinished

workload at that instant ($t + 1$) is represented as follows:

$$\begin{aligned} &U[t + 1, i] \\ &= U[t, i] + W[t, i] \\ &- P[t, i] \end{aligned} \quad (2)$$

The $W(t)$ statistics: workload is assumed to take values from a set containing non-negative finite values. However, the incoming workload is not influenced by the previous decisions and will take random values from the set W . W_{max} represents the maximum value that an incoming workload can bring. Thus, we have:

$$\begin{aligned} &0 \leq W(t) \\ &\leq W_{max} \end{aligned} \quad (3)$$

We assume that the incoming workload at each timeslot t will be greater than or equal to zero and less than or equal to W_{max} .

3.1.2 Brown Energy Model

Two commonly used sources to power data centers are brown energy and on-site renewable power generation (e.g., solar or wind) [51]. In this work, we considered brown power generation received from the grid. Let $p_i(t)$ denote the energy obtained from the grid for the data center i at a given time t .

$$\begin{aligned} &0 \leq p_i(t) \\ &\leq p_i^{max} \end{aligned} \quad (4)$$

Constraints (4) ensure that energy levels should always be non-negative and cannot exceed the maximum capacity.

We assume that peak power drawn from the grid at timeslot t must be greater than or equal to the maximum workload that can arrive at any timeslot t . That is:

$$\begin{aligned} &P_{peak} \\ &\geq W_{max} \end{aligned} \quad (5)$$

3.1.3 Electricity Price Model

$C(t)$ represents the per-unit cost of power drawn from the grid in time slot t . The cost mainly depends upon two factors—the amount of power drawn from the grid and the auxiliary state variable $S(t)$ [51]. For example, the per-unit cost will increase as the amount of power drawn from the grid increases. Similarly, the cost will vary with the state variable $S(t)$ that captures parameters such as time of day, the identity of the utility provider, etc., C_{min} representing the minimum per unit of electricity cost while C_{max} representing the maximum per unit of electricity cost. Therefore, we have:

$$\begin{aligned} &C_i^{min} \leq C_i(t) \\ &\leq C_i^{max} \end{aligned} \quad (6)$$

Constraints (6) ensure that per-unit costs should always be greater than or equal to the minimum cost and cannot exceed the maximum cost in any situation. However, C_i^{min} can be negative in some situations [52,53].

Optimization Problem

The main goal of the algorithm will be to minimize the overall cost of the operator by forwarding the incoming workload to a single data center among the various data centers located at different locations while trying to satisfy the deadline constraint of each workload. We capture this objective through the following optimization problem:

$$\min \sum_{t=1}^T \sum_{i=1}^N [P_i(t) * C_i(t)] \quad (7)$$

Subject to;

$$0 \leq W(t) \leq W_{max} \quad (7a)$$

$$\min_{arg i \in [1, N]} DC\_Cost[t, i] = W(t) \quad (7b)$$

$$P_{peak} \geq W_{max} \quad (7c)$$

$$0 \leq d_i(t) \leq D^{max} \quad \forall i \in [1, N] \quad (7d)$$

$$0 \leq p_i(t) \leq p_i^{max} \quad \forall i \in [1, N] \quad (7e)$$

Constraint (7a) ensures the lower bound of workload should be non-negative while the upper bound provides that the workload should not exceed the maximum allowed workload at timeslot t . (7b) ensures that at each time t , the incoming workload $W(t)$ assigned to that data center with the minor backlog and per-unit cost of electricity at that time. Constraint (7c) ensures that the peak power taken from the grid must be greater than or equal to the maximum workload that can arrive at any time t . Constraint (7d) ensures that the incoming workload is processed within the agreed deadline. (7e) ensures that the amount of energy received at each timeslot t will be within the specified lower and upper bounds.

Optimization Problem Solution

This section presents the solution to the optimization problem discussed in Section 3.2. A strict deadline algorithm based on the Lyapunov optimization technique is used in this paper to minimize the electricity cost by completing the workload within its deadline.

Strict Deadline Aware Algorithm (SDA)

To solve the optimization problem in (7), a Strict Deadline aware Algorithm (SDA) based on the Lyapunov optimization technique (Algorithm 1) is proposed. The proposed algorithm is comprised of two steps. In the first step, the workload arrived at the aggregator when instant t was assigned to that data center, which had a minor backlog plus the per-unit cost of electricity at that time slot. In the second step, at each data center, the amount of power to be drawn from the respective grid is decided while keeping in view the fulfillment of the deadline, reduction of electricity cost, and other constraints. The workings of the proposed algorithm are provided in Algorithm 1.

Algorithm 1: SDA to solve the optimization problem

Step 1: $U[1, i] = 0$

Step 2: **for** $t \leftarrow 1$ **to** T **do**

```
Step 3: Read  $W[t]$ 
Step 4:   for  $i \leftarrow 1$  to  $N$  do
Step 5:   Read  $C[t, i]$ 
Step 6:   Compute  $DC\_Cost[t, i] = U[t, i] + V * C[t, i]$ 
          i.e.  $V_{max} = \frac{U_{max}}{(C_{max} - C_{min})}$ 
Step 7:   Endfor
Step 8:   Find  $i^* = \min_{i \in [1, N]} DC\_Cost[t, i]$ 
Step 9:   Assign the incoming workload  $W[t]$  to the data center  $i^*$ 
Step 10:  Enclosed in for loop till the last line
Step 11:  Compute  $Q[t, i] = (U[t, i] + W[t, i]) - (V * C[t, i])$ 
Step 12:  if  $Q[t, i] \geq 0$ 
Step 13:   $P[t, i] = \min(power\_peak, W[t, i] + U[t, i])$ 
Step 14:  elseif (Any deadline == current slot)
Step 15:
 $P[t, i]$  = power required for the workloads whose deadline has reached
Step 16:  else
Step 17:   $P[t, i]$ 
Step 18:  Compute  $U[t + 1, i] = U[t, i] + W[t, i] - P[t, i]$ 
Step 19:  Endfor
```

A description of the algorithm: At instant 1, as no backlog exists, the initial value of the backlog variables of every data center, i.e., are set to zero. The remaining proposed algorithm is comprised of two parts. One is termed at the global level, while the second is at the local level.

Global Level: First, the workload $W(t)$ arrives at the aggregator at each instant. After that, the per-unit cost of electricity offered at each data center is read via $C[t, i]$. Each data center's cost for that instant is calculated based on backlog and per-unit cost. The data center with the lowest price is selected for forwarding the incoming workload. Before the workload is forwarded to the desired data center, the execution priority or scheduling location for the incoming workload is determined based on the user-assigned deadline. If the workload could be served within its specified deadline, the incoming workload would be accepted by default. The proposed algorithm tries to schedule the workload within its specified deadline. However, if the execution of the workload is not possible within its specified deadline, then the next possible scheduled time/ deadline is intimated to the user. If the user agrees to that deadline, the workload is accepted and entered into the system with the new accepted deadline. However, if the user disagrees with the offered deadline, the workload is dropped from the system. At this stage, the proposed algorithm is in a position to guarantee that the accepted workload will be executed 100% within its assigned deadline. The workload is then assigned to the scheduling location at that data center with the smallest cost value.

Local Level: At the local level, the following process is followed for each data center at each time instant: If the backlog is more significant than $VC[t, i]$, then the minimum of peak power or the power equals the backlog plus the current workload assigned at that instant to that data center plus whatever is acquired from the corresponding grid station. However, if the backlog is smaller or equal to $VC[t, i]$ and the deadline of some workloads has been reached, the power required for the workloads whose deadline has been reached is taken from the grid, and those workloads are executed. On the other hand, if the backlog is smaller or equal to the l to $VC[t, i]$, then no power is taken from the grid. After that, the backlog variables are updated for the next time instant.

Our problem setting significantly differs from other models proposed in the literature [54,55]. We have considered that each workload has a strict deadline, and the algorithm must try to complete maximum workloads within its deadline. To our knowledge, we are the first to propose a strict deadline-based workload execution.

Experiments and Evaluation

Experimental Setup

For experimental purposes, we considered a workload comprised of Wikipedia data for one month [21]. As the electricity prices vary hourly, the single time slot is set to one hour. The overall experiment is run for $T = 720$ time slots, equal to one month. Three data centers are assumed to be located in different geographical locations in Canada. Those locations are in Illinois, Utica, NY, and Ontario. Each data center consists of 15,000 servers.

Workload Description

The actual workload requests made at Wikipedia during September and October 2007, over 720 hours, are used in the experiment. The workload received during the abovementioned period is depicted in Fig. 2. The x-axis represents the time slot in an hour in the figure, while the y-axis represents the incoming workload in millions.

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

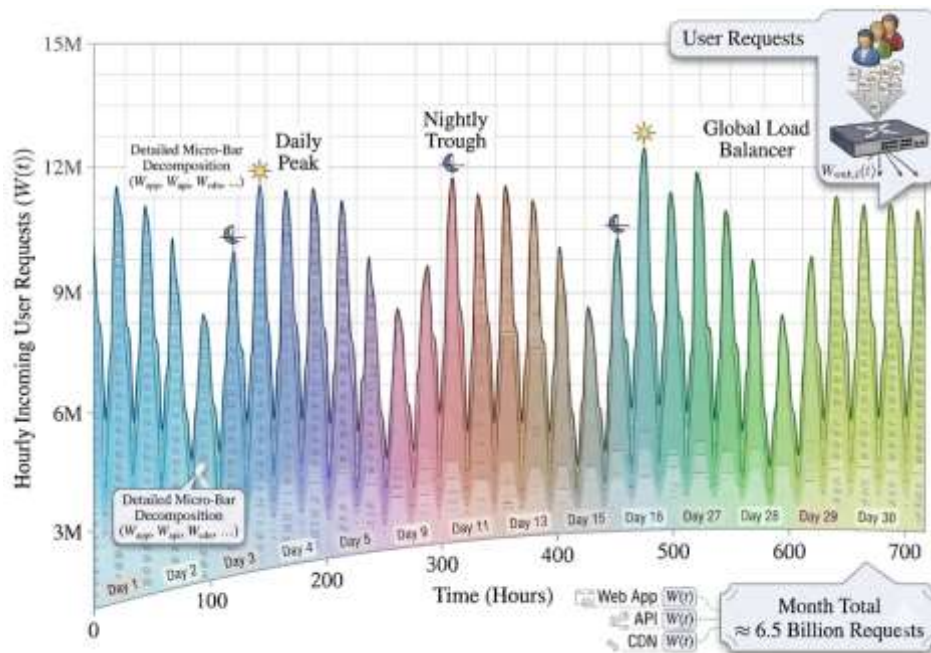


Figure 2: Hourly workload

Workload Delay

As we deal with a delay-tolerant workload, each workload arriving at a particular time slot is assigned a random delay between 1 and 24. Although the delay is randomly assigned for experimental purposes, the proposed algorithm will work perfectly on the actual delay given by the user. Fig. 3 shows the number of workloads assigned to a specific amount of delay. In the figure, the x-axis represents the delay range in an hour, while the y-axis represents the number of workloads. The minimum time to complete specific workloads is 1 hour, while some workloads could have a maximum delay of 24 hours.

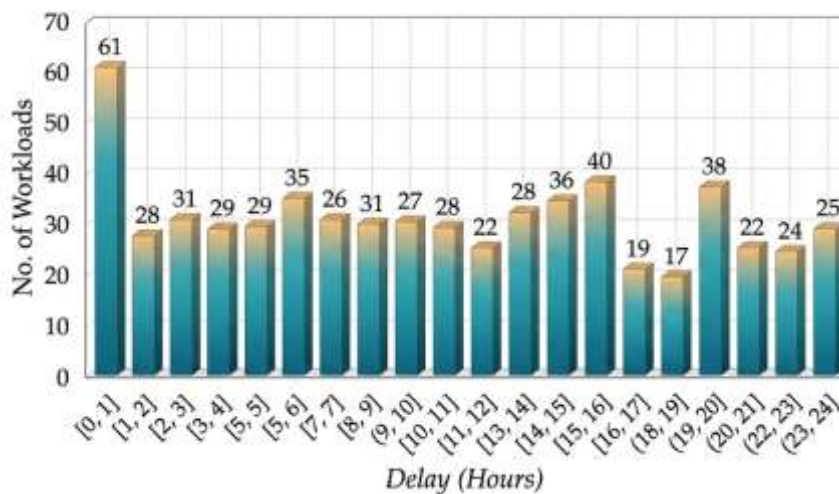


Figure 3: No. of workloads a specific delay is assigned

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

Electricity Prices

The dataset regarding electricity prices $C(t, i)$ for the three selected distributed data centers was obtained from the publicly available data at [21–24]. The electricity prices from September 18, 2017, to October 17, 2017, are depicted in Fig. 4. The electricity prices obtained are hourly and are in $\$/MWh$.

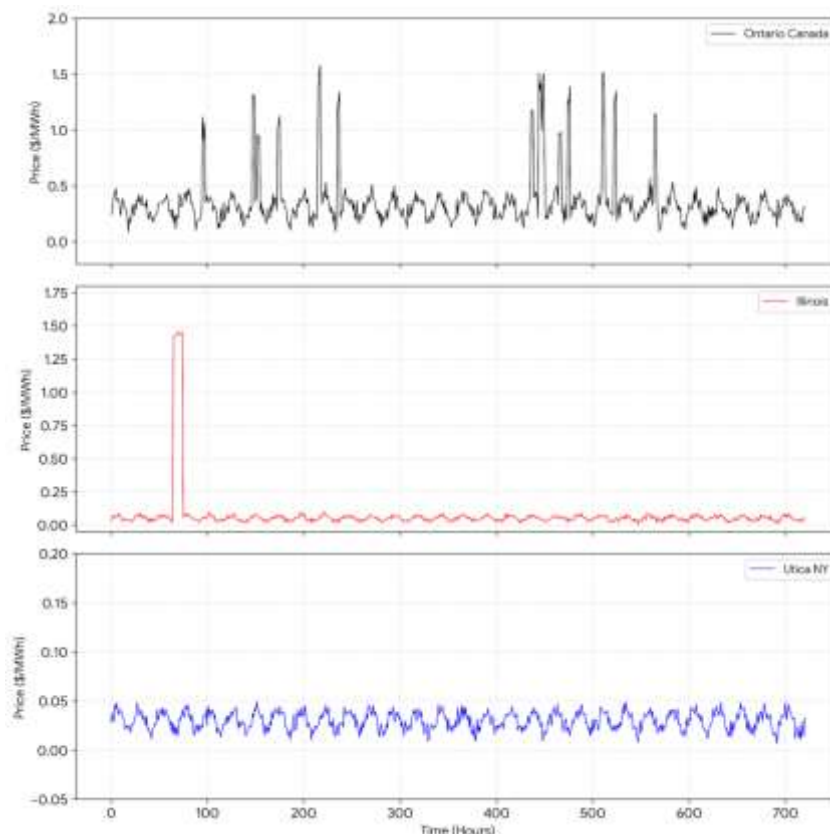


Figure 4: Hourly Electricity Prices

In Fig. 4, the x-axis represents the time slot in an hour. In contrast, the y-axis represents the electricity price in $\$/MWh$. It is worth mentioning that the electricity cost could be negative under certain conditions, which is also evident during some timeslots of the Ontario data center. There could be many reasons for the negative electricity prices. The demand could be higher than the availability of electricity, or the supplier may want to attract more customers than its competitors [20].

Benchmark Algorithms

Minimum Cost Geographical Load Balancing (MGLB)

In this approach, the workload arrives at the aggregator at the timeslot and is forwarded to that data center with the least negligible electricity cost at that specific instant. The data center will start executing the workload and continue the execution process till the workload is fully executed [25].

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

Random Geographical Load Balancing (RGLB)

In this approach, the workload arrives at the aggregator at the timeslot and is forwarded and executed at that randomly selected data center. This scheme is similar to the algorithm proposed in [26].

Round Robin (RR)

In this approach, the workload arrives at the aggregator at the timeslot t and is forwarded to the data center round-robin. The data center will start executing the workload and continue the execution process till the workload is fully executed [27].

Experimental Setup

This section discusses the performance outcomes of the benchmark algorithms and the proposed algorithm. The actual per-unit electricity prices and workload traces are used for experimental purposes.

Fig. 5 represents the daily electricity cost of the proposed algorithm and the benchmark algorithms. The figure shows that the round-robin technique resulted in the highest average electricity cost compared to other methods. This most increased cost is that while deciding to forward the workload to a specific data center, this technique does not consider the current offered electricity price or the current backlog. The random scheme is the second highest electricity cost technique because it forwards the incoming workload randomly without considering any factor. However, it is evident from the said figure that MGLB results in a nominal average electricity cost compared to other benchmark algorithms. This technique selects the data center with the smallest per unit electricity cost while making the decision, it results in the smallest average electricity cost compared to another benchmark algorithm. However, this algorithm only considers the current smallest cost among the data centers. It does not delay the execution of the workload to the instant when the electricity price may be low. MGLB is the best algorithm compared to other benchmark algorithms in the current experiment. MGLB is thus selected for comparison with the proposed algorithm.

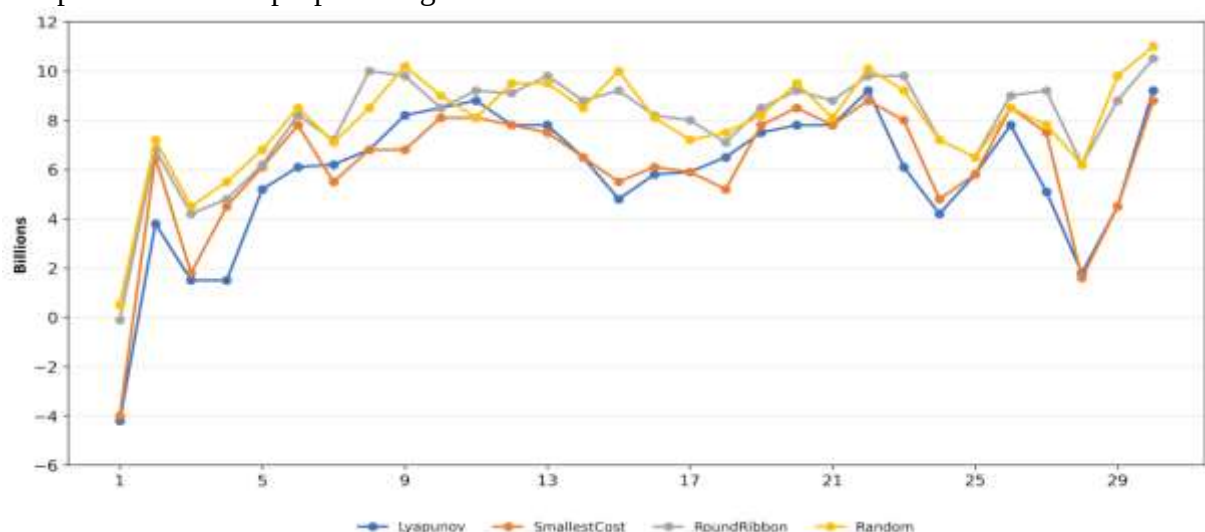


Figure 5: Average Daily Energy Cost of Algorithms

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

The results suggest that the proposed algorithm SDA is the best algorithm to minimize the average electricity cost of the data center's operator. Its best performance is precise because this algorithm selects the data center to forward the incoming workload with the smallest data center cost at that specific time. Secondly, it tries to execute the workload whenever the electricity prices are relatively low. It is also evident that the per-unit offered electricity price is negative during some timeslots at the data center in Ontario, Canada. There could be many reasons for the negative electricity prices. The demand is too low compared to the availability of electricity, or the supplier may want to attract more customers towards itself compared to its other competitors [28–32].

Table 2: Energy Cost Saving of SDA Algorithm

Comparison factor	Saving over MGLB	SDA Saving over RGLB	SDA Saving of SDA over RR
Energy cost	6%	27%	28%

Tab. 2 shows the percentage-wise cost saving of the proposed SDA algorithm compared to the benchmark algorithms, i.e., MGLB, RGLB, and RR. The table demonstrates that the proposed algorithm saves operating costs by 6%, 27% and 28% compared to MGLB, RGLB, and RR.

Fig. 6 reflects the proposed algorithm's average cost when varying the value of V . It is evident from the figure that as we increase the value of V , the average electricity cost decreases. However, when we increase the value of V too much, the electricity cost starts to rise. This is due to too much waiting for the algorithm to calculate the unit cost to become low enough. Therefore, as our proposed algorithm has to complete the workloads within its deadline, it executes the workloads at a high per-unit price to satisfy the deadline constraint.

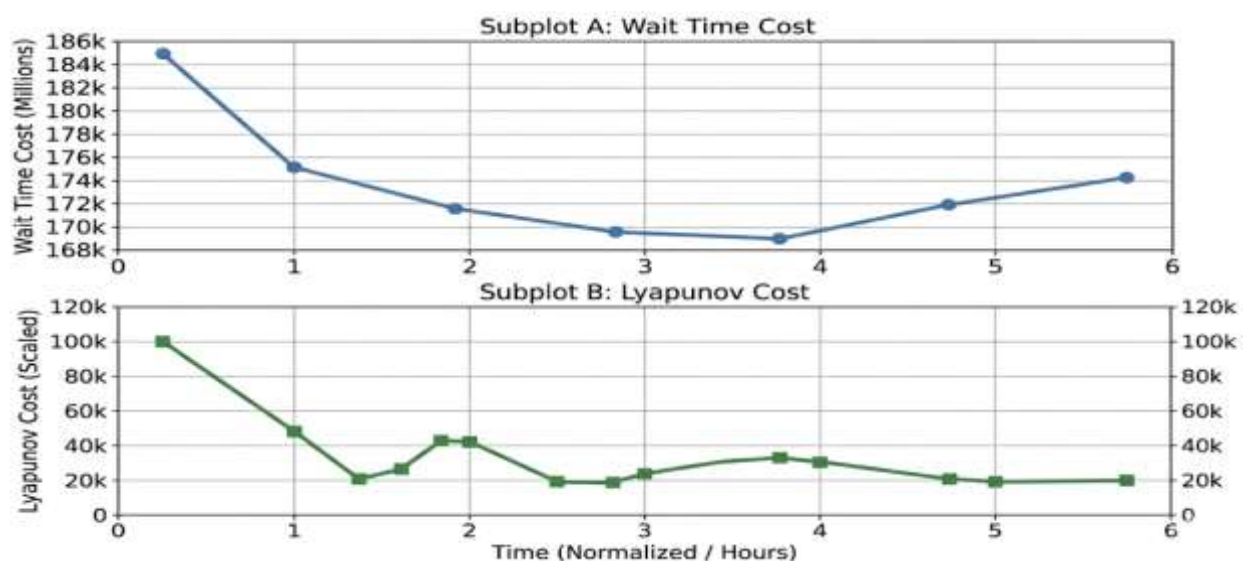


Figure 6: Average SDA cost vs. value of V

Conclusion and Future Work

Datacenter operators such as Google, Facebook, Amazon, etc., bear a significant portion of their operational cost in electricity consumption. Many researchers work on hardware and software-based techniques to minimize the electricity cost of those operators. An aspect of reducing the operator's electricity cost is exploiting the variation in per unit electricity cost, along with time and space. In this paper, we investigate the problem of minimizing the average electricity cost of geographically distributed data centers while respecting the user-agreed deadline for delay-tolerant workloads. We proposed an online algorithm based on the Lyapunov optimization technique, termed the Strict Deadline Algorithm (SDA). Our system is evaluated against the benchmark algorithms with the help of real-world workload traces and electricity prices. The result suggests that our proposed algorithm performed better in minimizing the electricity cost and completing all the workloads within its deadlines. In the future, we would like to investigate our proposed model with delay intolerant workload. Besides this, they use renewable sources of energy and energy storage devices to minimize the electricity cost of geographically distributed data centers.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- Hogade, N.; Pasricha, S.; Siegel, H.J. Energy and network aware workload management for geographically distributed data centers. *IEEE Trans. Sustain. Comput.* **2021**, *7*, 400–413.
- Jiang, D.; Wang, Y.; Lv, Z.; Wang, W.; Wang, H. An energy-efficient networking approach in cloud services for IIoT networks. *IEEE J. Sel. Areas Commun.* **2020**, *38*, 928–941.
- Shekhawat, V.S.; Gautam, A.; Thakrar, A. Datacenter workload classification and characterization: An empirical approach. In *Proceedings of the IEEE 13th International Conference on Industrial and Information Systems (ICIIS)*, Rupnagar, India, **2023**; 1–7.
- Kumari, A.; Tanwar, S.; Tyagi, S.; Kumar, N.; Obaidat, M.S.; et al. Fog computing for smart grid systems in the 5G environment: Challenges and solutions. *IEEE Wirel. Commun.* **2023**, *26*, 47–53.
- Albreem, M.A.; Sheikh, A.M.; Alsharif, M.H.; Jusoh, M.; Yasin, M.N.M. Green Internet of Things (GIoT): Applications, practices, awareness, and challenges. *IEEE Access* **2021**, *9*, 38833–38858.
- Chakravarthy, V.D.; Amutha, B. A novel software-defined networking approach for load balancing in data center networks. *Int. J. Commun. Syst.* **2022**, *35*, e4213.
- Ahmad, I.; Khalil, M.I.K.; Shah, S.A.A. Optimization-based workload distribution in geographically distributed data centers: A survey. *Int. J. Commun. Syst.* **2020**, *33*, e4453.

- Khalil, M.I.K.; Ahmad, I.; Almazroi, A.A. Energy efficient indivisible workload distribution in geographically distributed data centers. *IEEE Access* **2019**, *7*, 82672–82680.
- Jeba, J.A.; Roy, S.; Rashid, M.O.; Atik, S.T.; Whaiduzzaman, M. Towards green cloud computing an algorithmic approach for energy minimization in cloud data centers. In *Research Anthology on Architectures, Frameworks, and Integration Strategies for Distributed and Cloud Computing*; IGI Global: Hershey, PA, USA, **2021**; 846–872.
- Shakarami, A.; Ghobaei-Arani, M.; Shahidinejad, A.; Masdari, M.; Shakarami, H. Data replication schemes in cloud computing: A survey. *Cluster Comput.* **2021**, *24*, 2545–2579.
- Khalil, M.I.K.; Ahmad, I.; Shah, S.A.A.; Jan, S.; Khan, F.Q. Energy cost minimization for sustainable cloud computing using option pricing. *Sustain. Cities Soc.* **2020**, *63*, 102440.
- Javadpour, A.; Abadi, A.M.H.; Rezaei, S.; Zomorodian, M.; Rostami, A.S. Improving load balancing for data-duplication in big data cloud computing networks. *Cluster Comput.* **2022**, *25*, 2613–2631.
- Chong, F.T.; Heck, M.J.R.; Ranganathan, P.; Saleh, A.A.M.; Wassel, H.M.G. Data center energy efficiency: Improving energy efficiency in data centers beyond technology scaling. *IEEE Des. Test* **2013**, *31*, 93–104.
- Bird, S.; Aijit, A.; Othman, A.M.; Wenjin, H.; Kerop, J.; et al. Distributed (green) data centers: A new concept for energy, computing, and telecommunications. *Energy Sustain. Dev.* **2014**, *19*, 83–91.
- Koutitas, G.; Demestichas, P. Challenges for energy efficiency in local and regional data centers. *J. Green Eng.* **2010**, *1*, 1–32.
- Ernst, M.; Hogue, R.; Hollowell, C.; Strecker-Kellog, W.; Wong, A.; et al. Operating dedicated data centers—is it cost-effective? *J. Phys.* **2014**, *51*, 062053.
- Dhaya, R.; Kanthavel, R.; Venusamy, K. Dynamic secure and automated infrastructure for private cloud data center. *Ann. Oper. Res.* **2021**, 1–21.
- Guo, Y.; Gong, Y.; Fang, Y.; Khargonekar, P.P.; Geng, X. Energy and network aware workload management for sustainable data centers with thermal storage. *IEEE Trans. Parallel Distributed. Systems.* **2013**, *25*, 2030–2042.
- Lei, H.; Wang, R.; Zhang, T.; Liu, Y.; Zha, Y. A multi-objective co-evolutionary algorithm for energy-efficient scheduling on a green data center. *Comput. Oper. Res.* **2016**, *75*, 103–117.
- Yu, L.; Jiang, T.; Cao, Y. Energy cost minimization for distributed internet data centers in smart microgrids considering power outages. *IEEE Trans. Parallel Distrib. Syst.* **2014**, *26*, 120–130.
- Gu, C.; Li, Z.; Liu, C.; Huang, H. Planning for green cloud data centers using sustainable energy. In *Proceedings of the IEEE Symposium on Computers and Communication (ISCC)*, Messina, Italy, 2016; 804–809.
- Chen, T.; Marques, A.G.; Giannakis, G.B. DGLB: Distributed stochastic geographical load balancing over cloud networks. *IEEE Trans. Parallel Distrib. Syst.* **2016**, *28*, 1866–1880.

- Rahman, A.; Liu, X.; Kong, F. A survey on geographic load balancing-based data center power management in the smart grid environment. *IEEE Commun. Surv. Tutor.* **2013**, 16, 214–233.
- Shuja, J.; Gani, A.; Shamshirband, S.; Ahmad, R.W.; Bilal, K. Sustainable cloud data centers: A survey of enabling techniques and technologies. *Renew. Sustain. Energy Rev.* **2016**, 62, 195–214.
- Liu, Z.; Lin, M.; Wierman, A.; Low, S.H.; Andrew, L.L.H. Greening geographical load balancing. *ACM SIGMETRICS Perform. Eval. Rev.* **2011**, 39, 193–204.
- Lin, M.; Liu, Z.; Wierman, A.; Andrew, L.L.H. Online algorithms for geographical load balancing. In *Proceedings of the International Green Computing Conference (IGCC)*, San Jose, CA, USA, **2012**; 1–10.
- Abdelghany, M.A.; Mohsenian-Rad, H.; Alizadeh, M. Wholesale electricity pricing in the presence of geographical load balancing. In *Proceedings of the Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, **2017**; 653–658.
- Paul, D.; Zhong, W.-D.; Bose, S.K. Energy efficiency aware load distribution and electricity cost volatility control for cloud service providers. *J. Netw. Comput. Appl.* **2016**, 59, 185–197.
- Toosi, A.N.; Qu, C.; de Assunção, M.D.; Buyya, R. Renewable-aware geographical load balancing of web applications for sustainable data centers. *J. Netw. Comput. Appl.* **2017**, 83, 155–168.
- Lu, X.; Kong, F.; Liu, X.; Yin, J.; Xiang, Q.; et al. Bulk savings for bulk transfers: Minimizing the energy-cost for geo-distributed data centers. *IEEE Trans. Cloud Comput.* **2017**, 8, 73–85.
- Goiri, Í.; Haque, M.E.; Le, K.; Beauchea, R.; Nguyen, T.D.; et al. Matching renewable energy supply and demand in green datacenters. *Ad Hoc Netw.* **2015**, 25, 520–534.
- Qu, C.; Calheiros, R.N.; Buyya, R. Mitigating impact of short-term overload on multi-cloud web applications through geographical load balancing. *Concurr. Comput. Pract. Exp.* **2017**, 29, e4126.
- Saleh, A.Z.M.; Rozali, N.A.; Buja, A.G.; Jalil, K.A.; Ali, F.H.M.; et al. A method for web application vulnerabilities detection by using Boyer-Moore string matching algorithm. *Procedia Comput. Sci.* **2015**, 72, 112–121.
- Cantone, D.; Faro, S. Fast-Search: A new efficient variant of the Boyer-Moore string matching algorithm. In *Proceedings of the International Workshop on Experimental and Efficient Algorithms*, Berlin, Germany, 2003; 47–58.
- Peng, Y.; Kang, D.-K.; Al-Hazemi, F.; Youn, C.-H. Energy and QoS aware resource allocation for heterogeneous sustainable cloud datacenters. *Opt. Switch. Netw.* **2017**, 23, 225–240.
- Khalil, M.I.K.; Shah, S.A.A.; Taj, A.; Shiraz, M.; Alamri, B.; et al. Renewable-aware geographical load balancing using option pricing for energy cost minimization in data centers. *Processes* **2022**, 10, 1983.
- Hampton, H.; Foley, A. A review of current analytical methods, modelling tools and development frameworks applicable for future retail electricity market design. *Energy* **2022**, 260, 124861.

- Khan, F.; Jan, M.A.; Alam, M. Big data in healthcare: A survey. In *Applications of Intelligent Technologies in Healthcare*; Springer: Cham, Switzerland, 2019; 143–152.
- Qureshi, A.; Weber, R.; Balakrishnan, H.; Gutttag, J.; Maggs, B. Cutting the electric bill for internet-scale systems. In *Proceedings of the ACM SIGCOMM 2009 Conference on Data Communication*, New York, NY, USA, 2009; 123–134.
- Rao, L.; Liu, X.; Xie, L.; Liu, W. Coordinated energy cost management of distributed internet data centers in smart grid. *IEEE Trans. Smart Grid* **2011**, 3, 50–58.
- Khalil, M.I.K.; Shah, S.A.A.; Khan, I.A.; Hijji, M.; Shiraz, M.; et al. Energy cost minimization using string matching algorithm in geo-distributed data centers. *Comput. Mater. Contin.* **2023**, 75, 6305–6322.
- Khalil, M.I.K.; Mubeen, A.; Taj, A.; Jan, N.; Ahmad, S. Renewable and temperature aware load balancing for energy cost minimization in data centers. *J. Comput. Biomed. Inform.* **2023**, 6, 1–8.
- Luo, S.; Zhang, G.; Wu, C.; Khan, S.U.; Li, K. Boafft: Distributed deduplication for big data storage in the cloud. *IEEE Trans. Cloud Comput.* **2020**, 8, 1199–1211.
- Jawad, M.; Qureshi, M.B.; Khan, M.U.S.; Ali, S.M.; Mehmood, A.; et al. A robust optimization technique for energy cost minimization of cloud data centers. *IEEE Trans. Cloud Comput.* **2021**, 9, 447–460.
- Ali, S.M.; Jawad, M.; Khan, M.U.S.; Bilal, K.; Glower, J.; et al. An ancillary services model for data centers and power systems. *IEEE Trans. Cloud Comput.* **2020**, 8, 1176–1188.
- Sajid, S.; Jawad, M.; Qureshi, M.B.; Khan, M.U.S.; Ali, S.M.; Khan, S.U. A conditional-constraint optimization for joint energy management of data center and electric vehicle parking-lot. In *Proceedings of the 10th International Green and Sustainable Computing Conference (IGSC)*, Alexandria, VA, USA, October **2019**.
- Muhammad, D.; Ahmad, I.; Khalil, M.I.K.; Khalil, W., Ahmad, M.O. A generalized deep learning approach to seismic activity prediction. *Applied Sciences*, **2023**, 13(3), p. 1598.
- Khalil, T.; Jan, S.; Ghoniem, R.M.; Khalil, M.I.K. Multi-CNN fusion framework for predictive violence detection in animated movies. *CMC-Computers, Materials, and Continua*, **2025**.
- Rahman, Z.U.; Khalil, M.I.K.; Nawaz, A.; Jan, N.; Ahmad, S. Analysis and clustering of Pakistani music by lyrics: A study of Coke studio Pakistan. *Journal of Computing and Biomedical informatics*. **2024**, 7(1), 281–296.
- Khan, A.I.; Ullah, F.; Abrar, M.; Shah, S.; Taj, F.; Khalil, M.I.K. Ransomware early detection model using API-calls at runtime by random decision forests. *International Review of Basic and Applied Sciences*. **2022**, 10(2), 349–359.
- Khalil, M.I.K. Improve quality of service and secure communication in mobile adhoc networks through group key management. *International Review of Basic and Applied Sciences*. **2013**, 1(3), 107–115.
- Khalil, M.I.K.; Mubeen, A.; Taj, A.; Jan, A.; Ahmad, S. Renewable and temperature aware load balancing for energy cost minimization in data centers: A study of

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

- BRT, Peshawar. *Journal of Computing and Biomedical informatics*. **2023**, 6(1), 211–223.
- Khalil, M.I.K.; Ullah, A.; Taj, A.; Khan, I.A.; Ullah, F.; Taj, F, Shah, S. Analysis of critical risk factors affecting software quality: A study of software houses in Peshawar, Pakistan. *International Review of Basic and Applied Sciences*. **2022**, 10(2), 338–348.
- Akhtar, M. U.; Khalil, M.I.K. Link prediction techniques in complex networks. *International Review of Basic and Applied Sciences*. **2018**, 6(8), 60–64.
- Khalil, M.I.K.; Ali, F.; Shah, S.A.A.; Khan, I.A.; Alauddin, Taj, A. On Enabling Energy Efficiency in Geographically Distributed Data Center through Data Deduplication. *Annual Methodological Archive Research Review*. **2025**, 10(2), 338–348.