Strategic Placement of Data Centers for Economic Analysis: An Online Algorithm Approach

Christine Markarian and Claude Fachkha
College of Engineering and Information Technology Department, University of Dubai, Dubai, U.A.E.

Keywords: Data Center Location, Economic Growth, Data Analytics, Competitive Analysis, Online Algorithms.

Abstract: Governments worldwide have increasingly recognized the transformative potential of data analytics in economics, leading to the establishment of specialized research centers dedicated to economic analysis. These centers serve as hubs for experts to dissect economic indicators, inform policymaking, and foster sustainable growth. With data analytics playing a pivotal role in understanding economic trends and formulating policy responses, the strategic placement of data centers becomes crucial. In this paper, we address the strategic placement of data centers in urbanized environments within the framework of online algorithms. Online algorithms are designed to make sequential decisions without complete information about future inputs, making them suitable for dynamic urban environments. Specifically, we formulate the problem as the *Online Data Center Placement* problem (ODCP) and design a novel online algorithm for it. To gauge our algorithm's effectiveness, we use competitive analysis, a standard method for assessing online algorithms. This method compares our algorithm's solutions with those of the optimal offline solution. Our study aims to provide a systematic approach for informed decision-making, optimizing resource usage, and fostering economic growth.

1 INTRODUCTION

Governments globally have acknowledged the power of data analytics in economics. This led to the creation of research centers focused on analyzing economic data, forecasting trends, and guiding policymaking. These efforts highlight data analytics' vital role in shaping economic policy, fostering innovation, and promoting sustainable growth (Johnson et al., 2021).

These specialized research centers are hubs of economic analysis. They gather experts in statistics, econometrics, and data science. Together, they dissect economic indicators, understand market dynamics, and identify emerging trends. Institutions like central banks, finance ministries, and economic planning departments have set up dedicated research units. These units have expertise in data analytics. They support evidence-based policymaking and economic analysis.

The benefits of data analytics in economics are manifold. It helps governments gain valuable insights into the economy's health, make informed policy decisions, and respond effectively to economic challenges. Economic indicators like Gross domestic product (GDP) growth and inflation rates are analyzed to understand economic trends, identify areas of concern, and formulate appropriate policy responses. Data analytics also facilitates the assessment of policy interventions, the evaluation of program effectiveness, and the monitoring of progress towards economic goals (Awan et al., 2021).

Recent government initiatives have witnessed the establishment of new centers for data analytics in economics globally. Examples range from the Bank of England's Data and Statistics Division (Bank of England, 2024) to the United States (U.S.) Census Bureau's specialized units (United States Census Bureau, 2024), alongside institutions like the German Institute for Economic Research (DIW Berlin) (German Institute for Economic Research (DIW Berlin), 2024) and the National Institute of Economic and Social Research (NIESR) in the United Kingdom (UK) (National Institute of Economic and Social Research (NIESR), 2024). Moreover, governments in Singapore, Australia, and other countries have invested in data analytics capabilities to support economic research, policy analysis, and forecasting (Liu et al., 2023).

The focus of this paper centers around the strategic placement of data centers to harness the transformative potential of data analytics in economics. As

Markarian, C. and Fachkha, C.
Strategic Placement of Data Centers for Economic Analysis: An Online Algorithm Approach.
DOI: 10.5220/0012788500003756

In *Proceedings of the 13th International Conference on Data Science, Technology and Applications (DATA 2024)*, pages 425-433
ISBN: 978-989-758-707-8; ISSN: 2184-285X

governments worldwide increasingly recognize the importance of data-driven insights in shaping economic policy and fostering sustainable growth, the establishment of specialized research centers dedicated to analyzing economic data has become paramount.

The strategic placement of data centers presents challenges and objectives for governments. They aim to maximize the effectiveness and accessibility of data analytics capabilities while ensuring equitable access to resources across regions. This involves considering geographic distribution, population centers, partnerships, infrastructure, economic factors, and strategic priorities.

Geographic distribution is crucial for governments. They aim to ensure equitable access to data analytics resources across diverse regions, promoting inclusivity and regional development initiatives. Major urban areas serve as hubs of economic activity and innovation. They drive the placement of data centers to leverage existing infrastructure and talent pools. Collaborative initiatives with academic institutions and private-sector partners shape the placement of data centers. This amplifies the impact of data analytics initiatives.

Robust infrastructure and connectivity are crucial for data analytics platforms to work effectively, impacting where data centers are located Governments focus on places with good economic conditions and incentives to attract investment, spurring economic growth and job creation. Data center locations are chosen to support strategic goals like regional development, innovation clusters, and specific industry sectors, aiming for sustainable economic progress.

Online algorithms and *competitive analysis* are crucial in decision-making processes, notably in situations like placing research centers in urban areas (Borodin and El-Yaniv, 2005; Albers, 2003). Online algorithms decide sequentially without information about future inputs. They're vital in dynamic settings where real-time decisions are needed based on incomplete or uncertain data. In location problems, like data center placement, online algorithms help find the best locations as demands change over time (Borodin and El-Yaniv, 2005; Albers, 2003).

Competitive analysis uses the *competitive ratio* to measure how well online algorithms work compared to optimal offline ones. It gives insights into their effectiveness in real-world scenarios. The competitive ratio compares the cost of the online algorithm to that of an optimal offline solution. A ratio of 1 means the online algorithm matches the offline one's cost in the worst-case scenario (Borodin and El-Yaniv, 2005; Albers, 2003).

Achieving a competitive ratio of 1 is tough due to real-world uncertainties. In location problems, changing demands, resource limits, and geography affect online algorithm performance (Borodin and El-Yaniv, 2005; Albers, 2003). Analyzing the competitive ratio in data center placement helps understand how online algorithms handle dynamic decisions and uncertainties. It helps governments and organizations improve decision-making and resource allocation for urbanization and economic development (Borodin and El-Yaniv, 2005; Albers, 2003).

Using online algorithms for data placement, especially in urbanization contexts, is advantageous because they operate effectively under uncertainty. These algorithms can make decisions without knowing the future, which suits dynamic and unpredictable urban environments. While regrets may occur, evaluating them with competitive analysis guarantees performance.

2 OUR CONTRIBUTION

Data center placement in urban environments is a complex problem that requires careful consideration of many factors in order to achieve the best possible technological and economic outcomes. Let's take an example where a city administration has to choose where to locate 100 possible data centers within its urban area. Each data center's establishment incurs significant costs, covering initial construction, equipment procurement, and infrastructure development, averaging around €500,000 per center. Moreover, annual operational expenses, including utility bills, maintenance, and staffing, amount to approximately €50,000 per center. Additionally, connecting these data centers to 50 strategically located hubs introduces further financial complexity, with connectivity costs averaging €100,000 per connection. These costs fluctuate based on factors such as distance and technological requirements. Furthermore, each data center must handle incoming demands for data analytics services, incurring transportation and processing complexity costs. Transportation costs, associated with data movement to and from the centers, are estimated at €10,000 per demand-center pair, while processing complexity costs amount to approximately €20,000 per pair. The overarching objective is to minimize establishment, operation, connectivity, transportation, and processing complexity costs, ensuring efficient delivery of data analytics services.

In this paper, we tackle the problem of strategically positioning data centers in urbanized environments from the perspective of online algorithms. Specifically, we formulate the latter as the *Online*

Data Center Placement problem (ODCP), outlined below. We design an online algorithm for ODCP and assess its performance using the competitive analysis framework, widely recognized as the standard approach for evaluating online algorithms.

In this rapidly urbanizing world, the dynamic nature of urban environments, characterized by continuous population growth, evolving infrastructure needs, and shifting socio-economic dynamics, presents a complex challenge in strategically placing data centers. Our study aims to offer a systematic approach to navigate this complexity, enabling informed decisions that optimize resource utilization and stimulate economic growth.

3 PROBLEM DESCRIPTION: ONLINE DATA CENTER PLACEMENT PROBLEM (ODCP)

Given:

- A set $L = \{l_1, l_2, \ldots, l_m\}$ of potential locations where data centers can be established. This set represents the available locations where data centers can be built to serve the demands for data analytics services.
- A set $H = \{h_1, h_2, \ldots, h_k\}$ of hub locations. These locations serve as focal points for managing and coordinating the network of data centers, facilitating communication, data aggregation, and centralized decision-making.
- Each potential data center location l_i can be connected to one of the hubs in H, incurring a **connectivity cost** cc_{ih} associated with establishing and maintaining the network connection between the potential data center l_i and the hub h in H.
- A sequence $D = \{d_1, d_2, \ldots, d_n\}$ of demands representing the arrival of data analytics needs over time. These demands correspond to various factors such as geographical regions, population centers, economic hubs, and social or environmental contexts where data analytics services are required.
- **Establishment costs** ec_i associated with establishing a data center at location l_i. These costs represent the expenses involved in setting up a data center at each location l_i. They include initial construction costs, equipment procurement, and infrastructure development.
- **Operation costs** oc_i are expenses related to maintaining and running a data center at location l_i. They encompass ongoing expenditures necessary for efficient operations, including utility bills, equipment maintenance, staffing, security measures, and other recurring expenses essential for sustained functionality of the data center.
- **Transportation costs** tc_{ij} associated with each demand and data center pair. This cost represents the expenses associated with transporting data or resources to and from the data center to fulfill demand d_j. These costs include factors such as shipping, logistics, and distribution expenses required to move data or equipment to the designated data center and deliver services to customers or demands efficiently.
- **Processing complexity costs** pcc_{ij} associated with each demand and data center pair. This cost reflects the computational resources and expertise required to analyze and process the data associated with a particular demand at a specific data center.

4 OBJECTIVE

In the online setting, demands arrive sequentially and decisions regarding data center establishment, data center-hub connections, and demand assignment must be made without knowledge of future demands. Additionally, once decisions are made, they cannot be reversed. Upon the arrival of each demand, a decision must be made whether to assign it to an existing data center or to establish a new one and allocate the demand accordingly. Furthermore, once a data center is established, it must be promptly connected to one of the designated hubs to ensure network functionality. This can be more formally expressed as follows:

1. Each demand d_j must be assigned to one data center.
$$\sum_{i \in L} x_{ij} = 1 \quad \text{for all} \quad j \in D$$
2. Each data center l_i used for serving demands must be connected to one hub. U is used to denote the set of data centers to which demands are assigned.
$$\sum_{h \in H} y_{ih} = 1 \quad \text{for all} \quad i \in U$$

Minimizing the total sum of establishment, operation, connectivity, transportation, and processing complexity costs associated with setting up and maintaining the network of data centers, ensuring efficient delivery of data analytics services across diverse geographical regions and demand scenarios, is our objective. This can be more formally expressed as follows:

Minimize:

$$\sum_{i\in L} ec_i + \sum_{i\in L} oc_i + \sum_{i\in L}\sum_{h\in H} \mathrm{cc}_{ih} + \sum_{i\in L}\sum_{j\in D} tc_{ij} + \sum_{i\in L}\sum_{j\in D} pc_{ij}$$

The aim is to construct an online algorithm that has a competitive ratio as close to 1 as possible. In this pursuit, we adopt the oblivious adversary model for our online algorithm, wherein the adversary's actions are predetermined and independent of the algorithm's decisions. Below is a summary of the *Online Data Center Placement* problem (ODCP).

Online Data Center Placement Problem

Input:

- Set $L = \{l_1, l_2, \ldots, l_m\}$ of potential data center locations.
- Set $H = \{h_1, h_2, \ldots, h_k\}$ of hub locations.
- Sequence $D = \{d_1, d_2, \ldots, d_n\}$ of demands.
- Establishment costs ec_i.
- Operation costs oc_i.
- Connectivity costs cc_{ih}.
- Transportation costs tc_{ij}.
- Processing complexity costs pcc_{ij}.

Output:

- Demands arrive sequentially, prompting decisions made without knowledge of future demands.
- Upon each demand arrival d_j, a decision must be made to assign it to an existing data center or establish a new one.
- Once a data center is established, it must be promptly connected to one of the designated hubs.
- The goal is to minimize establishment, operation, connectivity, transportation, and processing complexity costs, aiming for a competitive ratio as close to 1 as possible.

5 RELATED WORK

The problem at hand, the *Online Data Center Placement* problem (ODCP), represents a broader form of the *Non-metric Online Facility Location* problem (NOCF). NOCF entails a scenario where a set of potential facility locations, a set of client locations (arriving sequentially over time), and a function representing the facility opening cost are given. The objective is to assign each client to an open facility while aiming to minimize the total assignment and facility opening costs. This broadness becomes evident when we consider a specific instance of ODCP. Setting the operation costs, the connectivity costs, the processing complexity costs, and the number of hubs to zero transforms the problem into NOCF. In this transformation, the transportation costs would correspond to assignment costs and establishment costs to facility opening costs.

NOCF, along with its variants, has garnered significant attention within the online algorithm community. Its lower bounds are followed by those for the *Online Set Cover* problem (OSC) (Alon et al., 2003; Korman, 2004). On the positive side, Alon *et al.* devised a randomized $O(\log m \log n)$-competitive online algorithm for NOCF., with m denoting the number of facility locations and n the number of client locations. Another approach involves reducing NOCF instances to OSC instances and employing a deterministic algorithm for OSC. This yields an $O((\log n + \log m) \cdot (\log n + \log\log m))$-competitive ratio. More recently, Bienkowski *et al.* (Bienkowski et al., 2021) introduced an online deterministic polynomial-time algorithm surpassing this bound, achieving an $O(\log m \cdot (\log n + \log\log m))$-competitive ratio. Other variants of NOCF have been explored in the context of service installation, service quality, and leasing (Markarian, 2021; Markarian, 2022; Markarian and auf der Heide, 2019).

It is important to note that the decisions in ODCP regarding data center establishment, hub connectivity, and demand assignment are interrelated. The choice to establish a new data center impacts connectivity costs, as it necessitates establishing connections with one of the designated hubs. Therefore, these decisions cannot be made independently but must be considered together to optimize the overall cost and performance of the network. Hence, while algorithms designed for NOCF may provide insights, they cannot be directly applied to ODCP due to the unique constraints and interdependencies inherent in data center placement and network optimization.

6 GRAPH-BASED FORMULATION FOR ODCP

In this section, we formulate ODCP as an edge-weighted graph problem, outlining the nodes, edges, and objective of the problem.

Nodes:

- We designate r as the root node.
- We pair each potential data center location l_i from the set $L = \{l_1, l_2, \ldots, l_m\}$ with two nodes. One is identified as the original data center node l_i, while the other is its replica denoted as l'_i. Here, the set $\mathcal{L}$ denotes the original data center nodes, while $\mathcal{L}'$ represents their corresponding replicas.
- We pair each hub h_i from the set $H = \{h_1, h_2, \ldots, h_k\}$ with a node, which we denote as the hub node h_i. Here, the set $\mathcal{H}$ represents the hub nodes.
- Whenever a demand d_i from the set $D = \{d_1, d_2, \ldots, d_n\}$ arrives, we create a node for it, denoted as the demand node d_i. Here, the set $\mathcal{D}$ represents the demand nodes.

Edges:

- An edge directed from each demand node to each data center node is added. The edges' weight is set to the sum of the associated transportation and processing complexity costs corresponding to the demand and data center.
- An edge directed from each demand node d_i to each data center node l_j is added. The edges' weight corresponds to the sum of transportation and processing complexity costs, denoted as tc_{ij} and pcc_{ij} respectively, for demand d_i and data center l_j.
- An edge directed from each data center node l_i to its replica l'_i is added. The edges' weight corresponds to the sum of the establishment and operation costs, denoted as ec_i and oc_i respectively, associated with the data center l_i.
- An edge directed from each replica data center node l'_i to each hub node h_j is added. The edges' weight corresponds to the connectivity cost associated with the data center and hub, denoted as cc_{ij}.
- An edge directed from each hub node h_i to the root node r is added and the weights of these edges are adjusted to zero.

Figure 1 presents an example of a graph generated from an input consisting of two demands, five data center locations, and two hubs.

Objective: The problem asks the following. Whenever a demand arises, the objective is to find, from the demand node to the root node r, a directed path. According to the problem formulation, this path comprises a data center node and a hub node. Once this path is determined, we can identify the associated hub with the hub node and the data center location with the data center node along the path. The costs associated with the solution are equivalent to the weights on the edges. Specifically, we pay the costs associated with the data center and the hub chosen along the path determined by the solution. This mapping allows us to derive a solution for the original non-graph problem, the *Online Data Center Placement* problem (ODCP).

This strategic mapping not only enables us to derive a solution for the original non-graph problem, the *Online Data Center Placement* problem (ODCP), but also ensures that the competitive ratio remains consistent. This is achieved by inherently maintaining the cost alignment within the formulation itself, thereby preserving parity between the graph-based problem and the original non-graph problem formulation.

7 ONLINE ALGORITHM DESIGN

We introduce, in this section, an online algorithm for the *Online Data Center Placement* problem (ODCP), utilizing the graph formulation outlined earlier.

An instance of ODCP consists of a set L of potential data center locations and a set H of hub locations, accompanied by their respective establishment, operation, and connectivity costs. The sequence D of demands unfolds incrementally as the algorithm advances. Each step introduces a new demand, along with its associated transportation and processing complexity costs.

The algorithm initiates by building the nodes, edges, and their associated weights for the data center and hub sets. Subsequently, as each demand is unveiled, the algorithm generates a node, incorporating its relevant edges and weights.

A demand $d_i \in D$ arrives. We represent the graph created by the aforementioned formulation as $G = (V, E)$. We refer to the graph problem variant of ODCP as $ODCP_g$. The algorithm associates each edge $e \in E$ with a fractional value that is set to 0 and increases gradually as the algorithm progresses. These fractions collectively form a fractional solution for $ODCP_g$. The algorithm primarily focuses on constructing a fractional solution for $ODCP_g$ and incrementally converting it into an integral solution upon the arrival of new demands. We let c_e denote the cost and f_e the fraction of edge e. For each demand that arrives, the algorithm outputs a collection of edges that form a path from the demand node d_i to the root node r. Once such a path is established, we can identify the corresponding hub and data center nodes along the path.

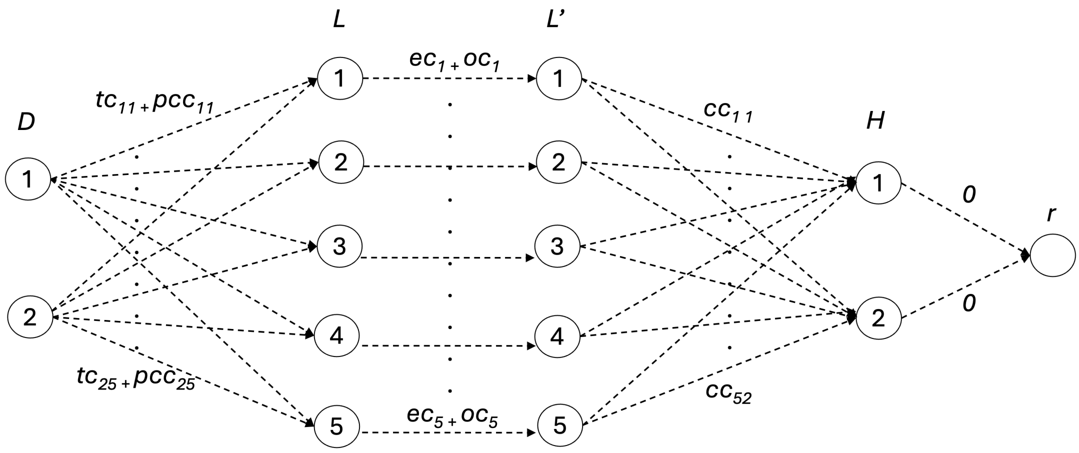

Figure 1: ODCP Instance - Graph Formulation.

The algorithm utilizes a stochastic decision-making procedure governed by a parameter α. This parameter is calculated as the smallest value within $2\lceil\log(n+1)\rceil$ independently selected random variables, with uniform distribution between 0 and 1, with n denoting the count of demands. Further clarification regarding this selection is provided in the competitive analysis segment.

The algorithm involves concepts related to network flows and connectivity, specifically focusing on the principles of maximum flow and minimum cut. The *maximum flow* between node u and node v in a graph represents the *minimum cut* from u to v—the smallest total fraction of edges that, if do not exist anymore, would disconnect u from v.

Input: $L = \{l_1, l_2, \ldots, l_m\}$, $H = \{h_1, h_2, \ldots, h_k\}$, $d_i \in D$, and the associated establishment, operation, connectivity, transportation, and processing complexity costs.
Output: Set $E' \subset E$ of edges of $G = (V, E)$ corresponding to the data centers, hubs, and demand-data center assignments forming the current solution for $ODCP_g$.

The algorithm works as follows:

1. As long as the maximum flow from d_i to r in G is below 1, form a minimum cut Q from d_i to r in G. Observe every edge $e \in Q$ and augment the following fraction:
$$f_e = f_e + \frac{f_e}{c_e} + \frac{1}{|Q| \cdot c_e}$$
2. Output every edge e if f_e exceeds α.
3. If the edges in the current solution do not create a feasible path for demand d_i, find a feasible path of minimum cost and add its edges to the solution.
4. Add the data center(s) and hub(s) associated with the edges outputted in the previous phase. Assign the demand to the data center associated with the solution path.

8 COMPETITIVE ANALYSIS AND PERFORMANCE EVALUATION

In the third and fourth phases of the algorithm, edges are acquired. In the third phase, selections are influenced by a stochastic process, while in the fourth phase, choices are tailored to ensure a feasible output.

Let $Optimal_{solution}$ represent the cost of the optimal offline solution and $frac$ signify the cost of the fractional solution formed by the algorithm.

Selections Based on Stochastic Process: Consider S' as the collection of edges acquired in the algorithm's third phase, with $C_{S'}$ denoting its anticipated cost. These edges are procured according to the stochastic process previously outlined. Let's designate l as an integer ranging from 1 to $2\lceil\log(n+1)\rceil$, and e as an edge. We define $X_{e,l}$ as the variable indicating whether e is selected by the algorithm through the stochastic process.

$$C_{S'} = \sum_{e \in S'} \sum_{l=1}^{2\lceil\log(n+1)\rceil} \text{Expectation}[X_{e,l}] \cdot c_e$$
$$= 2\lceil\log(n+1)\rceil \sum_{e \in S'} f_e c_e \quad (1)$$

Consider that the summation of $c_e f_e$ over all edges in S' is bounded above by the cost of the fractional solution. This comparison can be drawn against the optimal offline solution. The underlying concept is that whenever the algorithm increases a fraction, it does not surpass 2. Additionally, the total count of fraction increments can be gauged in relation to the cost of the optimal offline solution.

Each edge e in a minimum cut Q contributes to a fraction increase, quantified as $\left(\frac{f_e}{c_e} + \frac{1}{|Q| \cdot c_e}\right)$. The algorithm executes a fraction increase solely when the maximum flow is under 1. Thus, $\sum_{e \in Q} f_e < 1$ before any such increment. Consequently, the upper limit for each fraction increase is:

$$\sum_{e \in Q} \left(\frac{f_e}{c_e} + \frac{1}{|Q| \cdot c_e}\right) \cdot c_e < 2 \tag{2}$$

Ultimately, in each minimum cut Q, the algorithm is ensured to include an edge e from the optimal offline solution, as Q necessitates having an edge from every path by definition. Referring to the equation governing the fraction increase, once $O(\log |Q|)$ fraction increases occur, the fraction f_e for e reaches 1, and further increments are precluded since e won't appear in any subsequent minimum cut. The magnitude of any minimum cut is limited by m, denoting the number of facilities or the maximum available paths from d_i to the root r. Consequently, we are now able to constrain the fractional solution:

$$O(\log m \cdot Optimal_{solution}) \geq \text{frac} \tag{3}$$

Equations 1, 2, and 3 allow us to deduce an upper bound for the expected cost $C_{S'}$ of the edges purchased in the third phase of the algorithm:

$$O(\log(kn) \log m \cdot Optimal_{solution}) \geq C_{S'} \tag{4}$$

Choices Ensuring Feasibility: Consider S'' as the collection of edges acquired in the fourth phase of the algorithm, with $C_{S''}$ representing its anticipated cost. These edges are procured by the algorithm solely when a path hasn't been acquired via the stochastic process in the third phase. With each path purchase in this stage, the algorithm ensures that its cost doesn't surpass $Optimal_{solution}$, as it acquires a path of minimum cost.

Consider a specific demand d_i. Let Q_{j+1} denote a minimum cut created after the algorithm has procured a path from d_i to r and has finished the first phase. The probability of acquiring this path in the l-th trial, where $1 \leq l \leq 2\lceil \log(n+1) \rceil$, is:

$$\prod_{e \in Q_{j+1}} (1 - f_e) \leq e^{-\sum_{e \in Q_{j+1}} f_e}$$

$$e^{-\sum_{e \in Q_{j+1}} f_e} \leq \frac{1}{e}$$

It's notable that the last inequality is true because the algorithm guarantees that $\sum_{e \in Q_{j+1}} f_e \geq 1$ by the end of the first phase (as per the Max-flow min-cut theorem). The expected cost of acquiring the $(j+1)^{th}$ path across all $1 \leq l \leq 2\lceil \log(n+1) \rceil$ trials is less than $1/n^2 \cdot Optimal_{solution}$.

– (individual demand cost) Let's initiate by assessing the anticipated cost attributed to a single demand. Select a demand d_i. Assume Q_{j+1} as the minimum cut formed after the algorithm procures a path and concludes the first phase. The probability of acquiring the path for a single trial, denoted by $1 \leq l \leq 2\lceil \log(n+1) \rceil$, can be expressed as:

$$\prod_{e \in Q_{j+1}} (1 - f_e) \leq e^{-\sum_{e \in Q_{j+1}} f_e} \leq 1/e$$

– (cumulative cost of all demands) The total expected cost incurred by all n' incoming demands is bounded by:

$$n' \cdot 1/n^2 \cdot Optimal_{solution} \leq n \cdot 1/n^2 \cdot Optimal_{solution} = 1/n \cdot Optimal_{solution}$$

Consequently, the expected cost $C_{S''}$ of the edges purchased in the fourth phase of the algorithm is given by:

$$C_{S''} \leq 1/n \cdot Optimal_{solution} \tag{5}$$

Hence, we can infer the subsequent theorem.

Theorem 1. *For the Online Data Center Placement problem (ODCP), a randomized algorithm operates online with a competitive ratio of $O(\log n \log m)$. Here, m refers to the number of data center locations and n denotes the quantity of demands.*

9 NUMERICAL EXAMPLE

Consider an urbanized area with significant economic activity and technological infrastructure. In this scenario, there are 1000 potential locations suitable for data center deployment and 500 hub locations strategically positioned to ensure efficient network connectivity. With the increasing reliance on digital services in urban environments, there are 5000 demands for various digital applications, reflecting the diverse needs of the population.

Applying competitive analysis for the Online Data Center Placement problem (ODCP), we find that a randomized algorithm operates online with a competitive ratio of $O(\log n \log m)$, with m denoting the number of data center locations (1000) and n the quantity of demands (5000). Thus, in our numerical example, the competitive ratio is $O(\log 5000 \times \log 1000)$, which ensures the algorithm's efficiency and effectiveness in handling the substantial demands and data center locations in the urban environment.

Moreover, the scalability of the algorithm is evident in its ability to handle even larger numbers of demands and potential data center locations. No matter how many more demands arise or how many additional potential data center locations are identified, the algorithm's competitive ratio remains consistent, making it a highly scalable solution for data center placement in urbanized environments.

10 ALGORITHMIC TECHNIQUES AND KEY INSIGHTS

The techniques we use are commonly used to tackle problems within the realm of online algorithm design. Our algorithm utilizes network flow analysis principles to determine the path from each demand node to the root node in a graph representing the *Online Data Center Placement* problem. It employs a stochastic decision-making process introducing randomness to enhance decision flexibility. By constructing fractional solutions iteratively, the algorithm incrementally converts them into integral solutions, facilitating adaptability to evolving demands. Employing a greedy augmentation strategy, the algorithm gradually increases the flow along edges forming minimum cuts, aiming to approach the maximum flow limit efficiently. Additionally, it constructs minimum-cost paths when the current solution does not form a feasible solution for the current demand, ensuring cost-effective connectivity. These algorithmic techniques collectively enable the algorithm to efficiently solve the *Online Data Center Placement* problem.

Several mathematical methods, such as graph theory, stochastic processes, and probabilistic theory, are integrated in the evaluation of the algorithm for the *Online Data Center Placement* problem. The problem is modeled as a graph using the concepts of graph theory, where nodes stand in for data centers, hubs, and demand locations, and edges for the connections and expenses that exist between them. In the third stage of the algorithm, edge selection is guided by stochastic processes, which introduce randomness to adjust to changing conditions and optimize cost. This stochastic process is supported by probability theory, which allows the algorithm to make probabilistic decisions based on random variables and distributions. These mathematical methods are combined by the algorithm to produce a competitive ratio analysis, which sheds light on how well it functions in online scenarios.

We provide a comprehensive framework for data center placement in urban environments using our algorithmic approach, which has demonstrable benefits for cost-effectiveness, scalability, flexibility, and real-time decision-making as well as urban development planning. By leveraging online algorithms, our approach facilitates prompt responses to changing demands and emerging trends in dynamic urban environments, enhancing agility and adaptability in decision-making processes. Through competitive analysis, our algorithm provides cost-effective solutions relative to optimal offline algorithms, empowering governments and organizations to make informed decisions within budgetary constraints and resource allocations. With its scalable and flexible nature, our algorithm accommodates diverse demand levels, geographical distributions, and infrastructure requirements, ensuring adaptability to evolving urban landscapes and economic dynamics. Moreover, our algorithm contributes to evidence-based urban development planning by analyzing data-driven insights and economic indicators to strategically place data centers, fostering economic growth, innovation, and regional development in alignment with broader urban planning objectives and sustainability goals.

11 CONCLUSION AND FUTURE DIRECTIONS

In conclusion, our study sheds light on the critical role of data analytics in shaping economic policy and fostering sustainable growth, underscored by the strategic placement of data centers. Through the lens of online algorithms, we address the challenge of dynamically situating data centers in urbanized environments, considering factors such as geographic distribution, infrastructure, and economic priorities. Our novel approach, encapsulated in the Online Data Center Placement problem formulation and algorithm design, offers a systematic framework for decision-making amidst uncertainty and evolving demands.

One possible future direction is to investigate the application of our algorithm in data centers and urbanized environments, either in real or simulated environments. Practical testing allows for evaluation under real-world conditions, taking into account factors like data variability and constraints related to urban infrastructure, while competitive analysis offers insightful information about theoretical worst-case scenarios. We can learn more about the algorithm's performance in dynamic urban landscapes and its strengths and limitations in handling real-time data and decision-making in the context of data center placement strategies by conducting practical evaluations in urbanized environments.

Further research could expand our model to incorporate data security and infrastructure resilience. Infrastructure resilience is the process of strategically placing data centers to fortify urban infrastructure networks against external shocks such as cyberattacks and natural disasters. To reduce failures and increase redundancy, this involves determining key sites. Integrating data security goals also guarantees privacy requirements are followed and sensitive data is protected. By taking these aspects into account, data center location techniques are improved and help create safe and dependable urban data ecosystems, which are essential for reliable data-driven urban development projects.

Combining data center location methods with more general smart city initiatives to improve urban efficiency, sustainability, and quality of life is an intriguing new direction for future research. With this integration, there is a chance to use data analytics to improve public services, transit networks, urban infrastructure, and even data center locations. Cities may optimize the benefits of data-driven approaches to societal issues and create synergies by matching data center site decisions with smart city objectives.

REFERENCES

Albers, S. (2003). Online algorithms: a survey. *Mathematical Programming*, 97:3–26.

Alon, N., Awerbuch, B., and Azar, Y. (2003). The online set cover problem. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '03, page 100–105, New York, NY, USA. Association for Computing Machinery.

Awan, U., Shamim, S., Khan, Z., Zia, N. U., Shariq, S. M., and Khan, M. N. (2021). Big data analytics capability and decision-making: The role of data-driven insight on circular economy performance. *Technological Forecasting and Social Change*, 168:120766.

Bank of England (Accessed: 2024). Bank of England Statistics Research Datasets.

Bienkowski, M., Feldkord, B., and Schmidt, P. (2021). A nearly optimal deterministic online algorithm for non-metric facility location. In Bläser, M. and Monmege, B., editors, *38th International Symposium on Theoretical Aspects of Computer Science (STACS 2021)*, volume 187 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 14:1–14:17, Dagstuhl, Germany. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

Borodin, A. and El-Yaniv, R. (2005). *Online computation and competitive analysis*. cambridge university press.

German Institute for Economic Research (DIW Berlin) (Accessed: 2024). German Institute for Economic Research (DIW Berlin).

Johnson, M., Jain, R., Brennan-Tonetta, P., Swartz, E., Silver, D., Paolini, J., Mamonov, S., and Hill, C. (2021). Impact of big data and artificial intelligence on industry: developing a workforce roadmap for a data driven economy. *Global Journal of Flexible Systems Management*, 22(3):197–217.

Korman, S. (2004). On the use of randomization in the online set cover problem. *Weizmann Institute of Science*, 2.

Liu, R., Sun, W., and Hu, W. (2023). Placement of high availability geo-distributed data centers in emerging economies. *IEEE Transactions on Cloud Computing*.

Markarian, C. (2021). Online non-metric facility location with service installation costs. In *ICEIS (1)*, pages 737–743.

Markarian, C. (2022). Online non-metric facility location with service-quality costs. In *ICEIS (1)*, pages 616–622.

Markarian, C. and auf der Heide, F. M. (2019). Online algorithms for leasing vertex cover and leasing non-metric facility location. In *ICORES*, pages 315–321.

National Institute of Economic and Social Research (NIESR) (Accessed: 2024). National Institute of Economic and Social Research (NIESR).

United States Census Bureau (Accessed: 2024). United States Census Bureau Special Census Program.