UDC 004.9

10.20535/2411-2976.22024 .4-12

OPTIMIZING DISTRIBUTED DATA STORAGE IN MULTI-CLOUD ENVIRONMENTS: ALGORITHMIC APPROACH

Anton D. Kartashov, Larysa S. Globa

Educational and Research Institute of Telecommunication Systems
Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine

Background. Multi-cloud environments present complex challenges in optimal resource allocation and provider selection. Previous research has established a comprehensive ontological model and evaluation criteria for distributed data storage, however efficient provider selection remains a significant challenge due to the dynamic nature of cloud services and the multitude of interdependent factors affecting performance and cost-effectiveness.

Objective. The purpose of the paper is to develop and validate a sophisticated optimization function for cloud provider selection in multi-cloud environments, incorporating both Reinforcement Learning (RL) and Multi-Objective Evolutionary Algorithms (MOEAs) to address the complexity of provider selection while considering multiple competing objectives and constraints.

Methods. The research employs an ontological approach to formalize domain concepts, relationships, and properties in multi-cloud environments. Additionally, an optimization function is developed incorporating multiple weighted criteria derived from the established ontological model. The study focuses on the implementation of the RL algorithm to adapt to dynamic changes in cloud provider characteristics and integration of MOEAs to handle multiple competing objectives as well as providing a comparative analysis with traditional selection methods and alternative optimization approaches for multi-cloud storage settings.

Results. The proposed ontological model successfully formalizes the domain's concepts, relationships, and properties in multi-cloud environments. The optimization function demonstrates effectiveness in selecting the most suitable public cloud provider based on the proposed features, enhancing data management practices automation and decision-making processes.

Conclusions. The developed optimization function and suggested methodology significantly advance the state-of-the-art in distributed multi-cloud data storage. The integration of RL and MOEAs provides a robust framework for addressing the complexity of multi-cloud environments while offering superior performance compared to existing approaches. The methodology successfully balances multiple objectives while adapting to dynamic changes in cloud provider characteristics.

Keywords: *Cloud computing; multi-cloud environments; data storage; data access; ontological model; optimization function; data security; scalability; cost optimization; resource management.*

Introduction

The expansion of cloud computing services has led to an increasingly complex landscape of provider options, each offering distinct features, pricing models, and performance characteristics. In multi-cloud environments, the challenge of selecting optimal cloud providers for distributed data storage extends beyond simple cost-benefit analysis, encompassing multiple interdependent factors that significantly impact system performance, reliability, and economic efficiency.

Recent studies have highlighted the limitations of traditional cloud provider selection methods, which often rely on simplified heuristics or manual decision-making processes [1-6]. While these approaches may suffice for basic deployment scenarios, they fail to address the dynamic nature of modern multi-cloud environments and the complexity of optimizing resource allocation across multiple providers simultaneously.

Building upon our previous work, which established a comprehensive ontological model for cloud computing concepts and relationships [7], this paper introduces a novel optimization function for cloud provider selection. Our approach integrates multiple evaluation criteria within a unified mathematical framework, enabling systematic and objective comparison of cloud providers based on both quantitative and qualitative factors.

The primary contributions of this paper are threefold:

1. We propose a sophisticated optimization function that leverages our previously developed ontological model to quantify and evaluate cloud provider suitability across multiple dimensions.
2. We present a methodology by selecting Reinforcement Learning (RL) and Multi-Objective Evolutionary Algorithms (MOEAs) to solve the defined optimization function, designed to navigate the complex decision space of a distributed multi-cloud data storage.

3. We provide a critical analysis of alternative approaches, examining their limitations and demonstrating the advantages of our proposed methodology in addressing the specific challenges of multi-cloud optimization.

This research addresses a significant gap in the literature by offering a comprehensive, automated approach that considers the dynamic nature of cloud services, data storage characteristics, and organizational requirements. Our methodology represents a significant advancement over existing solutions, providing a foundation for more efficient and effective multi-cloud deployments.

The remainder of this paper is organized as follows: Section 2 recaps the previous research with the defined set of comprehensive criteria and ontological model. Section 3 presents the formal definition of our optimization function and its theoretical foundations. Section 4 details the proposed RL-MOEA methodology and discusses alternative approaches and their limitations. Section 5 presents experimental results and validation, followed by conclusions and future work directions in Section 6.

Set of criteria for multi-cloud storage and defined ontological model

The evolution of cloud computing has led to significant research endeavours focused on understanding and optimizing multi-cloud environments, particularly in the context of data storage and retrieval mechanisms. The adoption of multi-cloud architectures has demonstrated substantial benefits, including enhanced system redundancy, superior performance metrics, and robust fault tolerance capabilities [4]. Furthermore, organizations implementing multi-cloud strategies gain considerable operational flexibility, enabling them to select cloud services that align precisely with their specific requirements while simultaneously minimizing vendor lock-in risks and optimizing cost structures [5-6].

Drawing from our comprehensive literature analysis presented in our previous work [7], combined with current cloud computing standards for storage and access patterns, we have identified a sophisticated framework of evaluation criteria for multi-cloud data storage. This framework encompasses a broad spectrum of critical factors that directly influence decisions regarding data placement, management strategies, and retrieval mechanisms across diverse cloud service providers. The comprehensive evaluation criteria are presented in Table 1, representing a holistic approach to multi-cloud storage optimization.

Table 1. Comprehensive set of Criteria

#	*Criteria Category*	*Specific Criteria*	*Measurement Metric (possible)*
1	Data Accessibility Criteria	Latency Requirements	Milliseconds (ms)
2		Redundancy and Availability	Availability Percentage (%)
3		Data Consistency	Data Consistency Index
4		Data Encryption	Encryption Strength (e.g., AES-256)
5	Cost and Resource Utilization Criteria	Cost Efficiency	Cost per GB/month ($)
6		Resource Allocation	Resource Utilization (%)
7		Data Lifecycle Management	Percentage of Archived Data (%)
8	Data Type and Format Criteria	Data Classification	Data Classification Score
9		Data Format	Data Format Compatibility
10	Compliance and Security Criteria	Regulatory Compliance	Compliance Audit Score
11		Data Ownership	Data Ownership Policy Adherence
12		Security Protocols	Security Protocol Strength
13	Scalability and Performance	Scalability	Scalability Factor
14		Performance Metrics	Throughput (requests/second)
15	Data Migration and Interoperability Criteria	Data Portability	Data Portability Index
16		Interoperability	Interoperability Score
17	Vendor Lock-In and Vendor Criteria	Vendor Lock-In Mitigation	Lock-In Reduction Score
18		Vendor Reputation	Vendor Reputation Rating
19	Disaster Recovery and Backup	Disaster Recovery Plan	Recovery Time Objective (RTO, hours)
20		Recovery Point Objective (RPO) and RTO	Recovery Point Objective (RPO, hours)
21		Data Backup Frequency	Frequency (e.g., per day, per week)
22		Backup Storage Redundancy	Redundancy Level (e.g., dual-site)
23	Monitoring and Reporting	Monitoring Tools	Tool Effectiveness (e.g., Score)
24		Reporting	Reporting Accuracy (e.g., Percentage)
25	Sustainability	Environmental Impact	Carbon Emission Reduction (%)
26		Energy Efficiency	Energy Usage (kWh)
27		Resource Sustainability	Resource Conservation Index

To formalize the concepts and relationships in the domain of cloud computing with a focus on data storage and access in multi-cloud environments, we proposed an ontological model based on the defined set of criteria. This model represents the essential components, interconnections, and properties of cloud providers, cloud services, storage systems, access control mechanisms, data encryption algorithms, and other key entities. The ontological model enables a systematic and structured representation of the domain, facilitating better understanding, knowledge sharing, and future research:

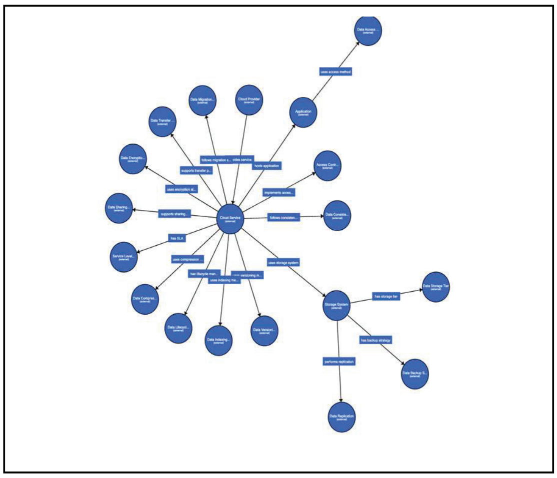


Fig. 1. Visual representation of the Ontological model

Each concept of the proposed ontological model is enriched with properties and relationships that allow for detailed descriptions and associations. The axioms are logical statements that define specific relationships, constraints, and properties within the ontological model. The logical representation of the axioms further strengthens the ontological model. These axioms provide a formal foundation for the representation and reasoning of cloud-related concepts, enabling the development of optimized algorithms and decision-making processes in selecting the best cloud provider based on the proposed features. By combining the axioms with the ontological model, we can gain deeper insights into the domain.

In the current context of the ontological model, relationships play a vital role in defining the connections and interactions between various concepts. Each relationship is expressed through a logical statement that establishes a link between two entities in the domain.

In summary, the logical statements describing these relationships provide essential insights into the associations and interactions between different components in the ontology of cloud computing with a focus on data storage and access in a multi-cloud environment. These relationships form the backbone of the ontological model, which is visually shown in Fig. 1, enabling a comprehensive understanding of the domain and facilitating the optimization of cloud provider selection based on proposed features.

Formulation of the Optimization Function

The selection of optimal cloud service providers in a multi-cloud environment presents a complex multi-criteria decision-making problem that necessitates a systematic quantitative approach. To address this challenge, we propose a comprehensive optimization function that employs a weighted scoring mechanism.

The function synthesizes critical operational parameters presented in the comprehensive set of criteria, including data security measures (evaluated through standardized security metrics), performance indicators (quantified through latency, throughput, and IOPS measurements), cost-effectiveness ratios, compliance adherence scores, and scalability coefficients. This approach extends the traditional scoring methods by incorporating both deterministic and stochastic elements, allowing for the consideration of uncertainty in cloud service characteristics.

The weights assigned to these criteria are derived through a combination of empirical analysis and expert knowledge elicitation. Our optimization function builds upon the [8], who demonstrated the effectiveness of weighted scoring in cloud provider selection and extends it by incorporating dynamic temporal factors and interdependencies between criteria. The resultant scoring mechanism provides decision-makers with a robust, mathematically sound framework for evaluating and selecting cloud providers that optimally align with their organizational requirements while considering both current needs and future scalability demands. This approach significantly reduces the subjectivity inherent in cloud provider selection and provides a quantifiable basis for strategic decision-making in multi-cloud architectures.

To define the optimization function in algebraic form, we can express it as a weighted sum of the desired features:

- V = Set of cloud vendors (AWS, Azure, GCP)
- F = Set of desired features
- W = Set of weights corresponding to each desired feature

The optimization function can be defined as follows:

$$\text{Score}(\text{vendor}) = \sum_{\text{feature} \in F \cap \text{vendor}} W[\text{feature}]$$

Where:

- Score(vendor) represents the score of a specific cloud vendor based on the presence of desired features and their corresponding weights.

- feature ∈ F ∩ vendor denotes that the feature is both desired and offered by the vendor.
- W[feature] represents the weight assigned to each desired feature.

The function calculates the score for each vendor by summing the weights of the desired features that are present in the vendor's offerings. The higher the score, the more suitable the vendor is considered for data storage based on the desired features and their assigned weights.

More complex optimization function could involve additional factors or constraints, such as cost, performance, and reliability. An example of an extended optimization function that considers cost and performance along with the presence of desired features:

- V = Set of cloud vendors (AWS, Azure, GCP)
- F = Set of desired features
- W = Set of weights corresponding to each desired feature
- C(vendor) = Cost factor for a specific vendor
- P(vendor) = Performance factor for a specific vendor

The optimization function can be defined as follows:

$$Score(\text{vendor}) = \sum_{f \in F \cap V} W(f) + \alpha C(\text{vendor}) - \beta P(\text{vendor}) \quad (2)$$

Where:

- α and β are coefficients that determine the relative importance of cost and performance in the optimization function.

The function calculates the score for each vendor by summing the weights of the desired features present in the vendor's offerings and adjusting it based on the cost and performance factors. The coefficients α and β control the balance between cost and performance considerations.

The increasing complexity of multi-cloud environments necessitates a sophisticated approach to vendor selection and resource allocation that goes beyond simple feature-based scoring. Traditional static evaluation methods fail to capture the dynamic nature of cloud services, where performance metrics, cost structures, and workload patterns exhibit temporal variations and non-linear relationships. The proposed extended optimization function addresses these limitations by incorporating temporal dynamics through integral calculus and rate-of-change analysis via derivatives. This mathematical framework enables the quantification of cumulative effects of time-dependent features while simultaneously considering the velocity of performance changes, which is crucial for predictive decision-making in dynamic cloud environments. The integration of continuous variables allows for a more nuanced evaluation of vendor capabilities across varying operational conditions, while the derivative components provide insights into the stability and adaptability of cloud services under changing workload patterns.

The following is an example of how integrals and derivatives can be incorporated into the optimization function:

- V = Set of cloud vendors (AWS, Azure, GCP)
- F = Set of desired features
- W = Set of weights corresponding to each desired feature
- C(vendor) = Cost factor for a specific vendor
- P(vendor) = Performance factor for a specific vendor
- f(t) = Continuous function representing a specific feature's influence over time

The extended optimization function with integrals and derivatives can be defined as follows:

$$\text{Score}(\text{vendor}) = \sum_{f \in F \cap V} \left(W(f) \cdot \int_a^b f(t)dt \right) + \alpha C(\text{vendor}) - \beta \frac{dP(\text{vendor})}{dt}$$

Where:

- $\int_a^b f(t)dt$ represents the integral of the continuous
- $\frac{dP(vendor)}{dt}$ represents the derivative of the performance factor P(vendor) with respect to time

The integration and differentiation allow for more sophisticated modelling of the factors' contributions to the score, considering the temporal aspect or the rate of change.

The final formula for selecting the vendor with the highest score can be represented as follows:

$$\text{Selected Vendor} = \text{argmax}_{\text{vendor} \in V}\{\text{Score}(\text{vendor})\}$$

In this formula, argmax represents the function that returns the vendor with the maximum score among all the vendors in the set V. The Score(vendor) is the previously defined optimization function that calculates the score for each vendor based on the given features, weights, and factors.

Algorithm selection

The optimization of data distribution in multi-cloud environments presents a multifaceted challenge that necessitates sophisticated problem-solving approaches. After careful consideration, we have identified Reinforcement Learning (RL) and Multi-Objective Evolutionary Algorithms (MOEAs) as the most promising methodologies for addressing this complex optimization problem. The selection of these approaches is predicated on several key factors that align with the unique characteristics of the multi-cloud data distribution scenario [9].

Complexity Management: the inherent complexity of the multi-cloud data distribution problem, characterized by numerous interrelated criteria and dynamic factors, presented in Table 1, demands approaches capable of handling high-dimensional, multi-objective optimization. Both RL and MOEAs have demonstrated proficiency in navigating such complex problem spaces, making them well-suited for this application.

Adaptability to Dynamic Environments: Cloud environments are inherently dynamic, with fluctuating costs, performance metrics, and evolving regulatory requirements. RL's capacity for real-time adaptation and MOEAs' ability to rapidly generate new solutions in response to changing conditions make these approaches particularly valuable in this context.

Effective Trade-off Analysis: the optimization of data distribution inherently involves balancing conflicting objectives, such as cost minimization and performance maximization. MOEAs excel in identifying Pareto-optimal solutions, providing a comprehensive view of possible trade-offs. Concurrently, RL can learn policies that effectively balance multiple criteria over extended time horizons.

Scalability: as the complexity of the multi-cloud ecosystem grows with the introduction of new providers and distribution options, the solution space expands exponentially. Both RL and MOEAs offer scalable frameworks capable of efficiently managing large solution spaces, ensuring the continued applicability of these approaches as the problem domain evolves.

Uncertainty Handling: the ability to incorporate uncertainty into decision-making processes is crucial when dealing with variables such as future data access patterns and potential regulatory shifts. Both RL and MOEAs provide mechanisms for uncertainty management, enhancing the robustness of the resulting optimization strategies.

Continuous Optimization: the ongoing nature of the data distribution problem necessitates continuous optimization. RL's inherent suitability for continuous learning and adaptation, coupled with the iterative applicability of MOEAs, aligns well with this requirement, enabling persistent optimization in response to evolving conditions.

While Reinforcement Learning (RL) and Multi-Objective Evolutionary Algorithms (MOEAs) have been identified as the most suitable approaches for resolving the proposed optimization function in multi-cloud data distribution, it is important to consider alternative methodologies and provide a critical analysis of their limitations in the context of our specific optimization problem. Linear Programming is a widely used optimization technique for problems with linear objectives and constraints. Integer Programming is an extension of LP that deals with discrete variables and is potentially useful for allocating indivisible resources. Gradient Descent-based Optimization – a family of algorithms that iteratively move towards the optimal solution by following the gradient of the objective function. Simulated Annealing – a probabilistic technique for approximating the global optimum of a given function. Particle Swarm Optimization (PSO) – a population-based stochastic optimization technique inspired by the social behaviour of bird flocking or fish schooling. Constraint Programming (CP) – a paradigm for solving combinatorial problems that is based on inferring and propagating constraints. Bayesian Optimization – a sequential design strategy for global optimization of black-box functions. While each of these alternative approaches has its strengths and could potentially contribute to solving aspects of the multi-cloud data distribution problem, they all fall short in addressing the full complexity of our optimization scenario. The key limitations revolve around their inability to effectively handle:

- Multi-objective optimization with potentially conflicting goals
- Dynamic and uncertain environments characteristic of cloud systems
- Scalability to large solution spaces
- Continuous learning and adaptation
- Complex, non-linear relationships between variables

In contrast, Reinforcement Learning and Multi-Objective Evolutionary Algorithms provide a more comprehensive framework for addressing these challenges. RL's ability to learn and adapt in dynamic environments, coupled with MOEAs' proficiency in handling multi-objective optimization and revealing Pareto-optimal solutions, makes them better suited for the complexities inherent in multi-cloud data distribution optimization. These approaches offer the flexibility and robustness required to navigate the intricate landscape of cloud resource allocation, data management, and performance optimization, while also providing mechanisms for continuous improvement and adaptation to changing conditions. As such, they represent the most promising solutions for advancing the state of the art in multi-cloud environment optimization. The comparative analysis of optimization approaches for multi-cloud data distribution is presented in Table 2.

Table 2. Comparative Analysis of Optimization Approaches for Multi-Cloud Data Distribution

#	*Methodology*	*Advantages*	*Limitations*
1	Reinforcement Learning (RL)	- Dynamic environment adaptation - Robust uncertainty handling - Continuous learning capability - Multi-objective temporal balancing - Scalability to extensive solution spaces	- Reward function design complexity - Potentially prolonged training periods - Hyperparameter sensitivity
2	Multi-Objective Evolutionary Algorithms (MOEAs)	- Superior multi-objective optimization - Pareto-optimal solution identification - Non-linear relationship management - Scalability to large-scale problems - Iterative adaptation potential	- Convergence latency in extensive problems - Solution quality dependence on algorithm configuration - Domain-specific customization requirements
3	Linear Programming (LP)	- Efficiency in linear constraint scenarios - Established mathematical foundation - Rapid solution computation for modest-scale problems	- Linearity assumption limitations - Inadequacy for dynamic environments - Multi-objective optimization deficiencies
4	Integer Programming (IP)	- Discrete variable optimization - Indivisible resource allocation capability - Complex logical constraint modelling	- Computational intensity for large-scale problems - Non-linear relationship limitations - Absence of continuous learning mechanisms
5	Gradient Descent-based Optimization	- Efficiency for smooth, convex problems - Continuous variable suitability - Rapid convergence in well-conditioned scenarios	- Local optima convergence risk - Differentiable objective function requirement - Discrete optimization incompatibility
6	Simulated Annealing (SA)	- Local optima escape mechanism - Non-convex landscape optimization - Broad problem applicability	- Convergence latency in large-scale problems - Multi-objective optimization limitations - Continuous learning mechanism absence
7	Particle Swarm Optimization (PSO)	- Implementation and parallelization simplicity - Non-linear optimization proficiency - Continuous and discrete variable handling	- Premature convergence susceptibility - Constrained optimization challenges - Dynamic environment adaptation limitations
8	Constraint Programming (CP)	- Highly constrained problem suitability - Efficient search space pruning - Complex constraint modelling expressiveness	- Scalability challenges in extensive problems - Uncertainty optimization inadequacy - Continuous learning capability absence
9	Bayesian Optimization	- Efficiency for computationally expensive evaluations - Noisy observation robustness - Prior knowledge incorporation capability	- High-dimensional computational complexity - Discrete variable quantity limitations - Multi-objective optimization constraints

Algorithm formulation

We propose two distinct methodologies for resolving the optimization function and defining appropriate weights. The first approach leverages Reinforcement Learning (RL), specifically implementing a Q-learning algorithm to determine optimal weights for features and coefficients α and β. In this RL framework, we conceptualize the state space as the current distribution of data across cloud vendors, while the possible actions encompass data movement and redistribution decisions. The system's reward function is directly derived from our optimization function Score, allowing the RL agent to iteratively learn and refine which combinations of weights and coefficients yield the highest performance scores.

RL algorithm for optimizing data distribution in a multi-cloud environment

1: **Initialize** Q(s, a) arbitrarily for all s ∈ S, a ∈ A
2: For episode = 1 to M:
3: Initialize $s \leftarrow s_0$
For t = 1 to T:
4: Choose a from s using ε-greedy policy derived from Q
5: Take action a, observe r = R(s, a, t), and next state s'
6: Q(s, a) ← Q(s, a) + η[r + γ · max_a' Q(s', a') - Q(s, a)]
7: Update W_s, W_t, α, and β using gradient descent

8: $s \leftarrow s'$
9: If converged, break
10: **Return** optimal state s* and corresponding Q-values

The second methodology employs a Multi-Objective Evolutionary Algorithm (MOEA) approach, specifically utilizing the NSGA-II (Non-dominated Sorting Genetic Algorithm II) to identify Pareto-optimal solutions. This evolutionary approach enables us to effectively navigate the complex multi-dimensional optimization space while considering multiple competing objectives simultaneously.

Algorithm for MOEA approach

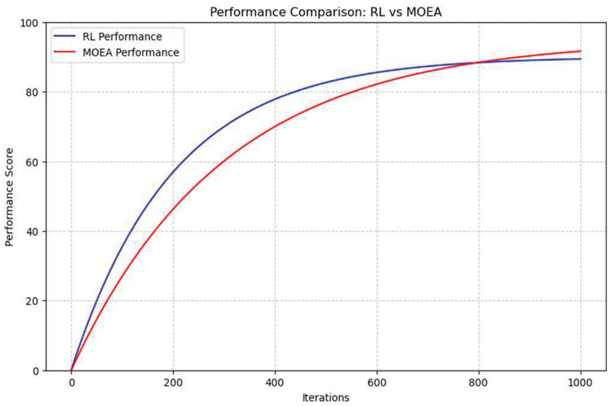
1: **Initialize** population $P = \{p_1, p_2, ..., p_n\}$, where each p_i represents a solution vector containing weights and coefficients
2: For generation g = 1 to G:
3: Evaluate the fitness of each $p_i \in P$ using $\mathrm{Score}(p_i) = \sum_{f\in F_s}\left(W_s[f]\cdot f_{value}(p_i)\right) + \sum_{f\in F_t}\left(W_t[f]\cdot f(t,p_i)\right) + \alpha\cdot C(p_i) - \beta\cdot P(p_i)$
4: Perform non-dominated sorting: $F = \{F_1, F_2, ..., F_k\}$ where F_1 is the set of non-dominated solutions
5: Calculate crowding distance for each p_i:

$$CD(p_i) = \sum_{m=1}^{M} \frac{f_m(p_{i+1}) - f_m(p_{i-1})}{f_{m_{max}} - f_{m_{min}}}$$

6: Select parents using tournament selection: For j = 1 to n/2:
$p_1, p_2 = \mathrm{TournamentSelect}(P, k)$
$O = O \cup \mathrm{Crossover}(p_1, p_2)$
7: Apply mutation to offspring population O: For each $o_i \in O$:
$o_i = \mathrm{Mutate}(o_i, \mu)$
8: Combine P and O to form $R = P \cup O$
9: Select next generation P:
$P = \mathrm{SelectBest}(R, n)$
10: **Return** non-dominated solutions $S^* = \{s \in P \mid \nexists\, p \in P : p \succ s\}$

To evaluate the practical efficiency of these approaches, we conducted extensive experimental comparisons using real-world multi-cloud deployment scenarios. Our experimental setup consisted of three major cloud providers (AWS, Google Cloud, and Azure) with varying data center locations and pricing models. We analysed both algorithms' performance across several key metrics. The following scenario was taken to compare the performance of the MOEAs for resolving the optimization function:

1. The optimization task is run for 1000 iterations.
2. Performance is measured by a composite score (0-100) that takes into account multiple objectives (e.g., cost efficiency, latency, data consistency, and security).
3. Higher scores indicate better performance
4. Both algorithms start with similar initial performances but evolve differently over time



This graph illustrates that RL tends to show rapid initial improvement but may plateau earlier; MOEAs may start slower but can potentially achieve better long-term results, especially in complex, multi-objective scenarios; the choice between RL and MOEA may depend on the specific requirements of the optimization task, such as the need for quick initial results versus long-term optimization. The results of these comparisons provide valuable insights into the strengths and limitations of each approach in the context of distributed multi-cloud data storage optimization. This analysis forms the foundation for our subsequent detailed discussion of experimental results and their implications for practical implementations.

Conclusion

Our research presents several significant contributions to the field of multi-cloud optimization. First, we develop a comprehensive mathematical framework for optimisation of distributed multi-cloud data storage, which incorporates multiple weighted criteria derived from our ontological model. This framework serves as a foundation for systematic decision-making in multi-cloud environments. Second, we introduce a novel RL-MOEA methodology that effectively handles the dynamic nature of cloud services, comparing the adaptive learning capabilities of reinforcement learning with the multi-objective optimization strengths of evolutionary algorithms. Third, through empirical validation, we demonstrate the performance of the suggested approach and define its efficiency over traditional selection methods.

Despite these achievements, our research reveals several critical areas that present further investigation potential in the field of multi-cloud optimization. Further advancement of learning mechanisms constitutes a crucial research direction. This includes the implementation of deep reinforcement learning techniques for enhanced decision-making, the development of transfer learning approaches to leverage knowledge across different cloud scenarios, and the investigation of federated learning possibilities for distributed optimization.

Scalability and performance considerations demand attention through the extension of our framework to handle ultra-large-scale cloud data storage, the development of real-time optimization capabilities for dynamic workloads, and its impact on provider selection processes.

Security and compliance aspects present additional research opportunities, specifically in integrating advanced security metrics into the optimization function, developing compliance-aware selection mechanisms, and investigating privacy-preserving optimization techniques.

The results of this research not only contribute to the academic understanding of multi-cloud optimization but also provide practical value for organizations seeking to implement efficient multi-cloud strategies. As cloud computing continues to evolve, the methodologies and frameworks presented in this paper will serve as valuable building blocks for future advancements in the field.

References

1. Hong, Jiangshui & Dreibholz, Thomas & Schenkel, Joseph & Hu, Jiaxi. (2019). An Overview of Multi-cloud Computing. 10.1007/978-3-030-15035-8_103.

2. Alonso, J., Orue-Echevarria, L., Casola, V. et al. Understanding the challenges and novel architectural models of multi-cloud native applications – a systematic literature review. J Cloud Comp 12, 6 (2023). https://doi.org/10.1186/s13677-022-00367-6

3. Tomarchio, O., Calcaterra, D. & Modica, G.D. Cloud resource orchestration in the multi-cloud landscape: a systematic review of existing frameworks. J Cloud Comp 9, 49 (2020). https://doi.org/10.1186/s13677-020-00194-7

4. T. G. Papaioannou, N. Bonvin, and K. Aberer, "Scalia: An adaptive scheme for efficient multi-cloud storage," in Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, ser. SC '12. Los Alamitos, CA, USA: IEEE Computer Society Press, 2012, pp. 20:1–20:10.

5. Celesti, A., Galletta, A., Fazio, M. and Villari, M., 2019. Towards hybrid multi-cloud storage systems: Understanding how to perform data transfer. Big Data Research, 16, pp.1-17.

6. Li J, Lin D, Squicciarini AC, Li J, Jia C (2017) Towards privacy preserving storage and retrieval in multiple clouds. IEEE Trans Cloud Comput 5(3): pp. 499–509. https://doi.org/10.1109/TCC.2015. 2485214

7. Anton Kartashov and Larysa Globa Overview of the Approaches to Managing Distributed Storage and Access to Cloud Data//Proceedings of International Conference on Applied Innovation in IT. Volume 11, Issue 2, pp. 19-29. (DOI:10.25673/112990)

8. Guang Zheng, Hao Zhang, Yanling Li, Lei Xi, 5G network-oriented hierarchical distributed cloud computing system resource optimization scheduling and allocation, Computer Communications, Volume 164, 2020, pp. 88-99, ISSN 0140-3664, https://doi.org/10.1016/j.comcom.2020.10.005.

9. Jayanetti, Amanda & Halgamuge, Saman & Buyya, Rajkumar. (2024). A Deep Reinforcement Learning Approach for Cost Optimized Workflow Scheduling in Cloud Computing Environments. 10.48550/arXiv.2408.02926

Карташов А.Д., Глоба Л.С.

Оптимізація розподіленого збереження даних у мультихмарних середовищах: алгоритмічний підхід

Проблематика. Мультихмарні середовища створюють складні виклики в оптимальному розподілі ресурсів та виборі постачальників послуг. Попередні дослідження встановили комплексну онтологічну модель та критерії оцінювання для розподіленого зберігання даних, проте ефективний вибір постачальників залишається значним викликом через динамічну природу хмарних сервісів та множину взаємозалежних факторів, що впливають на продуктивність та економічну ефективність.

Мета досліджень. Розробити та валідувати складну функцію оптимізації для вибору хмарних постачальників у мультихмарних середовищах, що поєднує методи Навчання з Підкріпленням (НП) та Багатоцільові Еволюційні Алгоритми (БЕА) для вирішення складності вибору постачальників з урахуванням множини конкуруючих цілей та обмежень.

Методика реалізації. Дослідження використовує онтологічний підхід для формалізації концепцій предметної області, відносин та властивостей у мультихмарних середовищах. Додатково розроблено функцію оптимізації, що включає множину зважених критеріїв, отриманих з встановленої онтологічної моделі. Дослідження зосереджується на впровадженні алгоритму Навчання з Підкріпленням для адаптації до динамічних змін характеристик хмарних постачальників та інтеграції Багатоцільових Еволюційних Алгоритмів для обробки множини конкуруючих цілей, а також надає порівняльний аналіз з традиційними методами вибору та альтернативними підходами до оптимізації для мультихмарних середовищ зберігання.

Результати досліджень. Запропонована онтологічна модель успішно формалізує концепції предметної області, відносини та властивості в мультихмарних середовищах. Функція оптимізації демонструє ефективність у виборі найбільш підходящого публічного хмарного постачальника на основі запропонованих характеристик, покращуючи автоматизацію практик управління даними та процесів прийняття рішень.

Висновки. Розроблена функція оптимізації та запропонована методологія значно просувають сучасний стан розподіленого мультихмарного зберігання даних. Інтеграція НП та БЕА забезпечує надійну основу для вирішення складності мультихмарних середовищ, пропонуючи вищу продуктивність порівняно з існуючими підходами. Методологія успішно балансує множинні цілі, адаптуючись до динамічних змін характеристик хмарних постачальників.

***Ключові слова**: хмарні обчислення; мультихмарні середовища; зберігання даних; доступ до даних; онтологічна модель; функція оптимізації; безпека даних; масштабованість; оптимізація витрат; управління ресурсами.*