Optimization of virtual machine placement in a university cloud considering information security requirements*

Bakhytzhan Akhmetov[1,†], Valerii Lakhno[2,*,†], Nurzhamal Oshanova[1,†] and Gulistan Yelubay[1,†]

[1] *Abai Kazakh National Pedagogical University, 13 Dostyk, 050026 Almaty, Kazakhstan*
[2] *National University of Life and Environmental Sciences of Ukraine, 15 Heroes Of Defense, 03041 Kyiv, Ukraine*

Abstract

A mathematical model is presented for determining the optimal number of computing nodes in a private cloud within a university infrastructure based on virtual desktop infrastructure (VDI) technology, talking into account the specific information security requirements of a cloud-oriented educational environment (COEE). Unlike existing models that mainly focus on analyzing resource availability—particularly CPU load and RAM capacity—the approach proposed in this work incorporates quantitative characteristics of the security level both from the side of computing nodes and virtual machines deployed in the university cloud. The proposed solution formulates the virtual machine placement problem as a multi-criteria optimization task that balances efficient resource utilization with adherence to the defined information security policies of the university's COEE. A distinctive feature of the model is its ability to account for dynamic changes in security requirements and resource loads over time. This development makes the model applicable to university cloud infrastructures, which are characterized by a high degree of variability, such as during academic and examination periods. To verify the model, a simulation was implemented that reflects realistic parameters of IT infrastructures of universities in the Republic of Kazakhstan, using Abai Kazakh National Pedagogical University as a case study. The results of the numerical experiment demonstrated the method's resilience to increasing demands and confirmed its applicability in designing and adaptively scaling secure private cloud solutions for educational institutions.

Keywords

university cloud, virtualization, cloud-oriented educational environment, information security, mathematical model, multi-criteria resource optimization

1. Introduction

The implementation of Virtual Desktop Infrastructure (VDI) based on private clouds in universities and other educational institutions (such as schools, colleges, etc.) enables more efficient use of shared computing resources. For universities in Kazakhstan (RK), this also provides the benefit of flexible scaling according to current user needs and the load on the university cloud infrastructure. It should be noted that the approach presented in this paper is relevant for universities in RK and research and educational institutions where the IT environment is characterized by dynamic workloads during examination periods, while the requirements for information security within the cloud infrastructure remain consistently high.

Given the variability of workloads on a university's cloud infrastructure (or, as referred to in [1, 2], the cloud-oriented educational environment—COEE) and the heterogeneous requirements of users, a key challenge is to estimate the required number of computing nodes (servers) to ensure stable and secure operation of the cloud platform. Previously proposed mathematical models for estimating the number of cluster nodes in VDI-based clouds [3, 4] are generally based on analyzing resource loads [5–7], assuming that the main efficiency criterion is the minimization of the number of physical servers given specific resource consumption parameters of virtual machines.

* *CSDP'2025: Cyber Security and Data Protection, July 31, 2025, Lviv, Ukraine*
* Corresponding author.
† These authors contributed equally.
bakhytzhan.akhmetov.54@mail.ru (B. Akhmetov); lva964@nubip.edu.ua (V. Lakhno); nurzhamal_o_t@mail.ru (N. Oshanova); elubaeva2019@list.ru (G. Yelubay)
0000-0001-5622-2233 (B. Akhmetov); 0000-0001-9695-4543 (V. Lakhno); 0000-0003-4728-3821 (N. Oshanova); 0000-0002-6272-0607 (G. Yelubay)

However, we argue that such approaches overlook a fundamental aspect of the COEE—information security (IS) of the deployed computations, including trust in physical nodes, requirements for isolation, access control policies, and resilience against internal and external threats to the university IT environment [8–11].

2. Problem statement

Under conditions of high load variability, seasonal activity (during examination periods and university admissions in Kazakhstan), and continuous changes in the composition and behavior of users, the rational scaling of cloud infrastructure becomes a key issue for the cloud-oriented educational environments (COEEs) of Kazakhstani universities. Specifically, it involves determining the required number of computing nodes that ensure the stable operation of the COEE.

Existing approaches [1, 2, 4, 6, 7] for estimating the necessary capacity of a VDI cluster typically rely solely on analyzing resource characteristics—such as CPU and RAM consumption during virtual machine (VM) deployment in a COEE—and are usually reduced to optimization problems focused on minimizing the number of servers under a fixed workload.

However, in a university environment where personal, scientific, and administrative data are processed, information security (IS) cannot be treated as an external or supplementary criterion. On the contrary, it must be integrated into the very logic of evaluating the architectural parameters of a university's COEE. Neglecting the requirements for the security level of computing nodes or failing to consider the need for isolation between different user groups will inevitably lead to information security incidents—such as data leaks, inter-faculty interference, and unauthorized access. Similar challenges have been noted in research on cybercrime detection in cloud environments using honeypots [12], where ensuring reliable isolation and trust in cloud nodes proved critical for preventing attacks. Furthermore, recent studies on Shadow IT risks [13], centralized secret data management in public cloud provisioning [14], and secure configuration repository efficiency [15] highlight that security-oriented resource allocation directly influences the resilience and scalability of cloud infrastructures.

Therefore, there is a need to develop a multi-criteria model in which the task of estimating the number of virtualization nodes and placing virtual machines (VMs) takes into account not only technical resources but also indicators of trust and security—both from the side of physical servers and the VMs hosted on them. Moreover, the model must be adaptable to changing operational conditions over time, support scaling scenarios, and enable predictive assessment of cluster behavior under peak loads and heightened information security requirements.

The objective of this study is to determine the minimum required number of computing nodes that ensures, for a university's Cloud-Oriented Educational Environment (COEE):

- correct placement of virtual machines (VMs) based on available resources (CPU, RAM, etc.);
- compliance with information security requirements (such as isolation and trust in nodes);
- adaptability to dynamic workloads and scalability of the private cloud (using a Kazakhstani university as a case study).

3. Methods and models

We introduce the following notation. For the university's OSS servers, respectively: M is the total number of available physical servers (nodes) of the university's public information system; R is a total number of resource types (for example, CPU, RAM, etc.); C_{jk} is the container; k a resource on the server $j \in \{1, ..., M\}$, $k \in \{1, ..., R\}$; $S_j \in [0,1]$ is the security level (trust) of the server j. That is, the higher, S_j. The safer the whole OOUS is.

And similarly for VM parameters, respectively: N is a number of virtual machines (VMs), which must be placed in the SAUCE of the university; $active(t) \subseteq \{1, ..., N\}$ multiple active VMs at a given

time; r_{jk} is the minimum required type resources for an INSTANCE at a given t; $a_j(t)$ is the current consumption CPU BM i at a moment in time t; $H_i \in [0,1]$ is the required security level for BM i at a moment in time t.

And additionally we use the following variables: $x_{ij}(t) \in \{0,1\}$ is a binary variable that takes the value 1 if the VM is hosted on the server at the time, 0 otherwise; $y_i(t) \in \{0,1\}$ is the server in use j at a moment in time t; $\alpha \in [0,1]$ is the redundancy level in case of security (aggressiveness coefficient in assessing the security of educational institutions from the university); $\lambda > 0$ is the penalty coefficient for non-compliance with the security of the environmental management system.

We use the following restrictions.

Each VM is hosted on the same server, i. e.

$$\sum_{j=1}^{M} x_{ij}(t) = 1, \forall i \in active(t). \tag{1}$$

Resource Constraints on Each Server in the COEE

For each resource

$$\sum_{i \in active(t)} x_{ij}(t) \cdot r_{jk} \le C_{jk}, \forall j, \forall k. \tag{2}$$

Then, for the CPU, we obtain

$$\sum_{i \in active(t)} x_{ij}(t) \cdot a_j(t) \le C_j^{CPU}, \forall j. \tag{3}$$

Then, unlike the works of [3, 7], we will introduce restrictions on the safety of the university's environmental management system (hard or soft approach).

A tough approach

$$x_{ij}(t) = 1 \Rightarrow S_j \ge R_i(t). \tag{4}$$

And a soft approach with the admission of a fine

$$x_{ij}(t) \cdot \max(0, R_i(t) - S_j) \ge 0. \tag{5}$$

Additionally, the minimum allowable node security can be set for the university's

$$\text{OUS} \quad S_j \ge S_{\min}.$$

Then let's consider two variants of the objective function.

Option 1—two-criteria task (safety and cost savings for the University's information security system)

$$\min\left\{\sum_{j=1}^{M} y_i(t) + \lambda \cdot \sum_{i,j} x_{ij}(t) \cdot \max(0, R_i(t) - S_j)\right\}, \tag{6}$$

where the first term is the number of servers involved (resource efficiency) in the university's COEE;

The second is the amount of fines for violating security requirements.

Option 2—Weighted criterion transformation

$$\min\left\{\beta \cdot \sum_{j} y_i(t) + (1-\beta) \cdot \sum_{i,j} x_{ij}(t) \cdot \max(0, R_i(t) - S_j)\right\}, \tag{7}$$

where $\beta -$ the degree of importance of the effectiveness of the environmental management system in comparison with safety.

Then we will calculate the lower estimate $M(t)$.

As in [3, 7], we will use the expansion of the Martello-Tossa estimate to take into account new subsets responsible for the university's information security and its scaling:

$$\alpha_1(t) = \left\{ i \in active(t) \middle| \alpha_i(t) > C_{CPU} - \alpha \right\} - \text{"large virtual machines"}; \tag{8}$$

$$\alpha_2(t) = \left\{ i \middle| C_{CPU} - \alpha \geq \alpha_i(t) > 0,5 \cdot C_{CPU} \right\}; \tag{9}$$

$$\alpha_3(t) = \left\{ i \middle| 0,5 \cdot C_{CPU} \geq \alpha_i(t) > \alpha \right\}. \tag{10}$$

Then, to estimate the lower bound, we use the following dependence:

$$M_1(\alpha, t) = \left| \alpha_1(t) + \alpha_2(t) \right| + \max\left(0, \frac{\sum\limits_{i \in \alpha_3(t)} \alpha_i(t) - \left(C_k \cdot \left| \alpha_2(t) \right| - \sum\limits_{i \in \alpha_2(t)} \alpha_i(t) \right)}{C_k} \right). \tag{11}$$

Or

$$M_2(\alpha, t) = \left| \alpha_1(t) + \alpha_2(t) \right| + \max\left(0, \alpha_3(t) - \sum_{i \in \alpha_2(t)} \left\lfloor \frac{C_k - \alpha_i(t)}{\alpha} \right\rfloor \right). \tag{12}$$

Then the final assessment

$$M(t) = \max_{\alpha \in [0,0.5]} \left(M_1(\alpha, t), M_2(\alpha, t) \right). \tag{13}$$

We also take into account the dynamic scaling of the university's COEE. Accordingly, the ability to evaluate the system during peak loads can be expressed as follows:

$$t^* = \arg\max_t \left(\sum_{i,j} x_{ij}(t) \cdot \alpha_i(t) \right), \quad M_{peak} = M\left(t^* \right). \tag{14}$$

As part of the research conducted under project IRN AP19678846, "Enhancing the Efficiency of Hybrid and Distance Learning Formats Through the Development of University Infrastructure in the Context of Digital Transformation," a mathematical model was developed and formalized to estimate the optimal number of computing nodes in a private virtualization cluster, aimed at implementing virtual desktop infrastructure (VDI) in the university environment.

Unlike existing approaches described in [3, 7], the novelty of the proposed model lies in the following: For the first time, new parameters have been introduced that take into account the information security of the university's Cloud-Oriented Educational Environment (COEE). Furthermore, the model introduces quantitative characteristics for both the security level of computing nodes and the security requirements of the virtual machines. This enables the model to account for both administrative and technical aspects of information security when planning the cloud architecture of a specific university.

Additionally, the model incorporates temporal dynamics of workload and security requirements. That is, it considers the time-varying structure of VM resource requests within the COEE, as well as the possibility of dynamically changing information security requirements. Lower and upper bounds for the required number of cluster nodes have been extended, and new estimation formulas have been developed based on an adaptation of the Martello-Toth method, allowing consideration of both the resource intensity of VMs and constraints related to the security level of physical nodes.

The multi—criteria optimization problem of VM placement in the university COEE has been further developed, and a quality function has been proposed. This function simultaneously includes the criterion of minimizing the number of physical nodes involved and a penalty for mismatch in security levels, thereby allowing a flexible balance between infrastructure efficiency and information security within the university's COEE.

Overall, the presented model provides more acceptable accuracy and relevance in solving the problem of scaling a private university cloud under conditions of heightened information security (IS) requirements. It is also suitable as a foundational framework for the synthesis of intelligent resource management systems within secure cloud environments of universities in the Republic of Kazakhstan.

4. The results obtained

To validate the developed mathematical model, a simulation of virtual machine (VM) placement in a university's private cloud was carried out, taking into account both resource requirements and information security (IS) constraints. The model was implemented in the Python environment and made it possible to analyze workload distribution and compliance with IS requirements under conditions of a limited number of servers in the university's Cloud-Oriented Educational Environment (COEE). The results are shown in Figure 1.

The simulation considered a cloud infrastructure consisting of 10 physical servers, each with a fixed CPU capacity (100 arbitrary units) and an assigned security level $S_j \in [0.6,1.0]$. Number of VMs accepted—50 ($N = 50$ with different base resource consumption CPU (in the range from 5 to 20 conventional units) and individual requirements for the security level of the hosting platform $H_j \in [0.4,0.9]$. A total of $T = 20$ discrete time steps were used to emulate the dynamic workload of the COEE servers.

At each time step, an attempt is made to place all VMs on the servers. The condition for successful placement is simultaneous compliance with the conditions: sufficiency of available CPU resources on the server; server security level compliance with VM requirements $S_j \geq R_i$.

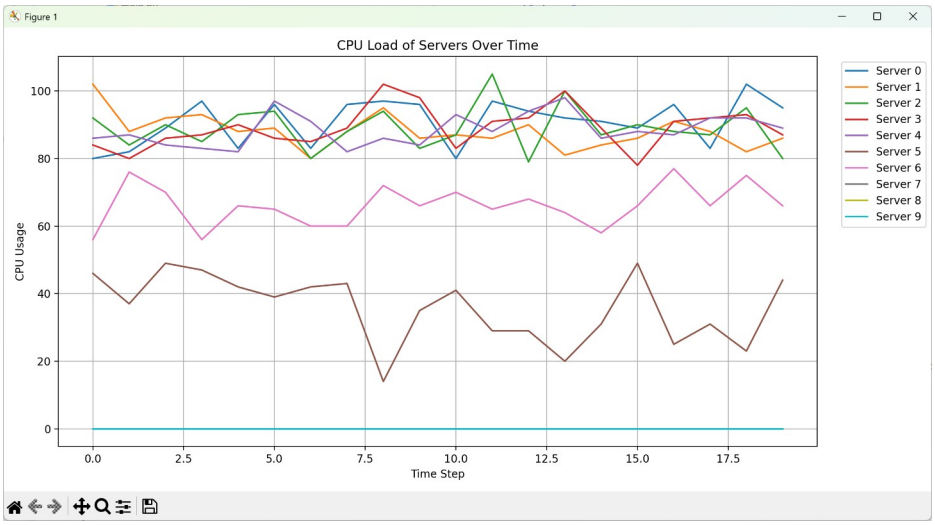

Figure 1: A graph of a computational experiment for simulating the load of the university's OSU servers over time

The simulation results are presented in Figure 1 as a graph showing CPU load per server over time. The analysis revealed a balanced distribution of workload across servers during periods of moderate activity. However, periodic load spikes indicate the difficulty of satisfying all requirements under a static COEE infrastructure. Significant differences in server load levels were observed, primarily due to filtering based on the security level criterion. During the simulation, the system was unable to place some VMs due to the simultaneous shortage of resources and insufficient trust levels in the available servers. Specifically, around 40 VM placement failures were recorded, mainly during time steps where certain VMs—particularly those with high security requirements—could not be allocated at any point throughout the simulation [16–18].

The graph in Figure 2 illustrates the trade-off between efficiency and security of virtual machines in the COEE at different values of the weight coefficient β, highlighting the prioritization of resource usage (efficiency) versus compliance with the COEE's information security requirements [19]. So when $\beta = 0$ model (1)–(14) it is exclusively focused on COEE information security, and accordingly, minimizes penalties for placing VMs on insufficiently protected servers, regardless of the number of active nodes. For $\beta = 1$ the model minimizes the number of servers involved, ignoring information security requirements. Values between 0 and 1 show varying degrees of compromise. That is, as β increases, the cost of resource efficiency decreases, but the risk of violating information security requirements increases. The obtained results were used to justify the selection of the optimal value of β based on the policy of Abai Kazakh National Pedagogical University (strict COEE information security policy of Abai KazNPU) $\beta \approx 0.3$; saving resources $\beta \approx 0.7$.

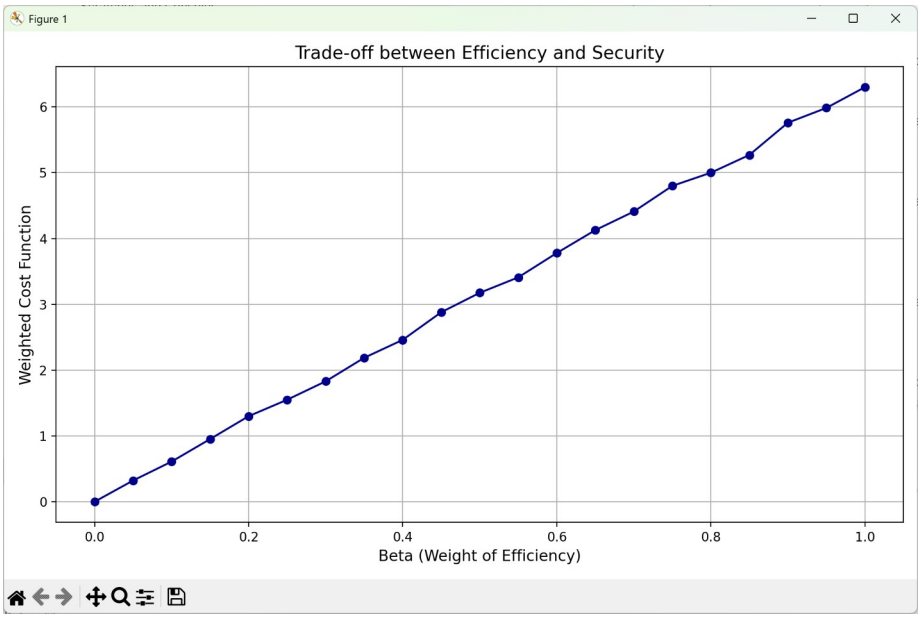

Figure 2: Compromise graph between efficiency and security of virtual machines in the COEE under different weight values β, based on data from Abai Kazakh National Pedagogical University

As a result, the conducted simulation confirms the necessity of considering information security when designing the architecture of a private university cloud. The model presented in the paper demonstrates that ignoring security requirements when estimating the number of required nodes leads to the risk of partial unavailability of computing resources, even when the total CPU capacity of the COEE servers is formally sufficient.

Conclusions

As part of the grant project IRN AP19678846 "Enhancing the Efficiency of Hybrid and Distance Learning Formats Through the Development of University Infrastructure in the Context of Digital Transformation", a mathematical model was developed to estimate the optimal number of computing nodes in a private university virtualization cloud, taking into account both technical constraints and information security requirements of the university's Cloud-Oriented Educational Environment (COEE).

Unlike existing approaches, the proposed model incorporates the trust level of COEE servers and the security requirements of virtual machines, enabling the formalization of information security (IS) risks and their inclusion in the objective function. This has made it possible to construct a multi-criteria optimization problem for VM placement, focused on achieving a balance between resource efficiency and COEE security in universities.

The simulation conducted in the Python programming environment demonstrated that ignoring IS factors leads to the failure to place certain VMs, even when the cluster's total computational capacity appears sufficient. The results confirm the relevance of integrating IS policies into planning and scaling models for university cloud infrastructure.

We believe the developed model is well-suited to serve as a foundation for designing intelligent cloud management systems in the educational environment, aimed at adaptive load balancing with consideration of information security requirements.

Acknowledgements

This work was carried out within the framework of the grant research project IRN AP19678846: "Enhancing the Efficiency of Hybrid and Distance Learning Formats Through the Development of University Infrastructure in the Context of Digital Transformation".

Declaration on Generative AI

While preparing this work, the authors used the AI programs Grammarly Pro to correct text grammar and Strike Plagiarism to search for possible plagiarism. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the publication's content.

References

[1] V. Bykov, M. Shyshkina, The Conceptual Basis of the University Cloud-based Learning and Research Environment Formation and Development in View of the Open Science Priorities, Inf. Technol. Learning Tools, 68(6) (2018) 1–19.

[2] O. Glazunova, M. Shyshkina, The Concept, Principles of Design and Implementation of the University Cloud-based Learning and Research Environment, arXiv, 2018. doi:10.48550/arXiv.1807.08560

[3] J. Park, H. Kim, Y. Jeong, Efficiency Sustainability Resource Visual Simulator for Clustered Desktop Virtualization based on Cloud Infrastructure, Sustainability, 6(11) (2014) 8079–8091.

[4] B. Akhmetov,et al., Estimation of the Required Number of Nodes of a University Cloud Virtualization Cluster, Int. J. Electrical & Comput. Eng. (2088-8708), 15(1) (2025).

[5] J. Zhang, et al., A Unified Algorithm for Virtual Desktops Placement in Distributed Cloud Computing, Math. Problems Eng. 2016(1) (2016) 9084370.

[6] Y. Huang, H. Xu, H. Gao, X. Ma, W. Hussain, SSUR: An Approach to Optimizing Virtual Machine Allocation Strategy based on User Requirements for Cloud Data Center, IEEE Transactions on Green Commun. Netw. 5(2) (2021) 670–681.

[7] M. Bichler, B. Speitkamp, A Mathematical Programming Approach for Server Consolidation Problems in VirtualizedData Centers, in: IEEE Transactions on Services Computing, 3, 2010.

[8] N. Gephart, B. Kuperman, Design of a Virtual Computer Lab Environment for Hands-on Information Security Exercises, J. Comput. Sci. Colleges, 26(1) (2010) 32-39.

[9] M. Alqahtani, Factors Affecting Cybersecurity Awareness among University Students, Appl. Sci. 12(5) (2022) 2589.

[10] J. Ulven, G. Wangen, A Systematic Review of Cybersecurity Risks in Higher Education, Future Internet, 13(2) (2021) 39.

[11] M. Haque, et al., Cybersecurity in Universities: An Evaluation Model, SN Computer Science, 4(5) (2023) 569.

[12] V. Susukailo, et al., Cybercrimes Investigation via Honeypots in Cloud Environments, in: Cybersecurity Providing in Information and Telecommunication Systems, 2923, 2021, 91–96.

[13] Y. Martseniuk, et al., Shadow IT Risk Analysis in Public Cloud Infrastructure, in: Cyber Security and Data Protection, 3800, 2024, 22–31.

[14] Y. Martseniuk, et al., Universal Centralized Secret Data Management for Automated Public Cloud Provisioning, in: Cybersecurity Providing in Information and Telecommunication Systems II, 3826, 2024, 72–81.

[15] Y. Martseniuk, et al., Research of the Centralized Configuration Repository Efficiency for Secure Cloud Service Infrastructure Management, in: Cybersecurity Providing in Information and Telecommunication Systems, 3991, 2025, 260–274.

[16] O. Mykhaylova, M. Korol, R. Kyrychok, Research and Analysis of Issues and Challenges in Ensuring Cyber Security in Cloud Computing, in: Cybersecurity Providing in Information and Telecommunication Systems II, vol. 3826, 2024, 30–39.

[17] Y. Martseniuk, et al., Automated Conformity Verification Concept for Cloud Security, in: Workshop on Cybersecurity Providing in Information and Telecommunication Systems, CPITS, vol. 3654 (2024) 25–37.

[18] V. Shapoval, et al., Automation of Data Management Processes in Cloud Storage, in: Workshop on Cybersecurity Providing in Information and Telecommunication Systems, CPITS, vol. 3654 (2024) 410–418.

[19] Q. Yao, Y. Wu, J. Gao, Research on Application of Cloud Desktop Virtualization for Computer Laboratories in Universities, in: IOP Conference Series: Materials Science and Engineering, 563(5), 2019, 052028.