International Journal of Electrical and Computer Engineering (IJECE)
Vol. 15, No. 1, February 2025, pp. 580~591
ISSN: 2088-8708, DOI: 10.11591/ijece.v15i1.pp580-591

Estimation of the required number of nodes of a university cloud virtualization cluster

Bakhytzhan Akhmetov[1], Valery Lakhno[2], Nurzhamal Oshanova[1], Zhuldyz Alimseitova[3], Madina Bereke[1], Nurgul Izbasova[4]
[1]Department of Informatics and Informatization of Education, Institute of Mathematics, Physics and Computer Science, Kazakh National Pedagogical University named after Abai, Almaty, Kazakhstan
[2]Department of Computer Systems, Networks and Cybersecurity, Faculty of Information Technology, National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine
[3]Department of Cybersecurity, Information Processing and Storage, Institute of Automation and Information Technology, Satbayev University, Almaty, Kazakhstan
[4]Department of Economics, Faculty of Business and Law, Caspian University of Technology and Engineering named after Sh. Yessenov, Aktau, Kazakhstan

Article Info

Article history:

Received Apr 23, 2024
Revised Aug 27, 2024
Accepted Sep 3, 2024

Keywords:

Cloud-based digital educational environment
Optimal placement model
University
Virtual machines
Virtualization

ABSTRACT

When designing a virtual desktop infrastructure (VDI) for a university or inter-university cloud, developers must overcome many complex technical challenges. One of these tasks is estimating the required number of virtualization cluster nodes. Such nodes host virtual machines for users. These virtual machines can be used by students and teachers to complete academic assignments or research work. Another task that arises in the VDI design process is the problem of algorithmizing the placement of virtual machines in a computer network. In this case, optimal placement of virtual machines will reduce the number of computer nodes without affecting functionality. And this, ultimately, helps to reduce the cost of such a solution, which is important for educational institutions. The article proposes a model for estimating the required number of virtualization cluster nodes. The proposed model is based on a combined approach, which involves jointly solving the problem of optimal packaging and finding the configuration of server platforms of a private university cloud using a genetic algorithm. The model introduced in this research is universal. It can be used in the design of university cloud systems for different purposes-for example, educational systems or inter-university scientific laboratory management systems.

Corresponding Author:

Zhuldyz Alimseitova
Department of Cybersecurity, Information Processing and Storage, Institute of Automation and Information Technology, Satbayev University
050000 Almaty, Kazakhstan
Email: zhuldyz_al@mail.ru

1. INTRODUCTION

For many years now, the world's leading universities have begun to use all kinds of cloud applications and technologies (CAP) to improve the efficiency of educational and scientific work. Cloud computing is a rapidly evolving computing paradigm. Its main goal is to free users of cloud resources from managing hardware and software. In this case, management tasks are transferred to the cloud service. Distributed and parallel cloud service systems consist of sets of virtual and interconnected personal computers (PCs) that can be dynamically provisioned and provided to users as more than one unified computing resource. One of the areas of widespread application of CAP has become the educational environment of universities, colleges and schools. The demand for CAP has especially increased during the COVID-19 coronavirus pandemic. As noted in [1], [2], one of the main conditions for improving the quality of training of specialists, increasing the level of their professional competence, and wider use of innovative technologies have become cloud-based digital educational environments (CBDEEs) of universities. The use of CBDEE tools contributes to the practical implementation of high-quality learning and professional development of students through the means of cloud-based open science systems.

The current pace of updating the necessary competencies, for example, in the field of technical university disciplines, often does not allow teachers to promptly update the laboratory equipment used and the corresponding software systems. Let us note that, for example, within the framework of the university educational process, the following problems of access to training materials for performing laboratory workshops in technical specialties quite often arise: i) Relatively low performance of workstations, which are mainly suitable only for general education subjects; and ii) Lack of opportunities for in-depth study of materials within laboratory workshops. In works [1]–[3], the experience of using CAP in training specialists was examined in detail and the feasibility of using such cloud service delivery models in training processes was determined: software as a service, platform as a service and infrastructure as a service based on the use of virtual machine technology.

There is no doubt that the use of CAP in training specialists in various fields has a number of advantages. The main ones include: i) For students: Universal availability of necessary electronic educational resources, mobility of programs and data, absence of significant software and hardware restrictions on the resources used, no need to administer software to achieve high performance when using programming systems, the ability to conduct non-destructive experiments in a virtualized hardware and software environment; ii) For teachers and staff: The ability to use flexible cloud-based resources (including for the purpose of developing tasks of different levels of complexity and resource consumption), the possibility of software unification in web-based operating systems, reducing the cost of administration and maintenance of university information technology (IT) infrastructure; extensive opportunities for modeling computer systems and networks, the ability to easily store and reuse virtual laboratories; iii) For administrators of computer systems and networks: Reduction of software licensing costs, removing restrictions on the hardware and software used due to virtualization technology; and iv) The ability to serve a potentially unlimited number of students, for example, within the framework of interuniversity courses, simplify and unify maintenance in the cloud.

As was shown in [4], [5] when transferring tasks characteristic of the educational process to cloud technologies, for example, when creating an inter-university private cloud, the university IT infrastructure is centralized and virtualized. With this approach, all computing resources (CR) can be collected into one computing cluster. It can then be divided into logical sections, *i.e.* virtual machines (VMs). After that, such VMs are provided to users as needed. Thus, the common hardware resources of a university or several universities can be used for different tasks, both curriculum and scientific work.

Wagh *et al.* [5] presented the results of a detailed review of the prospects for cloud computing in the field of education. The authors draw attention to the fact that many educational centers have already invested in their own cloud infrastructure instead of purchasing services from third parties. The authors focus on the fact that many educational institutions, from schools to universities, are busy searching for the best private cloud platform for education and scientific activities.

Khayyat *et al.* [6] presented a detailed overview of the capabilities of universities in the context of using CAP to organize the educational process and scientific research. The results of this work examine the CAP implementation process from the point of view of motivation and barriers to cloud computing adoption in universities. Yadegaridehkordi *et al.* [7] presented the results of research aimed at developing a model for introducing online tools for collaborative learning using CAP. The results obtained by the authors can become the basis for cloud providers targeting educational institutions.

In works [8]–[10], the authors touch upon some aspects of the economic feasibility of introducing cloud technologies into the educational process of universities, as well as interuniversity scientific cooperation. The emphasis is on reducing the complexity of the university IT infrastructure and increasing its manageability. According to the authors of the mentioned studies, the implementation of CAP will optimize the use of resources available to educational institutions. It is shown that savings can be achieved through the effective use of CR in universities, as well as centralized administration.

Zhang *et al.* [11] having analyzed the best practices of using cloud computing in universities and, based on personal experience in the field of application of information and communication technologies in the educational process of a university, identified the main CAPs to support the scientific and educational activities of the university. A number of researchers in their works [12], [13] note that the task of predicting the scale of use of server computing resources in the cloud infrastructure of an organization, or, for example, a university, is important and relevant. If there are not enough such CRs, then, accordingly, applications used in educational or scientific work will work slowly. In this case, users will encounter significant delays and problems accessing cloud services. If there is an excess of CR, then this leads to unjustified costs for cloud services on the part of the provider, since CR is used ineffectively. Understanding how CR will be used in an educational institution at the present time and in the future will make it possible to implement preventive measures aimed at balancing the load and timely scaling of cloud resources. And this, in turn, helps to ensure the smooth operation of the entire system built around CAP and improves the quality of customer service.

In study [14], the results of a comparative analysis of existing approaches and methods in the problem of short-term forecasting of the use of basic server resources by virtual machines of a cloud system are presented. The authors analyzed various models. In particular, the statistical model for analyzing time series auto-regressive integrated moving average (ARIMA) [15] homogeneous neural networks such as long short term memory network (LSTM) [15] gated recurrent unit (GRU) [16] as well as hybrid neural networks in the form of LSTM and convolutional neural network (CNN) [17] were considered. Speitkamp and Bichler [18] touched upon the issues of server consolidation. Moreover, the emphasis is on modeling the optimal distribution of source servers between physical target servers, taking into account real restrictions.

Averyanikhin *et al.* [19] consider the problems that arise when building cloud data centers based on virtual desktop infrastructure (VDI). In their study, the authors focused on assessing the optimal number of required virtualization cluster nodes. In this case, the efficiency criterion is taken as the basic criterion. The works [20], [21] touch upon such an important aspect of building an organization's cloud infrastructure as optimizing the placement of virtual machines on physical machines in cloud data centers. The authors explore the issues of reducing energy consumption and optimal use of computing resources, which together contribute to increasing the efficiency of cloud infrastructure.

The works [22], [23] consider the problem of selecting hardware for creating client workstations. The authors propose a model for selecting server platforms based on the amount of available and required random access memory (RAM) to accommodate a given number of VMs. And to solve the problem, a genetic algorithm is proposed.

However, we note that most authors pay little attention to this aspect when assessing the optimal number of required virtualization cluster nodes as an efficiency criterion. In particular, in conjunction with the choice of investment strategy in a particular CBDEE architecture of the university. This led to interest in conducting additional research in this direction.

The purpose of the study is to develop a model for estimating the required number of nodes in a private cloud virtualization cluster used in a cloud-based digital educational environment of universities. To achieve this goal, it is necessary to solve the following problems: i) To determine the optimal amount of RAM for virtual desktops in the university cloud, based on the specifics of educational and scientific work in the CBDEE of the university; and ii) To improve the model for assessing the required number of nodes of the university's CBDEE virtualization cluster, taking the efficiency of using virtual machines as an optimality criterion.

2. METHOD

For estimation of required number of nodes of the university's CBDEE virtualization cluster was used a model. For its development were used analytical methods. A genetic algorithm was used in solving the problem of selecting server equipment for forming the infrastructure of virtual users of a private university cloud.

2.1. Genetic algorithm for determining the amount of RAM of virtual desktops in a university cloud

Before assessing the needs for the nodes of the university cloud-based learning environment (CBLE) virtualization cluster, we will determine the requirements for the amount of RAM of virtual desktops in the university cloud. When planning a certain number of computing servers and VMs for a CBDEE university, the final cost of the solution is usually taken into account. In this case, this problem can be expressed as a cost minimization function with given constraints.

VM reserve on the CBDEE server is a static parameter. The VM request is a dynamically changing parameter. And it is the request that determines the real need for VM resources on the university's private cloud server. For example, VM types are used. The type of VM is determined by the size of the hard drive, the amount of RAM, and the number of processors and processor cores. VM types are designated as $T_1, \ldots, T_w$. Each user, as needed, *i.e.* upon request, will be allocated x_i resources of VM instances assigned to the type T_i.

One of the most important characteristics of the server platform (SRP) on which VMs are deployed is random access memory (hereinafter referred to as RAM). The simplest option for increasing the number of VMs can be implemented by adding RAM modules to the CBDEE server or servers. But the server has a limit on the number of available RAM slots. Therefore, it is better to immediately have an idea of the nature and needs of those tasks that will be implemented by students and teachers using VM.

In accordance with work [22] the objective function that allows solving the problem of optimizing the server infrastructure is presented as (1).

$$S = \sum_{l=1}^{q} \sum_{i=1}^{h} \left(co_i + \sum_{j=1}^{u} n_{ij} \cdot co_{zj} \right) \cdot cir_{il}, \quad (1)$$

where co_i is cost of the i^{th} server platform (hereinafter referred to as SRP); co_{zj} is cost of the j^{th} additional module for server RAM; u are number of memory types for server RAM (for example, by frequency, and generation number); h are number of university private cloud SRP; q are number of options for filling server RAM blocks; cir_{il} are number of RAM modules of j-type for x SRP; n_{ij} are number of SRPs in the private cloud. Here $q = \sum_{i=1}^{h} q_i = \sum_{i=1}^{h} \frac{(u+d_i)!}{d_i! u!}$, where d_i are number of slots for RAM of the i^{th} SRPs. Further optimization can be performed for the variables c_{zj} and cir_{il}, which are contained in expression (1).

In accordance with studies [22], [23], we write the limit on the amount of RAM for a specific SRP as (2) and (3).

$$\sum_{j=1}^{u} sram_j \cdot n_{ij} \le h_i, \quad i = 1 \ldots cir. \quad (2)$$

Limit on the number of added RAM modules:

$$\sum_{j=1}^{u} n_{ij} \le d_i, \quad i = 1 \ldots cir. \quad (3)$$

Limitation on the sufficiency of RAM for SRP, ensuring the functioning of the required number of VMs [23].

$$\sum_{i=1}^{h} \left(\frac{\sum_{j=1}^{u} sram_j \cdot n_{ij}}{V_{VM}} \right) \ge N_M, \quad (4)$$

where V_{VM} are amount of RAM allocated for one VM; N_M are required number of VMs. A limit that describes the integer nature of the current problem.

$$cir_{il}, n_{ij} \ge 0, \quad cir_{il}, n_{ij} - integer \quad (5)$$

In contrast to works [22], [23] we will consider the parameter $co_i -$ cost of the i^{th} SRP platform as a kind of analogue of investing in the server infrastructure of a private university cloud (or CBDEE in general). Such infrastructure includes the server/servers itself and the corresponding number of VMs that are deployed on SRP. Note that the choice of a rational investment strategy and the size of appropriate investments can be made based on the application of game theory [24]. So, in accordance with [24], in order to determine the cost (investment volumes in the server platform/CBDEE platforms) co_i in contrast to [22], [23] in the decision process, the strategies of stakeholders are first determined based on game theory. Such parties can, for example, be considered: i) administrators of the server infrastructure or administrators of the distance learning system (DLS), as shown in Figure 1. Administrators, due to their official tasks, strive to maximize the capabilities of the university cloud; and ii) financial administration of the university. Which, due to its tasks of cost optimization, seeks to minimize the cost of maintaining the cloud infrastructure of the university and IT in general.

The dynamics of changes in the resources of the parties are described in the form of a system of differential equations, given in detail in [24], [25]. We implemented the solution to the optimization problem described by expressions (1)-(5) using a genetic algorithm (GA). In the GA used, the population is the set of decisions during the selection of the SRP of a private university cloud. The key was considered to be the need for additional RAM when increasing the number of virtual desktops. Unlike the classic GA, which uses binary coding, this study used a coding list.

The list element contains the following information regarding the SRP CBDEE of the university, according to [22]: i) type of SRP; ii) a set of modules for RAM in accordance with the chosen investment strategy for CBDEE virtualization and existing restrictions, for example, financial [24]; iii) general RAM indicators (frequency and memory size); and iv) cost of a set of RAM for SRP. The number of genes in a chromosome (ch) was taken equal to the number of elements in the list of options for sets of RAM for scalable SRP. Expression (1) was used as a fitness function. Possible combinations of RAM sets of individual SRPs constitute a population (pop). At the same time, restrictions have been adopted on the number of minimum required parameters of the server (servers) in terms of memory volume and the total cost of RAM modules.

The initial population was created as follows. An entry number was randomly selected from the list of RAM sets for the server/servers (SRP) of the private university cloud. For example, you can use the roulette method. Next, add “1” to the gene that corresponds to this set.

Then we check that the chromosome (ch) meets constraints (3) and (4). We repeat the procedure until the required indicators for the necessary characteristics of the CR are achieved. We enter the numbers (NG) of chromosome generations (ch) into the data structure specially generated by the program. The size of the population depends on the number of chromosomes. For each chromosome (ch) in the population, a fitness estimate is performed. This procedure is performed by calculating the fitness function.

The lower the value of the fitness function, the higher the quality of chromosomes. The chromosome splitting point was chosen randomly. Selection was performed for each generation “Viable” individuals were selected based on constraint (4). Then ranking is performed according to the value of the fitness function (1).

The best individuals are transferred unchanged to the next generations. The calculations end when the specified number of generations is reached. As computational experiments have shown, the convergence of the algorithm is achieved for at least fifty generations.

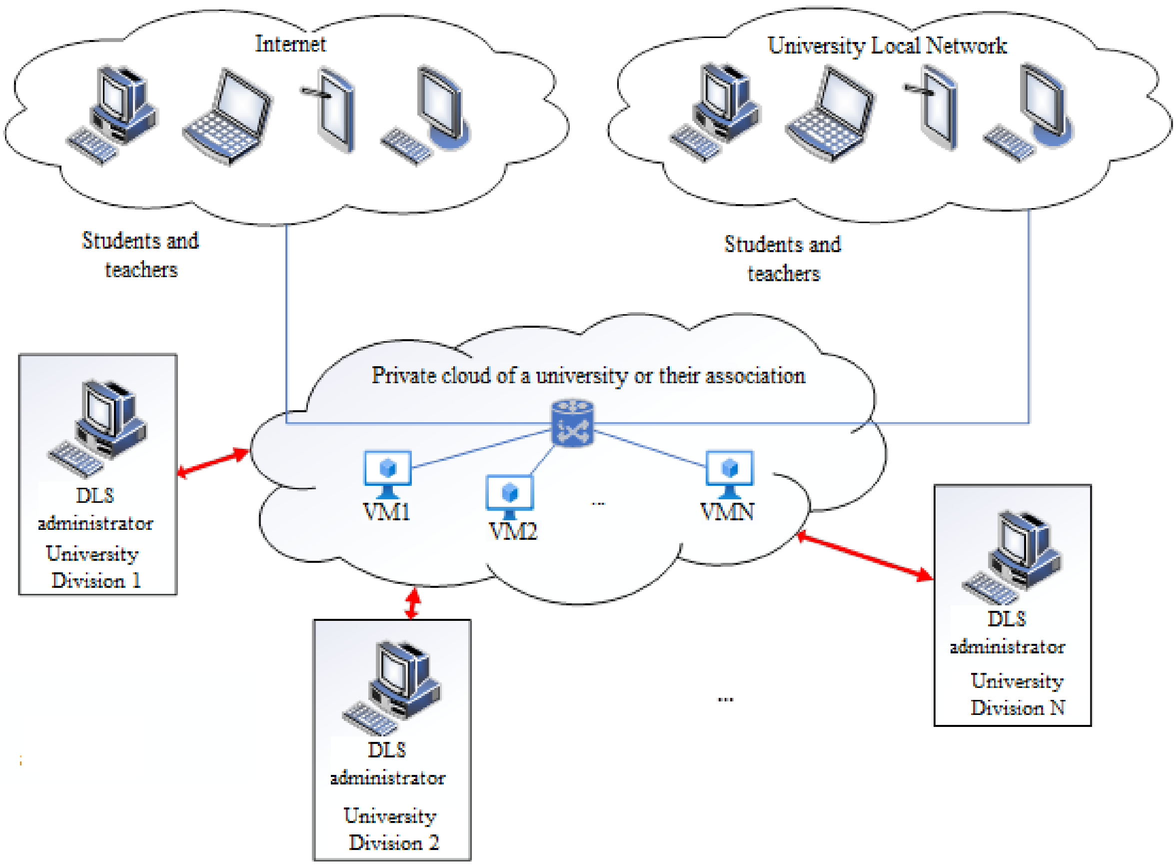

Figure 1. Typical architecture of a university cloud-based distance learning system (DLS)

2.2. Genetic algorithm for determining the amount of RAM of virtual desktops in a university cloud

To estimate the necessary resources for constructing a university private cloud, used, for example, to organize a distance learning system in Figure 1, we will create a mathematical model. The model set out in [18] is taken as a basis. However, let’s supplement this model with calculation data on the optimal number of cluster nodes. The criteria for optimality are the efficiency criterion and the chosen strategy for investing in the development of the university cloud, which is mentioned in [24], [25]. At the same time, based on the specifics of the university CBDEE, we will take into account that it is difficult to predict in advance the moment of deployment and destruction of the VM. Indeed, today the asynchronous organization of students’ completion of educational tasks is more relevant in the learning process. When each of them can, in the time allotted for performing work, independently decide when it is more convenient for him to deploy the VM and then destroy it, in Figure 1. An example of such system behavior is in Table 1, for 10 VMs of a private university cloud used in the DLS.

Table 1. Example of the VM state in DLS university private cloud

	VM state		
	Deployed VMs	Stoppable VMs	Shutdown (destroyed) VMs
VM number	2, 5, 9, 10	1, 3, 4	6, 7, 8

Moreover, all these actions occur at different times-t. As shown above, the hardware-software complex (HSC) of a private university cloud includes N computing servers (server platforms - SRP with RAM parameters previously calculated on the basis of a GA, paragraph 2.1). Each SRP can be described by a parameter Pa. This setting enables the basic server specifications (SRP). For example, such parameters include the amount of RAM, processor clock speed, and number of processor cores. In addition, each of the servers $j \in \{1...N\}$ has the capacity C_{jk} for the corresponding resource $k \in \{1...Pa\}$.

Each virtual machine (VM–vm) i, that is used in the educational process, scientific or administrative task of a CBDEE university requires a certain minimum amount of resources – r_{ik}. Moreover, the resource is allocated from the parameters k. For example, $k = 1$ a processor parameter, and $k = 2$ a RAM parameter. This is the so-called resource reserve. If the server cannot allocate the minimum required amount of resources according to the parameters k, then the VM will not be deployed. At a point in time t active VMs $Ac(t)$ can be conditionally divided into already placed VMs – $Pl(t)$, and those that need to be placed – $Rd(t)$. There are also switched off VMs – $Re(t())$. It can be written as (6).

$$Ac(t) = \left\{ {}^{Pl(t)}\!\big/_{Re(t)} \right\} \cup Rd(t) \tag{6}$$

VM placement in a private university cloud can be described by two variables. The first characterizes the number of virtual processors that are assigned to VMs of a certain type T_i. The second variable will describe the connection between the server and the VM. Then the variable $vm_{ij}(t) = 1$ if at the moment of time t the VM i, assigned to the type T_i is on the server j. We will assume that the CBDEE of the conditional university has a sufficient amount of virtualization complex. Then the equality is true.

$$\sum_{j=1}^{N} vm_{ij}(t) = 1,\ \forall i \in Ac(t). \tag{7}$$

Each type of VM resource must be guaranteed to be able to be hosted on the server. The total value of reserved VM resources should not exceed the server capacity for a specific resource. For example, the capacity of hard drives or RAM of all types of VMs cannot exceed the capacity of the server C_{jk}. This can be written as (8).

$$\sum_{j=1}^{N} vm_{ij}(t) = 1,\ \forall i \in Ac(t). \tag{8}$$

Next, you can move on to assessing the required resources of the entire private university cloud. When designing such a private university cloud, specialists will be able to predict the future load of the aircraft. As an initial data, you can take the number of cloud users. This is not difficult to do if you know the number of students, teachers and statistics on free students of a particular course. The latter is relevant when, for example, we are talking about disciplines of free choice of students. And, accordingly, an additional load may arise on the course with its assignments and VMs. In the simplest case, you can use static methods for predictive assessment. We believe that when designing such a private university cloud, the designer needs to place M of the VM on N servers (SRP). The total amount of resources of all VMs should not exceed the total amount of server resources.

Note that the problem of optimal resource allocation can be interpreted as a type of packaging problem in relation to virtualization technologies [26]. This problem can also be associated with the "container packing problem" or the "knapsack problem," *i.e.* NP-hard [27]. Algorithms for solving such problems are analyzed in more detail in [19]. In our study, we use the recommendations of [27] to determine the lower limit of the minimum number of servers in a private university cloud.

Then, to dynamically assess the capacity of the virtualization complex, we will add a time parameter. We believe that the computing servers in the university private cloud have the same characteristics. This is quite natural, since equipment of this class is purchased centrally. That is (9):

$$\mathrm{C}_{jk} = \mathrm{C}_k, \forall j \in 1..N, k = 1..Pa. \tag{9}$$

And a rational investment strategy that determines the choice of one option or another can be implemented, for example, using the calculations of [24], [25]. In which, in particular, as an example, they touch upon such an aspect of the search for a cost-effective IT infrastructure scheme for a university as connecting and ensuring secure access of the corporate network of an informatization object (IO) to an external network. In these works, the authors examined in detail various situations when the financial resources of investors (players) are aimed at solving the problem of providing secure access of the IO corporate network to an external network in the presence of alternative connection schemes. The results presented in [24], [25] allow, when designing a private university cloud, to effectively find rational investment options during the analysis of alternatives. To find a solution in [24], [25] the mathematical apparatus of a bilinear multi-stage quality game with several terminal surfaces with alternating moves of players is used.

Based on the above, we assume that the consumption of VM resources in the university's CBDEE may change over time. And indeed, for quality training, it is necessary to periodically update the software, which is becoming more and more resource-demanding. At the moment of time t we divide the set of VMs planned for deployment into subsets, which we denote as $a1(t), a2(t), a3(t)$. It is understood that $a1(t), a2(t), a3(t) -$ is, respectively, sets of VMs with high, medium and low load. Then, for example, for active VMs the following system of equations will be valid:

$$\begin{aligned} a1(t) &= \{i \in Ac(t) | a_i(t) > C_{CPU} - \mu\}; \\ a2(t) &= \{i \in Ac(t) | C_{CPU} - \mu \geq a_i(t) > 0.5 \cdot C_{CPU}\}; \\ a3(t) &= \{i \in Ac(t) | 0.5 \cdot C_{CPU} \geq a_i(t) > \mu\}; \end{aligned} \tag{10}$$

where C_{CPU} are total amount of private university cloud server resources by processor parameters, μ are parameter characterizing the maximum load of the processor (CPU) and SRP RAM– $0 \leq \mu \leq 0.5$. The latter is true based on the consideration that during virtualization it is not recommended to allocate more than 50% of the server's processor and RAM resources to VMs. The lower estimate of the optimal number of servers for a private university cloud can be expressed as (11) and the average lower bound for the optimal number of servers in a private university cloud looks as (12).

$$N_1(\mu, t) = |a_1(t)| + |a_2(t)| + max\left(0, \frac{\left\lceil \sum_{i \in a_3(t)} a_i(t) - \left(C_k \cdot |a_2(t)| - \sum_{i \in a_2(t)} a_i(t)\right) \right\rceil}{C_k}\right) \tag{11}$$

$$N(t) = max\{N_1(\mu, t), N_2(\mu, t), 0 \leq \mu \leq 0.5\} \tag{12}$$

To test the adequacy of the proposed model, a virtual computational experiment was carried out. The experiment was carried out on the basis of private cloud university structures - the National University of Life and Environmental Sciences of Ukraine (Kiev, Ukraine) and the Non-Profit Joint Stock Company "Kazakh National Pedagogical University named after Abai" (Almaty, Kazakhstan). A model for optimizing the number of nodes in a virtualization cluster of university clouds and specified educational institutions was tested, as shown in Figure 2. The CBDEE efficiency criterion was adopted as a criterion for optimizing the number of nodes in a university cloud virtualization cluster.

The model described by expressions (1) - (12) was implemented in the algorithmic language Python. As a result, the needs for the server infrastructure of private clouds at these universities were determined, based on the characteristics of specialist training and the planned teaching load. Next, taking the efficiency of organizing COLE of universities as the main criterion for optimality, the costs of creating a private cloud were compared with alternative options for organizing CBDEE of universities. To economically calculate costs, analytical data on prices for services of IT corporations were analyzed, in which the main characteristics were the number of processor cores, the amount of RAM, the amount of data storage, and the type of operating system (OS) for the IaaS topology [1], [2]. The simulation results are presented in Figure 3. Figure 3 shows the data obtained from comparing the dollar cost modeling results of a private university cloud with an infrastructure as a service (IaaS) model. The cost characteristics of a private university cloud, as shown above, are largely influenced by the optimization of the parameters of quantitative and qualitative (CPU and RAM) characteristics of CBDEE servers.

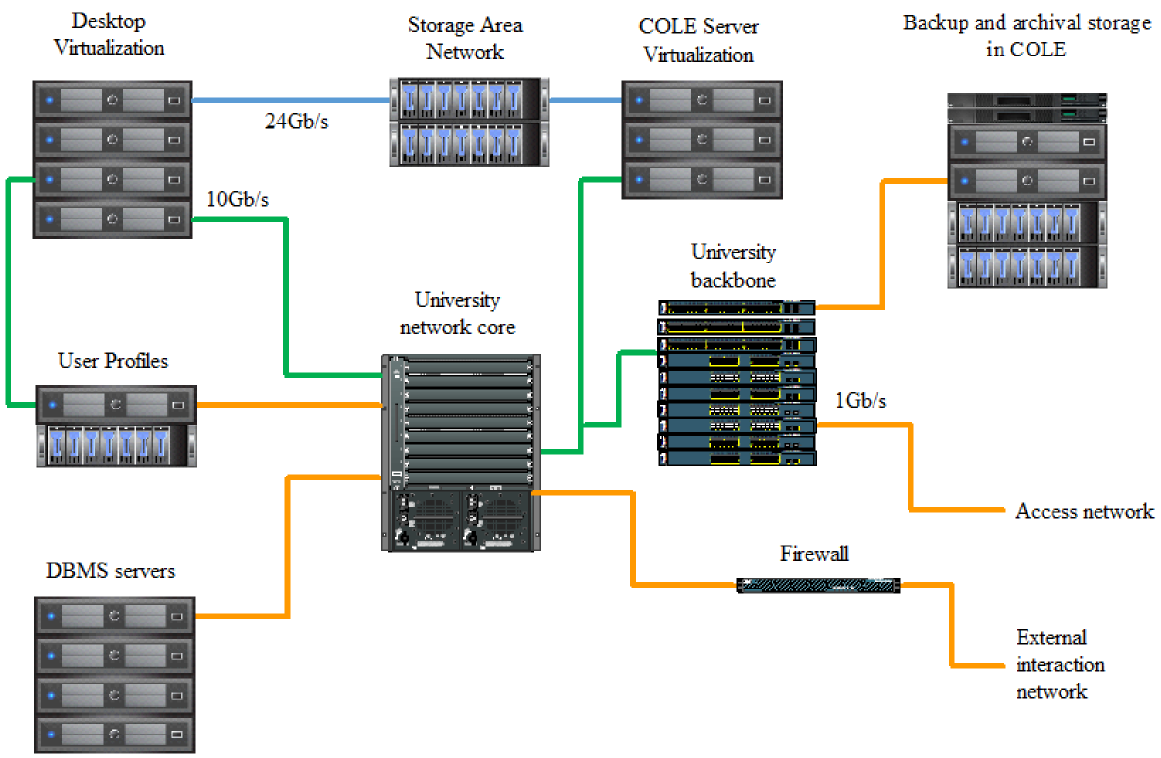

Figure 2. Block diagram of private university clouds

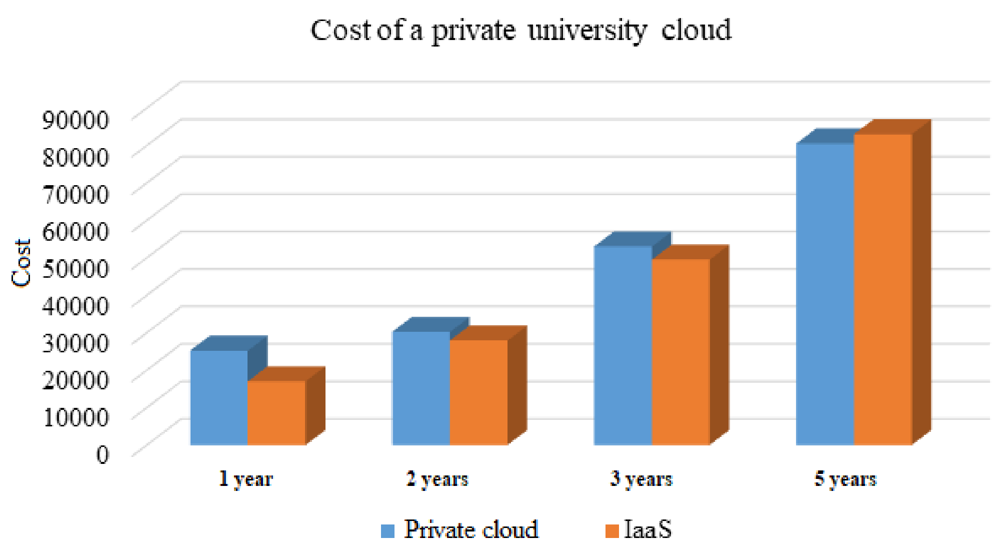

Figure 3. Results of computational experiments

3. RESULTS AND DISCUSSION

In the genetic algorithm used, the population is the set of decisions during the selection of the SRP of a private university cloud. The key was considered to be the need for additional RAM when increasing the number of virtual desktops. Unlike the classic GA, which uses binary coding, this study used a coding list. GA stages of work include: i) Data insertion: the initial population of possible SRP configurations is being inserted. SRP configurations contain data about platform type, RAM modules, RAM modules' characteristics, and cost. Also, parameters and limits (for example, budget) are being set; ii) Population initialization: initialization of SRP configurations' starting population. There are possible solutions for VM optimal placement; iii) Suitability assessment: each solution was assessed based on the suitability function.

Suitability functions take into account technical characteristics (performance, RAM capacity) and economical targets (cost); iv) Selection: selection of the most suitable solutions from the current population, using methods of tournament selection, roulette wheel selection, and others; v) Crossover: creation of new SRP configurations combining selected solutions. It allows exploring the search space, combining successful elements from different solutions; vi) Mutation: in the new configurations were added random modifications. It helped to increase search space exploration completeness and prevent premature convergence to a local optimum; and vii) Substitution: new configurations generation substitutes previous. This process repeats until a given maximum number of iterations is reached or an optimal solution is found.

The best solution from the last generation is an optimal configuration of SRP for the university cloud, considering all criteria and limits. Analysis of results shows that this model allows to significantly reduce expenses on IT infrastructure by optimizing usage of hardware resources and simplifying administration. We took analytical data on the cost of the IBM SoftLayer cloud as a basis. If we take into account the cost of maintaining the cloud by an administrator and software engineering staff, which is about $12,600, then the total annual cost is more than $16,500. With a total calculation of the costs of the corresponding infrastructure (server 2 units, data storage 2 units, network equipment - total cost about $10,000) for placement on own premises, their maintenance (administrator's salary, engineering staff - about $13,000) and utility costs (electricity, internet channel about $2,000), we get an amount of around $25,000. Carrying out similar calculations with a prolongation of 3–5 years, we can evaluate the most attractive model in terms of price for hosting an "academic cloud" as shown in Figure 3. In the PaaS model, all resources (software, allocation of structural elements of the system) are dictated by the vendor, that is, we are limited to the choice of what the IT company offers us. Therefore, we are not considering this model yet.

During the virtual computing experiment, the following was established. When conducting testing, dynamic changes in the indicators of resources consumed in the university cloud should be taken into account. This is due to the fact that users, depending on the specifics of educational or scientific work, can independently start and/or stop VMs in the common CBDEE resource pool. In this regard, in order to determine the required amount of resources, it is necessary to decide in what range the number of VMs running in the university cloud will vary. And also, how the VM's need for CBDEE resources will change.

During the virtual computing experiment, it became obvious that the introduction of cloud technologies and applications into the educational process and IT infrastructure of universities can significantly reduce the cost of its maintenance. The positive effect is primarily achieved by optimizing the use of hardware resources and simplifying the administration of the university IT infrastructure. The university cloud, consisting of virtual desktops (DaaS) of students in this case, will not be an exception, since it also allows to centralize the university's IT infrastructure, as well as create inter-university cloud structures that help to improve the quality of education. During the transition of the organization of the educational process from the use of conventional PCs to virtual desktops, there is a need to create an infrastructure of virtual desktops corresponding to the university profile (VDI).

The success of GA implementation was assessed by a virtual computing experiment on two university private clouds in Kazakhstan and Ukraine. The advantage of the GA is expense optimization. This model allows to significantly reduce expenses on IT infrastructure by optimizing usage of hardware resources and simplifying administration. A disadvantage of this model is the necessity of fine-tuning parameters like population size, crossover, and mutation probability. The results of this model highly depend on the initial population. Algorithm's efficiency on CBDEEs can be improved by fine-tuning initial configurations. Analysis shows that improvement of VM placement optimization models and cluster management algorithms for cost reduction and efficiency improvement can be a topic for future research.

4. CONCLUSION

It is shown that when designing a VDI university or interuniversity cloud, it is necessary to solve many complex technical problems. One of these tasks is the task of estimating the required number of virtualization cluster nodes on which user virtual machines (VMs) will be hosted. Another task that arises in the VDI design process is the problem of algorithmizing the placement of VMs in a computer network (CN). In this case, optimal placement of VMs will reduce the number of computer nodes.

The model is based on a combined approach. This approach provides a synergistic effect when jointly solving the problem of optimal packaging, a genetic algorithm in the course of solving the optimization problem of determining the amount of virtual desktops RAM in the university cloud. A game approach was also used to select a rational strategy for investing in CBDEE of universities. The model proposed in the work is universal and can be used in the design of cloud university platforms for various purposes, from the educational process to interuniversity scientific laboratories.

ACKNOWLEDGEMENTS

The research was carried out within the framework of the project of the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan No. AP19678846, improving the effectiveness of hybrid and distance learning forms based on the development of university infrastructure in the context of digital transformation.

REFERENCES

[1] E. N. G'ayratovich, "The theory of the use of cloud technologies in the implementation of hierarchical preparation of engineers," *Eurasian Research Bulletin*, vol. 7, pp. 18–21, 2022.

[2] M. Rasuleva and N. Alieva, "International conference on higher education teaching," in *International Conference on Higher Education Teaching*, 2023, pp. 4–11.

[3] M. A. Almaiah and A. Al-Khasawneh, "Investigating the main determinants of mobile cloud computing adoption in university campus," *Education and Information Technologies*, vol. 25, no. 4, pp. 3087–3107, Jul. 2020, doi: 10.1007/s10639-020-10120-8.

[4] T. Dong, Y. Ma, and L. Liu, "The application of cloud computing in universities' education information resources management," in *Information Engineering and Applications*, Springer London, 2012, pp. 938–945, doi: 10.1007/978-1-4471-2386-6_122.

[5] N. Wagh, V. Pawar, and K. Kharat, "Educational cloud framework – a literature review on finding better private cloud framework for educational hub," in *Microservices in Big Data Analytics*, Singapore: Springer Singapore, 2020, pp. 13–27, doi: 10.1007/978-981-15-0128-9_2.

[6] M. Khayyat, I. A. Elgendy, A. Muthanna, A. S. Alshahrani, S. Alharbi, and A. Koucheryavy, "Advanced deep learning-based computational offloading for multilevel vehicular edge-cloud computing networks," *IEEE Access*, vol. 8, pp. 137052–137062, 2020, doi: 10.1109/ACCESS.2020.3011705.

[7] E. Yadegaridehkordi, L. Shuib, M. Nilashi, and S. Asadi, "Decision to adopt online collaborative learning tools in higher education: a case of top Malaysian universities," *Education and Information Technologies*, vol. 24, no. 1, pp. 79–102, Jan. 2019, doi: 10.1007/s10639-018-9761-z.

[8] E. Alavi and M. C. Mohan, "An e-learning system architecture based on new business paradigm using cloud computing," *International Journal of Engineering Sciences & Research Technology*, vol. 2, no. 10, pp. 2990–2999, 2013.

[9] R. Buyya, C. S. Yeo, and S. Venugopal, "Market-oriented cloud computing: vision, hype, and reality for delivering it services as computing utilities," in *2008 10th IEEE International Conference on High Performance Computing and Communications*, Sep. 2008, pp. 5–13, doi: 10.1109/HPCC.2008.172.

[10] V. Chang, D. Bacigalupo, G. Wills, and D. De Roure, "A categorisation of cloud computing business models," in *2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing*, 2010, pp. 509–512, doi: 10.1109/CCGRID.2010.132.

[11] Q. Zhang, L. Cheng, and R. Boutaba, "Cloud computing: state-of-the-art and research challenges," *Journal of Internet Services and Applications*, vol. 1, no. 1, pp. 7–18, May 2010, doi: 10.1007/s13174-010-0007-6.

[12] H. Adamu, B. Mohammed, A. B. Maina, A. Cullen, H. Ugail, and I. Awan, "An approach to failure prediction in a cloud based environment," in *2017 IEEE 5th International Conference on Future Internet of Things and Cloud (FiCloud)*, Aug. 2017, pp. 191–197, doi: 10.1109/FiCloud.2017.56.

[13] V. R. Messias, J. C. Estrella, R. Ehlers, M. J. Santana, R. C. Santana, and S. Reiff-Marganiec, "Combining time series prediction models using genetic algorithm to autoscaling web applications hosted in the cloud infrastructure," *Neural Computing and Applications*, vol. 27, no. 8, pp. 2383–2406, Nov. 2016, doi: 10.1007/s00521-015-2133-3.

[14] S.-Y. Hsieh, C.-S. Liu, R. Buyya, and A. Y. Zomaya, "Utilization-prediction-aware virtual machine consolidation approach for energy-efficient cloud data centers," *Journal of Parallel and Distributed Computing*, vol. 139, pp. 99–109, May 2020, doi: 10.1016/j.jpdc.2019.12.014.

[15] D. Janardhanan and E. Barrett, "CPU workload forecasting of machines in data centers using LSTM recurrent neural networks and ARIMA models," in *2017 12th International Conference for Internet Technology and Secured Transactions (ICITST)*, Dec. 2017, pp. 55–60, doi: 10.23919/ICITST.2017.8356346.

[16] S. Gupta and D. A. Dinesh, "Resource usage prediction of cloud workloads using deep bidirectional long short term memory networks," in *2017 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*, Dec. 2017, pp. 1–6, doi: 10.1109/ANTS.2017.8384098.

[17] S. Ouhame, Y. Hadi, and A. Ullah, "An efficient forecasting approach for resource utilization in cloud data center using CNN-LSTM model," *Neural Computing and Applications*, vol. 33, no. 16, pp. 10043–10055, Aug. 2021, doi: 10.1007/s00521-021-05770-9.

[18] B. Speitkamp and M. Bichler, "A mathematical programming approach for server consolidation problems in virtualized data centers," *IEEE Transactions on Services Computing*, vol. 3, no. 4, pp. 266–278, Oct. 2010, doi: 10.1109/TSC.2010.25.

[19] A. E. Averyanikhin, A. V. Kotelnitsky, and K. A. Muraviev, "Method of calculation of optimum number of knots of the cluster of virtualization of the private cloud of virtual desktops by criterion of efficiency," *International Research Journal*, vol. 5, no. 47, pp. 6–13, 2016, doi: https://doi.org/10.18454/IRJ.2016.47.187.

[20] S. Gharehpasha, M. Masdari, and A. Jafarian, "Virtual machine placement in cloud data centers using a hybrid multi-verse optimization algorithm," *Artificial Intelligence Review*, vol. 54, no. 3, pp. 2221–2257, Mar. 2021, doi: 10.1007/s10462-020-09903-9.

[21] S. Gharehpasha, M. Masdari, and A. Jafarian, "Power efficient virtual machine placement in cloud data centers with a discrete and chaotic hybrid optimization algorithm," *Cluster Computing*, vol. 24, no. 2, pp. 1293–1315, Jun. 2021, doi: 10.1007/s10586-020-03187-y.

[22] D. K. Proskurin and K. A. Makovij, "The task of selecting server resources for the implementation of virtual desktop infrastructure," *Bulletin of the Voronezh State Technical University*, vol. 13, no. 4, pp. 26–32, 2017.

[23] D. K. Proskurin and K. A. Makovij, "Modified genetic algorithm for solving the problem of selecting server resources in building the infrastructure of virtual desktops," *Bulletin of the Voronezh State Technical University*, vol. 17, no. 3, pp. 46–51, Jul. 2021, doi: 10.36622/VSTU.2021.17.3.006.

[24] B. Akhmetov *et al.*, "A model for managing the procedure of continuous mutual financial investment in cybersecurity for the case with fuzzy information," in *Lecture Notes on Data Engineering and Communications Technologies*, Springer, Singapore, 2022, pp. 539–553, doi: 10.1007/978-981-16-6605-6_40.

[25] V. Lakhno, V. Malyukov, N. Mazur, L. Kuzmenko, B. Akhmetov, and V. Hrebeniuk, "Development of a model for decision support systems to control the process of investing in information technologies," *Eastern-European Journal of Enterprise Technologies*, vol. 1, no. 3, pp. 74–81, Feb. 2020, doi: 10.15587/1729-4061.2020.194531.
[26] D. S. Hochba, Ed., "Approximation algorithms for NP-hard problems," *ACM SIGACT News*, vol. 28, no. 2, pp. 40–52, Jun. 1997, doi: 10.1145/261342.571216.
[27] S. Martello and P. Toth, *Knapsack problems: algorithms and computer implementations*. USA: John Wiley & Sons, Inc., 1990.

BIOGRAPHIES OF AUTHORS

Bakhytzhan Akhmetov graduated from Moscow Higher Technical School named after N.E. Bauman in 1977 with a degree in 0608, electronic computer. In 1984, he also defended his Ph.D. thesis in the specialty 05.13.12, computer-aided design system. In 2007 he defended his doctoral dissertation at KazNTU named after K. I. Satpaev, specialty 05.13.10, management in social and economic systems. Professor on the specialty 05.13.00, informatics, computer engineering and management. Academician of the National Engineering Academy of the Republic of Kazakhstan. Passed all stages of a university career: from teacher to rector, scientific and pedagogical experience - 45 years. His scientific interests include information and communication technologies, artificial intelligence, and cybersecurity. He published more than 400 articles, including 40 articles indexed in Scopus. Hirsch index in the Scopus database-7. Prepared 2 candidates of science, 15 PhDs. Chairman of the National Scientific Council of the Republic of Kazakhstan on the priority information, communication and space technologies. Member of the editorial board of the Journals Information Protection, National Aviation University of Ukraine, Kyiv; Ural Radio Engineering Journal, Ural Federal University named after the first President of Russia B.N. Yeltsin, Yekaterinburg. He can contacted at email: bakhytzhan.akhmetov.54@mail.ru.

Valery Lakhno is a professor at the Department of Computer Systems, Networks and Cybersecurity at the National University of Life and Environmental Sciences of Ukraine, Ukraine. He holds a PhD with a specialization in cybersecurity. His areas of research include mathematical modeling, security of computer systems and networks, and intrusion recognition. He has filed a number of patents and industrial designs for his innovative ideas in the field of cybersecurity. He can be contacted at email: lva964@nubip.edu.ua.

Nurzhamal Oshanova is a Candidate of Pedagogical Sciences, associate professor at Abai Kazakh National Pedagogical University. Her scientific research focuses on the methodology of teaching computer science, digitalization of education, and AI in education university. She can be contacted at email: n.oshanova@abaiuniversity.edu.kz.

Zhuldyz Alimseitova is an associate professor at the Department of Cybersecurity, Information Processing and Storage at Satpayev University of Kazakhstan. She holds a PhD in computer engineering and software engineering. Her research areas are information and cybersecurity, biometrics, pattern recognition, cryptography, and fiber-optic sensors. She can be contacted at email: zhuldyz_al@mail.ru.

Madina Bereke doctoral student in computer science at the Kazakh National Pedagogical University named after Abai. Graduated from Al-Farabi Kazakh National University with a degree in computer engineering and software. Master of Technical Sciences. Her research areas are computer science and information systems, cloud technologies. She can contacted at email: mdbereke@gmail.com.

Nurgul Izbasova graduated from Al-Farabi Kazakh National University with a degree in information systems and completed her doctoral studies at Yasavi International Kazakh-Turkish University with a degree in management. She is the author of 10 articles, including 3 articles from the Scopus database. She can contacted at email: nurgul.izbassova@gmail.com.