

# Development of the Model of Dynamic Storage Distribution in Data Processing Centers

**Rashid G. Alakbarov**

Institute of Information Technology of ANAS, Baku, Azerbaijan  
Email: rashid@iit.ab.az

**Fahrad H. Pashaev, Mammad A. Hashimov**

Institute of Control Systems after Academician A. Huseynov of ANAS, Baku, Azerbaijan  
Institute of Information Technology of ANAS, Baku, Azerbaijan  
Email: pasha.farhad@gmail.com, m.hashimov@iit.ab.az

**Abstract**— The paper reviews dynamic distribution of storage resources among the users in data processing centers. The process of changing memory usage state was revealed to be the process of Markov. The paper proposes the development of stochastic model of the memory and computing usage distribution and the development of probability density functions over practical data. Parameters of probability density functions were defined with the help of stochastic model and practical data. The calculation of the developed model and the parameters of the probability density function is realized dynamically during the ongoing process. At the beginning of each time interval, it is forecasted that the process will be shifted to which state with which maximum probability. The adequacy of the previous forecasts is monitored. Note that, over the time, the quality of the forecast and the level of adequacy increases. The model is used in the virtualization of storage resources usage process and ensures the use of storage resources without wasting. Structure of visualization base is given. The base enables to monitor all stages of the process. Using monitoring base the issues can be resolved to analyze different aspects of the process. Recommendations are given on the use of obtained results.

**Index Terms**— Data Processing Center, Cloud Computing, Storage Capacity, Markov Process, Stochastic Model, Virtual Resource.

## I. INTRODUCTION

Nowadays, computing and storage resources of personal computers are not sufficient for the solution of complex problems requiring big computing and storage resources such as real time modeling of physical and chemical processes, nuclear reactions, global atmospheric processes, economic development in various fields of science, as well as Cryptography, Geology, development of new drugs. Supercomputers with high performance computing and big storage are widely used in the above-mentioned issues [1]. As a strategic product, the high price of supercomputers reduces its availability for many countries to be used in scientific and technical research. However, these countries have demand for big computing resources. On the other hand, computing and storage resources of the data processing centers connected to the computer networks are not used

effectively. Researches show, that only 60-70% of computing and storage resources of computers manufactured by giant companies (Intel, IBM, Google, etc.) are used effectively [2]. In this case, remaining unused computing and storage resources of data processing centers can be used to solve complex problems. Applying remote access to the data processing centers in daily practice with the help of high speed communication channels opens up new possibilities for the users. Now, the quantitative increase of opportunities of users to get information caused qualitative change in the organizational principles of distributed computing systems in the networks.

At present, research is conducted out for an effective use of computing and storage resources of data processing centers with the help of Cloud Computing. Such systems with big computing and storage resources are based on computer networks, provided with high-speed communication channels [3]. Cloud Computing enables organizations to use computing and storage resources of data processing centers more efficiently. The concept of Cloud Computing provides the development and utilization of infrastructure and software of computer technology in the network. With the help of this technology, the user data is stored and mined on Cloud Computing servers, at the same time, the results are viewed through browsers [4]. Cloud Computing enables to scale and use physical resources (e.g. processor, storage and disk space) through the internet. In this case, the data processing and storage processes are considered as a type of service. Cloud Computing allows the users to access powerful computing and storage resources, and at the same time, the user is not interested where these resources are located and installed [5,6]. The paper is dedicated to the optimal distribution of storage resources among the users with the help of above-mentioned service in the data processing centers. As it is mentioned above, distribution of memory and other computing resources according to the stable static division causes at least 30-40 percent of resources to remain unused. Although, potential users are in search of computational power in elsewhere or have to wait for a long time, in this case potential users could make profit from these unused

computing resources. Consequently, the Data Center can be deprived of serious income. Financial evaluation of the issue is not considered here. However, developed stochastic model is aimed at addressing the hardware and software issues.

The proposed model allows the data processing center to attract more users providing dynamical distribution of available storage and system resources among the users.

The structure of visualization base is given in the paper. The base records the monitoring stages of the use of computing power. Note that the base can be set in any of the Base Management Systems. The software based on the developed model realizes Base Management System functions itself.

The content of the article is organized as follows:

- State of the memory use in data center was set to change as Markov process;
- The characteristics of Markov process of the memory usage change were identified;
- Memory demand model was developed for data center;
- Using practical data a probability density function was proposed and its parameters were found;
- Approximate visualization process is described;
- The structure of visualization base is given;
- Recommendations were given for the use of obtained results.

## II. DEVELOPING THE MODEL OF DYNAMIC STORAGE DISTRIBUTION IN DATA PROCESSING CENTERS

Let's analyze the process of dynamic storage distribution among the users of the systems where Hypervisors are applied. The process of dynamic storage distribution is modeled as Markov process [7,10]. Note that Markov processes are applied to model the problems in telecommunication systems, variety of public service fields and generally in the fields of science and technology [11-13]. Let's suppose that  $M$  is the number of users specified in advance and suggesting storage need. In this case,  $m$ -th user ( $m \in [1, M]$ ) uses  $V_m$  amount of memory. Thus, an amount of memory required by users is as follows:

$$V_1, V_2, \dots, V_M. \quad (1)$$

If  $V_t$  denotes storage volume that is used instantly at any  $t$  time

$$V_{tmin} = 0, V_{tmax} = \sum_{m=1}^M V_m. \quad (2)$$

Nevertheless, in practice,

$V_{tmax} = \sum_{m=1}^M V_m$  is almost impossible. In the peak of storage use it is practically as follows:

$$V_{peak} \leq 0,7 * V_{tmax}. \quad (3)$$

This may lead to attracting additional users.

Each storage user can apply for storage at random moments of time, regardless of its physical identity and functionality. Therefore, the process of memory use is determined by a random  $V_t$  storage volume used instantly

at any time  $t$  and can be considered as Markov process. In this case, state space of the process is defined by the storage of different capacity in use. The process of transition from one state to another does not depend on the transition path. This transition depends on the current state, and it is one of the signs of Markov process. It is sometimes called the process without memory.

The second key feature of Markov process is the finite number of states. In our case, required number of different volumes (states) is finite. Thus,  $M$  number of users can apply for storage in a short time interval (time instant)  $0, 1, 2, \dots$ . Storage capacity defining the process state depends on these combinations. Obviously,  $n$  number of  $M$  users can be chosen from

$$C_M^n = \frac{M!}{n!(M-n)!} \text{ methods. That is, different combinations}$$

$C_M^n$  with  $n$  number of users can be set up from  $M$  number of users. If  $n \in [0; M]$  the number of all possible combinations can be as follows:

$$K = C_M^0 + C_M^1 + \dots + C_M^n + \dots + C_M^M \quad (4)$$

According to Newton binomial [14,15]:

$$\text{As } (1 + 1)^M = C_M^0 + C_M^1 + \dots + C_M^M, \quad (5)$$

$$K = C_M^0 + C_M^1 + \dots + C_M^M = (1 + 1)^M = 2^M. \quad (6)$$

In other words, the number of possible different states are  $K = 2^M$ , which is finite. Thus, Markov process is covered in this state. If  $V_n$  denotes the capacity appropriate to  $n$ -th ( $n \in [1, M]$ ) combination, the space of process states are determined by the following set:

$$\{ V_n \mid n \in [1, K] \}. \quad (7)$$

It should be noted, that if there are the volumes that cover the set  $|V_n - V_k| < \Delta V$  in  $\{ V_n \mid n \in [1, K] \}$  and which are equal or very close to each other, only one of them can be remained to reduce the number of  $K$ . For this purpose, it is possible to implement an algorithm similar to sieve of Eratosthenes [10]. The algorithm can be described in recursive way as follows:

- From  $V_r$  element each element is compared with  $r$ -th element for  $\forall r \in [1, K]$ .
- $V_j = V_r$  for random  $j > r$  the  $j$ -th element is excluded.  $K$  denotes remaining number of elements again. The process continues until  $r = K$  value is achieved.

On the other hand, in order to simplify the analysis of the results we can build the set  $\{ V_n \mid n \in [1, K] \}$  in increasing volumes. In this case,  $V_{m_1} > V_{m_2}$  is possible for random  $m_1 > m_2$ . The system states can be denoted by  $E_1, E_2, \dots, E_K$ .

As the all possible states can be numerated, this process is called random process with discrete state or discrete random process. Discrete random process can be simply shown in graphs. In this time, vertices of the graph show the states. And the edges of the graph show the transition from one state to another (Fig. 1).

Described feature of the process is that this process does not shift from one state to another one at pre-defined time, but at a random time. Such a process is called continuous-time random process.

To turn continuous-time random process into Markov process the period of remaining in any  $E_i$  state of the process should be changed by exponential law. Its distribution function is

$$F_{ij}(\tau) = 1 - e^{-\gamma_{ij}\tau} \quad (8)$$

Here  $\gamma_{ij}$  is a distribution parameter, and it characterizes the process transition frequency of from  $E_i$  state to  $E_j$  state. This parameter is quantitatively equal to the inverse value of the average period of remaining in the  $E_i$  state before shifting the process from  $E_i$  state to  $E_j$  state.

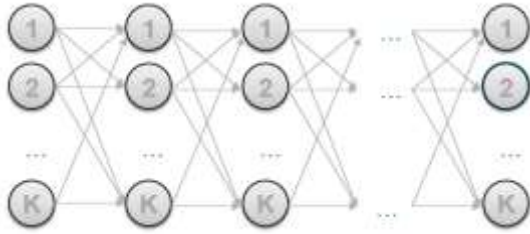


Fig. 1. Graph of process state shift

Let's try to calculate the probability of the followings:

If the process is already in state  $E_i$  for the period of  $\tau_0$ , then with what probability it shifts to  $E_j$  state at  $\Delta\tau$  period:

$$\begin{aligned} P_{ij}(\Delta\tau / \tau \geq \tau_0) &= P_r(\tau_0 \leq \tau \leq \tau_0 + \Delta\tau \mid \tau \geq \tau_0) \\ &= \frac{P_r(\tau_0 \leq \tau \leq \tau_0 + \Delta\tau)}{P_r(\tau \geq \tau_0)} = \frac{F(\tau_0 + \Delta\tau) - F(\tau_0)}{1 - F(\tau_0)} \\ &= \frac{1 - e^{-\gamma_{ij}(\tau_0 + \Delta\tau)} - [1 - e^{-\gamma_{ij}(\tau_0)}]}{1 - 1 + e^{-\gamma_{ij}\tau_0}} \\ &= \frac{e^{-\gamma_{ij}(\tau_0)} * e^{-\gamma_{ij}(\Delta\tau)} + e^{-\gamma_{ij}(\tau_0)}}{e^{-\gamma_{ij}(\tau_0)}} \\ &= \frac{e^{-\gamma_{ij}(\tau_0)}(1 - e^{-\gamma_{ij}(\Delta\tau)})}{e^{-\gamma_{ij}(\tau_0)}} = 1 - e^{-\gamma_{ij}(\Delta\tau)} \quad (9) \end{aligned}$$

It's obvious that transition probability depends on neither how the process shifts to  $E_i$  state, nor on how long it remains in  $E_i$  state. It depends on which state it is.

Another feature of exponential distribution is that if the parameter of the remaining period of the process in  $E_i$  state before transition from  $E_i$  state to  $E_j$  state is distributed by  $\gamma_{ij}$  exponential law, then the parameter of the time interval from the process transition at any instant time until to  $E_j$  state would be subject to the exponential distribution  $\gamma_{ij}$ .

So state and unstateal probabilities of transition of Markov process to  $E_j$  state are equal:

$$P_{ij}(\Delta\tau) = P_{ij}(\Delta\tau / \tau \geq \tau_0) = 1 - e^{-\gamma_{ij}(\Delta\tau)} \quad (10)$$

If we take  $\Delta\tau$  interval small enough then expression  $e^{-\gamma_{ij}(\Delta\tau)}$  can be lined up according to the power  $(\gamma_{ij}\Delta\tau)$ . If we drop powers with high composition we can get the following:

$$e^{-\gamma_{ij}(\Delta\tau)} \approx 1 - \gamma_{ij} * \Delta\tau \quad (11)$$

$$P_{ij}(\Delta\tau) = 1 - (1 - \gamma_{ij}\Delta\tau) = \gamma_{ij}\Delta\tau. \quad (12)$$

So the model of the need for storage in the Data Center by  $M$  number of users is developed as in Markov process. The process parameters are defined as follows:

1.  $E_1, E_2, \dots, E_K$  - state set.

As it's mentioned above, this set is finite and can be defined as  $K=2^M$ . Each state corresponds to regular order of storage volumes which can be required in different combinations.

2. Stochastic transition matrix. The matrix for continuous-time processes is an intensity matrix of the transition from one state to another. Intensity of transition from  $E_i$  state to  $E_j$  can be denoted by  $g_{ij}$ . This quantity is determined by the following limit:

$$g_{ij} = \lim_{\Delta\tau \rightarrow 0} \frac{P_{ij}(\Delta\tau)}{\Delta\tau} \quad (i, j = \overline{1, n}, i \neq j) \quad (13)$$

The limit shows, that probability of transition from  $E_i$  state to  $E_j$  for small enough  $\Delta\tau$  period is as follows:

$$P_{ij}(\Delta\tau) \approx g_{ij}\Delta\tau \quad (14)$$

If the transition intensity does not depend on time  $t$ , such Markov process is homogeneous. The transition intensity are given in the form of a square matrix

$$G = \begin{pmatrix} g_{11} & \dots & g_{1K} \\ \vdots & \ddots & \vdots \\ g_{K1} & \dots & g_{KK} \end{pmatrix} \quad (15)$$

The diagonal matrix elements  $(g_{ii})$  is found by the following terms:

$$\sum_{i=1}^K g_{ij} = 0 \quad (i = \overline{1, K}). \quad (16)$$

From here

$$g_{ii} = - \sum_{j=1}^K g_{ij} \quad (i, j = \overline{1, K}). \quad (17)$$

As shown above,  $P_{ij}(\Delta\tau) \approx \gamma_{ij}\Delta\tau$ , on the other hand, (13) taking into account

$$g_{ij} = \lim_{\Delta\tau \rightarrow 0} \frac{P_{ij}(\Delta\tau)}{\Delta\tau} = \gamma_{ij}, \quad (18)$$

intensity is a parameter of exponential distribution.

3. Initial probabilities are distribution of the process probability according to the states at the time of  $t = 0$ .

$$p_1^{(0)}, p_2^{(0)}, \dots, p_K^{(0)} \tag{19}$$

Here  $P_i^{(0)}$  is the probability of the process to be in  $E_i$  state at the time  $t=0$ . These probabilities can be assigned approximately as the result of initial observations. Nevertheless, if the modeled process has ergodic characteristics, it is not determined by the initial probabilities  $\{P_i^{(t)}\}$  at any instant time  $t$  and the initial probabilities are not needed anymore.

Ergodicity of the process can be defined in practice as follows:

$\{P_i^{(t_1)}\}$  and  $\{P_i^{(t_2)}\}$  probabilities are defined at different, enough long times  $t_1$  and  $t_2$ .

$$\sum_{i=1}^K (p_i^{(t_1)} - p_i^{(t_2)})^2$$

squared difference or

$$\sum (p_i^{(t)} - p_i^{(t+\Delta t)})^2$$

squared difference are calculated.

If  $\lim_{t \rightarrow \infty} \sum (p_i^{(t)} - p_i^{(t+\Delta t)})^2 \rightarrow 0$  state is covered, i.e. if the *difference* is approaching to zero as  $t$  grows, process is almost ergodic.

4. Probability vector is the key characteristics of the modeled process.

Let  $P_t = \{p_1^{(t)}, p_2^{(t)}, \dots, p_K^{(t)}\}$ . This vector gives complete information about the process at any time  $t$ : where  $0 \leq p_i^{(t)} \leq 1, \sum_{i=1}^K p_i^{(t)} = 1$  states are covered.

In continuous-time Markov processes, the following system is used to calculate these probabilities:

$$\begin{aligned} \frac{dp_i^{(t)}}{dt} &= \sum_{j=1}^K p_j^{(t)} * g_{ij} \\ p_j^{(0)} &= p_j^{(0)} \\ \dots & \\ \sum_{i=1}^K p_i &= 1 \end{aligned} \tag{20}$$

This system can be solved by approximate methods. If the process has ergodic features, it is a system of algebraic equations, which has only one solution.

$$\begin{cases} \sum_{i=1}^K p_i g_{i1} = 0 \\ \sum_{i=1}^K p_i g_{i2} = 0 \\ \dots \\ \sum_{i=1}^K p_i g_{iK} = 0 \\ \sum_{i=1}^K p_i = 1 \end{cases} \tag{21}$$

As the ergodic process shifts to stationary mode, it does not depend on time and it shifts to

$$\begin{aligned} P(t) &= (p_1^{(t)}, p_2^{(t)}, \dots, p_K^{(t)}) \\ &= (p, p_2, \dots, p_K) = P \end{aligned} \tag{22}$$

The vector (22) is the vector of transition probabilities of the process from the  $i$ -th state into another.

The success of the practical application of proposed model depends on the accurate monitoring in the Data

Center. For this purpose, if we develop and apply special software, definition of the current state of the process and accurate prediction of its future development will be achieved due to the values of transition matrix, and specification of the initial probabilities. In this case, the problems of storage distribution among new users can be solved reliably, as well.

As the analyzed process shows the quality of ergodicity over the time, the need for storage of different capacities is determined by a probability distribution  $P = (p_1, p_2, \dots, p_K)$ . Let's assume that there are 200 users totally. If the need for storage capacities varies, we could face  $2^{200}$  various circumstances. This number is calculated as follows:

If we denote

$$2^{200} = e^x, x = 200 * \ln(2)$$

Then, it can be calculated as

$$2^{200} = \exp(200 * \ln(2))$$

These features exist in most programming languages. Without terminating the generality, let's imagine for simplicity, that users need for the storage capacity is the same. In this case, there will be the need for totally 200 different storage capacities, and probability distribution of the storage usage will be as follows:

$$P = (p_1, p_2, \dots, p_{200}) \tag{23}$$

To imagine the outcome of the process in graphics, let us multiply each  $p_i$  by 200 and show in the graph (Fig. 2).

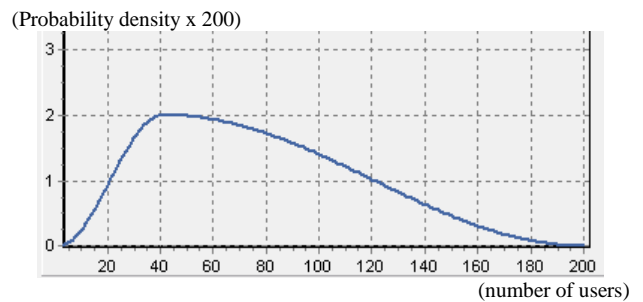


Fig. 2. Probability density of storage usage in Data Center

The discrete probability density at 200 points can be imitated as follows:

$$p(i) = \begin{cases} K * \left[ 1 + \sin\left(\frac{-\pi}{2} + (i-1) * \frac{\pi}{40}\right) \right] & i \in [1, 40] \\ K * \left[ 1 + \sin\left(\frac{\pi}{2} + (i-41) * \frac{\pi}{160}\right) \right] & i \in [41, 200] \end{cases} \tag{24}$$

If we assume  $K = 1$  and sum up obtained values to find  $K$  coefficient the outcome will be as follows:

$$\begin{aligned} &\sum_{i=1}^{40} \left[ 1 + \sin\left(\frac{-\pi}{2} + (i-1) * \frac{\pi}{40}\right) \right] + \\ &\sum_{i=41}^{200} \left[ 1 + \sin\left(\frac{\pi}{2} + (i-41) * \frac{\pi}{160}\right) \right] = 200. \end{aligned} \tag{25}$$

We have to get  $K * 200 = 1$ .

Therefore it shall be

$$K = \frac{1}{200} \text{ to get } \sum_{i=1}^{200} p(i) = 1. \quad (26)$$

Finally, we get the probability density function as follows:

$$p(i) = \begin{cases} \frac{1}{200} * [1 + \sin(\frac{-\pi}{2} + (i - 1) * \frac{\pi}{40})] & i \in [1, 40] \\ \frac{1}{200} * [1 + \sin(\frac{\pi}{2} + (i - 41) * \frac{\pi}{160})] & i \in [41, 200] \end{cases} \quad (27)$$

Obviously, if the number of users is in  $j \in [i + 1; i + N]$  the use of memory can be calculated as  $\sum_{j=i+1}^{i+N} p(i)$ .

Thus, we can calculate the possibilities of the request to the Data Center by the users of 1-10, 11-20, 21-30, 31-40 numbers in different combinations. It shows us the possibilities of the different numbers of user groups who may send requests. If  $i \in [0; 180]$  and if we calculate the probabilities  $\sum_{j=i+1}^{i+20} p(i)$ , we can get approximately the following graphic (Fig. 3). The graphic presents the percentage of user requests of various combinations.

As it is shown in the graph, this distribution differs from an ordinary one. It could be imagined well in advance. It is not excluded that in some values of the users number may be close to a normal distribution.

Described process could be presented by Gamma distribution with the distribution function

$$f(x, a, b) = a(ax)^{b-1}e^{-ax}/\Gamma(b) \quad (28)$$

or by generalized Gamma distribution with the distribution function  $f(x, a, b, c) = ac(ax)^{bc-1}e^{-ax^c}/\Gamma(b)$  (29) [16, 17].

Here  $\Gamma$  - is the gamma function. However, above mentioned distribution (27) was used, as it fits the experimental data best.

The 3rd graph is naturally similar to the distribution density graph. However, the values differ. The probability of the most requests for storage is 60 percent. But, the probability of the fact that 60 percent of user group would request at the same time is less than 40 percent. Time distribution of the storage usage can cause interesting outcomes [18].

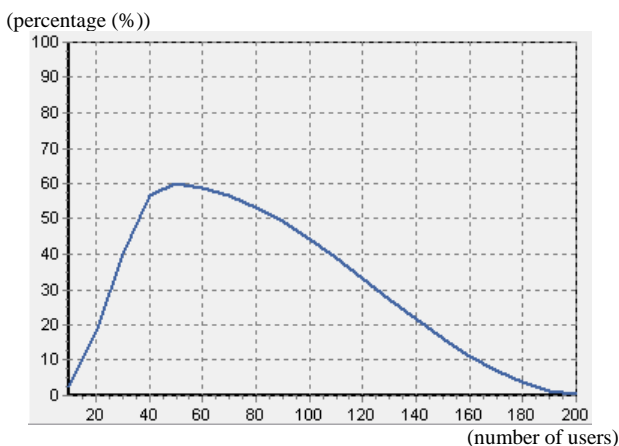


Fig. 3. Percentage of the user requests of various groups at the same time

A very important outcome is that the users mainly required only up to 60 percent of storage capacity that they can use. As we see, 40 % of user resource allocated in the storage systems remains unused.

At the beginning of each predefined time interval the followings can be defined online dynamically via proper software:

- state of the use of computing resources;
- the probability of the transition of the use of resources from one state into another in the current time interval;
- at the same time, the probability of the volume of memory resources to be given to the new potential users.

This, in its turn, may help you to prevent computing resources to remain unused and to use them efficiently.

State of the use of computing resources may be determined by online monitoring. If the work on the offered model has a certain date, then the vector (number) of transition probabilities can be easily determined. The maximum of the elements of this vector

$$\max\{p_1, p_2, \dots, p_K\} = P_{max} \quad (30)$$

shows the probability of the transition of the use of resources from one state into another in the current time interval. Ordinal number of the state equals to the index, which the maximum complies with. The index is denoted as  $j_{max}$ . In other words,

$$P_{j_{max}} = P_{max} \quad (31)$$

All above mentioned determines the algorithm to solve the problem of visualization of the process. Solving visualization issue the most important information units shall be delivered to the people on the computer screens without any obstacles. In addition, the principle of information integrity shall be followed. In other words, with the minimum density of it shall provide the full information to receive the monitoring of the course of the process [19,20]. Therefore, calculated transition probabilities shall also be displayed on the screen along with the sequence of process states. This shows how the adequacy of the forecast changes and increases over the time. States from which the process shifts in the course of the monitoring are stored at Data Center, and as the result, the following graph is set (fig. 4). In Figure 4  $i_k$  are the states of the monitoring process. Here,  $p_{ik, i(k+1)}$  is the probability of forecasted new state.

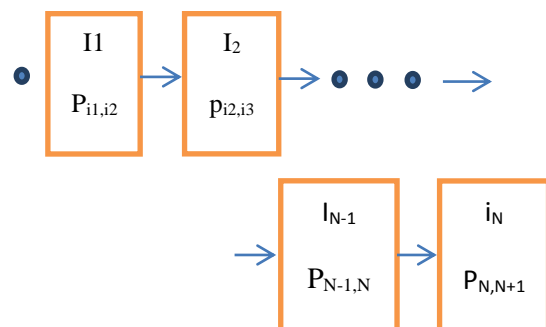


Fig. 4. Monitoring display.

The graph reveals how the process continues. At each new interval of the use of computing power in online mode the current state is included into the graph. Forecast decisions indicate the state of the use of computing power, which shifted from the current state, and the probability of this shift. The number of elements stored on the display can be selected. However, the number of default states provided on the display is set so that the course of the process can be imagined clearly.

In this case, demonstration and re-introduction of the states, which do not appear on the screen and including monitoring history are enabled.

Long-term monitoring results are stored in the database. The database has the following structure:

- Date, duration. These parameters indicate the date and duration of the information recording;
- State of computing power use (ordinal number or the number of users);
- Duration of the current state. This data is calculated when it shifts to the next state;
- Used computational power. This setting is denoted in the physical units or percentages;
- Unused computing power. This setting is denoted in the physical units or percentages;
- The following predicted state;
- Adequacy level of the forecast. It is set by the simple method, i.e. denoting each right forecast as 1 and each wrong forecast as 0;
- Forecasted probability of the transition into the next state.

The following actions are performed over the base:

- Writing in the base through software online;
- Reading in the base through software online;
- Reading and writing in the base through the operator's commands;
- Correcting the base through the operator's commands.

Information security is provided by administrative ways and by a system of randomly changing passwords for the base use.

Using monitoring base various operational characteristics of the Data Center can be defined, and the statistical analysis can be realized. As it is mentioned above, the results of the statistical analysis indicate state of the use of computing power and the change of the forecast quality in the course of monitoring. In addition, availability of the periodical changes or dependence on the seasons of the year can be determined during the utilization of computational power.

As a result, remained resources can be used by other users. The model is used in the virtualization of storage resources and prevents their wasting. The model provides that the storage resources allocated for any purpose is fully used and hold the space in the system as much as it is used. Allocated resource is not reserved and does not hold space in hard disks as long as it is not used. The model enables mining centers to use big storage resources efficiently. Thus, resources are allocated as much as it is used only without wasting. Such way of resource distribution is beneficial for both cloud provider

and the user. Accordingly, the user does not pay for reserved resource, but only for the actual resource use. And the provider reduces unnecessary purchase and installation of additional equipments; furthermore it can be able to offer the same service for lower price, which leads to greater user involvement in this type of services. It is obvious that the Data Center managers are able to attract new users without the risk.

### III. CONCLUSION

The paper reviews and dynamical distribution of storage resources among the users in data processing centers. The stochastic model of the dynamic distribution of storage resources is proposed. The model is used in the virtualization of storage resources and ensures the use of storage resources without any waste. The model provides that the storage resources allocated for any purpose hold the space in the system as much as they are used. Thus, resources are allocated as much as it is used only without wasting. It is beneficial for both cloud provider and the user. Accordingly, the user does not pay for reserved resource, but only for the actual resource use. And the provider reduces unnecessary purchase and installation of additional equipments; furthermore it can be able to offer the same service for lower price, which leads to greater user involvement in this type of services. Developed Monitoring database contributes to the solution of various issues related to the Data Center operation. In general, developed model and software can be used for the typical problem solution or different type of problems but with the analogical mathematical formalization.

### REFERENCES

- [1] Voevodin V.V., Voevodin VI.V. Parallel computing. St. Petersburg. "BHV – Petersburg", 2002.
- [2] Ilin Yu. IBM invests "Cloud Computing", - [www.pcnews.ru/news/ibm-300](http://www.pcnews.ru/news/ibm-300)
- [3] Alguliyev R.M., Alekperov R.K. Cloud Computing: Modern State, Problems and Prospects. Telecommunications and Radio Engineering, 2013, vol.72, no.3, pp.255-266
- [4] Marios D. Dikaiakos, George Pallis, Dimitrios Katsaros, Pankaj Mehra, Vakali Athena. Cloud Computing, Distributed Internet Computing for IT and Scientific Research // IEEE INTERNET COMPUTING - 2009. № 9. P. 10-13.
- [5] Keedong Yoo. Cloud Storage-based Intelligent Document Archiving for the Management of Big Data. International Journal of Information Technology & Computer Science (IJITCS) (ISSN No : 2091-1610) Volume 9 : IssueNo : 3 : Issue on May /June, 2013.
- [6] Rajendar Kandan, Mahendran Ellappan, Madhusudhana Rao S & Rajagopalan M.R. Dynamic Resource Provisioning Framework (DRPF) using Grid and Cloud Computing. International Journal of Information Technology & Computer Science (IJITCS) (<http://www.ijitcs.com>) (ISSN : 2091-1610), Volume No : 12, IssueNo : 3.
- [7] A.A. Markov. "Extension of the limit theorems of probability theory to a sum of variables connected in a

chain". Reprinted in Appendix B of: R. Howard. Dynamic Probabilistic Systems, volume 1: Markov Chains. John Wiley and Sons, 1971

- [8] Sean Meyn. Control Techniques for Complex Networks. Cambridge University Press. 2007, p.615.
- [9] S.P. Meyn and R.L. Tweedie. Markov Chains and Stochastic Stability. Springer-Verlag 1993, p 552.
- [10] Hemdi A. Taha, Introduction to Operations Research, 7th Edition. Hardcover, from English. / Hemdi A. Taha Publishing House "Williams", 2005. -912 p.
- [11] Fu and Koutras, M. J. C. V. Distribution theory of runs: A Markov chain approach // Journal of the American Statistical Association. V.89.-1996. P.1050-1058.
- [12] Balasubramanian K., Viveros R. and Balakrishnan N. Sooner and later waiting time problems for Markovian Bernoulli trials, statistically // Probab. Lett. - 1993. P.153-161.
- [13] Kolev N. and Minkov L. On the joint distribution of the successes and failures related to success runs of length  $k$  in the homogeneous Markov chain // Compt. Rendue Bulg. Acad. Sci., 48. -1995. V.9. 19-22.
- [14] Graham Ronald, Knuth Donald, Patashnik Oren. "(5) Binomial Coefficients" Concrete Mathematics (2nd ed.) // Addison Wesley. 1994. PP. 153-256. ISBN 0-201-55802-5. OCLC 17649857.
- [15] Shilov, G. E. (1977). Linear algebra. Dover Publications. ISBN 978-0-486-63518-7.
- [16] Cristian Walck. Hand-book on Statistical distributions for experimentalists. Practice Physics Group Fysikum, University of Stokholm, 2007, p.2002.
- [17] Gamma Function and Related Functions by Philip J. Davis in Handbook of Mathematical Functions, ed. M. Abramowitz and I. A. Stegun, Dover Publications, Inc., New York, 1965, p.253.
- [18] Alakbarov R.M., Guseinova A.A. Development of adaptive model for optimal load distribution in corporate networks. //Telecommunications. 2012. № 9.P.17-21.
- [19] Sutcliffe A.G., 1995, Human-Computer Interface Design (2nd Ed), MacMillan, pp. 274-276.
- [20] Dix et al, 1998, Human-Computer Interaction (2nd Ed), Prentice Hall, pp. 212-220.

#### Authors' Profiles



**Rashid G. Alakbarov** graduated from "Automation and Computer Engineering" faculty of Azerbaijan Polytechnic University named after C.Ildirim. He received his PhD degree in 2006 from Supreme Attestation Commission under the President of the Republic of Azerbaijan. His primary

research interests include various areas in cloud computing, data processing, computer networks, virtual computing, particularly in the area of distributed computing.

He is head of department at the Institute of Information Technology as of 2002. Since 2010, he has been leading the development of "AzScienceNet" infrastructure. In 2011, he was appointed a senior engineer of the institute by the decision of the Presidium of Azerbaijan National Academy of Sciences. He is the author of 41 scientific papers, including 5 inventions.

**Fahrad H. Pashayev** graduated from Azerbaijan State University. He received his PhD degree in 2011 from Supreme Attestation Commission under the President of the Republic of



Azerbaijan. His primary research interests include various areas in mathematical modeling, signal processing, software development. He is the author of 84 scientific papers.



**Mammad A. Hashimov** received his Master's degree in automation and control from Azerbaijan Technical University in Baku, Azerbaijan. He is PhD student of Institute of Information Technology of Azerbaijan National Academy of Sciences. His primary research interests include various areas in cloud computing, data processing,

computer networks, virtual computing, particularly in the area of cloud technology applications. He is the author of 2 journal scientific papers and 2 proceedings.

**How to cite this paper:** Rashid G. Alakbarov, Fahrad H. Pashaev, Mammad A. Hashimov, "Development of the Model of Dynamic Storage Distribution in Data Processing Centers", International Journal of Information Technology and Computer Science (IJITCS), vol.7, no.5, pp.18-24, 2015. DOI: 10.5815/ijitcs.2015.05.03