Semi-Markov Availability Models for an Infrastructure as a Service Cloud with Multiple Pools

Oleg Ivanchenko[1], Vyacheslav Kharchenko[2]

[1] University of Customs and Finance,
2/4 Dzerzhinskogo, 217 km, Dnepropetrovsk, Ukraine

`vmsu12@gmail.com`

[2] National Aerospace University "Kharkiv Aviation Institute"
17 Chkalova St., 61070, Kharkiv, Ukraine

`v_s_kharchenko@ukr.net`

Abstract. Solving of tasks for Cloud Computing is impossible without maintaining of high availability level of Infrastructure as a Service (IaaS) Cloud. Several large IaaS Cloud providers try to solve this problem by means of increasing number of physical machines (PMs) in multiple pools. However, migrations of available PMs from one pool to another and also repairs, diagnostic instances of failed physical machines are quite complex task for modeling of availability for an IaaS Cloud. In this paper, we show how we can build Semi-Markov availability models with discrete states and how we can use it in order to determine availability level for the IaaS Cloud with three pools.

Keywords. Infrastructure as a Service Cloud, three pools of physical machines, Semi-Markov availability models

Key Terms. Infrastructure, Mathematical Model, Development, Characteristic

1 Introduction

Nowadays Cloud Computing is one of the most widely used services in an enterprise environment. Therefore availability of cloud infrastructures is of paramount importance to improve quality of service (QoS) and development of cloud user's possibilities. Despite this fact, researchers of cloud infrastructure behavior, including availability and reliability analysis of respective components are still regarded as quite complex scientific direction for different kinds of modeling.

Large Infrastructure as a Service (IaaS) Cloud providers try to use multiple pools of physical machines (PMs) in order to maintain normal operation of cloud's components on quite a long period of time. However, with a larger number of PMs number of states for stochastic model of an IaaS Cloud also increase; the model ought to in-

clude a large number of parameters while still being tractable [1]. Some famous researchers in order to perform availability analysis of large side clouds had built interacting sub-models, before they started to build monolithic model for an IaaS Cloud [2]. They built interacting sub-models on based Markov models and used stochastic reward nets. At the same time these sub-models were built by them with using of their own software package SHARPE [3]. Earlier in paper [4] authors tried to build a continuous time Markov chain (CTMC) availability model, for an example with two PMs in each pool. Another researcher also proposed the availability model with the failure of PMs, repairing process and employment of cold PMs in case of failure in running machines [5].In this model PMs in each pool are modeled by a three-dimensional CTMC too. In spite of the fact that various authors used stochastic approach based Markov models to describe behavior of the physical machine pools, they couldn't get rigorous analytical expressions.

Focus of this paper is to build Semi-Markov availability models for the IaaS Cloud with three pools and different number of PMs in each pool.

2 Statement of the Researches Results

2.1 Metamodel for Availability Analysis of IaaS Cloud

Note that the architecture of an IaaS Cloud is not tied to a real cloud implementation [6]. Suppose that researchers have used a simple cloud infrastructure with certain number of PMs. To reduce power consumption, cooling and infrastructure costs, PMs are grouped into three pools such as: hot, warm and cold pools. Assume that hot pool consists from turned on and running PMs; warm pool contains turned on, although not ready physical machines; cold pool consists from turned off PMs. Moreover this architecture has certain number of virtual machines (VMs), which are deployed on PMs. Deployment of VMs on base PMs allows to reduce power consumption and to maintain enough high performance of the cloud implementation.

In difference from other, proposed concept for maintaining of availability for a cloud infrastructure bases on use of two additional systems, namely Technical State-Control System (TSCS)and Resource Provisioning Decision Engine (RPDE) [7].Our IaaS Cloud should be used TSCS, which is working in monitoring and diagnostic modes. In this case, these modes as regarded as an organization form of constant control of the significant parameters that the determinate not only the PMs performability, but also affect cloud infrastructure readiness to make effective intended use [8]. It's obviously, that monitoring and diagnostic sub-systems provide repair facilitated by information which is needed to repair and migration of PMs from one pool to other. As described in [6], RPDE tries to find a PM that can accept the job provisioning.

Figure 1 shows the portions of the taxonomy metamodel for availability analysis of IaaS Cloud. Researchers in order to deal with the complexities of metamodeling should work in the paradigm of four models, such as scalability, performance, flexibility (elasticity), power consumption. Each model ensures the overall metamodel by input parameters, namely initial number of PMs for each pool (scalability model), power consumption for each PM (power consumption model), management metrics

values, search rates (flexibility model) and failure rates, repair rates, migration rates, number of repair facilities (reliability model).In other words output parameters of these models are input parameters for meta model. At the same time values of design and temporal parameters of such models can be experimentally measured. The stages of meta modeling are colorfully shown by this figure. According to the illustrated Fig. 1, we will try to create analytical models with considering states and stochastic changing of all times failures, repairs and migrations of PMs.
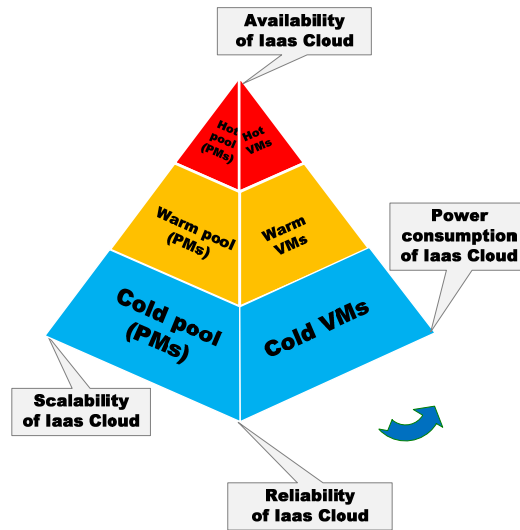

Fig. 1. Taxonomy metamodel for availability analysis of IaaS Cloud

On this basis, we will construct Semi-Markov model for availability analysis of an IaaS Cloud with three pools. Therefore it is proposed to describe various options of interactions of PMs at availability-model level.

2.2 Analytical Availability Models for an IaaS Cloud with Three Pools

Let's consider two interesting analytical models of an IaaS Cloud. Fig. 2 shows a Semi-Markov (SM) model for availability analysis of the IaaS Cloud with three pools (hot, warm and cold) and three PMs in each pool.

In our modeling we use the following assumptions and limitations.

- Hot, warm, and cold pools contain identical PMs [9]. If a hot PM fails the failed PM is replaced by available (non-failed) PM from warm or cold pools, respectively.
- We assume that periodic technical state control (CTS) of hot PMs is operated during a time interval, which lasts τ_c.
- To analyze the availability of the IaaS Cloud we also assume that all times to failure of all PMs are exponentially distributed. Typically, mean time to failure (MTTF) of warm PMs ($1/\lambda_w$) is higher than MTTF of hot PMs ($1/\lambda_h$) by a factor

of two to four [7]. At the same time MTTF of cold PMs is a very lower than $1/\lambda_w$. However, for process of SM modeling we will use only MTTF of hot PMs, considering quite high reliability level of warm and cold PMs.

- Moreover in real situations providers haven't enough time for repair of failed hot PMs, as well as they haven't enough number of repair facilities. Therefore we also assume that all times to repair are not exponentially distributed. In this occasion we have preferred to use Erlang-k distribution, where $k = 2, 3$ [10]. Parameter $1/\mu$ is mean time to repair (MTTR) of a PM.

- Available PMs can migrate from warm and cold pools to hot pool. We also assume that all times to migration (migration delays) of PMs are exponentially distributed. For modeling we have used mean time to migration (MTTM) of PMs from warm ($1/\gamma_{wh}$) and cold ($1/\gamma_{ch}$) pools to hot pool.

- The migrations of PMs to hot pool are implemented when providers can search non-failed warm or cold PMs with mean time to searches (MTTSs) $1/\delta_w$ and $1/\delta_c$.

- We consider that IaaS Cloud becomes unavailable when the SM model enters the state S_{15}.

Suppose that this infrastructure is operated during a time interval $t \in [0,T]$ and at the initial moment $t = 0$ the IaaS Cloud is ready for using (state S_0). The transition from state S_0 to state S_1 occurs at fixed nonrandom time $\tau_c < T$, where parameter T is operation time of IaaS Cloud between two periodic controls of technical state. The state S_1 is state of CTS. Note that the periodic CTS includes monitoring and diagnostic operations of hot PMs and conduct by means of using Technical State Control System. If third hot PM is available the SM model returns from state S_1 to state S_0.

Otherwise when the TSCS detects a failure, model goes to state S_2 with rate λ_h. In state failure of the third hot PM, model tries to search non-failed warm PM (transition from state S_2 to state S_3) with rate δ_w or cold PM (transition from state S_2 to state S_4) with rate δ_c. When warm or cold PMs are available, model transforms from state S_3 to state S_0 or from state S_4 to state S_0 respectively.

If the warm and cold pools are empty, repair facility tries to recover the failed hot PM, that is model goes from state S_2 to state S_0 with repair rate μ. When recovery the third failed PM is impossible, model transforms from state S_2 to state S_5 with overall failure rate $3\lambda_h$. It means that next steps of modeling as regards states of $S_5 - S_9$ for second hot PM and states of $S_{10} - S_{12}$ for first hot PM are repeated. Note that in this case we can maintain that transition from state S_7 to state S_{10} and transition from state S_{12} to state S_{15} are implemented with overall failure rates $2\lambda_h$ and λ_h respectively. We also consider that the model will transition from state S_{12} to state S_{15} when the last hot PM fails.
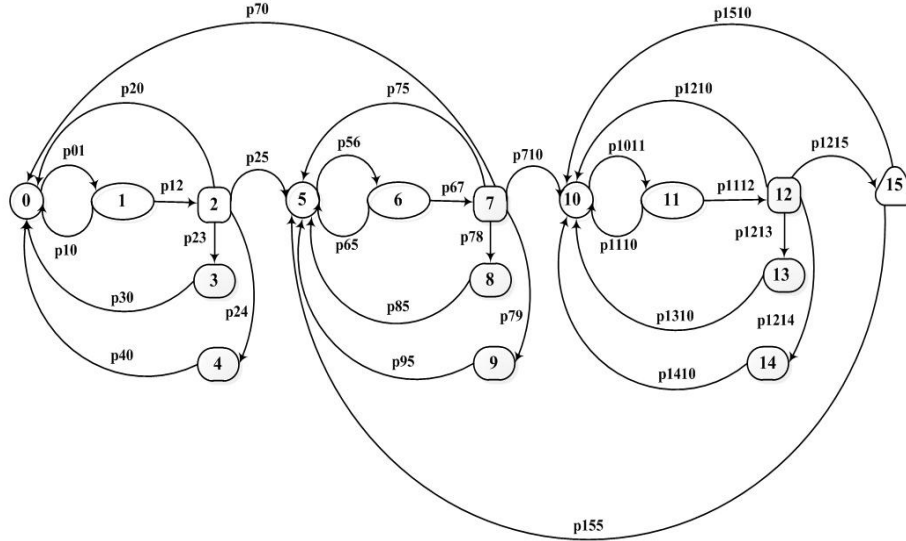
Fig. 2. SM model for the availability analysis of the IaaS Cloud with three PMs in each pool

To solve this task in the following we are inclined to use method of transformation of the SM models into embedded Markov chains [10]. For this type of models the transitions of process from state i to state j occur through unit time. Therefore the transitions of this SM process are interpreted as follows. Since CTS performs within fixed deterministic period of time T, consequently transition from state S_0 to state S_1 is given by:

$$Q_{01}(t) = \begin{cases} 0, t < T, \\ 1, t \geq T. \end{cases}$$

The transition from state S_1 to state S_0 is then given by:

$$Q_{10}(t) = \begin{cases} 0, t < \tau_c, \\ 1, t \geq \tau_c. \end{cases}$$

The other similar transitions can be got as follows:

$$Q_{56}(t) = Q_{1011}(t) = \begin{cases} 0, t < T, \\ 1, t \geq T, \end{cases}$$

$$Q_{65}(t) = Q_{1110}(t) = \begin{cases} 0, t < \tau_c, \\ 1, t \geq \tau_c. \end{cases}$$

At the same time, probabilities of sudden failures of hot PMs at random times for transitions from state S_1 to state S_2, from state S_6 to state S_7 and from state S_{11} to state S_{12} are given by:

$$Q_{12}(t) = Q_{67}(t) = Q_{1112}(t) = 1 - e^{-\lambda_h t} .$$

Implementations of transitions from state S_2 to state S_0, from state S_7 to state S_5, from state S_{12} to state S_{10} and from state S_7 to state S_0, from state S_{12} to state S_5, from state S_{15} to state S_{10} depend from time to repair of the hot PMs. Therefore in these cases, distribution functions of repair time are given by:

$$Q_{20}(t) = Q_{75}(t) = Q_{1210}(t) = 1 - (1 + \mu t)e^{-\mu t},$$

$$Q_{70}(t) = Q_{125}(t) = Q_{1510}(t) = 1 - \left(1 + \mu t + \frac{(\mu t)^2}{2}\right)e^{-\mu t}.$$

For our SM availability model, we assume that distribution functions of search time of non-failed warm and cold PMs respectively are given by:

$$Q_{23}(t) = Q_{78}(t) = Q_{1213}(t) = 1 - e^{-\delta_w t},$$

$$Q_{24}(t) = Q_{79}(t) = Q_{1214}(t) = 1 - e^{-\delta_c t}.$$

Similarly, distribution functions of migration time for warm and cold PMs respectively are given by:

$$Q_{30}(t) = Q_{85}(t) = Q_{1310}(t) = 1 - e^{-\gamma_{wh} t},$$

$$Q_{40}(t) = Q_{95}(t) = Q_{1410}(t) = 1 - e^{-\gamma_{ch} t}.$$

Then steady-state availability [10] of the cloud can be computed as

$$A = \pi_0 + \pi_5 + \pi_{10}, \tag{1}$$

where π_0, π_5, π_{10} are steady-state probabilities for states S_0, S_5, S_{10}.

On the other hand, the steady-state availability A (1) is given by [11]:

$$A = \lim_{t \to \infty} A(t),$$

where $A(t)$ – instantaneous availability of the cloud infrastructure.

In the overall case steady-state probabilities of SM availability model are given by:

$$\pi_0 = \frac{\overline{t_0}}{U}, \quad \pi_5 = \beta\frac{\overline{t_5}}{U}, \quad \pi_{10} = \alpha\beta\frac{\overline{t_{10}}}{U},$$

$$U = \overline{t_0} + \overline{t_1} + p_{12}\left(\overline{t_2} + p_{23}\overline{t_3} + p_{24}\overline{t_4}\right) + \beta\left[\overline{t_5} + \overline{t_6} + p_{67}\left(\overline{t_7} + p_{78}\overline{t_8} + p_{79}\overline{t_9}\right) + \right.$$
$$\left. + \alpha\left(\overline{t_{10}} + \overline{t_{11}}\right)\right] + \varepsilon\left(\overline{t_{12}} + p_{1213}\overline{t_{13}} + p_{1214}\overline{t_{14}} + p_{1215}\overline{t_{15}}\right),$$

where

$$\alpha = \frac{p_{67}p_{710}}{1 - p_{1110} - \xi p_{1112}}, \quad \beta = \frac{p_{12}p_{25}}{1 - p_{65} - \alpha p_{1112}p_{125} - v p_{67}},$$

$$\xi = p_{1210} + p_{1213} + p_{1214} + p_{1215}, \quad v = p_{75} - p_{78} - p_{79}, \quad \varepsilon = \alpha\beta p_{1112},$$

$$p_{12} = p_{67} = p_{1112} = 1 - e^{-\lambda_h \tau_c},$$

$$p_{23} = \delta_w \int_0^\infty (1 + \mu t) e^{-(3\lambda_h + \delta_w + \delta_c + \mu)t} dt \, , \quad p_{24} = \delta_c \int_0^\infty (1 + \mu t) e^{-(3\lambda_h + \delta_w + \delta_c + \mu)t} dt \, ,$$

$$p_{25} = \lambda_h \int_0^\infty (1 + \mu t) e^{-(3\lambda_h + \delta_w + \delta_c + \mu)t} dt \, ,$$

$$p_{75} = \frac{1}{2} \int_0^\infty \mu^2 t \left(\mu^2 t^2 + 2\mu t + 2 \right) \left(1 + \mu t + \frac{(\mu t)^2}{2} \right) e^{-(2\lambda_h + \delta_w + \delta_c + 2\mu)t} dt \, ,$$

$$p_{78} = \delta_w \int_0^\infty (1 + \mu t) \left(1 + \mu t + \frac{(\mu t)^2}{2} \right) e^{-(2\lambda_h + \delta_w + \delta_c + 2\mu)t} dt \, ,$$

$$p_{79} = \delta_c \int_0^\infty (1 + \mu t) \left(1 + \mu t + \frac{(\mu t)^2}{2} \right) e^{-(2\lambda_h + \delta_w + \delta_c + 2\mu)t} dt \, ,$$

$$p_{710} = \lambda_h \int_0^\infty (1 + \mu t) \left(1 + \mu t + \frac{(\mu t)^2}{2} \right) e^{-(2\lambda_h + \delta_w + \delta_c + 2\mu)t} dt \, , \quad p_{65} = p_{1110} = e^{-\lambda_h \tau_c} \, ,$$

$$p_{1210} = \frac{1}{2} \int_0^\infty \mu^2 t \left(\mu^2 t^2 + 2\mu t + 2 \right) \left(1 + \mu t + \frac{(\mu t)^2}{2} \right) e^{-(\lambda_h + \delta_w + \delta_c + 2\mu)t} dt \, ,$$

$$p_{1214} = \delta_c \int_0^\infty (1 + \mu t) \left(1 + \mu t + \frac{(\mu t)^2}{2} \right) e^{-(\lambda_h + \delta_w + \delta_c + 2\mu)t} dt \, ,$$

$$p_{1215} = \lambda_h \int_0^\infty (1 + \mu t) \left(1 + \mu t + \frac{(\mu t)^2}{2} \right) e^{-(\lambda_h + \delta_w + \delta_c + 2\mu)t} dt \, ,$$

$$p_{125} = 1 - p_{1210} - p_{1213} - p_{1214} - p_{1215} \, ,$$

$$\overline{t_0} = \overline{t_5} = \overline{t_{10}} = T \, , \quad \overline{t_2} = \frac{3\lambda_h + \delta_w + \delta_c + 2\mu}{(3\lambda_h + \delta_w + \delta_c + \mu)^2} \, , \quad \overline{t_3} = \overline{t_8} = \overline{t_{13}} = \frac{1}{\gamma_{wh}} \, , \quad \overline{t_4} = \overline{t_9} = \overline{t_{14}} = \frac{1}{\gamma_{ch}} \, ,$$

$$\overline{t_{15}} = \frac{3}{\mu} \, , \quad \overline{t_1} = \overline{t_6} = \overline{t_{11}} = \frac{1}{\lambda_h} \left(1 - e^{-\lambda_h \tau_c} \right),$$

$$\overline{t_7} = \int_0^\infty (1 + \mu t) \left(1 + \mu t + \frac{(\mu t)^2}{2} \right) e^{-(2\lambda_h + \delta_w + \delta_c + 2\mu)t} dt \, ,$$

$$\overline{t_{12}} = \int_0^\infty (1 + \mu t) \left(1 + \mu t + \frac{(\mu t)^2}{2} \right) e^{-(\lambda_h + \delta_w + \delta_c + 2\mu)t} dt \, .$$

Plots depending of steady-state availability A from failure rates λ_h of hot PMs and operation time T (repair rates μ are constant values) are shown in Fig. 3, Fig. 4. The values of steady-state availability A are greatly increased by means of increasing of repair rate μ and reduction of failure rate λ_h of hot PMs, as depicted in Fig. 3 and Fig. 4.
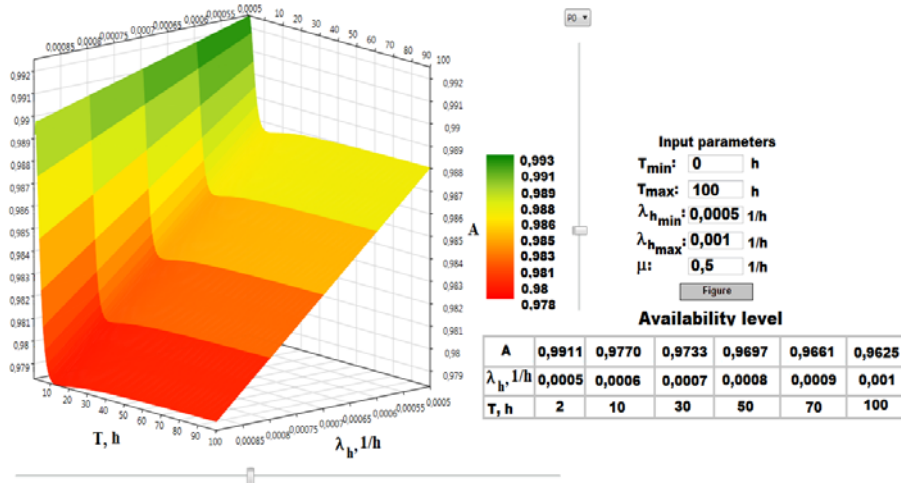
Fig. 3. Depending of steady-state availability $A(\lambda_h, T)$ for $T = 100$ h, $\mu = 0,5$ 1/h
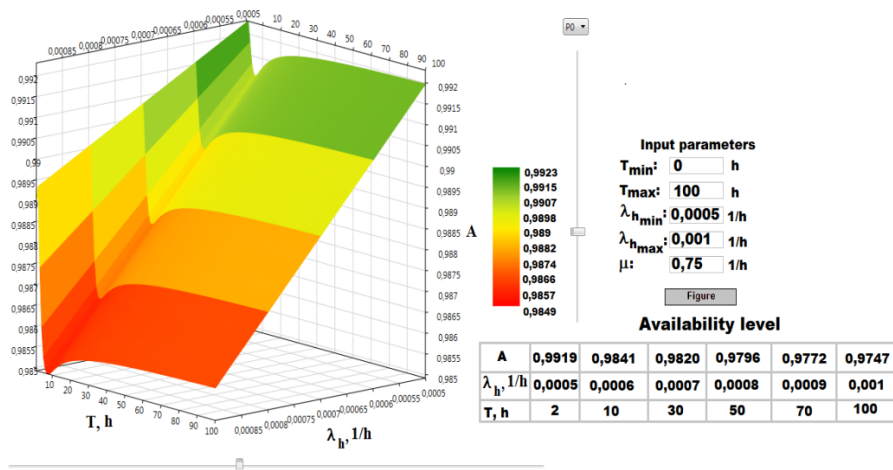
Fig. 4. Depending of steady-state availability $A(\lambda_h, T)$ for $T = 100$ h, $\mu = 0,75$ 1/h

Let's continue our researches by means of creation more scalable stochastic model for IaaS Cloud. Because with a larger number of PMs in a data center, the overall Cloud service availability increases, leading to lower cost of service downtime [7].Therefore within a unified methodological approach we will try to create an improved SM availability model of infrastructure with a larger number of PMs.

Additional researches have shown that IaaS Cloud providers wish to increase number of PMs in order to minimize downtime cost and damage business reputation [4], [6], [7]. Perhaps inspired by using stochastic approaches for solution various

serious tasks of determining the optimal PM capacity configuration of IaaS Cloud [6], we have been proposing next SM availability model.

Assume that our infrastructure contains similar three pools with ten PMs in each pool. This SM model for availability analysis of the IaaS Cloud is shown in Fig. 5. Also suppose that all times to failure of PMs are exponentially distributed and Erlang-k distribution, where $k = 2$ is general distribution for all times to repair. In spite of the fact, that both models are SMs models, we have to take into consideration some interesting features of their implementation.
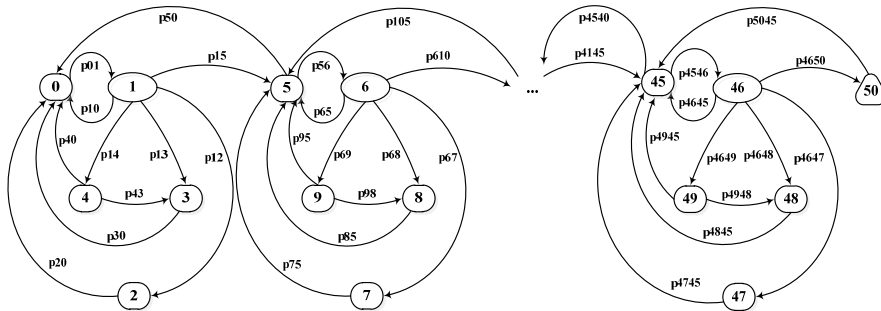
Fig. 5. SM model for the availability analysis of the IaaS Cloud with ten PMsineachpool

Unlike first SM model, second SM availability model of the IaaS Cloud includes modeling kernel from five states. The states S_0, S_1, S_5 for second model (Fig. 5) are the same as the first model (Fig. 2). But the difference between kernels of first model and second model is that states S_4, S_9 for first model are states of search of the cold PM, whilst these states for second model are states of failure of the warm PMs. For second model the following group assumptions can take place.

- Model contains hot, warm, and cold pools. Every pool consists of ten identical PMs[9]. If a hot PM fails the failed PM is replaced by available (non-failed) PM from warm or cold pools too.
- Upon failure of the warm PM, the failed PM is replaced by available (non-failed) PM from cold pool.
- We also assume that periodic technical state control (CTS) of hot PMs is operated during a time interval, which lasts τ_c.
- To analyze performance and availability of the IaaS Cloud we also assume that all times to failure of all hot and warm PMs are exponentially distributed.
- We also consider that all times to repair are not exponentially distributed. In this case we have used Erlang-k distribution, where $k = 2,3$ [10]. Parameter $1/\mu$ is mean time to repair (MTTR) of a PM.
- Cloud infrastructure becomes unavailable when the SM model enters the state S_{50}.

Therefore the transitions of modeling kernel for second SM model can be written as follows:

$$Q_{01}(t) = Q_{56}(t) = Q_{1011}(t) = Q_{1516}(t) = Q_{2021}(t) = Q_{2526}(t) = Q_{3031}(t) = Q_{3536}(t) =$$
$$= Q_{4041}(t) = Q_{4546}(t) = \begin{cases} 0, t < T, \\ 1, t \geq T, \end{cases}$$
$$Q_{10}(t) = Q_{65}(t) = Q_{1110}(t) = Q_{1615}(t) = Q_{2120}(t) = Q_{2625}(t) = Q_{3130}(t) = Q_{3635}(t) =$$
$$= Q_{4140}(t) = Q_{4645}(t) = \begin{cases} 0, t < \tau_c, \\ 1, t \geq \tau_c, \end{cases}$$

For other functions we can write the following:

$$Q_{12}(t) = Q_{67}(t) = Q_{1112}(t) = Q_{1617}(t) = Q_{2122}(t) = Q_{2627}(t) = Q_{3132}(t) = Q_{3637}(t) =$$
$$= Q_{4142}(t) = Q_{4647}(t) = 1 - e^{-\delta_w t},$$
$$Q_{13}(t) = Q_{68}(t) = Q_{1113}(t) = Q_{1618}(t) = Q_{2123}(t) = Q_{2628}(t) = Q_{3133}(t) = Q_{3638}(t) =$$
$$= Q_{4143}(t) = Q_{4648}(t) = 1 - e^{-\delta_c t},$$
$$Q_{14}(t) = Q_{69}(t) = Q_{1114}(t) = Q_{1619}(t) = Q_{2124}(t) = Q_{2629}(t) = Q_{3134}(t) = Q_{3639}(t) =$$
$$= Q_{4144}(t) = Q_{4649}(t) = 1 - e^{-\lambda_w t},$$
$$Q_{20}(t) = Q_{75}(t) = Q_{1210}(t) = Q_{1715}(t) = Q_{2220}(t) = Q_{2725}(t) = Q_{3230}(t) = Q_{3735}(t) =$$
$$= Q_{4240}(t) = Q_{4745}(t) = 1 - e^{-\gamma_{wh} t},$$
$$Q_{30}(t) = Q_{85}(t) = Q_{1310}(t) = Q_{1815}(t) = Q_{2320}(t) = Q_{2825}(t) = Q_{3330}(t) = Q_{3835}(t) =$$
$$= Q_{4340}(t) = Q_{4845}(t) = 1 - e^{-\gamma_{ch} t},$$
$$Q_{43}(t) = Q_{98}(t) = Q_{1413}(t) = Q_{1918}(t) = Q_{2423}(t) = Q_{2928}(t) = Q_{3433}(t) = Q_{3938}(t) =$$
$$= Q_{4443}(t) = Q_{4948}(t) = 1 - e^{-\delta_c t},$$
$$Q_{40}(t) = Q_{95}(t) = Q_{1410}(t) = Q_{1915}(t) = Q_{2420}(t) = Q_{2925}(t) = Q_{3430}(t) = Q_{3935}(t) =$$
$$= Q_{4440}(t) = Q_{4945}(t) = 1 - (1 + \mu t) e^{-\mu t},$$
$$Q_{50}(t) = Q_{105}(t) = Q_{1510}(t) = Q_{2015}(t) = Q_{2520}(t) = Q_{3025}(t) = Q_{3530}(t) = Q_{4035}(t) =$$
$$= Q_{4450}(t) = Q_{5045}(t) = 1 - \left(1 + \mu t + \frac{(\mu t)^2}{2} \right) e^{-\mu t},$$

$$Q_{15}(t) = 1 - e^{-\lambda_1 t}, \tag{2}$$
$$Q_{610}(t) = 1 - e^{-\lambda_2 t}, \tag{3}$$
$$Q_{1115}(t) = 1 - e^{-\lambda_3 t}, \tag{4}$$
$$Q_{1620}(t) = 1 - e^{-\lambda_4 t}, \tag{5}$$
$$Q_{2125}(t) = 1 - e^{-\lambda_5 t}, \tag{6}$$
$$Q_{2630}(t) = 1 - e^{-\lambda_6 t}, \tag{7}$$
$$Q_{3135}(t) = 1 - e^{-\lambda_7 t}, \tag{8}$$
$$Q_{3640}(t) = 1 - e^{-\lambda_8 t}, \tag{9}$$
$$Q_{4145}(t) = 1 - e^{-\lambda_9 t}, \tag{10}$$

$$Q_{4650}(t) = 1 - e^{-\lambda_{10}t}. \qquad (11)$$

We define the failure rates for $j = 1,2,...,n_h$ ($n_h = 10$) PMs nodes[9], [10]:

$$\lambda_j = (n_h - i)\lambda_0 , i = 0,1,...k \text{ (for } k = n_h - 1 \text{)}, \qquad (12)$$

where λ_0 – basic failure rate value for all PMs.

By replacing the λ_j expression (12) to the λ_s values in the equations (2), (3), …, (11), we will be finished description of second model.

As can be seen in Fig. 3 and Fig. 4 in case with three PMs in each pool, IaaS Cloud has quite high of availability level. Results of modeling for second SM availability model will get in the near future time. Overall feature for both SM models is identical modeling kernels.

3 Conclusions Statement of the Researches Results

Clearly, proposed stochastic approach based SM models gives opportunity to perform availability analysis of the Cloud Infrastructures with using of different modeling kernels. Thus, the contributions of this paper are the following.

If you wanted to make a deep availability and reliability analysis of the IaaS Cloud, for example, when this infrastructure is one of the most important components of Management System Critical Infrastructure, in particular during the accidents and disasters or other negative events, such as sudden or hidden failures, you would be able to use proposed SMs availability models. An additional advantage of these SMs models is that researches can use rigorous analytical expressions from this paper in order to determine availability and reliability values for the IaaS Cloud. Moreover you can use this stochastic approach in order to choose optimal architectures among the many various Cloud Infrastructures. Several optimization problems, including capacity planning, management of resources of Cloud Infrastructures can be solved using stochastic approach and SM models described in this paper.

References

1. Khazaei, H., Misic, J., Misic, V.B., Mohammadi, N.B.: Availability analysis of cloud computing centers. In: Globecom 2012 – Communications Software, Services and Multimedia Symposium (GC12 CSSM) (2012)
2. Longo, F., Ghosh, R., Naik, V.K., Trivedi, K.S.: A scalable availability model for infrastructure-as-a-service cloud. In: Proc. Int Conf on Dependable Systems and Networks, pp. 335--346 (2011)
3. Trivedi, K.S., and Sahner, R.: SHARPE at the age of twenty two. ACM Sigmetrics Performance Evaluation Review, 36(4), 52--57 (2009)
4. Ghosh, R., Trivedi, K.S., Naik, V.K., Kim, D.S.: End-to-end performability analysis for infrastructure-as-a-service cloud: An Interacting Stochastic Models Approach. In: IEEE PRDC, Tokyo (2010)

5. Khazaei, H: Performance modeling of cloud computing centers. Diss. The University of Manitoba (2013)
6. Ghosh, R., Longo, F., Xia, R., Naik, V., Trivedi, K.: Stochastic model driven capacity planning for an infrastructure-as-a-service cloud. IEEE Transactions on Services Computing, 7(4), 667--680 (2014)
7. Ghosh, R.: Scalable stochastic models for cloud services. Diss. Duke of University (2012)
8. Ivanchenko, O., Kharchenko, V., Skatkov, A.: Management of critical infrastructures Based on Technical Megastate. Int. J. "Information and Security", 28(1), 37--51 (2012)
9. Ghosh, R., Longo, F., Frattini, F., Russo, S., Trivedi, K.S.: Scalable analytics for IaaS cloud availability. IEEE Transactions on Cloud Computing, 2(1), 57--70 (2014)
10. Ivanchenko, O., Lovyagin, V., Maschenko, E., Skatkov, A., Shevchenko, V.: Distributed critical systems and infrastructures. National Aerospace University named after N. Zhukovsky "KhAI", Kharkiv, Ukraine (2013)
11. Abbadi, I.M.: Toward Trustworthy Clouds' Internet Scale Critical Infrastructure. In: Bao, F., Weng J. (eds.) Information Security Practice and Experience. LNCS-6672, pp. 71–82. Springer Verlag, Heidelberg (2011)