

УДК 62-192:519.2

О. В. ИВАНЧЕНКО¹, В. С. ХАРЧЕНКО²¹ *Университет таможенного дела и финансов, Украина*² *Национальный аэрокосмический университет им. Н. Е. Жуковского «ХАИ», Украина*

ПОЛУМАРКОВСКАЯ МОДЕЛЬ ГОТОВНОСТИ ОБЛАЧНОЙ ИНФРАСТРУКТУРЫ С ИНТЕГРИРОВАННОЙ СИСТЕМОЙ МОНИТОРИНГА ТЕХНИЧЕСКОГО СОСТОЯНИЯ

Для оценки функциональной готовности облачной инфраструктуры (ОИ) предложен вариант аналитико-стохастического моделирования процессов изменения технического состояния физических машин, которые объединяются в соответствующие подсистемы (пулы). В качестве базовой рассматривается полумарковская модель, которая составляет ядро модели верхнего уровня иерархии. Анализ и оценка технического состояния физических машин осуществляется с помощью встроенной системы мониторинга, работающей в режимах контроля и диагностирования технического состояния. В дальнейшем результаты моделирования могут быть использованы для определения оптимальной конфигурации ОИ.

Ключевые слова: облачная инфраструктура, полумарковская модель с вырожденными состояниями, контроль технического состояния, физические машины.

Постановка проблемы

Одним из перспективных направлений развития современной IT-индустрии является трансформация информационных технологий в сферу облачных вычислений. Для решения корпоративных задач различные облачные образования могут быть представлены в виде соответствующей инфраструктуры.

На сегодняшний день наиболее разветвленной и мощной облачной инфраструктурой (ОИ) обладают всемирно известные компании Google [1], IBM [2], Rackspace [3], Salesforce [4], Amazon [5]. Например, в ОИ Amazon EC2 вычислительный ресурс предоставляется по требованию пользователя, реализуемому через операционную систему, в виде определенного количества виртуальных машин, которые разворачиваются на базе облачных провайдерских информационных центров. Соответствующий сервис предоставляется пользователю в рамках действующего соглашения (SLA), в котором помимо требований к качеству услуг должны оговариваться требования к надежности, доступности и готовности ОИ. Известно [6], что из-за проблем с готовностью облачная инфраструктура Amazon за период времени чуть более одного года простаивает четыре часа тридцать восемь минут. Следовательно, поддержание и обеспечение требуемого уровня функциональной готовности ОИ, численное значение которого задается в метриках QoS, является актуальной научно-технической задачей. Для ее успешного решения предлагается использовать интегрированную систе-

му мониторинга (ИСМ) технического состояния (ТС) ОИ, которая может функционировать как в режимах контроля ТС, так и диагностирования. Оценка эффективности предлагаемой системы мониторинга может осуществляться на основе результатов моделирования.

Целью работы является разработка аналитико-стохастической модели готовности облачной инфраструктуры с интегрированной системой мониторинга и диагностирования технического состояния физических машин, входящих в состав ОИ.

Изложение основного материала

В качестве базовой рассмотрим аналитико-стохастическую модель готовности ОИ, граф состояний которой представлен на рис. 1. На графе (рис. 1) изображены ТС девяти физических машин (ФМ), которые входят в состав ОИ с пульной организацией функционирования [6,7]. Не привязываясь к конкретной архитектурной реализации, охарактеризуем процесс изменения ТС физических машин из состава рассматриваемой инфраструктуры.

Множество возможных состояний физических машин, которые объединены в три пула, записываются в виде $\{E\} = \{E_{i,j,k}\}$, где i, j, k – число работоспособных (РС) ФМ в горячем, теплом и холодном пулах, соответственно. Количество состояний представленной и аналогичных ей моделей с n -м числом РС ФМ определяется по формуле $(n+1)^2 + n$.

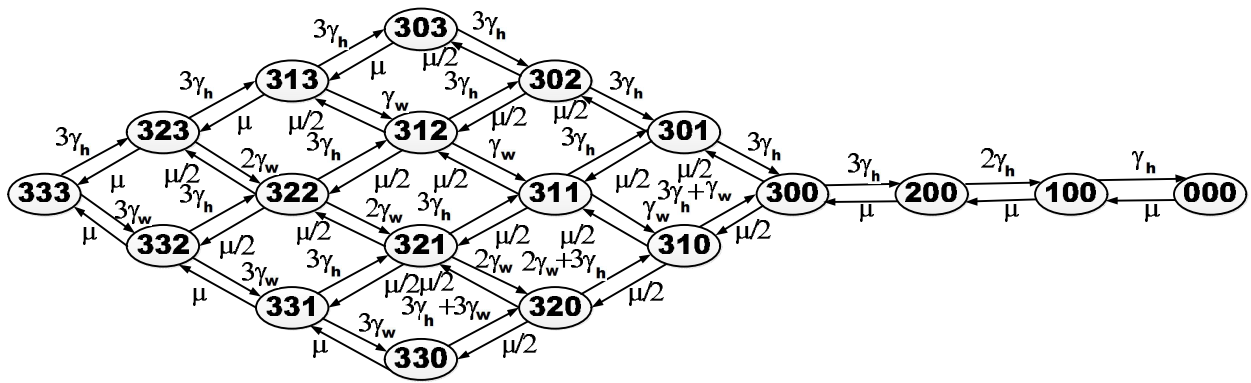


Рис. 1. Граф состояний аналитико-стохастической модели ОИ с тремя пулами физических машин ($n_h = n_w = n_c = 3$)

Для рассматриваемой модели (рис. 1) приняты следующие допущения [6]: 1) в начальный момент времени $t = 0$ работоспособны все девять ФМ; 2) с помощью ИСМ, функционирующей в режиме контроля и диагностирования ТС, в установленные моменты времени достоверно определяется техническое состояние ФМ горячего и теплого пулов; 3) ФМ холодного пула рассматриваются как абсолютно надежные и в начальный момент времени находятся в выключенном состоянии; 4) ФМ теплого пула включены и находятся в режиме ожидания применения по назначению; 5) при отказе ФМ горячего пула физические машины теплого и холодного пулов с установленной интенсивностью (т.е. с интенсивностью отказов машин горячего пула) заменяют отказавшие машины; 6) в случае отказа ФМ теплого пула они заменяются (с интенсивностью отказов машин теплого пула) на РС физические машины холодного пула; 7) ФМ теплого и холодного пулов после восстановления РС отказавших физических машин горячего (теплого) пула возвращаются (с интенсивностью восстановления машин горячего и теплого пулов) в исходные состояния.

Согласно представленного графа (рис. 1) замена отказавших ФМ горячего пула реализуется следующим образом: 1) переходы $(333) \rightarrow (323)$, $(323) \rightarrow (313)$, $(313) \rightarrow (303)$, $(332) \rightarrow (322)$, $(322) \rightarrow (312)$, $(312) \rightarrow (302)$, $(331) \rightarrow (321)$, $(321) \rightarrow (311)$, $(311) \rightarrow (301)$, $(330) \rightarrow (320)$, $(320) \rightarrow (310)$, $(310) \rightarrow (300)$ отображают миграцию ФМ теплого пула в горячий пул; 2) переходы $(303) \rightarrow (302)$, $(302) \rightarrow (301)$, $(301) \rightarrow (300)$ отображают миграцию ФМ холодного пула в горячий пул. Одновременно переходы $(333) \rightarrow (332)$, $(332) \rightarrow (331)$, $(331) \rightarrow (330)$, $(323) \rightarrow (322)$, $(322) \rightarrow (321)$, $(321) \rightarrow (320)$, $(313) \rightarrow (312)$, $(312) \rightarrow (311)$, $(311) \rightarrow (310)$ отображают

миграцию ФМ холодного пула в теплый пул при отказе машин теплого пула. В состоянии (300) остаются РС только ФМ горячего пула, отказ которых приводит к потере РС инфраструктуры в целом. С учетом описанного процесса изменения ТС коэффициент готовности (КГ) облачной инфраструктуры можно определить с помощью соотношения

$$K_{\Gamma}(t) = 1 - P_{0,0,0}(t), \quad (1)$$

где $P_{0,0,0}(t)$ – вероятность отказа всех ФМ.

Получить строгие аналитические выражения для указанной модели (рис. 1) достаточно сложно. Поэтому с целью упрощения было выполнено разложение узлов модели на дополнительные состояния, процесс изменения которых был рассмотрен как вырожденный полумарковский. После чего с учетом физики альтернирующего потока отказов-восстановлений ФМ и особенностей функционирования ИСМ была построена ключевая полумарковская модель (ПММ) нижнего уровня иерархии, являющаяся одновременно ядром рассматриваемого процесса аналитико-стохастического моделирования. Граф состояний для этой модели представлен на рис. 2.

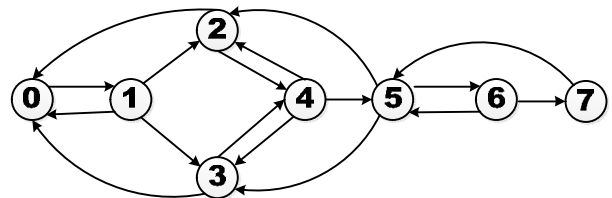


Рис. 2. Граф состояний ключевой ПММ

Ключевая ПММ (рис. 2) содержит по одной ФМ в каждом пуле, т.е. $n_h = n_w = n_c = 1$. Начальные состояния соответствуют $P_0(0) = 1, P_1(0) = 0$,

где $i = \overline{1,7}$. В соответствии с приведенным графом (рис. 2) будем полагать, что в процессе функционирования на протяжении интервала эксплуатации T физические машины могут находиться в следующих состояниях: 1) E_0 – РС все три ФМ, т.е. инфраструктура находится в состоянии готовности к использованию по назначению; 2) E_1, E_4, E_6 – состояния контроля ТС ФМ, который проводится с использованием ИСМ на протяжении детерминированного интервала времени, т.е. $t \in [0, \tau_c]$; 3) E_2 – состояние отказа «горячей» ФМ и замена ее на РС «теплую» ФМ с интенсивностью γ ; 4) E_3 – состояние отказа «теплой» ФМ и замена ее на РС «холодную» ФМ с интенсивностью γ ; 5) E_5 – состояние отказа «горячей» ФМ и замена ее на РС «холодную» или «теплую» ФМ с интенсивностью γ ; 6) E_7 – состояние отказа «горячей» ФМ, которое идентифицируется как потеря РС всей инфраструктуры. Восстановления РС отказавших ФМ (или возвращение мигрировавших ФМ в исходное состояние) выполняются с интенсивностями μ (переходы $2 \rightarrow 0, 3 \rightarrow 0$ и $7 \rightarrow 5$) и $\mu/2$ (переходы $5 \rightarrow 2$ и $5 \rightarrow 3$). Рассмотрим каким образом формируются составляющие матрицы $Q = |Q_{ij}(t)|$ независимых функций распределения времени пребывания ФМ в i -м состоянии перед переходом в j -е состояние, если бы данный выход был единственным.

Поскольку контроли ТС ФМ проводятся с установленной детерминированной периодичностью, то переходы $0 \rightarrow 1, 2 \rightarrow 4, 3 \rightarrow 4$ и $5 \rightarrow 6$ описываются

$$Q_{01}(t) = Q_{24}(t) = Q_{34}(t) = Q_{56}(t) = \begin{cases} 0, & t < T, \\ 1, & t \geq T. \end{cases} \quad (2)$$

Переходы $1 \rightarrow 0, 4 \rightarrow 2, 4 \rightarrow 3$ и $6 \rightarrow 5$, описывающие контроль ТС (КТС) соответствующих ФМ, происходят через неслучайное время τ_c , равное продолжительности КТС. Поэтому

$$Q_{10}(t) = Q_{42}(t) = Q_{43}(t) = Q_{65}(t) = \begin{cases} 0, & t < \tau_c, \\ 1, & t \geq \tau_c. \end{cases} \quad (3)$$

Переходы $1 \rightarrow 2, 1 \rightarrow 3, 4 \rightarrow 5$ и $6 \rightarrow 7$ из-за возникших в случайное время внезапных отказов ФМ горячего пула характеризуются вероятностями

$$Q_{12}(t) = Q_{13}(t) = Q_{45}(t) = Q_{67}(t) = 1 - e^{-\gamma t}. \quad (4)$$

Окончание восстановления и переходы $2 \rightarrow 0, 3 \rightarrow 0, 7 \rightarrow 5, 5 \rightarrow 2, 5 \rightarrow 3$ зависят от случайной продолжительности работы ремонтного подразделения пула, распределенной по закону Эрланга второго порядка, для которого

$$Q_{20}(t) = Q_{30}(t) = Q_{75}(t) = 1 - (1 + \mu_1 t) e^{-\mu_1 t}, \quad (5)$$

$$Q_{52}(t) = Q_{53}(t) = 1 - (1 + \mu_1 t) e^{-\mu_1 t}, \quad (6)$$

где $\mu_1 = \mu, \mu_2 = \mu/2$.

Выполнив нетривиальные расчеты, получим соотношения для расчета стационарного КГ $K_\Gamma(t)$ и соответствующих вероятностей $\pi_i(t)$, где $i = \overline{0,7}$, в виде

$$K_\Gamma(t) = \pi_0(t) + \pi_2(t) + \pi_3(t) + \pi_5(t), \quad (7)$$

$$\pi_0(t) = P_0 \frac{\bar{t}_0}{U}, \quad \pi_2(t) = P_2 \frac{\bar{t}_2}{U}, \quad \pi_3(t) = P_3 \frac{\bar{t}_3}{U}, \quad (8)$$

$$\pi_5(t) = P_5 \frac{\bar{t}_5}{U}, \quad U = \sum_{i=0}^7 P_i \bar{t}_i, \quad (9)$$

$$P_0 = P_1 = \frac{\alpha\beta}{2\alpha\beta + \xi_1 + \xi_2 + \xi_3[\beta + \eta]}, \quad (10)$$

$$P_2 = P_0 \frac{\xi_1}{\alpha\beta}, \quad P_3 = P_0 \frac{\xi_2}{\alpha\beta}, \quad P_4 = P_0 \frac{\xi_3}{\alpha\beta}, \quad (11)$$

$$P_5 = P_0 \frac{P_{20}\xi_3}{\alpha\beta}, \quad P_6 = P_0 \frac{P_{45}P_{56}\xi_3}{\alpha\beta}, \quad (12)$$

$$P_7 = 1 - \sum_{i=0}^6 P_i, \quad (13)$$

где $\beta = 1 - p_{56}(p_{65} + p_{67})$,

$$\alpha = \beta(1 - p_{24}p_{42}) - p_{34}(p_{43} + p_{45}p_{53}) - p_{24}p_{45}p_{52},$$

$$\xi_3 = p_{13}p_{34} + p_{12}p_{24}\beta,$$

$$\xi_2 = p_{45}p_{53}\xi_3 + \beta(\alpha p_{13} + p_{43}\xi_3),$$

$$\xi_1 = \beta(\alpha p_{12} + p_{42}\xi_3) + p_{45}p_{52}\xi_3,$$

$$\eta = p_{20}(1 + p_{56}(1 + p_{67})),$$

$$p_{10} = e^{-2\gamma\tau_c}, \quad p_{42} = p_{43} = p_{65} = e^{-\lambda\tau_c},$$

$$p_{12} = \frac{1}{2}(1 - e^{-2\gamma T}), \quad p_{13} = 1 - p_{10} - p_{12},$$

$$p_{24} = p_{34} = (1 + \mu_1 T) e^{-\mu_1 T},$$

$$p_{20} = p_{30} = 1 - p_{24}, \quad p_{45} = 1 - p_{42} - p_{43},$$

$$p_{52} = \frac{1}{2} \left[1 - (1 + \mu_2^2 T^2 + 2\mu_2 T) e^{-2\mu_2 T} \right],$$

$$p_{56} = \left[(1 + \mu_2 T) e^{-\mu_2 T} \right]^2,$$

$$p_{67} = 1 - p_{65}, \quad p_{53} = 1 - p_{52} - p_{56},$$

$$\bar{t}_0 = T, \quad \bar{t}_1 = \frac{1}{2\gamma}(1 - e^{-2\gamma\tau_c}), \quad \bar{t}_2 = \bar{t}_3 = \bar{t}_7 = \frac{2}{\mu_1},$$

$$\bar{t}_4 = \bar{t}_6 = \frac{1}{\gamma}(1 - e^{-\gamma\tau_c}),$$

$$\bar{t}_{51} = 1 - e^{-2\mu_2 T}, \quad \bar{t}_{52} = 1 - (1 + 2\mu_2 T) e^{-2\mu_2 T},$$

$$t_{53} = \frac{1 - [1 + 2\mu_2 T(1 + T)]e^{-2\mu_2 T}}{2},$$

$$t_5 = \frac{1}{2\mu_2} \left(t_{51} + t_{52} + t_{53} \right).$$

На рис. 3 представлен график зависимости $K_T(T)$, полученный с использованием соотношений (2) – (13).

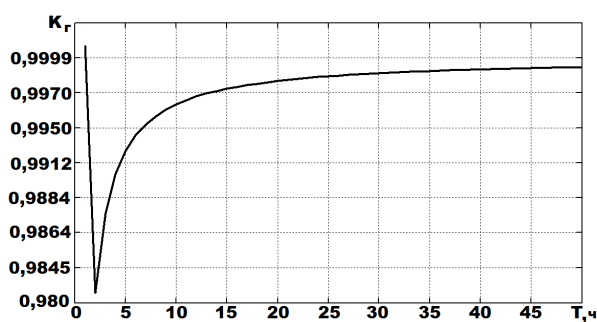


Рис. 3. Зависимость $K_T(T)$ для $T = 500$ ч,
 $\gamma = 0,002$ 1/ч, $\mu = 5$ 1/ч

Результаты моделирования (рис. 3) свидетельствуют о довольно высоком уровне функциональной готовности ОИ в условиях протекания внезапных отказов, что обеспечивается комплексной реализацией мероприятий по восстановлению работоспособности, рациональному распределению и резервированию ресурсного потенциала компонентных составляющих инфраструктуры. Установлено, что наибольшее влияние на изменение уровня надежности облачной инфраструктуры оказывает продолжительность ее использования по назначению.

Заключение

Анализ полученных результатов свидетельствует, что предложенная авторами аналитико-стохастическая модель (рис. 1), строится с использованием ядра ключевой полумарковской модели, сформированного состояниями E_0, \dots, E_7 (рис. 2). Узлы модели верхнего уровня иерархии могут быть впоследствии размножены в виде соответствующих звеньев ключевой модели. С увеличением в многофункциональных пулах облачной инфраструктуры числа физических машин пропорционально возрастает число звеньев, что позволяет адаптировать предлагаемую модель к изменению количественного состава элементов рассматриваемой ОИ.

Перспективы дальнейших исследований связаны с разработкой модельного ряда, учитывающего масштабируемость многофункционального взаимо-

действия резервируемых пулов; требования обеспечения гибкости управления компонентными составляющими инфраструктуры; с определением оптимальной конфигурации физических машин с целью минимизации стоимости облачного инфраструктурного хостинга; с целенаправленным использованием разрабатываемых полумарковских моделей для оценки эффективности предлагаемых облачных архитектурных решений.

Литература

1. *Google Apps for Work* [Электронный ресурс] / Google. – Режим доступа: <https://apps.google.com/index.html>. – 28.03.2016.
2. *IBM Global Technology Services* [Электронный ресурс] / IBM. – Режим доступа: <http://www-935.ibm.com/services/us/en/it-services/gts-it-service-home-page-1.html>. – 28.03. 2016.
3. *Rackspace Public Cloud* [Электронный ресурс] / Rackspace. – Режим доступа: <http://www.rackspace.com/cloud>. – 28.03. 2016.
4. *Salesforce. The fastest way to build apps for customers, employees, and partners* [Электронный ресурс] / Salesforce. – Режим доступа: <http://www.salesforce.com/servicecloud/overview/>. – 28.03. 2016.
5. *Amazon EC2 – Virtual Service Hosting* [Электронный ресурс] / Amazon EC2. – Режим доступа: http://aws.amazon.com/ec2/?nc1=h_ls. – 8.03.2016.
6. Ghosh, R. *Scalable Stochastic Models for Cloud Services* [Электронный ресурс] / R. Ghosh. – Режим доступа: http://dukespace.lib.duke.edu/dspace/bitstream/handle/10161/6110/Ghosh_duke_0066D_11619.pdf?sequence=1. – 29.03. 2016.
7. Khazaei, H. *Performance modeling of cloud computing centers* [Электронный ресурс] / H. Khazaei. – Режим доступа: http://mpace.lib.umanitoba.ca/xmlui/bitstream/handle/1993/16682/khazaei_hamzeh.pdf?sequence=1. – 29.03.2016.

References

1. *Google Apps for Work*, Google. Available at: <https://apps.google.com/index.html> (accessed 28.03. 2016).
2. *IBM Global Technology Services*, IBM. Available at: <http://www-935.ibm.com/services/us/en/it-services/gts-it-service-home-page-1.html> (accessed 28.03.2016).
3. *Rackspace Public Cloud*, Rackspace. Available at: <http://www.rackspace.com/cloud> (accessed 28.03. 2016).
4. *Salesforce. The fastest way to build apps for customers, employees, and partners*, Salesforce. Available at: <http://www.salesforce.com/servicecloud/overview/> (accessed 28.03. 2016).

5. Amazon EC2 – Virtual Service Hosting, *Amazon EC2*. Available at: http://aws.amazon.com/ec2/?nc1=h_ls (accessed 28.03.2016).

6. Ghosh, R. *Scalable Stochastic Models for Cloud Services*. Available at: <http://dukespace.lib.duke.edu/dspace/bitstream/handle/10161/6110/Ghos>

h_duke_0066D_11619.pdf?sequence=1 (accessed 29.03.2016).

7. Khazaei, H. *Performance modeling of cloud computing centers*. Available at: http://mspace.lib.umanitoba.ca/xmlui/bitstream/handle/1993/16682/khazaei_hamzeh.pdf?sequence=1 (accessed 29.03.2016).

Надійшла до редакції 29.03.2016, розглянута на редколегії 14.04.2016

НАПІВМАРКІВСЬКА МОДЕЛЬ ГОТОВНОСТІ ХМАРНОЇ ІНФРАСТРУКТУРИ З ІНТЕГРОВАНОЮ СИСТЕМОЮ МОНІТОРИНГУ ТЕХНІЧНОГО СТАНУ

О. В. Іванченко, В. С. Харченко

Для оцінки функціональної готовності хмарної інфраструктури (ХІ) запропоновано варіант аналітико-стохастичного моделювання процесів зміни технічного стану фізичних машин, які об'єднуються у відповідні підсистеми (пули). У якості базової розглядається напівмарківська модель, яка складає ядро моделі верхнього рівня ієрархії. Аналіз та оцінка технічного стану фізичних машин здійснюється за допомогою вбудованої системи моніторингу, яка працює в режимах контролю та діагностування технічного стану. В подальшому результати моделювання можуть бути використані для визначення оптимальної конфігурації ХІ.

Ключові слова: хмарна інфраструктура, напівмарківська модель з виродженими станами, контроль технічного стану, фізичні машини.

SEMI-MARKOV AVAILABILITY MODEL FOR CLOUD INFRASTRUCTURE WITH INTEGRATED TECHNICAL STATE CONTROL SYSTEM

O. V. Ivanchenko, V. S. Kharchenko

Option analytical and stochastic modeling of change of the technical state for determine physical machines availability which are organized in pools are proposed. Semi-Markov model as main model, which is kernel of high level hierarchy model is shown. Analyze and estimation of physical machines technical state can implement with help of integrated monitoring system, which are worked at control and diagnostic regimes of technical state. Further results of modeling could be to use in order to determine of the optimal cloud infrastructure configuration.

Key words: cloud infrastructure, Semi-Markov model with special states, control of technical state, physical machines.

Іванченко Олег Васильевич – канд. техн. наук, доцент, доцент кафедри інформаційних систем и технологій, Университет таможенного дела и финансов, Днепропетровск, Украина, e-mail: vmsu12@gmail.com.

Харченко Вячеслав Сергеевич – д-р техн. наук, профессор, зав. каф. компьютерных систем и сетей, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков, Украина, e-mail: v_s_kharchenko@ukr.net.

Ivanchenko Oleg Vasilevich – Candidate of Technical Science, Associate Professor of Department of Information Systems and Technologies, University of Customs and Finance, Dnipropetrovsk, Ukraine, e-mail: vmsu12@gmail.com.

Kharchenko Vyacheslav Sergeevich – Dr. Sc. in Engineering, Prof., Head of Department of Computer Systems and Networks, National Aerospace University “Kharkiv Aviation Institute”, Kharkiv, Ukraine, e-mail: v_s_kharchenko@ukr.net.