

**О. В. Иванченко**, кандидат технических наук, доцент кафедры информационных систем и технологий Академии таможенной службы Украины

**К. В. Смоктий**, кандидат экономических наук, доцент кафедры прикладной математики и теории систем управления Донецкого национального университета

**В. С. Харченко**, доктор технических наук, заведующий кафедрой компьютерных систем и сетей Национального аэрокосмического университета им. Н. Е. Жуковского "ХАИ"

**И. В. Блындюк**, студент Академии таможенной службы Украины

### СТОХАСТИЧЕСКАЯ МОДЕЛЬ НАДЕЖНОСТИ ИНФРАСТРУКТУРЫ КАК СЕРВИСА ОБЛАЧНЫХ ВЫЧИСЛЕНИЙ

*Современные информационные центры для развития своей деятельности стараются трансформировать традиционно предоставляемые сервисы в сферу облачных вычислений. Для достижения этой цели большая часть облачных провайдерских информационных центров увеличивает количество физических машин, что приводит к резкому снижению сервисного времени простоя. Однако с ростом числа физических машин усложняется процедура оценки уровня надежности инфраструктуры как сервиса облачных вычислений. Предлагается решить эту проблему путем построения стохастической полумарковской модели, которая, в отличие от известных, учитывает предысторию и многофункциональный характер построения облачной инфраструктуры.*

*Ключевые слова: инфраструктура как сервис облачных вычислений; таксономия поддержания работоспособного состояния облачной инфраструктуры; полумарковская модель надежности инфраструктуры с вырожденными состояниями.*

*Modern provider's data centers for development of its activity try to transform the traditional IT management into Cloud. Therefore most of all cloud providers' data centers increase number of physical machines. It is leading to lower cost of service downtime. However with a larger number of physical machines a normal procedure of determining the reliability level of an infrastructure as a service cloud tends to complicate. We propose to solve this problem by constructing a stochastic Semi-Markov model, that in contrast to the known, takes into account the background and multifunctional character build cloud infrastructure.*

*Key words: infrastructure as a service cloud; taxonomy maintain operability state of the cloud infrastructure; Semi-Markov reliability model of the infrastructure with special states.*

**Постановка проблемы.** На сегодняшний день одним из перспективных направлений практического приложения информационных технологий (ИТ) является сфера облачных вычислений. Применение облачных вычислений позволяет существенно снизить стоимость ИТ-услуг, предоставляемых в корпоративной среде [1].

© О. В. Иванченко, К. В. Смоктий, В. С. Харченко, И. В. Блындюк, 2014

Известно, что фактическая конфигурация инфраструктуры как сервиса облачных вычислений (IaaS Cloud) предусматривает использование объединенного ресурса совокупности центральных процессоров (CPU), оперативной памяти (RAM), сетевого оборудования и устройств хранения с расширенным дисковым пространством. Поэтому разработчики IaaS Cloud стараются развернуть сеть из физических машин (ФМ) с оптимально возможной конфигурацией для организации работы как можно большего количества виртуальных машин (ВМ), без использования которых нормальное функционирование облачной инфраструктуры невозможно.

Рост числа ВМ сопровождается увеличением количества используемых ФМ, что способствует улучшению возможностей по предоставлению облачного сервиса. Это в значительной степени достигается за счет снижения сервисного времени простоя. Однако увеличение числа ФМ порождает проблемы, связанные с обеспечением надежности IaaS Cloud; приводит к росту энергетических затрат и стоимости оборудования для охлаждения физических машин [2].

**Анализ последних исследований и публикаций.** Наиболее распространенными среди IaaS Cloud являются инфраструктуры, построенные на основе использования сервисов Amazon EC2 [3] и IBM Cloud [4]. Эти сервисы предоставляют пользователям резервируемые вычислительные ресурсы, доступные через разветвленную сеть облачных провайдерских центров (Cloud ЦЦ) [5]. Причем вычислительные ресурсы предоставляются Cloud ЦЦ в дистанционном режиме (т. е. в аутсорсинговом) через Internet путем создания определенного количества виртуальных машин. Такой сервис, в отличие от базового web-сервиса, анализ которого выполнен в [6; 7], обеспечивает более широкий спектр услуг с высоким уровнем автоматизации реализуемых процессов. Например [8], если несколько лет назад какая-либо компания с целью обеспечения эффективной работы своего бизнеса хотела создать свое офисное бизнес-приложение (web-сайт), то она была вынуждена закупать дорогостоящие серверы и другое оборудование локального контроля. В настоящее время с появлением облачной инфраструктуры подобная задача решается значительно проще, а именно: компания обращается к представителю IaaS Cloud о предоставлении соответствующего оборудования и программного обеспечения (ПО); после этого компания – владелец IaaS Cloud за определенную арендную плату через облачный провайдерский центр обеспечивает заказчика серверами, устройствами хранения и различным сетевым оборудованием. Обмен информацией осуществляется через Internet в соответствии со схемой, представленной на рис. 1.

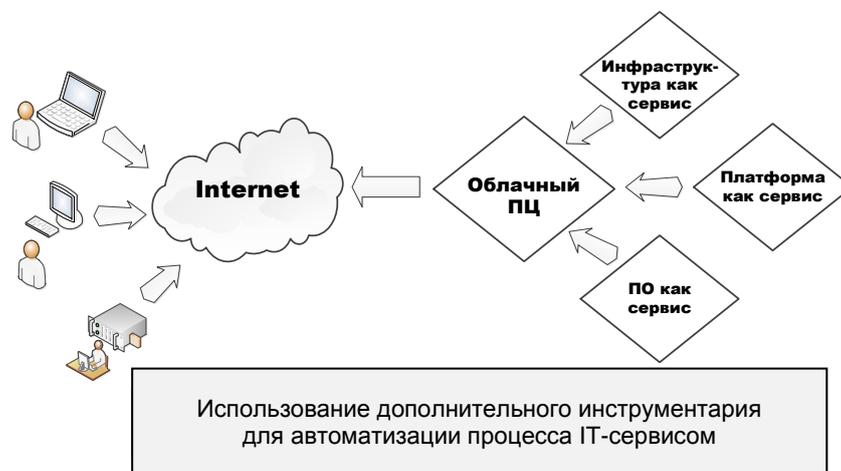


Рис. 1. Схема предоставления IT-сервиса в режиме аутсорсингового обеспечения

С точки зрения пользователя, наличие или отсутствие сервиса, а также его быстродействие являются важнейшими показателями качества обслуживания. Поэтому для количественной оценки качества обслуживания и уровня работоспособности сервиса используются метрики QoS (quality-of-service). Показатели работоспособности сервиса определяют с учетом как альтернирующего потока отказов-восстановлений, так и очереди на предоставление соответствующего ресурса (ресурсного резервирования) [9; 10]. Такая модель учитывает возможности по восстановлению вычислительного ресурса, формированию очереди на ресурсное резервирование и относится к классическим моделям теории массового обслуживания [11]. Однако наряду с достоинствами существенным недостатком модели QoS является необходимость проведения большого числа экспериментов в нагруженном состоянии для оценки надежности и анализа последствий ресурсных отказов соответствующего облачного сервиса. В качестве альтернативы указанной модели предлагается использовать стохастическую модель, которая, несмотря на рост масштабов и сложности архитектуры Cloud ПЦ, имеет достаточно низкую стоимость реализации для относительно большого числа оцениваемых параметров. На рис. 2 представлена диаграмма пошагового стохастического моделирования процессов обслуживания заявок и ресурсного резервирования.



Рис. 2. Диаграмма пошагового моделирования процессов обслуживания заявок и ресурсного резервирования

Согласно представленной диаграмме (рис. 2) моделирование осуществляется в два этапа [12]: на первом этапе строятся модели обслуживания заявок для различных вариантов резервирования облачного сервиса; на втором – определяются значения соответствующих показателей эффективности для всех возможных вариантов обслуживания заявок и различных вариантов резервирования. Существенным недостатком такой модели является относительно большое количество вариантов резервирования, что затрудняет анализ полученных результатов.

Для устранения недостатков предыдущей модели в [2; 13] предложено использовать стохастический подход, основанный на моделировании взаимодействия между резервируемыми подсистемами (пулами) рассматриваемой инфраструктуры. Фундаментальной основой

разработанного подхода является представление модели в виде однородной непрерывной марковской цепи с экспоненциально распределенными временными интервалами наработок между отказами, восстановлениями и миграциями [1]. На наш взгляд, допущение об экспоненциальном распределении всех временных интервалов [13] является слишком общим и не соответствует действительности, поскольку время восстановления существенно зависит как от числа ремонтных бригад (подразделений) в каждом пуле, так и от квалификации их персонала. В этих условиях можно предположить, что распределение времени восстановления ФМ каждого пула имеет дискретный характер. Поэтому мы ввели допущение о показательном-степенном распределении интервалов восстановления, которое соответствует распределению Эрланга [14]. Дальнейшие рассуждения будут базироваться на представлении процесса изменения технических состояний ФМ как полумарковского [15–17].

**Цель статьи.** Следовательно, определение уровня надежности инфраструктуры как сервиса облачных вычислений на основе стохастического представления процессов многофункционального резервирования основных подсистем, что минимизирует стоимость инфраструктурного хостинга и поддерживает работоспособность системы в целом, является актуальной и важной задачей. Исходя из этого, необходимо построить полумарковскую модель надежности IaaS Cloud с вырожденными состояниями, учитывающую взаимодействие между резервными пулами физических машин, входящих в состав облачной инфраструктуры.

**Изложение основного материала.** С целью детализации классификационных признаков построим таксономическую схему поддержания работоспособного состояния (РС) IaaS Cloud, руководствуясь отдельными положениями и результатами исследований, изложенными в [1; 2; 5; 12; 13].

Известно [13; 18], что для снижения общего времени простоя виртуальных машин физические машины группируются в три сервисных пула, а именно: горячий пул, состоящий только из работающих ФМ; теплый пул, в состав которого входят включенные, но частично готовые к использованию ФМ; в состав холодного пула входят только выключенные ФМ. Поддержание инфраструктуры в РС обеспечивается путем выполнения следующего комплекса мероприятий:

- 1) по контролю информационно-технического состояния компонентов IaaS Cloud;
- 2) по резервированию сервиса, предоставляемого IaaS Cloud, посредством “миграции” физических машин из одного пула в другой, что позволяет существенно снизить время простоя ФМ вследствие возникновения сбоев, отказов программного обеспечения, технологического оборудования и сетевого сервиса;
- 3) по восстановлению работоспособности ФМ с использованием соответствующих ремонтного оборудования и сервиса.

На формальном уровне учет этих факторов позволяет получить теоретическое представление объекта исследований в виде

$$TM = CrEIS, CrSS, CrEIF, CrEIR, CrSF, CrSR, \quad (1)$$

где  $IaaS_{SCS} = IaaS_{SCS}_i \prod_{i=1}^N$  – множество технических состояний, в которых облачная инфраструктура (ОИ, т. е. инфраструктура как сервис облачных вычислений) выполняет заданные функции;  $PhMS = \{PhMS_j\}_{j=1}^M$  – подмножество технических состояний физических машин, влияющих на выполнение ОИ заданных функций, т. е.  $PhMS \subset IaaS_{SCS}$ ;  $IaaS_{SCF} = \{IaaS_{SCF}\}_{p=1}^Q$  – множество состояний отказов, неисправностей ОИ;  $IaaS_{SCR} = IaaS_{SCR}_d \prod_{d=1}^U$  – множество состояний,

---

в которых реализуется процесс восстановления облачной инфраструктуры, влияющих на выполнение ОИ заданных функций;  $PhMM = PhMM_{g, g=1}^H$  – множество состояний “миграций” ФМ из одного пула в другой;  $PhMSR = PhMSR_z_{z=1}^V$  – множество состояний, в которых реализуется процесс восстановления физических машин, влияющих на выполнение ОИ заданных функций.

Представленная таким образом теоретико-множественная модель в сочетании с классификационными признаками, детальным анализом внутренних, внешних объективных и субъективных факторов негативного влияния на процессы функционирования инфраструктуры является базовой основой для построения таксономической схемы (ТКС) поддержания работоспособного состояния IaaS Cloud. Кроме того, в разрабатываемой таксономии должны наглядно демонстрироваться расширенные возможности по контролю информационно-технического состояния и принятию решения на ресурсное резервирование облачной инфраструктуры, что достигается посредством применения одноименной системы. Исходя из приведенных предпосылок, была построена указанная ТКС (рис. 3).

Как следует из рис. 3, ядром ТКС является система контроля информационно-технического состояния (КИТС) и принятия решения на ресурсное резервирование облачной инфраструктуры (т. е. IaaS Cloud), которая включает:

- 1) собственно систему контроля информационно-технического состояния (СКИТС);
- 2) устройство принятия решения на ресурсное резервирование (УПРНР).

Проанализируем работу СКИТС и УПРНР с позиций оценки их возможностей в обеспечении и поддержании работоспособного состояния инфраструктуры как сервиса облачных вычислений.

В соответствии с представленной таксономией (рис. 3) КИТС функционирует в течение детерминированного интервала времени  $\tau_{КИТС}$ , в ходе которого определяется информационно-техническое состояние  $n_h$  физических машин горячего пула (или горячих ФМ). При обнаружении отказа одной из горячих ФМ она отправляется в ремонт, после чего КИТС с интенсивностью  $\rho_{bhw}$  (или  $\rho_{bhc}$ ) подключает теплый (или холодный) пул для дальнейшей реализации процесса “миграции” предоставляемого сервиса, что позволяет обеспечить резервирование отказавшей физической машины с минимальным временем задержки. Теплый и холодный пулы содержат  $n_w, n_c$  физических машин соответственно. Если по результатам КИТС установлено, что все  $n_h$  горячих ФМ отказали, а теплый, холодный пулы не имеют в наличии свободных физических машин, то инфраструктура признается неработоспособной и выполняется полное восстановление ее РС. Одновременно можно предположить, что если для  $i$ -й ФМ выполняется условие  $1 \leq i \leq n_h$ , то производится частичное восстановление работоспособности IaaS Cloud. Далее в работу вступает УПРНР, с помощью которого определяется свободный ресурс (т. е. свободные физические машины) в теплом или холодном пулах и осуществляется замена  $k$ -й отказавшей горячей физической машины на работоспособную ФМ теплого пула. Если в теплом пуле отсутствуют свободные физические машины, то УПРНР подключает работоспособную ФМ холодного пула, предварительно “разогревая” ее до состояния дальнейшего функционального использования. Фактически это означает, что выполняется аппаратное включение холодной ФМ и дальнейшая ее подготовка к использованию по назначению с последующей заменой  $k$ -й отказавшей горячей физической машины.

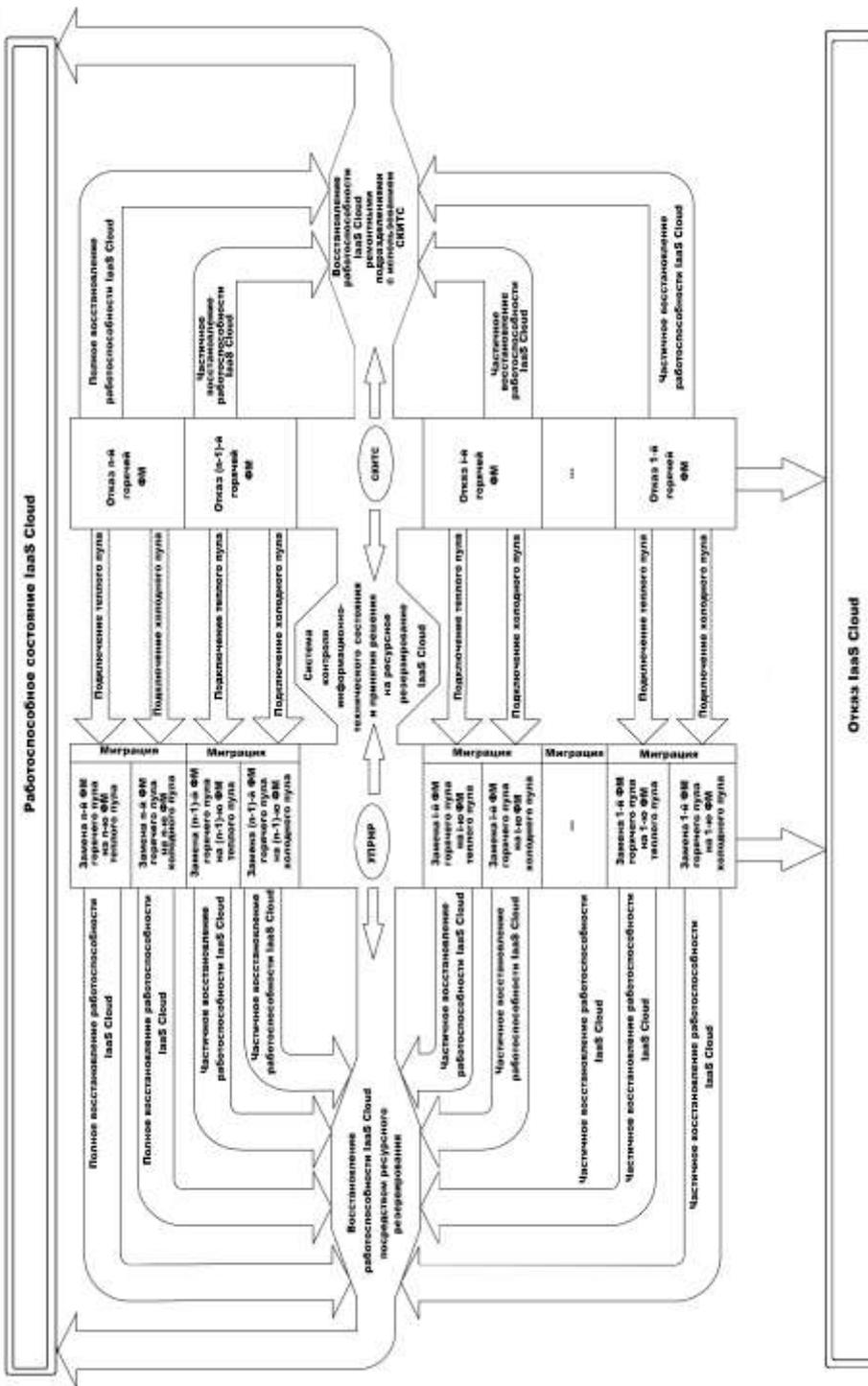


Рис. 3. Таксономия поддержания работоспособности состояния IaaS Cloud

---

Как вывод, осуществляемая таким образом совместная работа СКИТС и УПНР по своевременному обнаружению как функциональных, так и ресурсных отказов физических машин должна обеспечить поддержание требуемого уровня функциональной готовности рассматриваемой инфраструктуры. Исходя из этого, перейдем к следующему этапу построения стохастической модели надежности IaaS Cloud, которую будем использовать для получения оценок стационарного коэффициента готовности (КГ) инфраструктуры как сервиса облачных вычислений. Сфокусируем наши усилия на построении полумарковской модели надежности IaaS Cloud, учитывающей особенности протекания отказов, восстановления работоспособного состояния и ресурсного резервирования физических машин рассматриваемой инфраструктуры. Для моделирования будем использовать базовую облачную архитектуру, представленную в [1].

Анализ известных публикаций в сфере применения по назначению IaaS Cloud свидетельствует о целесообразности введения следующих ограничений и допущений:

1) под отказом физических машин будем подразумевать отказы программного обеспечения, компьютерного или сетевого оборудования [19]. В разрабатываемой модели общий эффект от возникновения указанных типов отказов учитывается с помощью показателя средней наработки между отказами [20];

2) горячий, теплый, холодный пулы содержат идентичные ФМ [1];

3) в инфраструктуре возникают внезапные отказы, которые своевременно обнаруживаются СКИТС;

4) интенсивности отказов горячих ФМ выше, чем теплых физических машин, однако интенсивность отказов холодных ФМ значительно ниже, чем теплых физических машин. Логично предположить, что средняя наработка между отказами теплых ФМ в 2–4 раза превышает аналогичный показатель для горячих физических машин [2];

5) наработки между отказами ФМ распределены по экспоненциальному закону во время, как продолжительности интервалов восстановления физических машин распределены по закону Эрланга 2–3 порядка. Более высокий порядок закона Эрланга определяется тем, что при возникновении внезапного отказа на интервале времени между ближайшими КИТС облачная инфраструктура более уязвима. Это обстоятельство в смысле ухудшения надежности выражается в увеличении среднего времени ее восстановления;

6) предполагается, что время подключения и “миграции” ФМ распределено по экспоненциальному закону; возможными являются “миграции” ФМ из теплого и холодного пулов в горячий;

7) с установленной периодичностью на рассматриваемом интервале эксплуатации  $T$  облачной инфраструктуры проводится контроль информационно-технического состояния физических машин горячего пула продолжительностью  $\tau_{h_s}$ , где  $s = \overline{1, m}$ ,  $m$  – количество работоспособных ФМ горячего пула.

В качестве примера рассмотрим стохастическую полумарковскую модель (ПММ) IaaS Cloud с многофункциональными пулами (горячим, теплым, холодным), каждый из которых содержит три ФМ, т. е.  $n_h = n_w = n_c = 3$ . При построении модели будем исходить из того, что полумарковский процесс задан графом состояний (рис. 4), т. е. состояниями  $i = \overline{0, 15}$  из множества  $E$  ( $i \in E$ ), и возможными переходами  $i \rightarrow j$ . Начальные состояния соответствуют  $P_0 \rightarrow 1, P_i \rightarrow 0$ , где  $i = \overline{1, 15}$ .

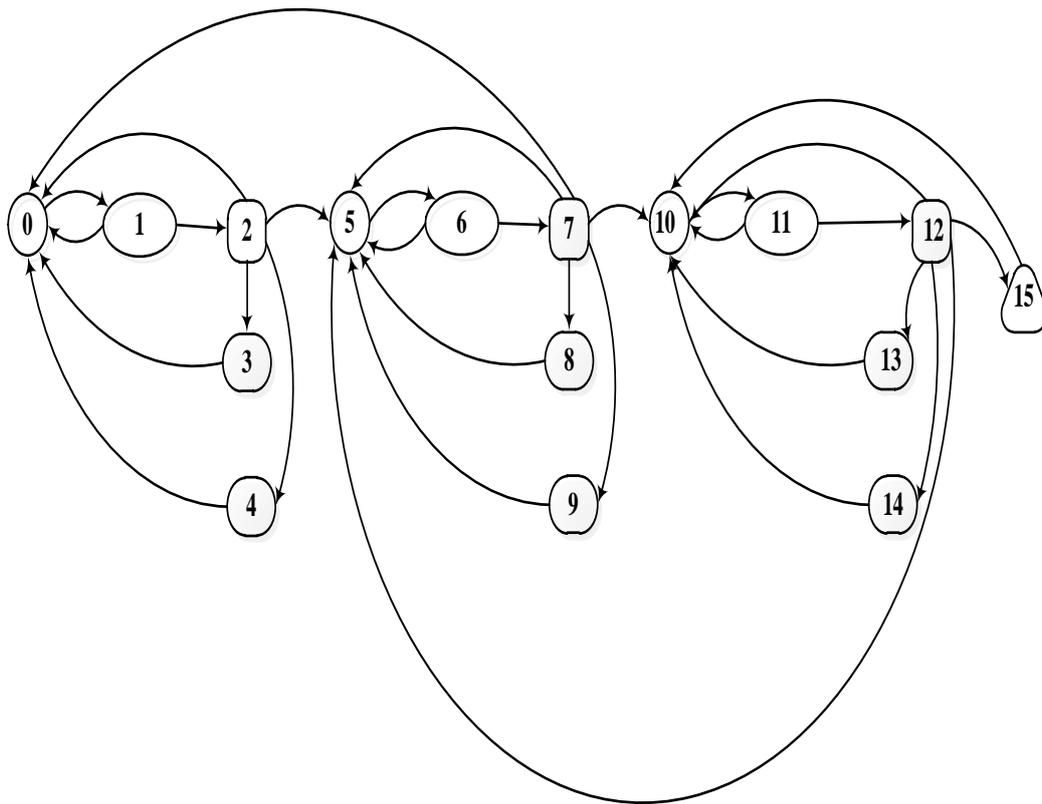


Рис. 4. Граф состояний ПММ надежности IaaS Cloud с многофункциональными пулами  
 $(n_h = n_w = n_c = 3)$

Согласно приведенному графу (рис. 4) будем полагать, что в процессе функционирования на протяжении интервала эксплуатации  $T$  инфраструктура как сервис облачных вычислений может находиться в следующих состояниях:

- 1)  $E_0$  – PC IaaS Cloud, т. е. инфраструктура находится в состоянии готовности к использованию по назначению (все  $n_h = 3$  горячие ФМ PC);
- 2)  $E_1$  – IaaS Cloud в состоянии КИТС ( $n_h = 3$ ), в ходе которого возникают внезапные отказы;
- 3)  $E_2$  – отказ третьей ФМ горячего пула (вторая и первая ФМ горячего пула PC);
- 4)  $E_3$  – состояние подключения ФМ теплого пула в случае отказа третьей физической машины горячего пула;
- 5)  $E_4$  – состояние подключения ФМ холодного пула в случае отказа третьей физической машины горячего пула и при отсутствии свободных физических машин теплого пула;

- 
- 6)  $E_5$  – частичная РС IaaS Cloud соответствует состоянию, когда отказала третья ФМ, а  $n_h = 2$  горячие ФМ РС;
  - 7)  $E_6$  – IaaS Cloud в состоянии КИТС ( $n_h = 2$ ), в ходе которого возникают внезапные и ложные отказы;
  - 8)  $E_7$  – отказ второй ФМ горячего пула (первая ФМ горячего пула РС);
  - 9)  $E_8$  – состояние подключения ФМ теплого пула в случае отказа второй физической машины горячего пула;
  - 10)  $E_9$  – состояние подключения ФМ холодного пула в случае отказа второй физической машины горячего пула и при отсутствии свободных физических машин теплого пула;
  - 11)  $E_{10}$  – частичная РС IaaS Cloud соответствует состоянию, когда отказала вторая ФМ, а  $n_h = 1$  горячая ФМ РС;
  - 12)  $E_{11}$  – IaaS Cloud в состоянии КИТС ( $n_h = 1$ ), в ходе которого возникают внезапные и ложные отказы;
  - 13)  $E_{12}$  – отказ первой ФМ горячего пула;
  - 14)  $E_{13}$  – состояние подключения ФМ теплого пула в случае отказа первой физической машины горячего пула;
  - 15)  $E_{14}$  – состояние подключения ФМ холодного пула в случае отказа первой физической машины горячего пула и при отсутствии свободных физических машин теплого пула;
  - 16)  $E_{15}$  – состояние отказа IaaS Cloud, т. е. инфраструктура не готова к использованию по назначению (все  $n_h = 3$  горячие ФМ находятся в состоянии отказа; в составе теплого и холодного пулов нет свободных физических машин).

В соответствии с принятой организацией использования IaaS Cloud по назначению возможны следующие переходы: из состояния  $0 \rightarrow 1(01)$ ; из состояния  $1 \rightarrow 0(10)$  и  $2(12)$ ; из состояния  $2 \rightarrow 0(20)$ ,  $3(23)$ ,  $4(24)$  и  $5(25)$ ; из состояния  $3 \rightarrow 0(30)$ ; из состояния  $4 \rightarrow 0(40)$ ; из состояния  $5 \rightarrow 6(56)$ ; из состояния  $6 \rightarrow 5(65)$  и  $7(67)$ ; из состояния  $7 \rightarrow 0(70)$ ,  $5(75)$ ,  $8(78)$ ,  $9(79)$  и  $10(710)$ ; из состояния  $8 \rightarrow 5(85)$ ; из состояния  $9 \rightarrow 5(95)$ ; из состояния  $10 \rightarrow 11(1011)$ ; из состояния  $11 \rightarrow 10(1110)$  и  $12(1112)$ ; из состояния  $12 \rightarrow 5(125)$ ,  $10(1210)$ ,  $13(1213)$ ,  $14(1214)$  и  $15(1215)$ ; из состояния  $13 \rightarrow 10(1310)$ ; из состояния  $14 \rightarrow 10(1410)$ ; из состояния  $15 \rightarrow 10(1510)$ , т. е. всего тридцать переходов. Следовательно, матрица  $Q = |Q_{ij}(t)|$  независимых функций распределения времени пребывания IaaS Cloud в  $i$ -м состоянии перед переходом в  $j$ -е состояние, если бы данный выход был единственным, должна включать тридцать ненулевых составляющих. В табл. 1 представлены значения входных параметров модели, соответствующие ее определенному состоянию и реализуемому переходу. Рассмотрим каким образом формируются составляющие матрицы  $Q = |Q_{ij}(t)|$ .

## Входные параметры модели

Переход $ij$	Параметр		
	Символ	Описание	Величина
01, 56, 1011	$t$	Период проведения КИТС ФМ ( $n_h = 3$ – переход 01; $n_h = 2$ – переход 56; $n_h = 1$ – переход 1011)	Задается из условия $t \leq T$ (как правило, от одного до нескольких часов)
10	$\tau_{h_3}$	Продолжительность КИТС ФМ ( $n_h = 3$ )	30 сек.
65	$\tau_{h_2}$	Продолжительность КИТС ФМ ( $n_h = 2$ )	30 сек.
1110	$\tau_{h_1}$	Продолжительность КИТС ФМ ( $n_h = 1$ )	30 сек.
12, 67, 1112	$\lambda_0$	Интенсивность отказов ФМ горячего пула (базовое значение задается как $\lambda_0 = \lambda_h$ )	В диапазоне от 0,00 085 до 0,0005 1/ч
20, 70, 75, 125, 1210, 1510	$\mu$	Интенсивность восстановления физической машины горячего пула	В диапазоне от 0,5 до 0,75 1/ч
23, 78, 1213	$\rho_1$	Интенсивность подключения ФМ теплого пула ( $n_w = 3$ – переход 23; $n_w = 2$ – переход 78; $n_w = 1$ – переход 1213)	В диапазоне от 0,00 017 до 0,0001 1/ч
24, 79, 1214	$\rho_2$	Интенсивность подключения ФМ холодного пула ( $n_c = 3$ – переход 24; $n_c = 2$ – переход 79; $n_c = 1$ – переход 1214)	В диапазоне от 0,000 085 до 0,00 005 1/ч
30, 85, 1310	$\gamma_1$	Интенсивность “миграций” ФМ из теплого пула в горячий пул ( $n_w = 3$ – переход 30; $n_w = 2$ – переход 85; $n_w = 1$ – переход 1310)	В диапазоне от 0,000 425 до 0,00 025 1/ч
40, 95, 1410	$\gamma_2$	Интенсивность “миграций” ФМ из холодного пула в горячий пул ( $n_c = 3$ – переход 40; $n_c = 2$ – переход 95; $n_c = 1$ – переход 1410)	В диапазоне от 0,0 000 425 до 0,000 025 1/ч
25	$\lambda_1$	Интенсивность отказа облачной инфраструктуры (частичная потеря РС вследствие отказа третьей ФМ горячего пула), т. е. происходит деградация уровня надежности IaaS Cloud	Задается из условия $\lambda_1 = \lambda_h n_h$
710	$\lambda_2$	Интенсивность отказа облачной инфраструктуры (частичная потеря РС вследствие отказа второй ФМ горячего пула)	Задается из условия $\lambda_1 = \lambda_h n_h - 1$
1215	$\lambda_1$	Интенсивность полного отказа облачной инфраструктуры вследствие отказа первой ФМ горячего пула (т. е. отказ всех трех ФМ горячего пула)	Задается из условия $\lambda_1 = \lambda_h n_h - 2$

Поскольку КИТС инфраструктуры проводится с установленной детерминированной периодичностью, то переход 01 описывается

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

Переход из состояния 1 (КИТС) работоспособной IaaS Cloud в состояние 0 (РС) готовности к использованию по назначению при отсутствии отказов в ходе контроля информационно-технического состояния происходит через неслучайное время  $\tau_{h_3}$ , равное продолжительности КИТС. Поэтому

$$Q_{10} = \begin{cases} 0, t < \tau_{h_3}, \\ 1, t \geq \tau_{h_3}. \end{cases} \quad (3)$$

Аналогично описываются переходы 56 (65) и 1011 (1110), т. е.

$$Q_{56} = \begin{cases} 0, t < T, \\ 1, t \geq T, \end{cases} \quad (4)$$

$$Q_{65} = \begin{cases} 0, t < \tau_{h_2}, \\ 1, t \geq \tau_{h_2}, \end{cases} \quad (5)$$

$$Q_{1011} = \begin{cases} 0, t < T, \\ 1, t \geq T, \end{cases} \quad (6)$$

$$Q_{1110} = \begin{cases} 0, t < \tau_{h_1}, \\ 1, t \geq \tau_{h_1}. \end{cases} \quad (7)$$

Переходы 12, 67, 1112 из-за возникших в случайное время внезапных отказов ФМ горячего пула характеризуются вероятностями [15]

$$Q_{12}(t) = 1 - e^{-\lambda_0 t}; \quad Q_{67}(t) = 1 - e^{-\lambda_0 t}; \quad Q_{1112}(t) = 1 - e^{-\lambda_0 t}. \quad (8)$$

Окончание восстановления и переход из состояний 2, 7, 12 (отказ третьей, второй, первой ФМ, соответственно) в состояния 0, 5, 10 (полное и первый вариант частичного восстановления РС IaaS Cloud, соответственно) зависят от случайной продолжительности работы ремонтного подразделения горячего пула, распределенной по закону Эрланга второго порядка, для которого

$$Q_{20}(t) = 1 - \mu t e^{-\mu t}; \quad Q_{75}(t) = 1 - \mu t e^{-\mu t}; \quad Q_{1210}(t) = 1 - \mu t e^{-\mu t}. \quad (9)$$

Второй вариант полного и частичных восстановлений PC IaaS Cloud описывается переходами 70, 125, 1510 с функциями распределений времени восстановления (соответствуют закону Эрланга третьего порядка), которые записываются в виде следующих соотношений:

$$Q_{70}(t) = 1 - \left(1 + \mu t + \frac{\mu^2 t^2}{2}\right) e^{-\mu t}; \quad Q_{125}(t) = 1 - \left(1 + \mu t + \frac{\mu^2 t^2}{2}\right) e^{-\mu t}; \quad Q_{1510}(t) = 1 - \left(1 + \mu t + \frac{\mu^2 t^2}{2}\right) e^{-\mu t}. \quad (10)$$

При отказе физических машин горячего пула подключение ФМ теплого пула (переходы 23, 78, 1213) осуществляется в течение случайного времени, распределенного по экспоненциальному закону с функциями распределений

$$Q_{23}(t) = 1 - e^{-\rho_1 t}; \quad Q_{78}(t) = 1 - e^{-\rho_1 t}; \quad Q_{1213}(t) = 1 - e^{-\rho_1 t}. \quad (11)$$

Аналогично записываются функции распределения времени подключения ФМ холодного пула (переходы 24, 79, 1214), т. е.

$$Q_{24}(t) = 1 - e^{-\rho_2 t}; \quad Q_{79}(t) = 1 - e^{-\rho_2 t}; \quad Q_{1214}(t) = 1 - e^{-\rho_2 t}. \quad (12)$$

Функции распределения времени “миграций” в горячий пул облачной инфраструктуры записываются следующим образом: для ФМ теплого пула

$$Q_{30}(t) = 1 - e^{-\gamma_1 t}; \quad Q_{85}(t) = 1 - e^{-\gamma_1 t}; \quad Q_{1310}(t) = 1 - e^{-\gamma_1 t}; \quad (13)$$

для ФМ холодного пула

$$Q_{40}(t) = 1 - e^{-\gamma_2 t}; \quad Q_{95}(t) = 1 - e^{-\gamma_2 t}; \quad Q_{1410}(t) = 1 - e^{-\gamma_2 t}. \quad (14)$$

Выполнив нетривиальные расчеты в соответствии с методикой, изложенной в [15–17], получим соотношения для расчета стационарного КГ  $P_0 \bar{K}_\Gamma$  и соответствующих вероятностей  $P_i \bar{K}_\Gamma$ , где  $i = \overline{1,15}$ , в виде

$$K_\Gamma = P_0 \bar{K}_\Gamma \frac{\bar{t}_0}{U}, \quad (15)$$

$$P_1(t) = \frac{\bar{t}_1}{U}, \quad P_2 \bar{K}_\Gamma = p_{12} \frac{\bar{t}_2}{U}, \quad (16)$$

$$P_3 \bar{K}_\Gamma = p_{12} p_{23} \frac{\bar{t}_3}{U}, \quad P_4 \bar{K}_\Gamma = p_{12} p_{24} \frac{\bar{t}_4}{U}, \quad (17)$$

$$P_5 \bar{\mathbf{A}} \bar{\mathbf{U}} \bar{\mathbf{t}}_5, P_6 \bar{\mathbf{A}} \bar{\mathbf{U}} \bar{\mathbf{t}}_6, P_7 \bar{\mathbf{A}} \bar{\mathbf{U}} \bar{\mathbf{t}}_7, \quad (18)$$

$$P_8 \bar{\mathbf{A}} \bar{\mathbf{U}} \bar{\mathbf{t}}_8, P_9 \bar{\mathbf{A}} \bar{\mathbf{U}} \bar{\mathbf{t}}_9, \quad (19)$$

$$P_{10} \bar{\mathbf{A}} \bar{\mathbf{U}} \bar{\mathbf{t}}_{10}, P_{11} \bar{\mathbf{A}} \bar{\mathbf{U}} \bar{\mathbf{t}}_{11}, \quad (20)$$

$$P_{12} \bar{\mathbf{A}} \bar{\mathbf{U}} \bar{\mathbf{t}}_{12}, P_{13} \bar{\mathbf{A}} \bar{\mathbf{U}} \bar{\mathbf{t}}_{13}, P_{14} \bar{\mathbf{A}} \bar{\mathbf{U}} \bar{\mathbf{t}}_{14}, P_{15} \bar{\mathbf{A}} \bar{\mathbf{U}} \bar{\mathbf{t}}_{15}, \quad (21)$$

$$U = \bar{t}_0 + \bar{t}_1 + p_{12} \bar{\mathbf{A}} \bar{\mathbf{U}} \bar{\mathbf{t}}_2 + p_{23} \bar{\mathbf{A}} \bar{\mathbf{U}} \bar{\mathbf{t}}_3 + p_{24} \bar{\mathbf{A}} \bar{\mathbf{U}} \bar{\mathbf{t}}_4 + \beta \bar{\mathbf{A}} \bar{\mathbf{U}} \bar{\mathbf{t}}_5 + \bar{t}_6 + p_{67} \bar{\mathbf{A}} \bar{\mathbf{U}} \bar{\mathbf{t}}_7 + p_{78} \bar{\mathbf{A}} \bar{\mathbf{U}} \bar{\mathbf{t}}_8 + p_{79} \bar{\mathbf{A}} \bar{\mathbf{U}} \bar{\mathbf{t}}_9 + \alpha \bar{\mathbf{A}} \bar{\mathbf{U}} \bar{\mathbf{t}}_{10} + \bar{t}_{11} + \varepsilon \bar{\mathbf{A}} \bar{\mathbf{U}} \bar{\mathbf{t}}_{12} + p_{1213} \bar{\mathbf{A}} \bar{\mathbf{U}} \bar{\mathbf{t}}_{13} + p_{1214} \bar{\mathbf{A}} \bar{\mathbf{U}} \bar{\mathbf{t}}_{14} + p_{1215} \bar{\mathbf{A}} \bar{\mathbf{U}} \bar{\mathbf{t}}_{15}, \quad (22)$$

где

$$\alpha = \frac{P_{67} P_{710}}{1 - P_{1110} - \zeta P_{1112}}, \quad \beta = \frac{P_{12} P_{25}}{1 - P_{65} - \alpha P_{1112} P_{125} - \nu P_{67}},$$

$$\zeta = P_{1210} + P_{1213} + P_{1214} + P_{1215}, \quad \nu = P_{75} - P_{78} - P_{79}, \quad \varepsilon = \alpha \beta P_{1112},$$

$$p_{12} = 1 - e^{-\lambda_0 \tau_{h3}}, \quad p_{23} = \rho_1 \int_0^\infty \mathbf{1} + \mu t \bar{e}^{-\mathbf{A}_1 + \rho_1 + \rho_2 + \mu \bar{\mathbf{t}}} dt, \quad p_{24} = \rho_2 \int_0^\infty \mathbf{1} + \mu t \bar{e}^{-\mathbf{A}_1 + \rho_1 + \rho_2 + \mu \bar{\mathbf{t}}} dt,$$

$$p_{25} = \lambda_1 \int_0^\infty \mathbf{1} + \mu t \bar{e}^{-\mathbf{A}_1 + \rho_1 + \rho_2 + \mu \bar{\mathbf{t}}} dt, \quad p_{65} = e^{-\lambda_0 \tau_{h2}}, \quad p_{67} = 1 - e^{-\lambda_0 \tau_{h2}},$$

$$p_{75} = \frac{1}{2} \int_0^\infty \mu^2 t \mathbf{A}^2 t^2 + 2\mu t + 2 \left( 1 + \mu t + \frac{\mu t^2}{2} \right) e^{-\mathbf{A}_2 + \rho_1 + \rho_2 + 2\mu \bar{\mathbf{t}}} dt,$$

$$p_{78} = \rho_1 \int_0^\infty \mathbf{1} + \mu t \left( 1 + \mu t + \frac{\mu t^2}{2} \right) e^{-\mathbf{A}_2 + \rho_1 + \rho_2 + 2\mu \bar{\mathbf{t}}} dt,$$

$$p_{79} = \rho_2 \int_0^\infty \mathbf{1} + \mu t \left( 1 + \mu t + \frac{\mu t^2}{2} \right) e^{-\mathbf{A}_2 + \rho_1 + \rho_2 + 2\mu \bar{\mathbf{t}}} dt,$$

$$p_{710} = \lambda_2 \int_0^\infty \mathbf{1} + \mu t \left( 1 + \mu t + \frac{\mu t^2}{2} \right) e^{-\mathbf{A}_2 + \rho_1 + \rho_2 + 2\mu \bar{\mathbf{t}}} dt,$$

$$p_{25} = \lambda_1 \int_0^\infty \mathbf{1} + \mu t \bar{e}^{-\mathbf{A}_1 + \rho_1 + \rho_2 + \mu \bar{\mathbf{t}}} dt, \quad p_{65} = e^{-\lambda_0 \tau_{h2}}, \quad p_{67} = 1 - e^{-\lambda_0 \tau_{h2}},$$

$$p_{1110} = e^{-\lambda_0 \tau_{h1}}, \quad p_{1112} = 1 - e^{-\lambda_0 \tau_{h1}},$$

$$P_{1210} = \frac{1}{2} \int_0^{\infty} \mu^2 t \left( \lambda_0^2 t^2 + 2\mu t + 2 \left( 1 + \mu t + \frac{\lambda_0^2 t^2}{2} \right) e^{-\lambda_0^2 t} \right) e^{-\lambda_0^2 t} dt,$$

$$P_{1214} = \rho_2 \int_0^{\infty} \lambda_0 + \mu t \left( 1 + \mu t + \frac{\lambda_0^2 t^2}{2} \right) e^{-\lambda_0^2 t} dt,$$

$$P_{1215} = \lambda_3 \int_0^{\infty} \lambda_0 + \mu t \left( 1 + \mu t + \frac{\lambda_0^2 t^2}{2} \right) e^{-\lambda_0^2 t} dt, \quad P_{125} = 1 - P_{1210} - P_{1213} - P_{1214} - P_{1215},$$

$$\bar{t}_0 = \bar{t}_5 = \bar{t}_{10} = T, \quad \bar{t}_2 = \frac{\lambda_0 + \rho_1 + \rho_2 + 2\mu}{\lambda_0 + \rho_1 + \rho_2 + \mu}, \quad \bar{t}_3 = \bar{t}_8 = \bar{t}_{13} = \frac{1}{\gamma_1}, \quad \bar{t}_4 = \bar{t}_9 = \bar{t}_{14} = \frac{1}{\gamma_2}, \quad \bar{t}_{15} = \frac{3}{\mu},$$

$$\bar{t}_1 = \frac{1}{\lambda_0} \left( 1 - e^{-\lambda_0 \tau_{h3}} \right), \quad \bar{t}_6 = \frac{1}{\lambda_0} \left( 1 - e^{-\lambda_0 \tau_{h2}} \right), \quad \bar{t}_{11} = \frac{1}{\lambda_0} \left( 1 - e^{-\lambda_0 \tau_{h1}} \right),$$

$$\bar{t}_7 = \int_0^{\infty} \lambda_0 + \mu t \left( 1 + \mu t + \frac{\lambda_0^2 t^2}{2} \right) e^{-\lambda_0^2 t} dt,$$

$$\bar{t}_{12} = \int_0^{\infty} \lambda_0 + \mu t \left( 1 + \mu t + \frac{\lambda_0^2 t^2}{2} \right) e^{-\lambda_0^2 t} dt.$$

На рис. 5–10 изображены графики зависимости  $K_{\Gamma}, \lambda, T$ , полученные для исходных данных, представленных в табл. 1.

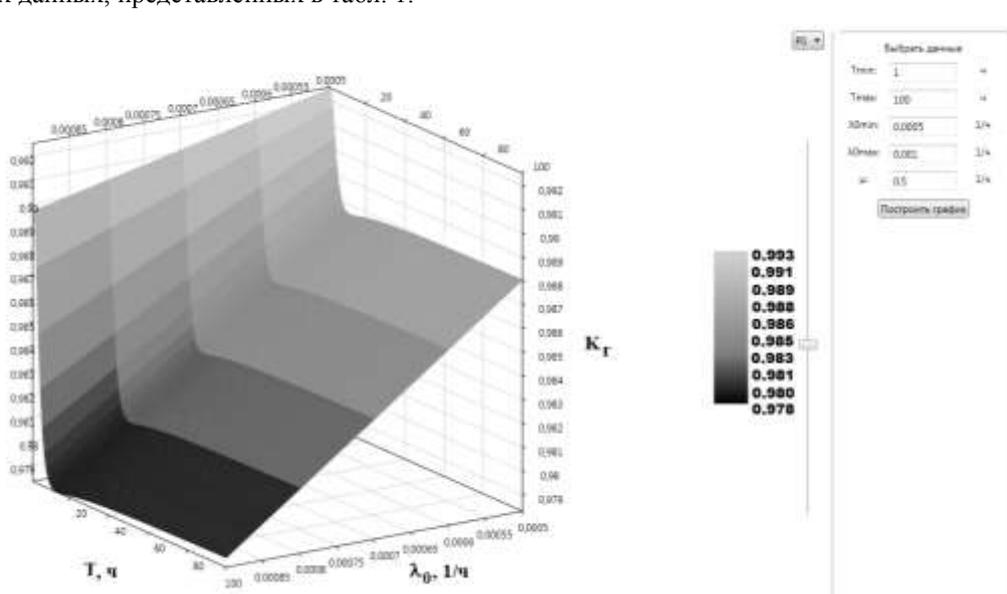


Рис. 5. Зависимость  $K_{\Gamma}, \lambda, T$  для  $T=100$ ч,  $\mu=0,5$  1/ч

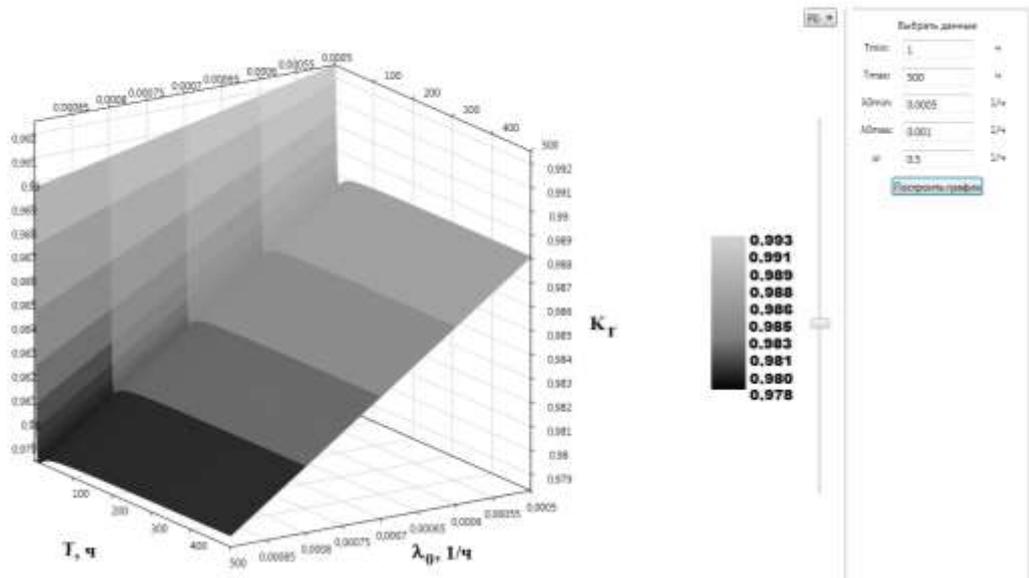


Рис. 6. Зависимость  $K_T$   $\lambda_0, T$  для  $T = 500$  ч,  $\mu = 0,5$  1/ч

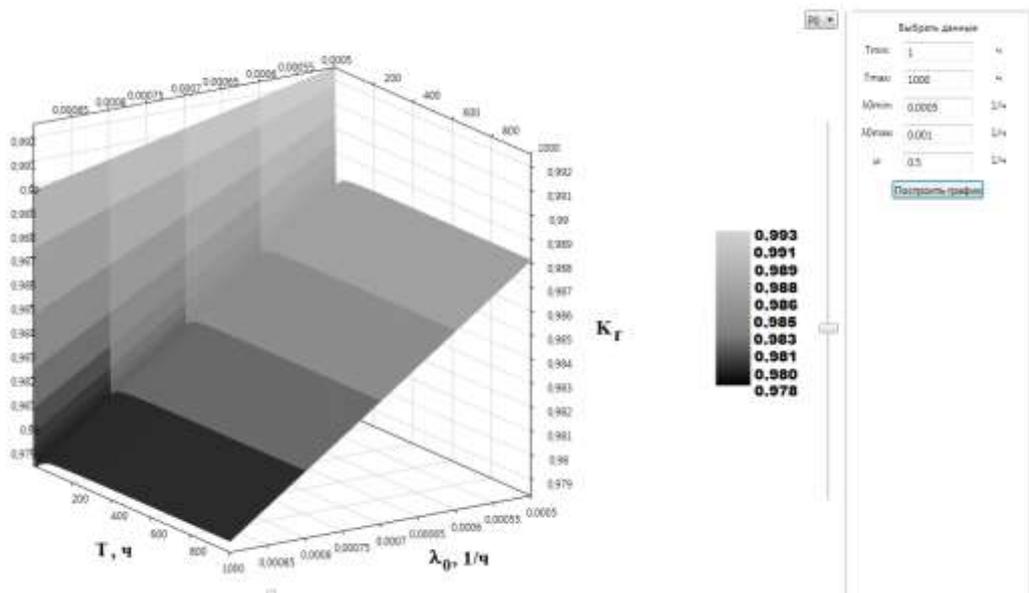


Рис. 7. Зависимость  $K_T$   $\lambda_0, T$  для  $T = 1000$  ч,  $\mu = 0,5$  1/ч

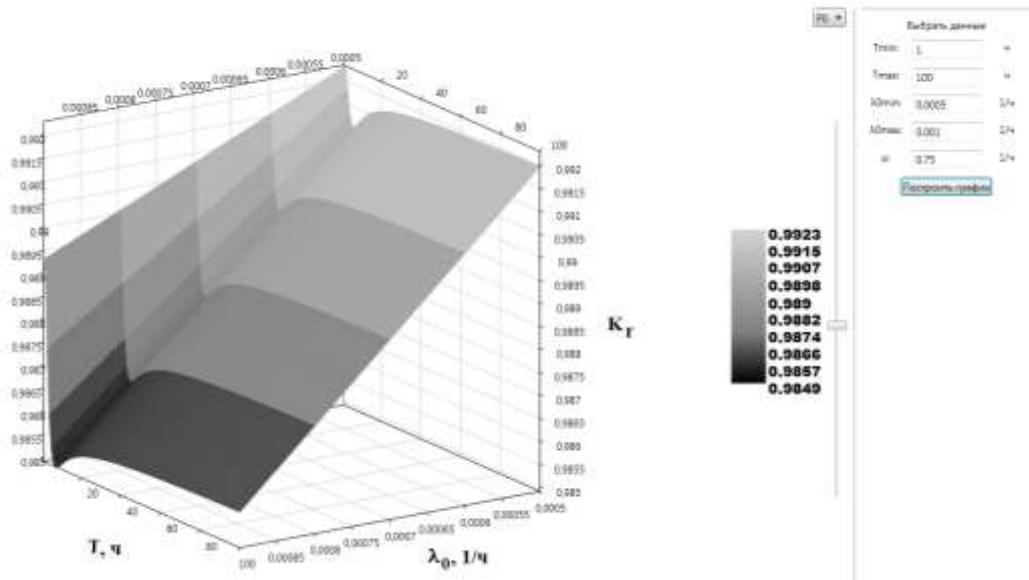


Рис. 8. Зависимость  $K_T$  от  $\lambda_0, T$  для  $T = 100$  ч,  $\mu = 0,75$  1/ч

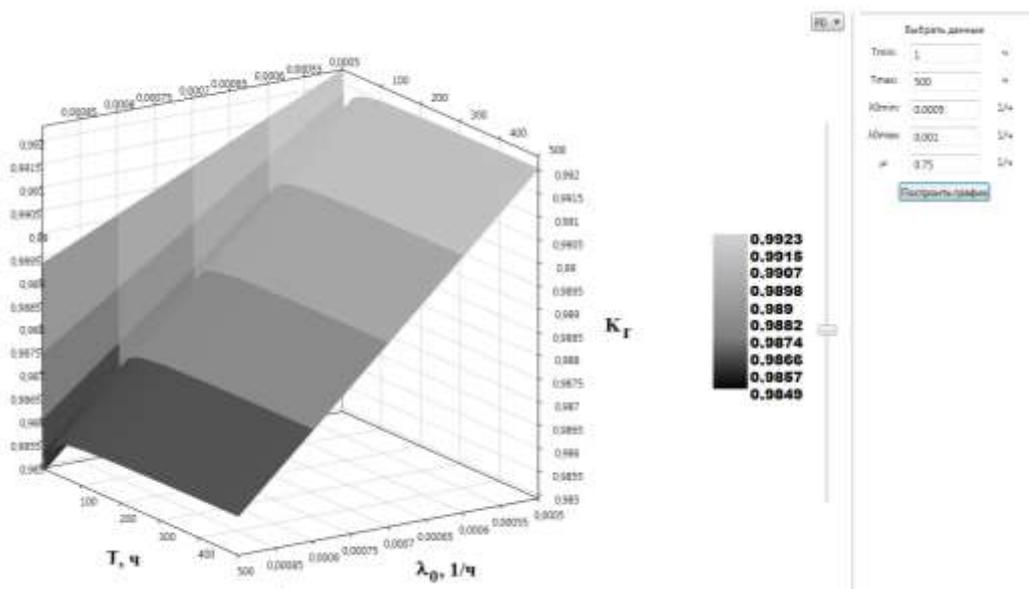


Рис. 9. Зависимость  $K_T$  от  $\lambda_0, T$  для  $T = 500$  ч,  $\mu = 0,75$  1/ч

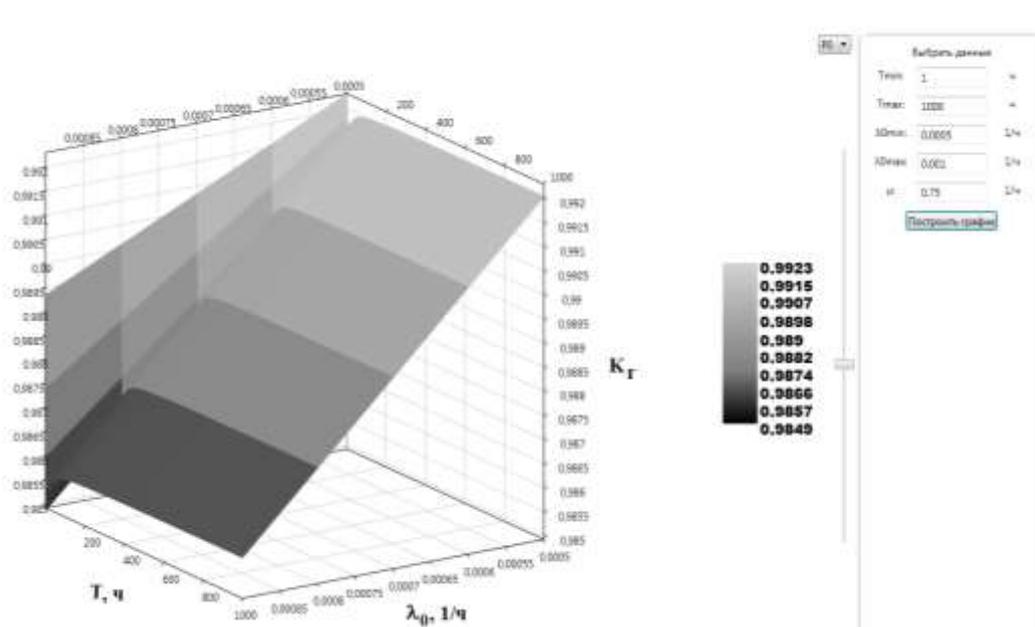


Рис. 10. Зависимость  $K_T$  от  $T$  и  $\lambda_0$  для  $T = 1000$  ч,  $\mu = 0,75$  1/ч

Результаты моделирования (рис. 5–10) свидетельствуют о довольно высоком уровне функциональной готовности IaaS Cloud в условиях протекания внезапных отказов, что обеспечивается комплексной реализацией мероприятий по восстановлению работоспособности, рациональному распределению и резервированию ресурсного потенциала компонентных составляющих инфраструктуры как сервиса облачных вычислений. Установлено, что наибольшее влияние на изменение уровня надежности облачной инфраструктуры оказывает продолжительность ее использования по назначению.

**Выводы из данного исследования и- перспективы дальнейших разведок в данном направлении.** Не составляет труда заметить, что предложенная авторами стохастическая модель (рис. 4) строится с использованием ядра, сформированного состояниями  $E_0, \dots, E_4$ , которые в последствии размножаются в виде звеньев до состояний  $E_5, \dots, E_{14}$  и финального состояния  $E_{15}$ . С увеличением в многофункциональных пулах облачной инфраструктуры числа физических машин пропорционально возрастает количество звеньев, что позволяет адаптировать предлагаемую модель к изменению количественного состава элементов рассматриваемой IaaS Cloud.

Перспективы дальнейших исследований связаны с разработкой модельного ряда, учитывающего масштабность многофункционального взаимодействия резервируемых пулов; 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. Кроме того, предложенный стохастический подход может быть применен для оценки жесткости разрабатываемых полумарковских моделей надежности инфраструктуры как сервиса облачных вычислений с ресурсным резервированием ее компонентных составляющих.

---

**Список использованных источников:**

1. Scalable Analytics for IaaS Cloud Availability [Электронный ресурс] / R. Ghosh, F. Longo, F. Frattini and oth. // IEEE Transactions On Cloud Computing. – 2014. – Vol. 2, no. 1. – P. 57–70. – Режим доступа : [http://www.researchgate.net/profile/Kishor\\_Trivedi2/publication/261923705Scalable\\_Analytics\\_for\\_IaaS\\_Cloud\\_Availability/links/0f31753648d22cc28f000000.pdf](http://www.researchgate.net/profile/Kishor_Trivedi2/publication/261923705Scalable_Analytics_for_IaaS_Cloud_Availability/links/0f31753648d22cc28f000000.pdf)
2. Stochastic Model Driven Capacity Planning for an Infrastructure-as-a-Service Cloud / R. Ghosh, F. Longo, R. Xia and oth. // IEEE Transactions On Services Computing. – 2013. – Vol. 7, no. 4. – P. 667–680.
3. Amazon EC2 [Электронный ресурс]. – Режим доступа : <http://www.aws.amazon.com/ec2>
4. IBM SmartCloud Enterprise [Электронный ресурс]. – Режим доступа : <http://www-935.ibm.com/services/us/en/cloud-enterprise/index.html>
5. Ghosh R. Modeling and Performance Analysis of Large Scale IaaS Clouds [Электронный ресурс] / R. Ghosh, F. Longo, Vijay K. Naik, Kishor S. Trivedi // Future Generation Computer Systems. – 2013. – Vol. 29. – P. 1216–1234. – Режим доступа : <http://www.mdslab.unime.it/documents/FGCS2013.pdf>
6. Martinello M. Web Service Availability: Impact of Error Recovery and Traffic Model / M. Martinello // Journal of Reliability Engineering and System Safety. – 2005. – Vol. 89, no. 1. – P. 6–16.
7. Sato N. Stochastic Modeling of Composite Web Services for Closed-form Analysis of their Bottlenecks / N. Sato, K. S. Trivedi // Service-Oriented Computing – ICSOC. – 2007. – P. 107–118.
8. Sean L. Cloud 101: What do IaaS, PaaS and SaaS companies should do? [Электронный ресурс] / Sean L. – November 14, 2011. – Режим доступа : <http://www.venturebeat.com/2011/11/14/cloud-iaas-paas-saas>
9. Meyer J. F. One Valuating the Performability of Degradable Computing Systems / J. F. Meyer // IEEE Transactions on Computers. – 1980. – Vol. 29, no.8. – P. 720–731.
10. Trivedi K. S. Probability and Statistics with Reliability, Queuing and Computer Science Applications / Trivedi K. S. / John Wiley and Sons, 2001. – 356 p.
11. Dragovic B. Mathematical Models of Multiserver Queuing System for Dynamic Performance Evaluation in Port [Электронный ресурс] / B. Dragovic, Nam-Kyu Park, Nenad D. Zrnica, R. Mestrovic // Hindawi Publishing Corporation, Mathematical Problems in Engineering. – 2012. – Режим доступа : <http://www.computer.org/csdl/trans/td/2013/05/ttd2013050849-abs.html>
12. Ghosh R. Quantifying Resiliency of IaaS Cloud [Электронный ресурс] / R. Ghosh, F. Longo, Vijay K. Naik, K. S. Trivedi // In RACOS Workshop. – 2010. Режим доступа : <http://www.ieeexplore.ieee.org/xpl/login.jsp>
13. Ghosh R. End-to-End Performability Analysis for Infrastructure-as-a-Service Cloud: An Interacting Stochastic Models Approach [Электронный ресурс] / R. Ghosh, Kishor S. Trivedi, Vijay K. Naik, Dong S. Kim // Dependable Computing (PRDC), 2010 IEEE 16th Pacific Rim International Symposium on. – 2010. – P. 125–132. – Режим доступа : [http://www.researchgate.net/profile/Kishor\\_Trivedi2/publication/220700061\\_End-to-End\\_Performability\\_Analysis\\_for\\_Infrastructure-as-a-Service\\_Cloud\\_An\\_Interacting\\_Stochastic\\_Models\\_Approach/links/54233a4d0cf238c6ea6e3632.pdf](http://www.researchgate.net/profile/Kishor_Trivedi2/publication/220700061_End-to-End_Performability_Analysis_for_Infrastructure-as-a-Service_Cloud_An_Interacting_Stochastic_Models_Approach/links/54233a4d0cf238c6ea6e3632.pdf)
14. Острейковский В. А. Теория надежности / Острейковский В. А. – М. : Высшая школа, 2003. – 463 с.
15. Волков Л. И. Управление эксплуатацией летательных комплексов / Волков Л. И. – М. : Высшая школа, 1981. – 368 с.
16. Безопасность критических инфраструктур: математические и инженерные методы анализа и обеспечения / под ред. В. С. Харченко ; Министерство образования и науки Украины, Национальный аэрокосмический университет им. Н. Е. Жуковского “ХАИ”, 2011. – 641 с.

- 
17. Информационные технологии для критических инфраструктур / под ред. А. В. Скаткова. – Севастополь : СевНТУ, 2012. – 306 с.
18. Trivedi K. S. Optimization of IaaS Cloud including Performance, Availability, Power Analysis [Электронный ресурс] / K. S. Trivedi. – Networking 2014, Trondheim, Norway. – June 2. – 2014. – Режим доступа : <http://www.networking2014.item.ntnu.no/pdfs/K2-KishorTrivedi-IFIP-Networking-2014.pdf>
19. Dai Y.-S. Cloud Service Reliability: Modeling and Analysis, Proc. IEEE Pacific Rim Int'l Symp [Электронный ресурс] / Y-S. Dai, B. Yang, J. Dongarra, G. Zhang // 15th IEEE Pacific Rim International Symposium on Dependable Computing (PRDC). – 2009. – Режим доступа : <ftp://www.ftp.radiomaryja.pl.eu.org/vol/rzml/netlib/utk/people/JackDongarra/PAPERS-2012-01-24/Cloud-Shaun-Jack.pdf>
20. Lanus M. Hierarchical Composition and Aggregation State-Based Availability and Performability Models / M. Lanus, L. Yin, K. S. Trivedi // IEEE Trans. Reliability. – 2003. – Vol. 52, no. 1. – P. 44–52.