

УДК 004.942

О.В. Яновская

*Национальный аэрокосмический университет имени Н.Е. Жуковского «ХАИ», Харьков*

## МОДЕЛИ ДОСТУПНОСТИ СЕРВИСОВ РАСПРЕДЕЛЕННЫХ ОБЛАЧНЫХ СИСТЕМ

*Рассмотрены основные проблемы, возникающие при разработке и исследовании моделей доступности сервисов распределенных облачных систем. Предложены полумарковские модели доступности сервисов распределенных облачных систем в зависимости от вида ресурса, типа частей ресурса, принципа построения модели, наличия резервного сервера и других критериев. Выполнен структурный и параметрический синтез предложенных моделей. Проведен анализ полученных значений коэффициента доступности составного ресурса на основе характеристик полумарковских процессов.*

**Ключевые слова:** *распределенная облачная архитектура, одноранговые сети, полумарковские процессы, облачные сервисы, доступность, готовность, надежность.*

### Введение

**Постановка проблемы.** Одноранговая распределенная облачная архитектура [1] реализует концепцию распределения ресурсов в облачной сети с децентрализованной структурой, объединяя преимущества технологий GRID, облачных вычислений и пиринговых сетей. Особенностью предложенной модели является самоорганизация процесса репликации ресурсов средствами рабочих станций узлов-участников системы: данный процесс не требует вмешательства администратора или сторонних механизмов. Независимо от роли, любой узел-участник одноранговой распределенной облачной системы получает высокоскоростной доступ ко всем ресурсам системы благодаря возможности получить ресурс или его часть от ближайшего узла-участника, который имеет копию данного ресурса.

Актуальной задачей является разработка и исследование моделей доступности сервисов распределенной облачной системы для определения целесообразности ее внедрения. Основными проблемами при разработке и исследовании моделей такого типа облачной архитектуры является неопределенность ее структуры и топологии, отсутствие полноценной достоверной информации о параметрах рабочих станций и других устройств, которые являются основными элементами архитектуры и одновременно узлами репликации, сложность предсказания поведения системы в зависимости от степени популярности ресурса, а также динамичность интенсивности появления и отключения узлов репликации, и других параметров модели.

Для решения поставленной задачи необходимо выполнить анализ последних исследований и публикаций в данной предметной области, структурный и параметрический синтез моделей, а также представить результаты моделирования.

**Анализ последних исследований и публикаций.** Анализ источников [2, 3] показывает, что наиболее распространенным является формальный метод оценивания готовности и доступности облачной системы, согласно которому коэффициент готовности определяется следующим отношением:

$$K_g = \text{MTBF} / (\text{MTBF} + \text{MTTR}),$$

где MTBF – среднее время работы системы до отказа, MTTR – среднее время ремонта, необходимое для восстановления системы.

В ряде исследований [2 – 4] коэффициенты доступности, готовности и надежности облачных систем определяются на основе структурных схем надежности, деревьев отказов, марковских цепей и сетей Петри. В работах [5, 6] предложены марковские модели для оценки готовности облачной архитектуры на различных уровнях иерархической структуры виртуального ЦОД. В исследованиях [7, 8] проводится анализ и моделирование надежности облачных сервисов и GRID-систем на основе марковских процессов и систем массового обслуживания. Однако существующие модели в явном виде не могут быть применимы к анализируемой распределенной облачной архитектуре, поскольку не учитывают особенностей пассивной репликации.

Таким образом, **целью данной работы** является разработка моделей и анализ доступности сервисов одноранговой распределенной облачной системы с учетом поведения узлов репликации и процесса распределения ресурсов.

### Структурный синтез

На начальном этапе разработки модели доступности облачного сервиса распределенной облачной архитектуры необходимо описать состояния, в которых может находиться система, проанализировать ее поведение, после чего выделить те состоя-

ния, которые будут влиять на доступность сервиса [9]. Для анализа доступности сервисов одноранговой распределенной облачной системы разработано пять базовых моделей, которые отличаются структурой и входными параметрами. Выбор типа модели зависит от типа ресурса, типа частей ресурса, принципа построения модели, наличия резервного сервера. Выходным параметром всех базовых моделей является коэффициент доступности сервиса, который определяет вероятность того, что в произвольный момент времени облачный сервис окажется доступным для конечного пользователя с удовлетворительным временем отклика.

Список входных параметров:  $N$  – количество пользовательских узлов репликации;  $\lambda_{off}$  – интенсивность отключения пользовательских узлов репликации;  $\mu_{on}$  – интенсивность включения пользовательских узлов репликации;  $\omega_n$  – интенсивность появления новых пользователей, запрашивающих ресурс;  $\lambda_r$  – интенсивность получения ответа на запрос пользователя, запрашивающего ресурс,  $k$  – количество частей ресурса;  $\lambda_0$  – интенсивность отказов узла вла-

дельца ресурса;  $\mu_0$  – интенсивность восстановлений узла владельца ресурса;  $\lambda_j$ , где  $j = 1, \dots, k$  – интенсивность получения ответа на запрос пользователя, запрашивающего  $j$ -ую часть ресурса.  $g_i$ , где  $i = 1, \dots, k$  – интенсивность выполнения репликации  $i$ -й части ресурса с удовлетворительной степенью;  $d_i$ , где  $i = 1, 2, \dots, k$  – интенсивность снижения степени репликации  $i$ -й части ресурса до неудовлетворительной;  $R$  – интенсивность выполнения репликации всех частей ресурса с удовлетворительной степенью;  $D$  – интенсивность снижения степени репликации всех частей ресурса до неудовлетворительной;  $\lambda_s$  – интенсивность отказов резервного сервера;  $\mu_s$  – интенсивность восстановлений резервного сервера.

Для анализа моделей необходимо провести структурный и параметрический синтез.

В самом простом случае, когда ресурс является цельным (например, веб-ресурс на базе концепции “single page web site”), а все узлы – равноправными (без выделенной роли узла-владельца ресурса), граф состояний и переходов, описывающий поведение системы, представлен на рис. 1.

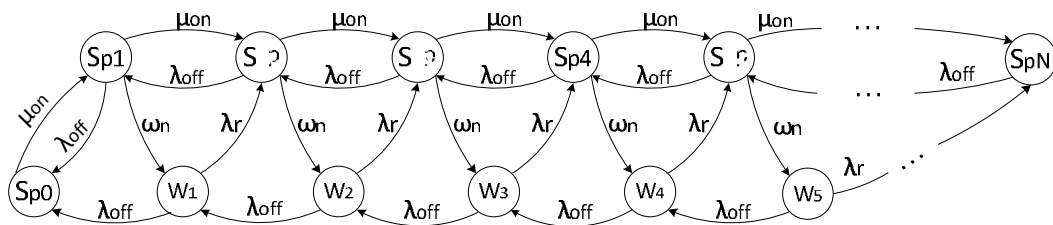


Рис. 1. Модель доступности простого (цельного) ресурса

Пространство состояний системы:  $Sp_i$ , где  $i = 0, 1, \dots, N$  – состояния, соответствующие количеству доступных узлов репликации  $i$  в системе;  $W_i$ , где  $i = 1, \dots, N$  – промежуточные состояния, соответствующие получению ответа на запрос от одного из доступных  $i$  узлов для нового пользователя, запрашивающего ресурс.

При расчете показателя доступности будем считать работоспособными такие состояния, в которых время ответа на запрос не превышает некоторого порогового значения, например, 1,5 секунды. Данная характеристика определяется требованиями к системе. Такая модель достаточно полно описывает поведение системы согласно принципу появления и исчезновения пользовательских узлов репликации в случае, если ресурс состоит из одной части. Однако в ряде других случаев ресурс является комбинацией из нескольких частей (например, нескольких веб-страниц) и пользователи могут запрашивать части ресурса отдельно. В этом случае граф состояний и переходов, построенный по аналогичному принципу, является более громоздким, поскольку содержит состояния, соответствующие учету всех комбинации различных частей ресурса, которые находятся на пользовательских узлах (рис. 2).

Интенсивность обращения и получения ответов на запрос для различных частей ресурса будет разным. Отдельно стоит выделить роль узла владельца ресурса, который будет обеспечивать постоянный доступ ко всем частям ресурса. Ниже описано пространство состояний системы:

$S_1$  – начальное состояние системы, при котором доступен только один пользовательский узел владельца ресурса;  $S_0$  – состояние отказа узла владельца ресурса, в котором не доступен ни один узел репликации;  $1, 0, \dots, 0 - 0, \dots, 0, N$  – состояния, соответствующие всем возможным комбинациям  $k$  частей ресурса по  $N$  пользовательским узлам репликации;  $W_0$  – промежуточное состояние, соответствующее появлению первого пользователя, запрашивающего какую-либо часть ресурса;  $W_{ij}$ ,  $i = 1, \dots, N$ ;  $j = 1, \dots, k$  – промежуточные состояния, соответствующие получению ответа на запрос от одного из доступных  $i$  узлов для новых пользователей, запрашивающих какую-либо часть ресурса.

Основной проблемой расчета характеристик системы на основе данной модели является лавинообразный рост количества состояний, который зависит от числа пользовательских узлов-участников и частей ресурса. С учетом того, что количество узлов-

участников распределенной облачной системы, которые связываются между собой посредством сети Интернет, может достигать порядка сотен тысяч пользователей, задача расчета характеристик системы на основе данной модели будет являться достаточно трудоемкой. Однако при небольших значениях количества узлов и частей ресурса подход на основе данной модели применим. Для устранения недостатков применения данной модели предлагается модель доступности облачного сервиса, учитывающая количество пользовательских узлов как параметр (рис. 3). В данной модели также выделяется роль узла владельца ресурса, который будет обеспечивать постоянный доступ ко всем частям ресурса. Будем считать, что ресурс состоит из  $k$  равных по размеру частей. Под удовлетворительной степенью репликации части ресурса будем понимать такую степень репликации, при которой время ответа на запрос данной части ресурса не превышает некоторое пороговое значение, определяемое требованиями к системе. Под удовлетворительной степенью репликации ресурса будем понимать такую степень репликации, при которой время ответа на запрос всех частей ресурса не превышает некоторое пороговое значение, определяемое требованиями к системе. Значение среднего времени ответа на запрос определяется количеством доступ-

ных узлов репликации и их параметрами, характеристиками каналов связи.

Пространство состояний системы:  $S_0$  – начальное состояние системы, в котором доступен только один пользовательский узел владельца ресурса;  $S_{0i}$ , где  $i=1, \dots, k$  – состояния отказа узла владельца ресурса, в которых  $i$  частей ресурса доступны с удовлетворительной степенью репликации;  $1, 12, \dots$  – состояния, в которых первая, первая и вторая, и т.д. частей ресурса доступны с удовлетворительной степенью репликации.

При расчете показателя доступности будем учитывать такие состояния, в которых степень репликации удовлетворительна для всех частей ресурса. Очевидным преимуществом перехода к данной модели является существенное снижение размерности задачи определения вероятностей состояний, поскольку общее количество состояний системы в этом случае зависит только от числа частей ресурса.

Если сделать допущение, что все части ресурса одинаковы по размеру, и средние времена выполнения репликации с удовлетворительной степенью и снижения степени репликации до неудовлетворительной для всех частей ресурса одинаковы, данная модель преобразуется к более простому виду (рис. 4).



Рис. 2. Модель доступности составного ресурса

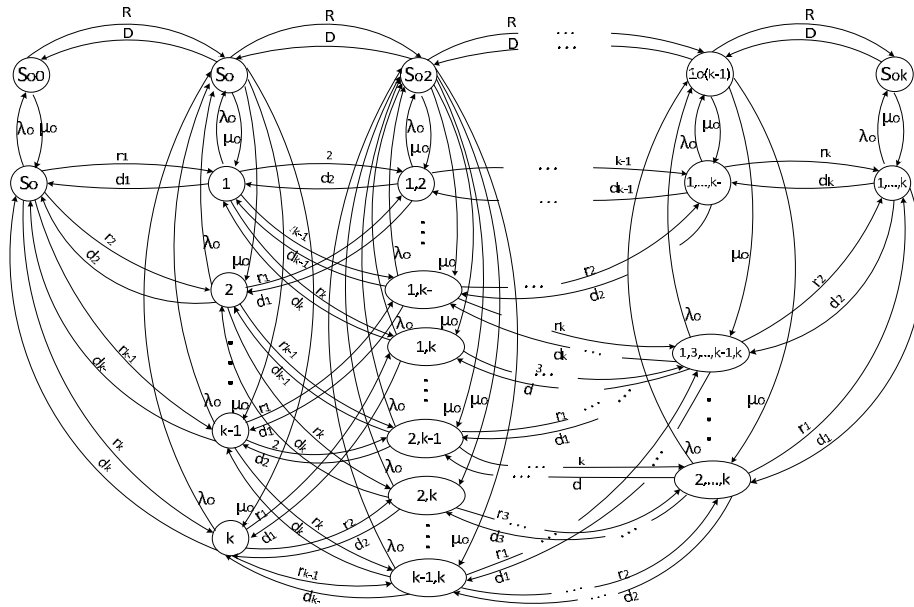


Рис. 3. Модель доступности составного ресурса из k разнотипных частей

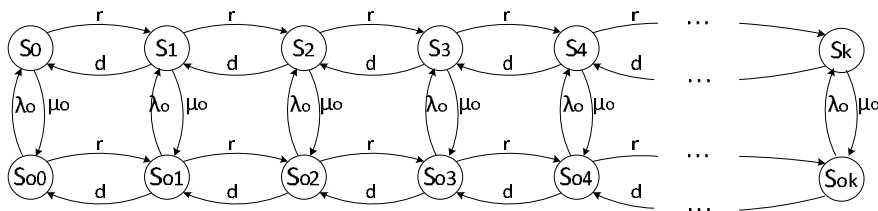


Рис. 4. Модель доступности составного ресурса из k однотипных частей

На рис. 4:  $S_0$  – начальное состояние системы, в котором доступен только один пользовательский узел владельца ресурса;  $S_{0i}$ , где  $i = 1, \dots, k$  – состояния отказа узла владельца ресурса, при которых  $i$  частей ресурса доступны с удовлетворительной степенью репликации;  $S_i$ , где  $i = 1, \dots, k$  – состояния, в которых  $i$  частей ресурса доступны с удовлетворительной степенью репликации.

В качестве одного из методов повышения готовности сервисов распределенной облачной архитектуры используется внедрение резервного always-on сервера, который будет находиться в режиме горячего резерва 24/7 и поддерживать уровень доступности сервиса в случае, когда степень репликации неудовлетворительна. В этом случае резервный сервер является еще одним основным элементом системы, что необходимо учесть при построении модели (рис. 5). Состояниями модели являются:  $S_0$  – начальное состояние системы, в котором доступны только пользовательский узел владельца ресурса и резервный сервер;  $S_j$ , где  $j = 1, \dots, k$  – состояния, в которых в котором доступны пользовательский узел владельца ресурса, резервный сервер и  $i$  частей ресурса с удовлетворительной степенью репликации;  $S_{0,i=j}$ , где  $j = 1, \dots, k$  – состояния отказа узла владельца ресурса, при которых доступны резервный сервер и  $i$  частей ресурса с удовлетворительной степенью репликации;  $S_{s,i=j}$ , где  $j = 1, \dots, k$  – состоя-

ния отказа резервного сервера, при которых доступны пользовательский узел владельца ресурса и  $i$  частей ресурса с удовлетворительной степенью репликации;  $S_{0,s,i=j}$ , где  $j = 0, \dots, k$  – состояния отказа резервного сервера и узла владельца ресурса, при которых  $i$  частей ресурса доступны с удовлетворительной степенью репликации.

### Параметрический синтез

Для расчета характеристик предложенных моделей, поведение которых описано в соответствии с представленными графами состояний и переходов, необходимо определить законы распределения, которым подчиняются времена пребывания системы в каждом состоянии до перехода в новое, после чего применить подходящий математический аппарат.

Согласно [10], отказы и восстановления – случайные события, подчиняющиеся экспоненциальному закону распределения, поэтому время пребывания во всех работоспособных состояниях до перехода в состояния отказа, а также время восстановления в работоспособное состояние являются случайными величинами, распределенными по экспоненциальному закону с параметрами  $\lambda_0$  и  $\mu_0$  соответственно. Параметры  $\lambda_0$  и  $\mu_0$  являются детерминированными величинами и определяются по технической спецификации оборудования исходя из условий эксплуатации системы.

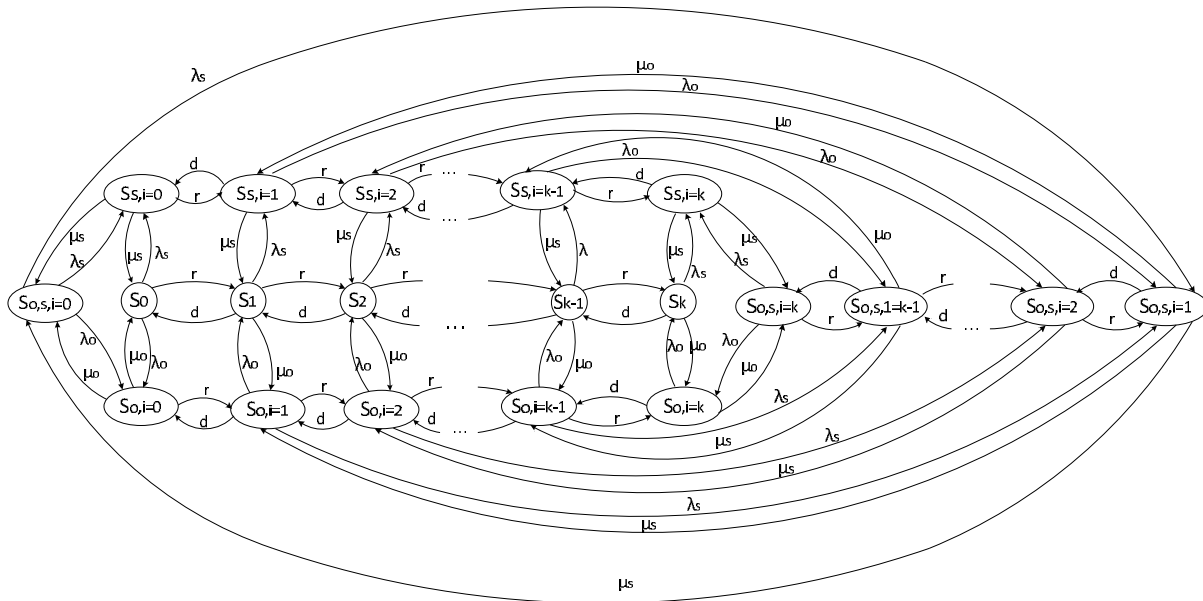


Рис. 5. Модель доступности составного ресурса при условии наличия резервного сервера

Согласно ряду исследований [11 – 13] установлено, что взаимодействие, связанное с использованием сервис-ориентированных интернет-приложений и ресурсов пиринговых файлообменных сетей описывается степенным законом распределения и, в том числе, одним из его вариантов – законом Зипфа [13]. В результате многочисленных экспериментальных наблюдений [12] было обнаружено, что степенной закон распределения применим ко многим процессам, происходящим в Интернет и сервис-ориентированных компьютерных системах: изменение размера очереди при обращении к базе данных; изменение количества посетителей веб-сайта, количества гиперссылок на веб-странице, размера веб-объектов; поведение пользователей социальных сетей; рост популярности ресурсов в пиринговых распределенных файлообменных сетях. Так как процессы выполнения репликации с удовлетворительной степенью и снижения степени репликации до неудовлетворительной фактически отображают поведение пользователей, выполняющих запросы к интернет-сервису в зависимости от его популярности, поведение системы до соответствующих переходов также подчиняется степенному закону распределения с параметрами  $r$  и  $a$  соответственно. Таким образом, специфика предложенных моделей такова, что не все функции распределения времени пребывания в  $i$ -м состоянии перед переходом в  $j$ -е состояние подчиняются экспоненциальному закону. Проведенный анализ дает возможность сделать заключение, что при расчете характеристик системы целесообразно использовать математический аппарат полумарковских процессов (ПМП) [14].

С учетом принятых допущений проанализируем доступность сервиса на основе последней модели. Пусть ресурс состоит из трех равных частей, а

средние времена выполнения репликации с удовлетворительной степенью и средние времена снижения степени репликации до неудовлетворительной для всех частей ресурса одинаковы (рис. 6).

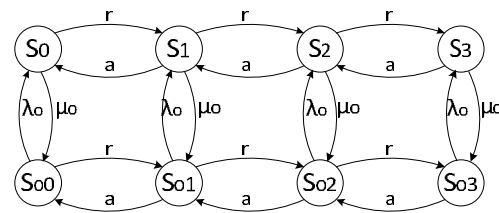


Рис. 6. Модель доступности составного ресурса из трех однотипных частей

В начальный момент времени  $t=0$  система находится в состоянии  $S_0$ , когда доступен только один пользовательский узел – рабочая станция владельца ресурса. В случае отказа рабочей станции владельца ресурса выполняется переход в состояние  $So_0$  с интенсивностью  $\lambda_o=1/t_{MTBF}$  – величины обратной среднему времени наработки на отказ  $t_{MTBF}$ , и сервис перестает быть доступным. Обратный переход происходит с интенсивностью восстановления  $\mu_o=1/t_{MTTR}$ , обратной среднему времени ремонта (восстановления)  $t_{MTTR}$  рабочей станции владельца ресурса. Время пребывания системы до перехода в состояние отказа и в обратном направлении распределено по экспоненциальному закону с параметрами  $\lambda_o$  и  $\mu_o$  соответственно. В начальном состоянии  $S_0$  система проводит случайное время, распределенное по степенному закону с параметром  $r$  до перехода в состояние  $S_1$ , когда степень репликации одной из частей ресурса становится удовлетворительной, при этом среднее время, за которое может быть достигнута удовлетворительная степень репликации части ресурса составляет  $tr$ . Обратный переход из состояния  $S_1$  в состояние  $S_0$

виконується в случайні моменти часу, також мають степене розподіл з параметром  $a$ . Далі йдуть послідовні переходи в стани S2, S3 і в зворотному напрямку описуються аналогічним чином. З кожного з станів S1, S2, S3 можливі переходи в стани відмови вузла власника ресурсу So1, So2, So3 відповідно, а також зворотні переходи в случайні моменти часу, які розподілені за експоненціальним законом так, як описано вище. Відповідно до наведеного опису отримати матрицю  $Q = \{Q_{ij}(t)\}$  незалежних функцій розподілу часу перебування в  $i$ -му стані перед переходом в  $j$ -е стан, якщо б даний вихід був єдиним:

$$\begin{aligned} Q_{0-0}(t) &= 1 - e^{-\lambda_0 t}; & Q_{00-0}(t) &= 1 - e^{-\mu_0 t}; \\ Q_{1-01}(t) &= 1 - e^{-\lambda_0 t}; & Q_{01-1}(t) &= 1 - e^{-\mu_0 t}; \\ Q_{0-1}(t) &= 1 - C \cdot t^{-r}; & Q_{1-0}(t) &= B \cdot t^{-a}; \\ Q_{1-2}(t) &= 1 - C \cdot t^{-r}; & Q_{2-1}(t) &= B \cdot t^{-a}; \\ Q_{00-01}(t) &= 1 - C \cdot t^{-r}; & Q_{01-00}(t) &= B \cdot t^{-a}; \\ Q_{01-02}(t) &= 1 - C \cdot t^{-r}; & Q_{02-01}(t) &= B \cdot t^{-a}; \\ Q_{2-3}(t) &= 1 - C \cdot t^{-r}; & Q_{3-2}(t) &= B \cdot t^{-a}; \\ Q_{3-03}(t) &= 1 - e^{-\lambda_0 t}; & Q_{03-3}(t) &= 1 - e^{-\mu_0 t}; \\ Q_{2-02}(t) &= 1 - e^{-\lambda_0 t}; & Q_{02-2}(t) &= 1 - e^{-\mu_0 t}; \\ Q_{02-03}(t) &= 1 - C \cdot t^{-r}; & Q_{03-02}(t) &= B \cdot t^{-a}. \end{aligned}$$

На основі цих вихідних даних знайдемо доступність сервісу, яку можна виразити через характеристики ПМП такі, як ймовірності переходу  $p_{ij}$  в момент скачка, а також безумовні математичні очікування  $t \square_i$  часів перебування системи в кожному стані. Відповідно до [14], ймовірність переходу  $p_{ij}(t)$  за час  $t$  з стану  $i$  в стан  $j$  залежить від наявності інших напрямків переходу  $k \neq j$ , характеризує функціями  $Q_{ik}(t)$ . Тоді ймовірність виходу з стану  $S_i$  за час  $\tau$  по напрямку  $k \neq j$  визначається як

$$P_{\text{nout}_i} = \prod_{k \neq j} [1 - Q_{ik}(\tau)].$$

Ймовірність переходу по напрямку  $j$  в околі моменту  $\tau$  дорівнює  $dQ_{ik}(\tau)$ . Тоді шукана ймовірність переходу з стану  $S_i$  в стан  $S_j$ :

$$p_{ij}(t) = \int_0^t \prod_{k \neq j} [1 - Q_{ik}(\tau)] dQ_{ij}(\tau).$$

ПМП, розглядається тільки в моменти переходу, визначається матрицею  $\{p_{ij}\}$ , і, відповідно, зводиться до марковської ланцюга, в якій історія процесу до потрапляння в поточний стан не впливає на його подальше поведінку [14]. Такий марковський процес, що знаходиться в напівмарковському, називається вкладеною марковською ланцюгом (ВМЦ). Стационарні ймовірності станів ПМП виражаються через стационарні ймовірності станів вкладеної в розглядається ПМП мар-

ковської ланцюга і визначаються для розглядається моделі з системи рівнянь [14]:

$$\begin{cases} \pi_0 = \pi_{00} \cdot p_{00-0} + \pi_1 \cdot p_{1-0}; \\ \pi_1 = \pi_0 \cdot p_{0-1} + \pi_{01} \cdot p_{01-1} + \pi_2 \cdot p_{2-1}; \\ \pi_2 = \pi_1 \cdot p_{1-2} + \pi_{02} \cdot p_{02-2} + \pi_3 \cdot p_{3-2}; \\ \pi_3 = \pi_2 \cdot p_{2-3} + \pi_{03} \cdot p_{03-3}; \\ \pi_{00} = \pi_0 \cdot p_{0-00} + \pi_{01} \cdot p_{01-00}; \\ \pi_{01} = \pi_{00} \cdot p_{00-01} + \pi_1 \cdot p_{1-01} + \pi_{02} \cdot p_{02-01}; \\ \pi_{02} = 1 - \pi_0 - \pi_{00} - \pi_1 - \pi_{01} - \pi_2 - \pi_3 - \pi_{03}; \\ \pi_{03} = \pi_{02} \cdot p_{02-03} + \pi_3 \cdot p_{3-03}. \end{cases}$$

Відповідно до [14], ймовірності досліджуємого ПМП визначаються виразом  $P_i = \pi_i \cdot \bar{t}_i / \bar{t}$ , де  $\bar{t}_i$  – математичне очікування часу одного переходу,  $\bar{t} = \sum_{i=1}^n \pi_i \cdot \bar{t}_i$ .

Середнє час нароботки на відмову  $t_{\text{MTBF}}$  робочої станції власника ресурсу відповідає даним з технічної документації про середнє час роботи до відмови персонального комп'ютера відповідно до технічними умовами ТУ У 05837085.001-97. Час  $t_{\text{MTTR}}$  вибрано на основі експертної оцінки часу, необхідного для ремонту/заміни компонентів ПК системним адміністратором або технічним спеціалістом. Параметри  $C, B, r, a, \tau_p$  вибрані на основі вимірювань, опублікованих в [11]. Значення вхідних параметрів моделі зведені в табл. 1:

Таблиця 1

Вхідні параметри

Параметр	Значення	Параметр	Значення
$t_{\text{MTBF}}, \text{ч}$	10000	B	1
$t_{\text{MTTR}}, \text{ч}$	2	C	1
$\tau_p, \text{ч}$	72	r	2
$\lambda_0, \text{ч}^{-1}$	0,0001	a	0,1
$\mu_0, \text{ч}^{-1}$	0,5		

## Результати

Для виконання розрахунків був використаний пакет прикладних програм Maple версії 13. З урахуванням прийнятих допущень при вибраному наборі вхідних параметрів доступність сервісу  $K_a$  визначається як сума ймовірностей станів, в яких доступні всі частини ресурсу:

$$K_a = P_0 + P_1 + P_2 + P_3 + P_{03}.$$

В залежності від вимог до системи в суму можуть бути включені тільки ті ймовірності станів, в яких час відповіді на запит визначених частин ресурсу не перевищує певного порогового значення. Результати моделювання представлені на графіку нижче (рис. 7).

## Висновки

Аналіз отриманих результатів дає можливість зробити висновок, що функція доступності сервісу має провал на початку часу експлуатації.

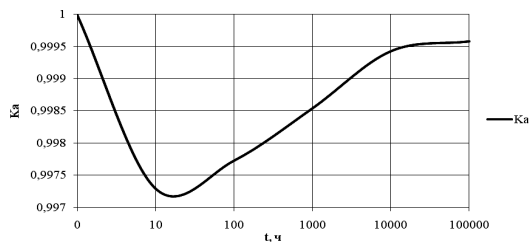


Рис. 7. Результати моделювання

Это объясняется тем, что для произвольных распределений могут наблюдаться провалы функции готовности, в том числе, ниже ее стационарного значения, а в общем случае функция готовности может иметь колебательный характер. Таким образом, в отличие от экспоненциального случая, когда функция готовности является монотонно убывающей, степенной закон распределения времени пребывания системы в состояниях полумарковской модели доступности сервиса одноранговой распределенной облачной системы обуславливает немонотонный характер функции доступности. Следовательно, доступность сервисов такой системы для небольшого времени эксплуатации меньше, чем при ее длительной эксплуатации. Этот факт необходимо учитывать на практике, так как время наступления стационарного режима может значительно увеличиваться.

Важным направлением дальнейших исследований является определение количества узлов репликации, достаточного для автономной реализации и стабильной работы сервиса на базе одноранговой распределенной облачной инфраструктуры без сервера.

## Список литературы

1. Yanovskaya O. The Concept of Green Cloud Infrastructure Based on Distributed Computing and Hardware Accelerator within FPGA as a Service [Text] / O. Yanovskaya, M. Yanovsky, V. Kharchenko // EWDTS-2014 (East-West Design and Test Symposium 2014). Proc. of IEEE East-West Design & Test Symposium (EWDTS 2014). – P. 45-48.
2. Gosh R. Modeling and performance analysis of large scale IaaS Clouds [Text] / R. Ghosh, F. Longo, V.K. Naik, K.S. Trivedi // Future Generation Computer Systems. – 2013. – Vol. 29 (5). – P. 1216-1234.

3. Dong S.K. Availability Modeling and Analysis of a Virtualized System / S.K. Dong, F. Machida, K.S. Trivedi. // 15th IEEE Pacific Rim International Symposium on Dependable Computing PRDC '09. – 2009. – P. 365-371.

4. Dependability Modeling and Analysis for the Virtual Data Center of Cloud Computing [Text] / B. Wei, C. Lin, X. Kong // IEEE International Conference on High Performance Computing and Communications. – 2011. – P. 784-789.

5. Resilient hosting in a continuous available virtualized environment [Text] / R.R. Scadden, R.J. Bogdany; J.W. Clifford, H.D. Pearthree // IBM Systems Journal / IBM. – Atlanta, 2008. – Vol. 47(4). – P. 535-548.

6. Thein T. Availability Analysis of Application Servers Using Software Rejuvenation and Virtualization [Text] / T. Thein, J. S. Park // Journal of Computer Science and Technology / Institute of Computing Technology Beijing. – Beijing, 2009. – Vol. 24(2). – P. 339 – 346.

7. Dai Y.S. Cloud Service Reliability: Modeling and Analysis [Electronic resource] / Y.S. Dai, B. Yang, J. Dongarra, G. Zhang. – Access mode: <http://www.netlib.org/utk/people/JackDongarra/PAPERS/>.

8. Dai Y.S. Reliability of grid service systems [Text] / Y.S. Dai, M. Xie, K.L. Poh. // Computers & Industrial Engineering / Pergamon Press. – Tarrytown, 2006. – Vol. 50(1-2). – P. 130 – 147.

9. Харченко В.С. Разработка марковских моделей надежности компьютерной сети информационно-управляющей системы с использованием FME(C)A-таблиц [Текст] / В.С. Харченко, И.Э. Комари // Системы обработки информации. – X: ХУПС, 2007. – Вып. 1 (59). – С. 144-147.

10. Яковлев А.В. Надежность информационных систем [Текст] / А.В. Яковлев. – Муром, 2004. – 63 с.

11. A Tale of the Tails: Power-laws in Internet Measurements [Text] / A. Mahanti, N. Carlsson, A. Mahanti et al. // IEEE Network. – 2013. – P. 59-64.

12. Power-Law Distributions in Empirical Data [Text] / A. Clauset, C.R. Shalizi, and M.E.J. Newman // SIAM Review. – 2009. – Vol. 51, no. 4, – P. 661-703.

13. Web Caching and Zipf-like Distributions: Evidence and Implications [Text] / L. Breslau, P. Cao, L. Fan et al. // IEEE INFOCOM Conference, 1999.

14. Волков Л.И. Управление эксплуатацией летательных комплексов [Текст] / Л.И. Волков // М.: Высшая школа, 1981. – 368 с.

Поступила в редколлегию 11.12.2015

**Рецензент:** д-р техн. наук проф. А.В. Горбенко, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков.

## МОДЕЛІ ДОСТУПНОСТІ СЕРВІСІВ РОЗПОДІЛЕНИХ ХМАРНИХ СИСТЕМ

О.В. Яновська

Розглянуто основні проблеми, що виникають при розробленні й дослідженні моделей доступності сервісів розподілених хмарних систем. Запропоновано напівмарковські моделі доступності сервісів розподілених хмарних систем в залежності від виду ресурсу, типу частин ресурсу, принципу побудови моделі, наявності резервного сервера та інших критеріїв. Виконано структурний і параметричний синтез запропонованих моделей. Проведено аналіз отриманих значень коефіцієнта доступності складеного ресурсу на основі характеристик напівмарковських процесів.

**Ключові слова:** розподілена хмарна архітектура, однорангові мережі, напівмарковські процеси, хмарні сервіси, доступність, готовність, надійність.

## SERVICE ACCESSIBILITY MODELS OF DISTRIBUTED CLOUD SYSTEMS

O.V. Yanovskaya

The paper discusses the main problems related to development and research of service availability models of distributed cloud systems. Semi-Markov service availability models of distributed cloud system depending on resource type, principle of creating the model, presence of backup always-on server and other criteria are proposed. Structural and parametric synthesis of the proposed models is performed. The obtained accessibility coefficient value of the composite resource based on the characteristics of Semi-Markov processes is analyzed.

**Keywords:** distributed cloud architecture, peer-to-peer networks, Semi-Markov processes, cloud services, accessibility, availability, reliability.