

УДК 004.75.05

А.В. Горбенко

Национальный аэрокосмический университет им. Н.Е. Жуковского "ХАИ", Харьков

МОДЕЛИ ЗАКОНОВ РАСПРЕДЕЛЕНИЯ ВРЕМЕНИ ОБСЛУЖИВАНИЯ РЕЗЕРВИРОВАННЫХ СЕРВИС-ОРИЕНТИРОВАННЫХ СИСТЕМ

В статье предложены модели законов дискретных и непрерывных распределения случайного времени обслуживания резервированных сервис-ориентированных систем. Рассмотрены два базовых метода резервирования: с ожиданием ответа от всех резервных сервисов или только первого полученного. Полученные результаты являются основой для оценки среднего времени обслуживания резервированных сервис-ориентированных систем при известных законах распределения случайного времени обслуживания резервируемых сервисов.

Ключевые слова: *web-сервисы, резервированные сервис-ориентированные системы, время обслуживания, закон распределения.*

Введение

Результатом эволюции распределенных систем обработки информации, парадигмы компонентно-интегрированного проектирования и Интернет-технологий является концепция сервис-ориентированной архитектуры (Service Oriented Architecture, SOA) [1] для создания распределенных информационно-вычислительных систем, компонентами которых являются Web-сервисы.

Web-сервисы – это новый вид Интернет-приложений, которые обеспечивают динамическое взаимодействие разнородных систем на основе использования общих стандартов представления информации и сетевых протоколов [2]. Т.е., Web-сервис представляет собой программное приложение или информационный ресурс, доступ к которому осуществляется удаленно, используя стандартные протоколы Интернет.

Важным является то, что Web-сервисы позволяют получать доступ к программным компонентам, независимо от того, как эти компоненты реализованы, где и на какой платформе они расположены, кем предоставляются и какие физические или информационные ресурсы используют. Web-сервисы поддерживают интеграцию и взаимодействие с другими программными компонентами с помощью XML-подобных сообщений. Областью применения сервис-ориентированной архитектуры и технологий Web-сервисов являются системы электронной коммерции и бизнеса (e-commerce), Интернет-банкинга (e-banking), теле-медицины (e-health), науки (e-science), глобальные системы распределенного хранения, поиска и обработки больших объемов информации (grid-системы), например, информационные системы мониторинга окружающей среды, системы оперативной обработки результатов космического зондирования и прочие.

Надежность и производительность распределенных информационно-вычислительных систем и компонентов (программных, аппаратных, сетевых),

а также их устойчивость к влиянию внешней среды является определяющим фактором надежности и безопасности широкого спектра приложений критического и бизнес-критического применения, электронной коммерции и современной науки.

Одной из наиболее актуальных задач для сервис-ориентированных систем является оценка производительности и времени обслуживания. Об этом свидетельствуют приоритетные направления финансирования научных исследований международными и европейскими грантами, прежде всего FP6 и FP7 и многочисленные публикации в ведущих научных журналах, например [3 – 5].

Распределенный характер взаимодействия компонентов сервис-ориентированных систем, неопределенность их характеристик, а также многочисленные воздействующие факторы обуславливают случайный характер задержек, возникающих в таких системах. Это затрудняет оценку временных характеристик обслуживания в сервис-ориентированных системах, особенно в случае использования различных методов резервирования [6, 7].

Целью статьи является оценка среднего времени обслуживания резервированных сервис-ориентированных систем на основе определения комбинированного закона распределения времени обслуживания резервированных сервис-ориентированных систем при известных законах распределения для резервируемых сервисов.

1. Методы резервирования web-сервисов

Для повышения надежности и производительности сервис-ориентированных систем в работах [6, 7] предложены различные методы резервирования. Из них наибольший практический интерес представляют следующие методы:

1) метод параллельного выполнения резервных сервисов с ожиданием ответа от всех сервисов (ПВОВ) и формированием результата по принципу

большинства – метод мажоритарного резервирования, направленный на повышение достоверности результата обслуживания;

2) метод параллельного выполнения резервных сервисов с ожидания только первого ответа (ПВОП) – метод резервирования, предназначенный для повышения готовности и снижения времени обслуживания.

В первом случае среднее время ожидания ответа от всех сервисов будет выше среднего времени обслуживания каждого из них из-за того, что время обслуживания каждого из сервисов является случайной величиной [5]. Таким образом, повышение достоверности при использовании метода резервирования ПВОВ одновременно с повышением достоверности обслуживания приводит к увеличению времени обслуживания.

В то же время метод ПВОП не позволяет повысить достоверность обслуживания, поскольку отправляет пользователю первый полученный результат не дожидаясь получения оставшихся результатов.

2. Среднее время обслуживания web-сервисов

Случайный характер времени обслуживания web-сервисов может быть описан с помощью соответствующего закона распределения случайной величины. Например, в [5] представлены результаты определения адекватного закона распределения времени обслуживания web-сервисов на основе сбора и обработки статистических данных их вызова.

При известном законе распределения времени обслуживания сервиса f_X среднее время обслуживания может быть рассчитано, используя известное выражение для вычисления математического ожидания случайной величины:

$$M[X] = \int_0^{\infty} x \cdot f_X(x) dx .$$

Очевидно, что в случае резервирования сервисов среднее время обслуживания будет зависеть от используемого метода резервирования, а также количества резервных сервисов.

3. Оценка времени обслуживания методов резервирования при использовании дискретных законов распределения

Рассмотрим оценку математического ожидания дискретного закона распределения случайной величины x . Пусть имеется два сервиса S_1 и S_2 , для каждого из которых известен ряд распределения случайного времени обслуживания X_1 и X_2 (табл. 1). Тогда математическое ожидание будет равно

$$M[X] = \sum_{i=1}^n x_i \cdot p_i \quad \text{при условии, что} \quad \sum_{i=1}^n p_i = 1 .$$

Для рассматриваемого примера величина математического ожидания первого сервиса будет равна 190 мс, а второго – 170 мс.

Таблица 1

Ряд распределения случайного времени обслуживания сервисов S_1 и S_2

$x_{i,j}$, мс	100	200	300
$p_{1,j}$	0,3	0,5	0,2
$p_{2,j}$	0,5	0,3	0,2

Проанализируем, каким образом изменится среднее время обслуживания для при использовании метода резервирования ПВОВ. Время ожидания обслуживания при таком способе вызова сервисов будет равно 100 мс при условии, что время выполнения каждого из них также равно 100 мс. Таким образом, совместная вероятность этих событий $P(100)$ равна произведению вероятностей $p_{1,1}$ и $p_{2,1}$: $P(100) = 0,3 \cdot 0,5 = 0,15$. Время ожидания обслуживания, равное 200 мс, будет возможным в следующих случаях:

- 1) время выполнения двух сервисов равно 200 мс;
- 2) время выполнения первого сервиса равно 200 мс, в то время как второго – 100 мс;
- 3) время выполнения второго сервиса равно 200 мс, в то время как первого – 100 мс.

В результате для вероятности времени обслуживания, равного 200 мс, справедливым будет выражение: $P(200) = p_{1,2} \cdot p_{2,2} + p_{1,2} \cdot p_{2,1} + p_{1,1} \cdot p_{2,2}$. Таким образом, $P(200) = 0,5 \cdot 0,3 + 0,5 \cdot 0,5 + 0,3 \cdot 0,3 = 0,49$.

Наконец, время ожидания обслуживания, равное 300 мс, будет возможным в следующих случаях:

- 1) время выполнения двух сервисов равно 300 мс;
- 2) время выполнения первого сервиса равно 300 мс, в то время как второго – любое меньшее, т.е. 100 мс или 200 мс;
- 3) время выполнения второго сервиса равно 300 мс, в то время как первого – любое меньшее, т.е. 100 мс или 200 мс.

В результате для вероятности времени обслуживания, равного 300 мс, справедливым будет выражение: $P(300) = p_{1,3} \cdot p_{2,3} + p_{1,3} \cdot (p_{2,1} + p_{2,2}) + p_{2,3} \cdot (p_{1,1} + p_{1,2})$, т.е. $P(300) = 0,2 \cdot 0,2 + 0,2 \cdot (0,5 + 0,3) + 0,2 \cdot (0,3 + 0,5) = 0,36$.

Проверка полученных результатов, основанная на том, что сумма вероятностей всех возможных исходов, а именно $P(100)$, $P(200)$ и $P(300)$, должна быть равна единице, подтверждает правильность рассуждений: $P(100) + P(200) + P(300) = 0,15 + 0,49 + 0,36 = 1$.

Аналогичным образом проанализируем закономерность изменения среднего времени обслуживания для метода резервирования ПВОП.

Время ожидания обслуживания при таком способе вызова сервисов будет равно 300 мс при условии, что время выполнения каждого из них также равно 300 мс. Таким образом, совместная вероятность этих событий $P(300)$ равна произведению вероятностей $p_{1,3}$ и $p_{2,3}$: $P(300) = 0,2 \cdot 0,2 = 0,04$.

Время ожидания обслуживания, равное 200 мс, будет возможным в следующих случаях:

- 1) время выполнения двух сервисов равно 200 мс;
- 2) время выполнения первого сервиса равно 200 мс, в то время как второго – 300 мс;

3) время выполнения второго сервиса равно 200 мс, в то время как первого – 300 мс.

В результате для вероятности времени обслуживания, равного 200 мс, справедливым будет выражение: $P(200) = p_{1,2} \cdot p_{2,2} + p_{1,2} \cdot p_{2,3} + p_{1,3} \cdot p_{2,2}$. Таким образом, $P(200) = 0,5 \cdot 0,3 + 0,5 \cdot 0,2 + 0,3 \cdot 0,2 = 0,31$.

Наконец, время ожидания обслуживания, равное 100 мс, будет возможным в следующих случаях:

1) время выполнения двух сервисов равно 100 мс;

2) время выполнения первого сервиса равно 100 мс, в то время как второго – любое большее, т.е. 200 мс или 300 мс;

3) время выполнения второго сервиса равно 100 мс, в то время как первого – любое большее, т.е. 200 мс или 300 мс.

В результате для вероятности времени обслуживания, равного 100 мс, справедливым будет выражение: $P(100) = p_{1,1} \cdot p_{2,1} + p_{1,1} \cdot (p_{2,2} + p_{2,3}) + p_{2,1} \cdot (p_{1,2} + p_{1,3})$.

В результате $P(100) = 0,3 \cdot 0,5 + 0,3 \cdot (0,3 + 0,2) + 0,5 \cdot (0,5 + 0,2) = 0,65$.

Проверка полученных результатов, основанная на том, что сумма вероятностей всех возможных исходов, а именно $P(100)$, $P(200)$ и $P(300)$, должна быть равна единице, подтверждает правильность рассуждений: $P(100) + P(200) + P(300) = 0,65 + 0,31 + 0,04 = 1$.

Результаты анализа рядов распределения случайного времени обслуживания при надежностно-компонентной интеграции двух сервисов для разных шаблонов выполнения представлены на рис. 1.

Среднее время обслуживания резервированной сервис-ориентированной системы, соответствующее полученным рядам распределения случайного времени обслуживания, будет равно: 221 мс – в случае метода ПВОВ; 139 мс – в случае метода ПВОП.

Полученные результаты согласуются с ожидаемыми, а также результатами имитационного моделирования, представленными в работе [6].

На основе проведенного анализа могут быть получены общие аналитические выражения для определения ряда распределения времени обслуживания для рассматриваемых методов резервирования при известных дискретных законах распределения времени обслуживания сервисов.

При использовании метода ПВОВ, элемент ряда распределения общего времени обслуживания определяется по формуле:

$$P_i = p_{1,i} \cdot \sum_{j=1}^i p_{2,j} + p_{2,i} \cdot \sum_{j=1}^i p_{1,j} - p_{1,i} \cdot p_{2,i} \quad (1)$$

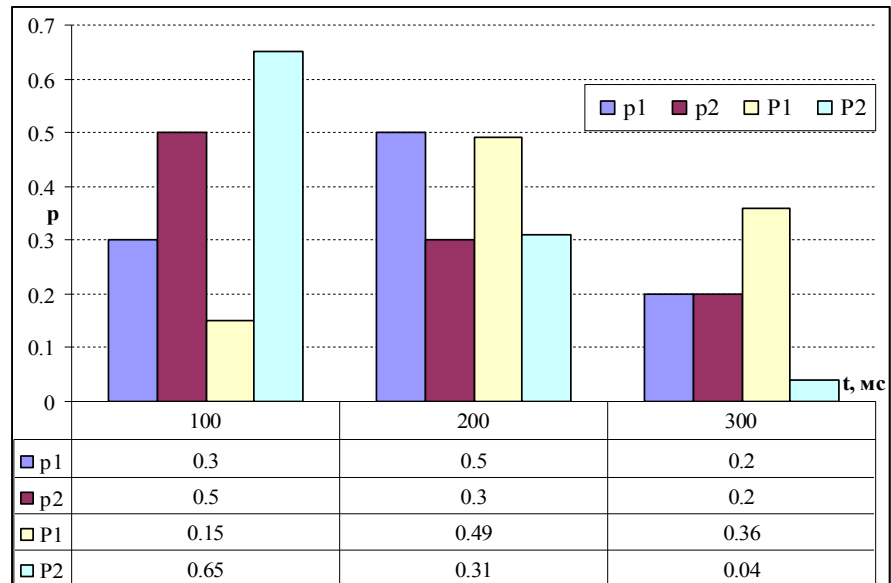


Рис. 1. Ряд распределения случайного времени обслуживания: p_1 – сервиса S_1 ; p_2 – сервиса S_2 ; P_1 – резервирование S_1 и S_2 при использовании метода ПВОВ; P_2 – резервирование S_1 и S_2 при использовании метода ПВОП

В случае использования метода ПВОП для вычисления значения элемент i -го ряда распределения времени обслуживания используется выражение:

$$P_i = p_{1,i} \cdot \sum_{j=i}^m p_{2,j} + p_{2,i} \cdot \sum_{j=i}^m p_{1,j} - p_{1,i} \cdot p_{2,i} \quad (2)$$

где m – количество элементов дискретного ряда распределения времени обслуживания.

В общем случае, для резервирования n функционально-эквивалентных сервисов при использовании метода с параллельным выполнением и ожиданием всех ответов было получено выражение:

$$P_i = \sum_{k=1}^n \left(p_{k,i} \cdot \prod_{j=1}^n \sum_{q=1}^i p_{j,q} / \sum_{q=1}^i p_{k,q} \right) - (n-1) \cdot \prod_{k=1}^n p_{k,i}$$

При использовании метода резервирования n сервисов с параллельным выполнением и ожиданием только первого было получено выражение:

$$P_i = \sum_{k=1}^n \left(p_{k,i} \cdot \prod_{j=1}^n \sum_{q=1}^m p_{j,q} / \sum_{q=1}^m p_{k,q} \right) - (n-1) \cdot \prod_{k=1}^n p_{k,i}$$

4. Оценка времени обслуживания методов резервирования при использовании непрерывных законов распределения

Ключевым моментом для определения среднего времени обслуживания резервированных сервис-ориентированных систем при известных законах распределения времени обслуживания резервируемых и резервных сервисов является определение результирующей (композиционной) функции плотности распределения. Для получения непрерывной композиционной функции распределения выполним анализ выражений (1) и (2) с учетом того, что при переходе

от дискретного закона распределения случайной величины к непрерывному, ряд распределения времени обслуживания сервисов заменяется функцией плотности распределения, сумма последовательных элементов ряда – интегралом, а вероятность конкретного значения случайной величины равна нулю.

В результате для функции плотности распределения времени обслуживания при использовании метода резервирования двух сервисов S_1 и S_2 с ожиданием всех ответов получено выражение:

$$f_{\text{ПВОВ}}(x) = f_1(x) \cdot \int_0^x f_2(t) dt + f_2(x) \cdot \int_0^x f_1(t) dt = \\ = f_1(x) \cdot F_2(x) + f_2(x) \cdot F_1(x),$$

где $f_1(x)$, $f_2(x)$ – функции плотности распределения вероятности случайного времени обслуживания сервисов S_1 и S_2 соответственно; $F_1(x)$, $F_2(x)$ – функции распределения случайного времени обслуживания сервисов S_1 и S_2 .

По аналогии, для шаблона параллельного выполнения с ожиданием только первого ответа имеем:

$$f_{\text{ПВОП}}(x) = f_1(x) \cdot \int_x^{\infty} f_2(t) dt + f_2(x) \cdot \int_x^{\infty} f_1(t) dt = \\ = f_1(x) \cdot (1 - F_2(x)) + f_2(x) \cdot (1 - F_1(x)).$$

В общем случае, при использовании n резервных сервисов, участвующих в надежность-компонентной интеграции, полученные выражения могут быть преобразованы к следующему виду:

$$f_{\text{ПВОВ}}(x) = \sum_{i=1}^n \left(f_i(x) \cdot \prod_{j=1}^n F_j(x) / F_i(x) \right), \quad (3)$$

$$f_{\text{ПВОП}}(x) = \sum_{i=1}^n \left(f_i(x) \cdot \prod_{j=1}^n (1 - F_j(x)) / (1 - F_i(x)) \right). \quad (4)$$

Среднее время обслуживания при использовании того или иного шаблона резервирования может быть получено на основе известного выражения для расчета математического ожидания при известном законе распределения случайной величины.

5. Пример оценки среднего времени обслуживания резервированных сервис-ориентированных систем

Рассмотрим практический пример оценки среднего времени обслуживания для шаблонов резервирования с ожиданием и без, если известны законы распределения целевых сервисов.

Пусть имеется два диверсных сервиса S_1 и S_2 , для каждого из которых случайное время обслуживания подчиняется экспоненциальному закону распределения $f_1(t) = f_2(t) = \mu \cdot e^{-\mu \cdot t}$ с параметрами $\mu_1 = \mu_2 = \mu = 0.01$, что соответствует математическому ожиданию времени обслуживания каждого сервиса 100 мс.

Тогда, с учетом (3) для функции плотности распределения вероятности времени обслуживания при использовании метода резервирования ПВОВ:

$$f_{\text{ПВОВ}}(x) = \mu_1 \cdot e^{-\mu_1 \cdot t} \cdot (1 - e^{-\mu_2 \cdot t}) + \\ + \mu_2 \cdot e^{-\mu_2 \cdot t} \cdot (1 - e^{-\mu_1 \cdot t}) = 2 \cdot \mu \cdot e^{-\mu \cdot t} \cdot (1 - e^{-\mu \cdot t}),$$

а при использовании метода ПВОП:

$$f_{\text{ПВОП}}(x) = \mu_1 \cdot e^{-\mu_1 \cdot t} \cdot e^{-\mu_2 \cdot t} + \mu_2 \cdot e^{-\mu_2 \cdot t} \cdot e^{-\mu_1 \cdot t} = \\ = 2 \cdot \mu \cdot e^{-2 \cdot \mu \cdot t}.$$

Проверка показывает, что интеграл от нуля до бесконечности от функций $f_{\text{ПВОВ}}(x)$ и $f_{\text{ПВОП}}(x)$ равен единице, что свидетельствует о корректности полученных функций плотности распределения. Анализ графиков этих функций, представленных на рис. 2, также подтверждают, что полученные теоретические результаты согласуются с логическим заключением о том, что использование метода резервирования ПВОВ приводит к увеличению среднего времени обслуживания, а использование метода ПВОП – к уменьшению. Применяя известную формулу для расчета математического ожидания случайной величины к функциям $f_{\text{ПВОВ}}(x)$ и $f_{\text{ПВОП}}(x)$ в среде математического моделирования MathCAD были получены следующие результаты:

- среднее время резервирования сервисов S_1 и S_2 при использовании метода ПВОВ составляет 150 мс;
- среднее время резервирования сервисов S_1 и S_2 при использовании метода ПВОП составляет 50 мс.

Заключение

Оценка и обеспечение гарантированности комплексных *сервис-ориентированных систем* (СОС) существенно усложнена при условиях их динамического построения и интеграции или в случаях, когда компоненты этих систем (т.е. Web-сервисы) динамично заменяют друг друга выполняя тоже самую или альтернативную функциональность, но имея неизвестные или не до конца определенные характеристики гарантированности. Эта неопределенность выражается в непредвиденности времени отзыва Web-сервисов, сложности диагностирования первопричины возникающих отказов, а также прогнозирования отказов по общим причинам, отсутствию информации об особенностях реализации Web-сервисов, и т.п.

Как следствие, усложняется задача выбора между имеющимися альтернативными компонентами. Во-вторых, становится невозможным оптимальное использование традиционных методов резервирования, а также оценка выигрыша от их применения.

В этом контексте в статье предложены модели композитных законов распределения для резервированных сервис-ориентированных систем. Эти модели позволяют прогнозировать среднее время обслуживания при использовании методов резервирования с ожиданием ответа от всех сервисов и ожиданием только первого ответа.

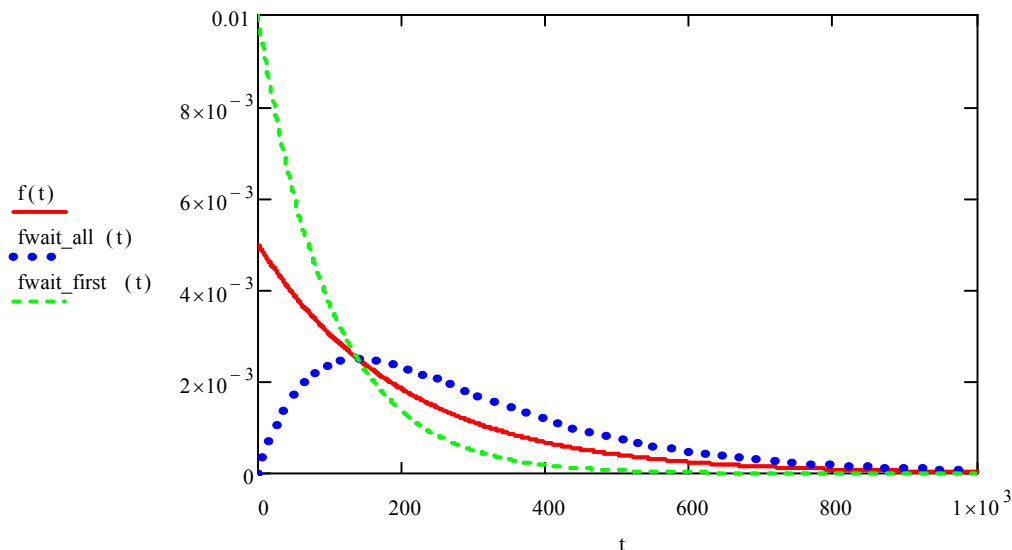


Рис. 2. Графік функцій щільності розподілу часу обслуговування: $f(t)$ – сервісів S_1 і S_2 ; $f_{\text{wait_all}}(x)$ – при резервуванні S_1 і S_2 з використанням методу ПВОВ; $f_{\text{wait_first}}(x)$ – при резервуванні S_1 і S_2 з використанням методу ПВОП

В першому випадку підвищення доступності і достовірності призводить до зниження продуктивності резервованої сервіс-орієнтованої системи, в той час як другою з розглянутих методів резервування дозволяється зменшити час обслуговування без підвищення достовірності повернутого результату.

Приведені приклади підтверджують вірність запропонованих аналітичних виражень (3) і (4), а також можливість їх практичного використання для оцінки часу обслуговування резервованих сервіс-орієнтованих систем при використанні методів резервування ПВОВ і ПВОП.

Список літератури

1. MacKenzie C.M. *OASIS Reference Model for Service Oriented Architecture. Ver. 1.0* / C.M. MacKenzie, K. Laskey, F. McCabe, etc. – Burlington: OASIS, 2006. – 31 p.
2. Хаббибуллин І.Ш. *Разработка Web-служб средствами Java* / И.Ш. Хаббибуллин. – СПб.: БХВ-Петербург, 2003. – 400 с.
3. Gonczy, L. *Dependability Evaluation of Web Service-Based Processes* / L. Gonczy, S. Chiaradonna, F. di Gian

domenico // *Proc. European Performance Engineering Workshop (EPEW'2006)*. – Bertinoro (Italy), 2006. – P. 166–180.

4. Menascé D.A. *Response-Time Analysis of Composite Web Services* / D.A. Menascé // *IEEE Internet Computing*. – 2004. – №1,2. – P. 90–94.

5. *Real Distribution of Response Time Instability in Service-Oriented Architecture* / A. Gorbenko, V. Kharchenko, S. Mamutov, O. Tarasyuk, Yu. Chen, A. Romanovsky // *Proc. 29th IEEE International Symposium on Reliable Distributed Systems (SRDS'2010)*. – Delhi (India), 2010. – P. 92–99.

6. Gorbenko A. *Using Inherent Service Redundancy and Diversity to Ensure Web Services Dependability*. / A. Gorbenko, V. Kharchenko, A. Romanovsky. In M. Butler, C. Jones, A. Romanovsky, E. Troubitsyna (Eds.) *“Methods, Models and Tools for Fault Tolerance”*, LNCS 5454, Springer, 2009, pp. 324–341.

7. Zheng, Z. *A QoS-Aware Fault Tolerant Middleware for Dependable Service Composition* / Z. Zheng, M. Lyu // *Proc. Int. Conf. On Dependable Systems and Networks (DSN'2009)*. – Lisbon (Portugal), 2009. – P. 239–248.

Поступила в редакцію 1.12.2011

Рецензент: д-р техн. наук, проф. В.С. Харченко, Національний аерокосмічний університет ім. Н.Е. Жуковського «ХАІ», Харків.

МОДЕЛІ ЗАКОНІВ РОЗПОДІЛУ ЧАСУ ОБСЛУГОВУВАННЯ РЕЗЕРВОВАНИХ СЕРВІС-ОРІЄНТОВАНИХ СИСТЕМ

А.В. Горбенко

У статті запропоновано моделі законів дискретних та безперервних розподілу випадкового часу обслуговування резервованих сервіс-орієнтованих систем. Розглянуто два базових методи резервування: з очікуванням відповіді від усіх резервних сервісів або тільки першої отриманої. Отримані результати є основою для оцінки середнього часу обслуговування резервованих сервіс-орієнтованих систем при відомих законах розподілу випадкового часу обслуговування резервних сервісів.

Ключові слова: web-сервіси, резервовані сервіс-орієнтовані системи, час обслуговування, закон розподілу.

MODELS OF SERVICING TIME DISTRIBUTION LAWS FOR REDUNDANT SERVICE-ORIENTED SYSTEMS

A.V. Gorbenko

Models of discrete and continuous distribution laws of random servicing time for redundant service-oriented systems are discussed in the paper. There are two basic redundancy techniques used: method of parallel execution of redundant services with waiting of all responses and method of parallel execution of redundant services with waiting only the first arrived response. The results obtained provide a foundation for average servicing time assessment of redundant service-oriented systems when response time distribution laws are known for all redundant services.

Keywords: web-services, redundant service-oriented systems, servicing time, distribution laws.