

УДК 004.75.05

А.В. ГОРБЕНКО, О.М. ТАРАСЮК

*Национальный аэрокосмический университет им. Н.Е. Жуковского "ХАИ", Украина*

## МЕТОДИКА МИНИМИЗАЦИИ ВРЕМЕНИ ОБСЛУЖИВАНИЯ СЕРВИС-ОРИЕНТИРОВАННЫХ СИСТЕМ

*В статье предложена методика минимизации времени обслуживания сервис-ориентированных систем, построенных на основе параллельно-последовательной интеграции Web-сервисов, с учетом надежностных ограничений. Методика базируется на использовании модели обслуживания Web-сервисов, устанавливающей соотношения между вероятностью обслуживания, средним временем обслуживания и максимальным временем ожидания – тайм-аутом, который используется в качестве параметра оптимизации.*

**Ключевые слова:** web-сервисы, сервис-ориентированные системы, время обслуживания, минимизация.

### Введение

Концепция *сервис-ориентированной архитектуры* (СОА) определяет современный подход к созданию распределенных информационных систем путём комбинации удаленных слабо-связанных программных и информационных компонентов – Web-сервисов, взаимодействующих на основе некоторого платформенно-независимого и языково-независимого протокола [1]. В основе СОА лежат принципы многократного использования функциональных программных компонентов, ликвидации дублирования, унификации типовых операционных процессов и их кросс-платформенной интеграции.

Компоненты сервис-ориентированной системы (СОС) могут быть распределены по разным узлам глобальной компьютерной сети Интернет, а их интеграция выполняется на основе последовательного или параллельного вызова.

Однако отсутствие достаточной информации о деталях реализации и нефункциональных характеристиках Web-компонентов, которые принимают участие в интеграции, их принадлежность к разным административным доменам, распределенный асинхронный характер взаимодействия и нестабильность Интернет-среды создают значительные проблемы в прогнозировании показателей гарантоспособности и производительности сервис-ориентированных систем [2]. Вследствие этого, потребители Web-услуг и пользователи сервис-ориентированных систем не могут быть уверенными в определенном времени обслуживания, готовности, безотказности и достоверности. Таким образом, неопределенность характеристик гарантоспособности и производительности является основным вызовом для научных работников и разработчиков таких систем [2, 3].

Существующие экспериментальные исследования [4] свидетельствуют о том, что характеристики производительности Web-сервисов – компонентов сервис-ориентированных систем могут изменяться случайным образом в достаточно широких пределах. Неопределенность времени обслуживания Web-сервисов может быть описана с помощью так называемых *heavy-tailed* законов распределения случайной величины с тяжелым (весомым) «хвостом», к которым относится распределение Вейбулла [5].

Глобально-распределенный характер взаимодействия и отсутствие информации о деталях реализации web-сервиса не гарантируют получение ответа от web-сервиса или же то, что ответ обязательно будет получен за некоторое фиксированное время.

Поэтому при построении распределенных информационных систем распространенным решением является использование таймеров, ограничивающих ожидание ответа от удаленной системы или компонента. Такие таймеры используются как на прикладном уровне, так и входят в реализацию различных коммуникационных протоколов, TCP, HTTP, SOAP и др. В результате, при функционировании сервис-ориентированных систем возможна ситуация, когда за установленное время ожидания от удаленного web-сервиса не будет получен ответ.

Целью статьи является исследование вероятностно-временной взаимосвязи между различными вариантами обслуживания при использовании Web-сервисов, а также разработка методики оптимизации параметров ожидания обслуживания – тайм-аутов при создании сервис-ориентированных систем на основе параллельно-последовательной интеграции Web-сервисов для минимизации времени обслуживания при установленных надежностных ограничениях.

## 1. Вероятностно-временная модель обслуживания web-сервисов

В случае вероятностной неопределенности время обслуживания, которое является случайной величиной, описывается некоторым законом распределения  $f_i(t)$ .

С учетом того, что на практике время ожидания ответа от Web-сервиса всегда ограничено, возможны два варианта обслуживания, характеризующиеся вероятностями:

$p^{ok}$  – вероятность получения за время ожидания результат обслуживания (вероятность обслуживания);

$p^{to}$  – вероятность того, что за установленное время ожидания от сервиса не будет получен ответ (вероятность необслуживания за время ожидания).

Взаимосвязь вероятностей  $p^{ok}$  и  $p^{to}$  иллюстрируется на рис. 1.

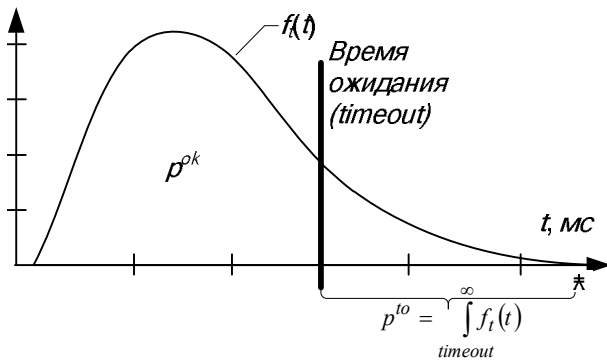


Рис. 1. Вероятностно-временная взаимосвязь результатов обслуживания web-сервиса

Очевидно, что сумма этих вероятностей равна единице, а их отношение зависит от установленного времени ожидания – тайм-аута:

$$p^{ok} = \int_0^{timeout} f_t(t) dt = 1 - p^{to}.$$

При использовании тайм-аута среднее время обслуживания  $t^{av}$  является математическим ожиданием усеченного закона распределения, которое может быть вычислено с помощью выражения [6]:

$$M^{trunc}[X] = \frac{\int_A^B x \cdot f_X(x) dx}{\int_A^B f_X(x) dx} = \frac{\int_A^B x \cdot f_X(x) dx}{F_X(B) - F_X(A)},$$

где  $A$  – левая граница усеченного закона распределения, которая в нашем случае равна нулю, а  $B$  – правая граница усечения, равная установленному значению тайм-аута.

## 2. Оптимизационная задача минимизации времени обслуживания последовательной интеграции web-сервисов

Как следует из анализа графика на рис. 1, оптимальным решением задачи минимизации времени обслуживания без дополнительных ограничений будет значение тайм-аута, стремящееся к нулю, при котором время обслуживания также будет приближаться к нулю. Однако в этом случае вероятность необслуживания будет стремиться к единице, что, очевидно, не является желаемым результатом.

Таким образом, в качестве ограничения для задачи минимизации времени обслуживания необходимо использовать минимально-допустимое значение вероятности обслуживания  $P^{req}$ .

В случае сервис-ориентированной системы, основанной на *последовательной интеграции* нескольких Web-сервисов, вероятность корректного обслуживания  $P^{ok}_{COC}$  определяется как произведение вероятностей корректного обслуживания отдельных сервисов, а среднее время обслуживания  $T^{av}_{COC}$  является суммой их среднего времени обслуживания. Если надежность СОС превышает требуемое значение, то существует принципиальная возможность минимизировать среднее время обслуживания СОС на основе выбора оптимальных значений тайм-аута для сервисов, входящих в состав СОС.

В результате, постановка оптимизационной задачи минимизации времени обслуживания при последовательной интеграции  $n$  сервисов в общем виде может быть сформулирована следующим образом:

$$\begin{aligned} f(X) &= \sum_{i=1}^n t_i^{av}(x_i) \rightarrow \min, \\ g(X) &= \prod_{i=1}^n P_i^{ok}(x_i) \geq P^{req}, \\ x_i &\geq 0, \end{aligned} \quad (1)$$

где  $f(X)$  – целевая функция оптимизации – среднее время обслуживания, являющееся суммой среднего времени выполнения всех последовательно-интегрируемых сервисов;  $g(X)$  – ограничение надежности (вероятности обслуживания) для СОС, являющееся произведением вероятностей корректного обслуживания входящих в неё сервисов;  $x_i$  – искомое оптимальное время ожидания ответа от  $i$ -го сервиса (тайм-аут).

Данная оптимизационная задача согласуется с задачей нахождения оптимального размера очередей коммуникационных устройств в сетях передачи информации при заданном ограничении на вероятность потери заявок с целью минимизации времени

ожидания в очереди [7].

Для решения оптимизационной задачи должны быть детализированы функции среднего времени обслуживания индивидуальных сервисов, а также вероятности обслуживания на основе конкретного закона распределения времени обслуживания.

Рассматриваемая оптимизационная задача относится к классу задач нелинейного программирования с ограничениями в виде неравенств. Общим подходом к решению таких задач является использование условий Куна-Таккера, развивающих метод множителей Лагранжа. Более эффективным способом решения является метод сепарабельного программирования, применимый в случае, если целевая функция и ограничения являются сепарабельными, т.е. могут быть представлены в виде суммы функций одной переменной. В нашем случае целевая функция является непосредственно сепарабельной в то время, как ограничение на надежность сопряженно-компонитной цепочки может быть приведено к сепарабельному виду путем замены произведения вероятностей корректного обслуживания индивидуальных сервисов суммой логарифмов этих значений.

### 3. Оптимизационная задача минимизации времени обслуживания параллельной интеграции web-сервисов

Время обслуживания сервис-ориентированной системы, использующей параллельную интеграцию Web-сервисов, определяется сервисом, имеющим наибольшее время обслуживания по сравнению с другими.

В то же время вероятность обслуживания СОС рассчитывается как произведение вероятностей об-

служивания всех сервисов, так же как и в случае их последовательной интеграции.

Очевидно, что в первую очередь должно быть оптимизировано значение тайм-аута сервиса, имеющего максимальное среднее время обслуживания. При этом нет необходимости поиска глобального минимума среднего времени обслуживания в том случае, если его значение будет меньше среднего времени обслуживания сервиса, второго по длительности. Тогда на втором этапе должна быть решена задача одновременной минимизации времени обслуживания уже двух Web-сервисов при установленном надежностном ограничении и т.д.

Предлагаемый подход к решению задачи минимизации времени обслуживания СОС при использовании параллельной интеграции Web-сервисов проиллюстрирован на рис. 2. На его основе может быть сформулирована общая методика минимизации времени обслуживания СОС при использовании параллельной интеграции Web-сервисов, базирующаяся на последовательном выполнении следующих операций.

1. Расчет и сравнение вероятности обслуживания сервис-ориентированной системы  $P^{ok}_{COC}$  с требуемым значением  $P^{req}_{COC}$ . Если полученное значение выше требуемого, то выполняются все последующие действия. В противном случае, минимизация среднего времени обслуживания  $T^{av}_{COC}$  сервис-ориентированной системы невозможна.

2. Расчет среднего времени обслуживания для каждого из параллельно-интегрируемых сервисов  $S_i$ , входящих в состав СОС, и их ранжирование по убыванию этой характеристики:  $COC = \langle S_i \rangle_{i=1}^n$ . В результате сервис, имеющий максимальное значение времени обслуживания будет иметь индекс, равный

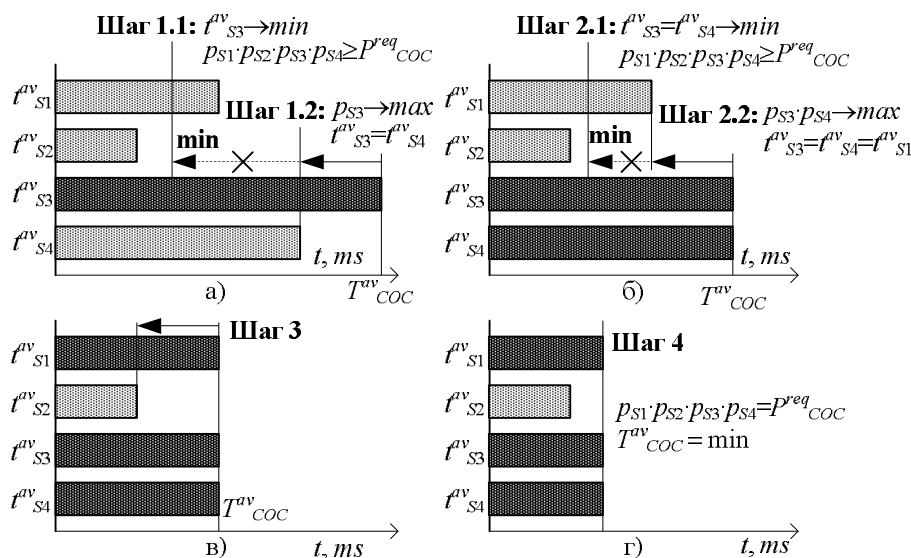


Рис. 2. Последовательность минимизации времени обслуживания параллельно-интегрированных сервисов

единице –  $S_1$ .

3. Определение «избытка» надежности СОС и вычисление граничного значения, до которого может быть уменьшена вероятность обслуживания для сервиса с максимальным временем обслуживания, при котором еще соблюдается требование к надежности всей СОС:

$$P_{S_1}^{ok'} = P_{COC}^{req} / \prod_{i=2}^n P_{S_i}^{ok}.$$

4. Вычисление оптимального значения тайм-аута для минимизации среднего времени обслуживания сервиса  $S_1$  при заданном ограничении на вероятность корректного обслуживания  $P_{S_1}^{ok'}$ .

5. Если полученное среднее время обслуживания сервиса  $S_1 - t_{S_1}^{av}$  больше или равно среднему времени обслуживания следующего, по величине этого времени, сервиса  $S_2$  ( $t_{S_1}^{av} \geq t_{S_2}^{av}$ ), то полученное решение является окончательным, а среднее время выполнения СОС равно  $T_{COC}^{av} = t_{S_1}^{av}$ .

6. В противном случае ( $t_{S_1}^{av} < t_{S_2}^{av}$ ) вычисляется оптимальное значение тайм-аута для максимизации вероятности корректного обслуживания сервиса  $S_1$  при заданном ограничении на среднее время обслуживания, равном значению среднего времени обслуживания следующего, по величине этого времени, сервиса ( $t_{S_1}^{av} = t_{S_2}^{av}$ ):

7. Определение «избытка» надежности СОС и вычисление граничного значения до которого может быть уменьшено произведение вероятностей обслуживания двух сервисов с максимальным временем обслуживания  $S_1$  и  $S_2$ , при котором еще соблюдается требование к надежности всей СОС:

$$\prod_{j=2}^n P_{S_j}^{ok'} = \frac{P_{COC}^{req}}{\prod_{i=3}^n P_{S_i}^{ok}}.$$

8. Постановка и решение задачи одновременной минимизации среднего времени обслуживания для сервисов  $S_1$  и  $S_2$  при заданном ограничении на вероятность обслуживания  $P_{S_1}^{ok'} \cdot P_{S_2}^{ok'}$ , а также с учетом дополнительного условия  $t_{S_1}^{av} = t_{S_2}^{av}$ . Для этого постановка задачи, предложенная в подразделе 2, должна быть расширена за счет введения дополнительного ограничения следующим образом:

$$f(X) = t_1^{av}(x_1) + t_2^{av}(x_2) \rightarrow \min,$$

$$g_1(X) = P_1^{ok}(x_1) \cdot P_2^{ok}(x_2) = P_{COC}^{req} / \prod_{i=3}^n P_{S_i}^{ok},$$

$$g_2(X) = t_1^{av}(x_1) = t_2^{av}(x_2)$$

$$x_1 \geq 0, x_2 \geq 0.$$

После решения этой оптимизационной задачи

выполняется переход к пункту 5, где выполняется анализ результата оптимизации и принимается решение о завершении или итеративном продолжении процесса оптимизации. В последнем случае выполняется соответствующее дополнение функции оптимизации и ограничений решаемых в п. 5 и п. 6 оптимизационных задач.

## Заключение

В общем случае сервис-ориентированная система представляет собой параллельно-последовательную интеграцию Web-сервисов.

Решение задачи минимизации времени обслуживания такой СОС может базироваться на методике, предложенной в разделе 3 с учетом того, что каждый из параллельно-интегрируемых сервисов может быть заменен цепочкой последовательно-выполняемых сервисов.

В результате, общая методика минимизации времени обслуживания СОС включает в себя постановку и решение двух типов оптимизационных задач:

1) задача *групповой равномерной* минимизации среднего времени обслуживания для последовательных цепочек сервисов, имеющих равное наибольшее среднее время обслуживания, при заданном ограничении на вероятность обслуживания. Эта задача может быть сформулирована следующим образом:

$$f(X) = \sum_{i=1}^r \sum_{j=1}^{m_i} t_{i,j}^{av}(x_{i,j}) \rightarrow \min,$$

$$g_1(X) = \prod_{i=1}^r \prod_{j=1}^{m_i} P_{i,j}^{ok}(x_{i,j}) = P_{COC}^{req} / \prod_{i=r+1}^n \prod_{j=1}^{m_i} P_{i,j}^{ok},$$

$$g_2(X) = \forall i, k \leq r: \sum_{j=1}^{m_i} t_{i,j}^{av}(x_{i,j}) = \sum_{j=1}^{m_k} t_{k,j}^{av}(x_{k,j}),$$

$$x_{i,j} \geq 0, x_{k,j} \geq 0, r \leq n,$$

где  $r$  – количество параллельных цепочек сервисов;  $m_i$  – количество последовательных сервисов в  $i$ -ой, параллельно-выполняемой цепочке.

2) задача *индивидуальной* максимизации вероятности обслуживания для параллельных цепочек сервисов, имеющих равное наибольшее среднее время обслуживания. В качестве ограничения устанавливается значение среднего времени обслуживания следующей по продолжительности обслуживания цепочки:

$$\forall i \leq r: \begin{cases} f(X) = \prod_{j=1}^{m_i} P_{i,j}^{ok}(x_{i,j}) \rightarrow \max, \\ g(X) = \sum_{j=1}^{m_i} t_{i,j}^{av}(x_{i,j}) = \sum_{j=1}^{m_{r+1}} t_{r+1,j}^{av}, x_{i,j} \geq 0. \end{cases}$$

Таким образом, в статье предложена методика минимизации времени обслуживания сервис-ориентированных систем на основе постановки и решения задач оптимизации тайм-аутов Web-сервисов, участвующих в параллельно-последовательной интеграции.

### Литература

1. OASIS Reference Model for Service Oriented Architecture. Ver. 1.0 [Text] / C.M. MacKenzie, K. Laskey, F. McCabe, etc. – Burlington: OASIS, 2006. – 31 p.
2. Gonczy, L. Dependability Evaluation of Web Service-Based Processes [Text] / L. Gonczy, S. Chiaradonna, F. di Giandomenico // European Performance Engineering Workshop – EPEW'2006: conference proceedings. – Bertinoro (Italy), 2006. – P. 166 – 180.
3. Zheng, Z. A QoS-Aware Fault Tolerant Middleware for Dependable Service Composition [Text] / Z. Zheng, M. Lyu // Int. Conf. On Dependable Systems

and Networks. – DSN'2009: conference proceedings. – Lisbon (Portugal), 2009. – P. 239 – 248.

4. Gorbenko, A. Instability analysis of de-lays contributing to Web Service response time [Text] A. Gorbenko // Радіоелектронні і комп'ютерні системи. – 2010. – № 6 (47). – С. 63 – 67.

5. Real Distribution of Response Time Instability in Service-Oriented Architecture [Text] / A. Gorbenko, V. Kharchenko, S. Mamutov, et al. // 29th IEEE Int. Symposium on Reliable Distributed Systems – SRDS'2010: conference proceedings. – Delhi (India), 2010. – P. 92 – 99.

6. Olive, D.J. Applied Robust Statistics [Text] / D.J. Olive. – Illinois: Southern Illinois University Press, 2008. – 571 p.

7. Dhamdhere, A. Buffer Sizing for Congested Internet Links [Text] / A. Dhamdhere, H. Jiang, C. Dovrolis // Annual IEEE Int. Conference on Computer Communications – INFOCOM'2005: conference proceedings. – 2005. – Vol. 2. – P. 1072 – 1083.

Поступила в редакцію 15.03.2012

**Рецензент:** д-р техн. наук, проф., зав. каф. комп'ютерних систем і мереж В.С. Харченко, Национальний аерокосмічний університет ім. Н.Е. Жуковського «ХАІ», Харків.

### МЕТОДИКА МІНІМІЗАЦІЇ ЧАСУ ОБСЛУГОВУВАННЯ СЕРВІС-ОРІЄНТОВАНИХ СИСТЕМ

*А.В. Горбенко, О.М. Тарасюк*

У статті запропоновано методику мінімізації часу обслуговування сервіс-орієнтованих систем, що побудовані на основі паралельно-последовної інтеграції Web-сервісів, з урахуванням надійнісних обмежень. Методика базується на використанні моделі обслуговування, що встановлює співвідношення між імовірністю обслуговування, середнім часом обслуговування та максимальним часом очікування – тайм-аутом, який використовується в якості параметра оптимізації.

**Ключові слова:** web-сервіси, сервіс-орієнтовані системи, час обслуговування, мінімізація.

### SERVICING TIME MINIMIZATION TECHNIQUE FOR SERVICE-ORIENTED SYSTEMS

*A.V. Gorbenko, O.M. Tarasyuk*

A technique of servicing time minimization for service-oriented systems which are built by parallel-sequential integration of Web-services in case of dependability limitation is discussed in the paper. The technique uses a servicing model which defines a relationship between probability of successful servicing, servicing time and maximal waiting time – time-out. Time-out of individual Web-services is used as an optimization parameter.

**Keywords:** web-services, service-oriented systems, servicing time, minimization.

**Горбенко Анатолий Викторович** – канд. техн. наук, доцент, доцент кафедри комп'ютерних систем і мереж Национального аерокосмічного університета ім. Н.Е. Жуковського «ХАІ», Харків, Україна, e-mail: A.Gorbenko@csac.khai.edu.

**Тарасюк Ольга Михайловна** – канд. техн. наук, доцент, доцент кафедри комп'ютерних систем і мереж Национального аерокосмічного університета ім. Н.Е. Жуковського «ХАІ», Харків, Україна, e-mail: O.Tarasyuk@csac.khai.edu.