

ОЦЕНИВАНИЕ ПРОИЗВОДИТЕЛЬНОСТИ GRID-СИСТЕМ С ПОМОЩЬЮ ОПЕРАЦИОННОГО АНАЛИЗА

Введение. За последние 30 лет производительность всегда была одним из важных свойств разрабатываемых систем, в том числе компьютерных [1]. Сегодня задача оценивания производительности grid-системы и ее компонентов – довольно актуальна [2–4]. Относительно функционирования grid-системы необходимы оценки некоторых рабочих параметров, существенно влияющих на качество обслуживания пользователей (Qos, Quality of Service) [5, 6].

Поскольку при выполнении задач в grid-системе для пользователя важно то, чтобы его задача с высокой степенью вероятности гарантированно выполнялась, чтобы задача выполнялась в нужные сроки и при этом в системе были доступными затребованные ресурсы и сервисы. Поэтому для grid-систем особенно актуальной является задача обеспечения нужного качества обслуживания. Для этого нужны устоявшиеся количественные меры функционирования системы либо необходимо их оценивать.

Поскольку grid-система состоит из многих вычислительных узлов, на которых возможен запуск разных составляющих модулей одной комплексной задачи, то следует оценивать производительность как системы в целом, так и ее отдельных компонентов. Дальше производительность grid-системы как сложной распределенной системы оценена с использованием средств операционного анализа [7, 8].

Рассмотрен вопрос применения операционного анализа для задачи оценивания производительности grid-системы. В качестве исследуемой системы рассмотрен Украинский академический grid-сегмент. Для оценки выбраны два параметра: пропускная способность и время отклика системы.

© С.И. Лавренюк, А.Н. Лавренюк,
Ю.А. Грипич, 2010

Выбор параметров производительности. Наиболее важные характеристики (параметры), используемые для оценивания производительности:

- время отклика (response time) – характеристика производительности системы с точки зрения пользователя;
- пропускная способность (throughput) – определяет количество задач, выполненных за единицу времени;
- готовность (availability) – определяет интервал времени, на протяжении которого система находится в рабочем состоянии (т. е. доступна для выполнения задач пользователя);
- стоимость (cost) – характеризует производительность в форме соотношения «цена / производительность».

Учитывая взаимосвязь производительности с качеством обслуживания, приведенный список можно дополнить характеристиками: надежности (reliability), безопасности (security), масштабируемости (scalability), расширяемости (extensibility).

В качестве параметров производительности предлагается использовать пропускную способность (X) и время отклика (R). Эти параметры влияют на количество и качество выполнения задач grid-системой, поэтому оценка этих параметров имеет практическое значение.

Оценка предельных значений параметров производительности. Для моделирования разных аспектов grid-систем используют прежде всего аналитические или стохастические модели. Аналитические модели дают возможность получить явные зависимости в общем, что связывают искомые характеристики с начальными условиями, параметрами и переменными системы.

Операционный анализ как метод аналитического моделирования обеспечивает удобный подход к оцениванию производительности компьютерных систем [8–9]. Для оценки предельных значений параметров производительности использован именно этот подход. Он помогает получить информацию о загруженности компьютерной системы на основе базовых показателей [10]:

T – временной интервал наблюдения за системой;

A – число запросов, сделанных на протяжении времени T ;

C – число обработанных запросов на протяжении времени T ;

B – интервал времени, на протяжении которого запрос обрабатывался ресурсом, в интервале времени наблюдения.

Пусть система состоит из $k = 1, \dots, n$ ресурсов, каждый из которых характеризует параметры:

X_k – средняя пропускная способность (производительность) ресурса k ;

S_k – среднее время обработки запроса ресурсом k ;

V_k – среднее число посещений запросом ресурса k ;

D_k – среднее время, необходимое ресурсу k для обработки запроса;

U_k – коэффициент использования ресурса k ;

R_k – среднее время ответа на запрос ресурса k ;

N – среднее количество запросов в сети массового обслуживания (СМО);

λ – средняя частота получения запросов ресурсом СМО.

В общем случае, для пропускной способности X можно получить выражение (1) [11]:

$$\frac{N}{ND + Z} \leq X(N) \leq \min\left(\frac{1}{D_{\max}}, \frac{N}{D + Z}\right). \quad (1)$$

Введя время ответа R , получаем следующее выражение:

$$\max\left(D_{\max}, \frac{D + Z}{N}\right) \leq \frac{R(N) + Z}{N} \leq \frac{ND + Z}{N} \Rightarrow \quad (2)$$

$$\max(ND_{\max} - Z, D) \leq R(N) \leq ND. \quad (3)$$

Поэтому, по формулам (1) и (3) можно оценить предельные значения параметров производительности отдельных grid-ресурсов, а также всей системы в целом.

Оценивание производительности grid-системы на примере сегмента Украинского академического grid. В качестве объекта исследования для проверки работоспособности предложенного подхода была взята часть инфраструктуры Украинского академического grid-сегмента – 12 узлов.

Архитектура исследуемой grid-системы показана на рис. 1.



РИС. 1. Архитектура исследуемой grid-системы

Данные для моделирования взяты из журналов мониторинга УАГ, а для получения данных о выполнении задач проведен активный эксперимент, в котором, однозначно определено, способен ли конкретный компонент (grid-сервис) предоставлять и обрабатывать необходимую информацию в допустимом интервале времени.

Поскольку эксперимент проведен на реально работающей системе, то производительность оценена при незначительной

нагрузке на систему. Это дало возможность системе функционировать в штатном режиме, а задачи равномерно поступали на каждый вычислительный ресурс.

Для получения данных о работе grid-инфраструктуры Украинского академического сегмента проведен активный эксперимент с запуском задач на узлах и сборе данных о состоянии узлов и качестве выполнения задач. Для проведения активного эксперимента по оценке доступности узлов и качества выполнения задач в grid-инфраструктуре использованы многоуровневые сценарии. При необходимости тестирования всех сервисов, развернутых в grid-инфраструктуре, эти сценарии можно расширить (см. [12]).

На основе полученных данных можно вычислить значение базовых параметров в терминах операционного анализа (ОА) [9, 10, 13]. В терминах сетей массового обслуживания (СМО) [1, 14] формализованную модель можно представить как на рис. 2.

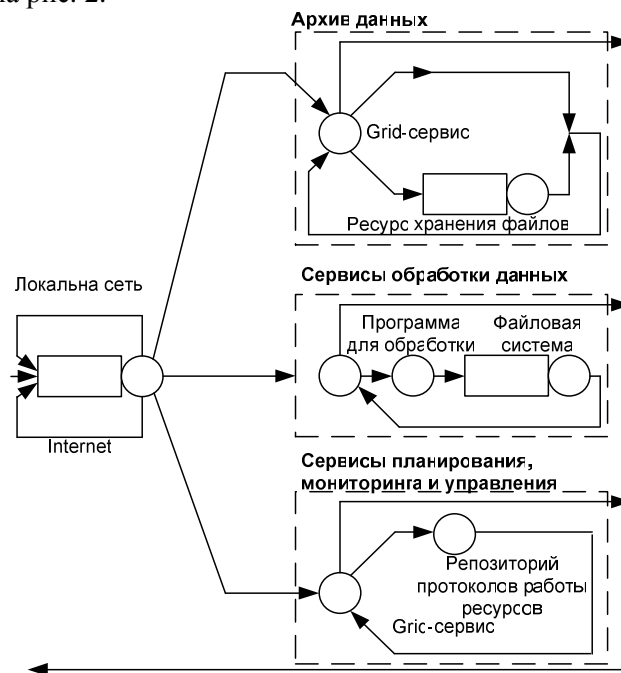


РИС. 2. Пример формализации системы в терминах теории массового обслуживания

Пусть Z – среднее время обдумывания при работе пользователей с системой в целом. Введем обозначения таких потребностей в обслуживании:

D_1 – ресурсом № 1;

D_2 – ресурсом № 2;

D_n – ресурсом № n ;

D – системой в целом. Предполагается, что система является совокупностью взаимозависимых ресурсов, а это значение – усреднение D по всем ресурсам;

D_{\max} – ресурсом, для которого потребность в обслуживании принимает максимальное значение;

D_{FS} – хранилищем файлов.

Данные показатели для отдельных узлов представлены на рис. 3.

При декомпозиции большой задачи на отдельные подзадачи, и при выполнении каждой подзадачи на отдельном узле получено следующие показатели. На протяжении времени $T = 2592000$ с, в систему послано 7773 запроса, со средним интервалом в 3600 с. Общее количество принятых и правильно обработанных 3125 запросов (за время наблюдения за системой).

На основании данных экспериментов и использования вышеприведенных формул (1)–(3) возможно рассчитать показатели потребности в обслуживании для каждого из grid-узлов системы.



РИС. 3. Показатели потребности в обслуживании для отдельных узлов

Частота запросов насыщения $\lambda = 0.002105$.

Максимальное значение потребности в обслуживании $D_{\max} = 475.15$.

Среднее время размышления при работе пользователей с системой в целом $Z = 720.55$.

Общее количество узлов, что принимали участие в эксперименте – 12.

В таблице приведены данные по отдельным узлам.

ТАБЛИЦА

$D1$	$D2$	$D3$	$D4$	$D5$	$D6$	$D7$	$D8$	$D9$
58.04	120.49	14.19	459.29	105.58	334.97	37.01	12.52	90.29
$D10$	$D11$	$D12$	D_{\max}	D	Z			
73.27	61.80	475.15	475.15	1842.61	720.55			

Используя формулы (1) и (3), получаем следующие неравенства для оценивания выбранных параметров производительности системы:

Для пропускной способности

$$\frac{N}{1842.60 * N + 720.55} \leq X_{\text{сист}}(N) \leq \min\left(\frac{1}{475.15}, \frac{N}{1842.60 + 720.55}\right).$$

Для времени ответа $\max(N475.15 - 720.55, 1842.60) \leq R(N) \leq N1842.60$.

Из вышеприведенных данных видно, что первым (по критерию) «узким местом» системы является ресурс 12, поскольку значение потребности обслуживания у него максимальное. Второе «узкое место» системы – ресурс 6, поскольку значение потребности обслуживания у него приближается к максимальному. Для увеличения производительности системы в первую очередь необходимо ликвидировать первое и второе «узкое место».

Зависимость указанных параметров от количества задач показано на рис. 4. Рассматривая каждый из ресурсов отдельно (следующий уровень детализации системы), аналогично можно определить «узкие места» конкретного ресурса.

Зависимость производительности от количества задач

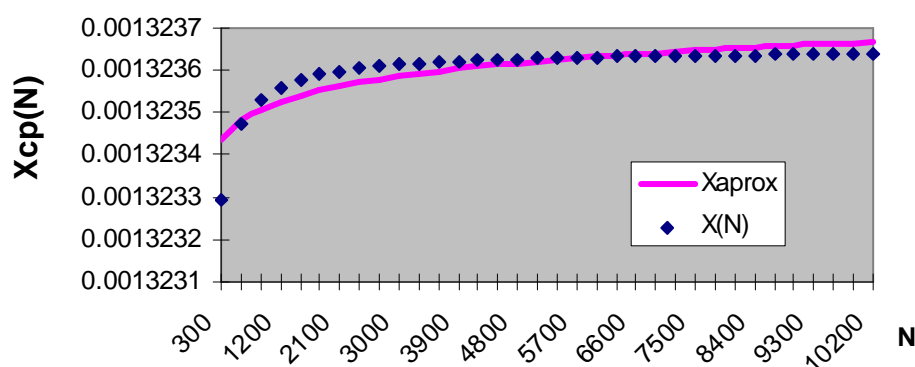


РИС. 4. График зависимости производительности системы от количества задач

Аппроксимацию выполнено по алгоритму наименьших квадратов для логарифмической функции:

$$f(x) = a \ln(x) + c$$

$$a = \frac{\sum_{i=1}^n X(N_i) * \ln(N_i) - \sum_{i=1}^n \ln(N_i) * \sum_{i=1}^n X(N_i)}{n * \sum_{i=1}^n \ln^2(N_i) - \left(\sum_{i=1}^n \ln(N_i)\right)^2},$$

$$c = \frac{\sum_{i=1}^n \ln^2(N_i) * \sum_{i=1}^n X(N_i) - \sum_{i=1}^n \ln(N_i) * \sum_{i=1}^n X(N_i) * \ln(N_i)}{n * \sum_{i=1}^n \ln^2(N_i) - \left(\sum_{i=1}^n \ln(N_i)\right)^2},$$

где n – количество точек.

В результате функция производительности для данного сегмента grid-системы имеет вид $X(N) = 6.54544 * 10^{-8} * \ln(N) + 0.001323062$.

Выводы. В статье приведены результаты использования одного из подходов к оцениванию производительности компьютерных систем. Для предварительной оценки производительности grid-системы предложено использовать

аппарат операционного анализа совместно с теорией сетей массового обслуживания. Заметим, что для детального анализа производительности целесообразно использовать имитационное моделирование [9, 10]. Именно такой подход помогает получить качественную и количественную оценку производительности системы относительно практики. Об этом свидетельствуют результаты использования метода для анализа реальной grid-системы.

С.І. Лавренюк, А.М. Лавренюк, Ю.А. Гріпич

ОЦІНЮВАННЯ ПРОДУКТИВНОСТІ GRID-СИСТЕМ З ДОПОМОГОЮ ОПЕРАЦІЙНОГО АНАЛІЗУ

Розглянуто питання застосування операційного аналізу для задачі оцінювання продуктивності grid-систем. В якості досліджуваної системи розглянуто Український академічний grid-сегмент. Для оцінювання вибрано два параметри: пропускна здатність та час відгуку системи.

S. Lavreniuk, A. Lavreniuk, Yu. Gripich

PERFORMANCE ASSESSMENT OF GRID- SYSTEMS USING OPERATIONAL ANALYSIS

The question of operational analysis for the problem of estimating the performance of the Grid-system is investigated. The Ukrainian Academic Grid-segment is under consideration. The throughput and response time of the system are chosen for evaluation.

1. Менаске Д., Алмейда В. Производительность Web-служб // Анализ, оценка и планирование. – Киев: ДиаСофт, 2003. – 480 с.
2. Серієнко І.В. Про основні напрямки створення інтелектуальних інформаційних технологій // Системні дослідження та інформаційні технології. – 2002. – № 1. – С. 39–64.
3. Grid – нова інформаційно-обчислювальна технологія для науки / А.Г. Загородній, Г.М. Зінов'єв, Є.С. Мартинов, С.Я. Свистунов, В.М. Шадуря // Вісник НАН України. – 2005. – № 6. – С. 17–25.
4. Zheng Yan and Silke Holtmanns. «Trust Modeling and Management: from Social Trust to Digital Trust», book chapter of Computer Security, Privacy and Politics: Current Issues, Challenges and Solutions, IGI Global, 2007.
5. Qin Liu, Xiaohua Jia, Chanle Wu. Optimal precomputation for mapping service level agreements in grid computing, Future Generation Computer Systems, October, 2008. – 24, N 8. – P. 849–859.
6. Шелестов А.Ю., Лавренюк С.І. Моделирование работы grid-системы в условиях неопределенности // Материалы тр. конф. «Моделирование и компьютерная графика», 7–9 апреля 2009. – Донецк: ДонНТУ, 2009.
7. Menascé D.A. Computer system lifecycle. – <http://cs.gmu.edu/~menasce/cs672/slides/>.
8. Operational Analysis: Fundamental Laws – <http://www.owl.net.rice.edu>
9. Шелестов А.Ю. Структурно-функциональный анализ элементов grid-систем // Проблемы управления и информатики. – 2007. – № 5. – С. 25–32.
10. Куссуль Н.Н., Шелестов А.Ю. Grid-системы для задач исследования Земли. Архитектура, модели и технологии. – Киев: Наук. думка, 2008. – 452 с.

11. *Куссуль Н.Н., Шелестов А.Ю., Лобунец А.Г.* Применение методов операционного анализа для оценки производительности grid-систем // Кибернетика и вычислительная техника. – 2004. – Вып. 144. – С. 3–19.
12. *Лавренюк С.И., Лавренюк А.Н., Грипич Ю.А.* Построение базы данных мониторинга состояния узлов grid-инфраструктуры // Материалы тр. 6-й Междунар. конф. «Теоретические и прикладные аспекты построения программных систем». – ТААПСД' 2009. – Киев, 9–10 декабря 2009. – С. 124–131.
13. *Кунцевич В.М.* Управление в условиях неопределенности: гарантированные результаты в задачах управления и идентификации. – Киев: Наук. думка, 2006. – 206 с.
14. *Клейнрок Л.* Вычислительные системы с очередями / Пер. с англ. Под ред. Б.С. Цыбакова. – М.: Мир, 1979. – 600 с.

Получено 02.12.2009

Об авторах:

Лавренюк Сергей Иванович,
младший научный сотрудник
Института кибернетики имени В.М. Глушкова НАН Украины,

Лавренюк Алла Николаевна,
кандидат технических наук,
доцент Национального технического университета Украины
«Киевский политехнический институт»,

Грипич Юлия Анатольевна,
младший научный сотрудник
Института космических исследований НАН Украины и НКА Украины.