

программного обеспечения и их оценка (SQuaRE). Модели качества систем и программного обеспечения. – Введ. 2011-03-03. – Женева, 2011.

СИСТЕМА КРИТЕРИЕВ ОЦЕНКИ ЭФФЕКТИВНОСТИ ПРОГРАММНЫХ СРЕДСТВ ОРГАНИЗАЦИИ РАСПРЕДЕЛЕННЫХ ВЫЧИСЛЕНИЙ

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Царевич Д.Ю.

Бахтизин В.В. – к.т.н., доцент

В связи с интенсивным развитием сфер человеческой деятельности, в том числе научной, экономической, медицинской, растут потребности общества в вычислительных ресурсах, необходимых для обработки информации. Одним из возможных способов удовлетворения данных потребностей является использование распределенных вычислений (РВ).

Для оценки и сравнения параметров программных средств организации РВ в работе формулируется система критериев оценки эффективности ПС организации РВ, включающая следующие критерии [1]:

1. Производительность – ключевой критерий, характеризуется отношением показателей производительности ПС организации РВ: максимального из полученных практически к максимально возможному теоретически.

2. Масштабируемость – способность ПС организации РВ выполнять работу, пропорциональную числу вычислительных узлов (ВУ) без потери производительности при росте числа ВУ.

3. Безопасность – описывает эффективность механизмов защиты информации, вычислительных узлов и процессов ПС организации РВ от несанкционированного доступа.

4. Отказоустойчивость – характеризуется соотношением промежутка времени, в течение которого ПС организации РВ находилось в состоянии работы к общему промежутку времени, в течение которого производилось измерение.

5. Прозрачность – характеризует способность ПС скрывать свою распределенную природу, т.е. распределение процессов и ресурсов по множеству ВУ, и способность представляться для пользователей и разработчиков приложений в виде единой централизованной компьютерной системы.

В качестве критерия оценки *производительности* ПС организации РВ выбрано отношение значений двух используемых вариантов измерения производительности – максимальной и пиковой [2].

Пиковая производительность (P_{peak} , TFLOPS) – теоретический предел производительности (выражаемый через операции с плавающей точкой) для оцениваемого ПС, находится с помощью формулы (1.1):

$$P_{peak} = \sum_{i=1}^N (F_i * P_i * I_{tick} / C), \quad (1.1)$$

где F_i – тактовая частота процессора i ВУ, МГц.

P_i – число процессоров в i ВУ.

I_{tick} – количество инструкций с плавающей запятой на такт (4 – для процессоров Core2; 8 – для процессоров Intel с AVX).

C – константа, $C = 10^6$.

N – количество ВУ, используемых ПС.

Максимальная производительность – максимальная производительность ПС, достигаемая при решении практических задач.

Критерий оценки производительности P ПС организации РВ определяется как соотношение значений максимальной и пиковой производительностей по формуле (1.2):

$$P = P_{max} / P_{peak} * 100\%, \quad (1.2)$$

где P_{peak} – пиковая производительность, TFLOPS.

P_{max} – максимальная производительность, TFLOPS.

Рассматривая *масштабируемость* как способность оцениваемого ПС выполнять работу, пропорциональную числу вычислительных узлов (ВУ), критерий масштабируемости определяется с помощью критерия оценки параллельной эффективности [2].

Ускорение (англ. «speedup»), получаемое при использовании распределенного алгоритма (РА) для P ВУ, по сравнению с последовательным вариантом выполнения вычислений, определяется величиной K по формуле (1.3):

$$K = T_S / T_N, \quad (1.3)$$

где T_S – время, затраченное на решение задачи последовательным алгоритмом на одном ВУ

T_N – время, затраченное на решение той же задачи при использовании распределенного алгоритма, выполненного на N ВУ.

N – количество ВУ, используемых ПС.

Параллельная эффективность E определяется как отношение ускорения к количеству используемых ПС вычислительных узлов по формуле (1.4):

$$E = K / N * 100\%, \quad (1.4)$$

где K – ускорение, получаемое при использовании распределенного алгоритма

N – количество ВУ, используемых ПС.

В качестве основы для разработки критерия оценки *безопасности* ПС организации РВ выбран стандарт Department of Defense Trusted Computer System Evaluation Criteria, TCSEC, DoD 5200.28-STD [3].

В соответствии с вышеуказанным стандартом, критерии определения безопасности ПС делятся на 4 раздела: D, C, B и A, из которых наивысшей безопасностью обладает раздел A. Каждый раздел представляет собой значительные отличия в доверии индивидуальным пользователям или организациям. Разделы C, B и A иерархически разбиты на серии подразделов, называемые классами: C1, C2, B1, B2, B3 и A1. Каждый раздел и класс расширяет или дополняет требования, указанные в предшествующем разделе или классе.

Таким образом, критерий оценки безопасности S определяется как принадлежность ПС организации РВ к определенному классу вышеуказанного стандарта в соответствии с формулой (1.5):

$$S = S_{class}, \quad (1.5)$$

где S_{class} – класс безопасности по стандарту Department of Defense Trusted Computer System Evaluation Criteria, TCSEC, DoD 5200.28-STD [3].

Значение критерия оценки безопасности S постоянно для ПС, т.е. не зависит от конкретной решаемой вычислительной проблемы.

Критерий оценки *отказоустойчивости* R определяется как соотношение промежутка времени, в течение которого ПС организации РВ находилось в состоянии работы к общему промежутку времени, в течение которого производилось измерение и рассчитывается по формуле (1.6):

$$R = T_{ff} / T_{total} * 100\%, \quad (1.6)$$

где T_{ff} – промежуток времени, в течение которого ПС находилось в состоянии работы.

T_{total} – общий промежуток времени, в течение которого производилось измерение.

Для определения критерия оценки *прозрачности* ПС организации РВ из стандартов эталонной модели для распределенной обработки в открытых системах Reference Model for Open Distributed Processing (RM-ODP) выбраны семь наиболее важных типов прозрачности [4]: прозрачность доступа (англ. access transparency), прозрачность местоположения (англ. location transparency), прозрачность перемещения (англ. migration transparency), прозрачность смены местоположения (англ. relocation transparency), прозрачность репликации (англ. replication transparency), прозрачность одновременного доступа (англ. concurrency transparency), прозрачность отказов (англ. failure transparency).

Критерий прозрачности O определяется как соотношение количества поддерживаемых ПС организации РВ типов прозрачности к общему количеству наиболее важных типов прозрачности по формуле (1.7):

$$O = O_{supp} / O_{total} * 100\%, \quad (1.7)$$

где O_{supp} – количество поддерживаемых ПС типов прозрачности.

O_{total} – общее количество наиболее важных типов прозрачности, $O_{total} = 7$.

Значение критерия оценки прозрачности O постоянно для ПС, т.е. не зависит от конкретной решаемой вычислительной проблемы.

Предлагаемая система критериев оценки эффективности ПС организации РВ может применяться для анализа эффективности различных ПС организации РВ, а также выбора ПС с более высокими значениями приоритетных критериев для решаемой распределенной вычислительной задачи.

Список использованных источников:

1. Coulouris, G. Distributed Systems: Concepts and Design (5th Edition). Coulouris G. [and others]. – Boston: Addison-Wesley, 2011. – 1008 p.
2. Черемсинов, Д.И. Проектирование и анализ параллелизма в процессах и программах / Д.И. Черемсинов. – Минск: Беларус. навука, 2011. – 300 с.
3. 5200.28-STD Trusted Computer System Evaluation Criteria // National Security Institute [Электронный ресурс]. – 2015. – Mode of access: <http://csrc.nist.gov/publications/history/dod85.pdf>. – Date of access: 18.02.2016.
4. Farooqi, K. ISO Reference Model of Open Distributed Processing (RM-ODP) // Institute For Enterprise Architecture Developments [Электронный ресурс]. – 2015. – Режим доступа: <http://www.enterprise-architecture.info/Images/Document-s/RM-ODP2.pdf>. – Дата доступа: 15.01.2016.

СЕМАНТИЧЕСКАЯ НАВИГАЦИЯ В БОЛЬШИХ ИНФОРМАЦИОННЫХ МАССИВАХ

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Чеушев К. В.

Глухова Л. А. – канд. техн. наук, доцент

Предложено системное решение для актуальной задачи навигации в больших информационных массивах, основанное на применении комплекса аналитических функций, создающего возможность целенаправленного движения пользователя в пространстве информационного ресурса, направляемого не априори установленными ссылками, а его текущими потребностями и интересами. Выделена обеспечивающая семантическую навигацию операциональная триада «структурирование – экстрагирование – поиск подобного», элементы которой собираются в цепи и каскады, реализующие все необходимые аналитические функции, такие как сложные высокоточные формы поиска, классификация, кластеризация, автоматическое выделение ключевых термов, построение сниппетов, аннотаций, дайджестов и т.д., позволяющих пользователю удовлетворить сложные информационные потребности, итерационно двигаясь с помощью аналитической системы к искомому результату.

Экспоненциальный рост доступных в сетях информационных ресурсов (ИР) создаёт заманчивую перспективу их использования в профессиональных, образовательных и других целях. Гигантские ИР сами по себе ещё не достаточное условие доступности и качества знаний, с ростом масштабов ИР увеличивается и усложняется проблема нахождения элементов ИР, удовлетворяющих конкретной информационной потребности пользователя (ИПП), критериям полноты, точности и качества. Острая ещё пару десятилетий назад проблема информационного голода трансформировалась в информационное перенасыщение, когда мы практически уверены в том, что в пределах нашей доступности нужная информация есть, но слабо представляем, как её получить, отобрать, сократить объём подлежащего «ручному» анализу до приемлемого. Раньше основные надежды возлагались на то, что информация перед помещением в состав ИР будет переработана человеком, должным образом структурирована и снабжена семантическими и другими метками, но сегодня всё чаще на это рассчитывать уже нельзя – объёмы новых ИР велики настолько, что качественно их обработать и всесторонне структурировать люди уже не в состоянии, но и в любом случае структурирование по одним критериям может не коррелировать с критериями конкретной ИПП. В текстовых ресурсах разметка для обеспечения гиперссылочной навигации уже перестала достаточно полно обеспечивать возможности доступа к искомой информации, а полнотекстовый поиск зачастую порождает огромный «шум», что вынуждает либо проводить очень ёмкую последующую обработку найденного, либо ограничиться использованием малого количества найденных документов, сомневаясь в их качестве и полноте удовлетворения ИПП. Велика и возрастает актуальность и востребованность так называемой семантической навигации – развитого способа ориентирования в огромных информационных ресурсах с опорой не на априори расставленные гиперссылки, а на интеллектуальные автоматические инструменты.

Под семантической навигацией понимается процесс движения аналитика или просто пользователя в информационном пространстве ИР, при котором он имеет возможность итерационно корректировать свои цели и интересы и переходить в соответствии с ними к другим подобластям ИР, стартуя с какой-то позиции, возможно семантически даже отдалённой от искомого результата (т.е. допускается нечёткое начальное представление цели). Отличие от навигации в классическом понимании (по априори установленным ссылкам) состоит в том, что направление движение в пространстве ИР не определено заранее, оно обеспечивается функциональными аналитическими механизмами непосредственно в момент взаимодействия пользователя с информационно-аналитической системой.