

УДК 681.324

С.В. Минухин

Харьковский национальный экономический университет, Харьков

МОДЕЛИРОВАНИЕ И АНАЛИЗ ВЛИЯНИЯ ОСНОВАННЫХ НА МЕТОДЕ ПОКРЫТИЯ АЛГОРИТМОВ ПЛАНИРОВАНИЯ ЗАДАНИЙ И РАСПРЕДЕЛЕНИЯ РЕСУРСОВ НА ПРОИЗВОДИТЕЛЬНОСТЬ ГЕТЕРОГЕННЫХ РАСПРЕДЕЛЕННЫХ СИСТЕМ

В работе исследуется подход к планированию заданий и ресурсов гетерогенной распределённой системы на основе метода покрытия. Рассмотрены основные положения подхода, математическая модель и проведен сравнительный анализ эффективности эвристических и точного алгоритмов решения задачи о покрытии для решения задач планирования. Приведены результаты компьютерных экспериментов и анализ в соответствии с разработанными сценариями моделирования, показавшими преимущества предлагаемых алгоритмов при условии изменения параметров модели при фиксированном количестве и интенсивности потоков заданий.

Ключевые слова: алгоритм, наименьшее покрытие, гетерогенная распределенная система, планирование, производительность.

Введение

Одним из направлений развития современных информационно-коммуникационных технологий является организация распределённых вычислительных систем, совместно использующих ресурсы различных организаций. Это направление предполагает реализацию идеи распределённого метакомпьютинга – создание высокопроизводительных кластеров, позволяющих решать разнообразные задачи в таких сферах как физика высоких энергий, молекулярная биология, решения разряженных систем линейных уравнений большой размерности и т.д., основной особенностью которых является потребность в значительном количестве свободных и доступных в требуемые промежутки времени ресурсов. Система управления такими системами предполагает наличие одновременно двух механизмов для ее реализации – механизма планирования заданий и механизма планирования распределения ресурсов между заданиями. Отметим, что в литературе эти понятия различаются следующим образом: под планированием заданий и распределением ресурсов понимают нахождение такого соответствия заданий –ресурсы (matching), которое отражает требования пользователей к реализации запроса на выполнение задания (QoS) и позволяет максимизировать критерий пользовательской удовлетворённости. С другой стороны, эффективность работы распределённой системы характеризуется ее системной производительностью, которая определяется уровнем использования имеющихся ресурсов и определяется на основе коэффициента использования ресурсов. Таким образом, эти механизмы должны не только взаимодействовать между собой, но и обеспечивать при этом требуемый уровень эффективности работы распределенной системы в целом.

Существующая на настоящий момент времени классификация методов планирования заданий включает статический и динамический режимы. В динамическом режиме задания поступают в обработку непосредственно после их представления для решения при помощи системы представления заданий (submission system), при этом используются пакетный (packet mode) или одиночный (immediately) режимы. В пакетном режиме задания формируются в так называемые пакеты (metatask, Bag of Tasks) в определенные промежутки времени, называемые интервалом (временем) планирования (scheduling interval) или событием назначения (mapping event). При этом требуется выбрать один из следующих методов определения времени планирования: использовать априори задаваемую величину периодичности планирования (промежутка времени между двумя последовательными событиями назначения) или решать задачу планирования в момент времени, когда пакет заданий достигнет априори задаваемого размера. В обоих случаях необходимо использовать информацию о следующих характеристиках распределённой среды – интенсивность потока поступающих в систему заданий, их гетерогенность, количество и производительность используемых ресурсов, ожидаемое время выполнения заданий на различных ресурсах и ряд других.

Для оценки эффективности работы системы используются следующие метрики: makespan (время завершения выполнения самого последнего задания в системе), flowtime (response time) – время ответа системы или время выполнения всех заданий, поступивших в обработку в момент времени планирования, и коэффициент использования, отражающий системную производительность (system performance) – характеризует уровень использования (загрузки) ресурсов с учетом спланированных на этот

ресурс заданий за время планирования всех заданий глобальной очереди. Выбор ресурсов осуществляется при этом на основе эвристик, для каждого задания определяется (назначается) только один ресурс и в случае, например, невысокой (низкой или средней) гетерогенности кластера возникает конкуренция, которая приводит к необходимости разработки других эвристик, позволяющих устранить этот недостаток.

В данной работе развивается подход, предложенный и исследованный в работе [1]. В соответствии с рассмотренными методами планирования заданий и планирования распределения ресурсов он базируется на пакетном режиме планирования заданий, и пакетном методе распределения ресурсов, для выбора которых используется задача о наименьшем покрытии.

Целью данной работы является моделирование различных алгоритмов решения задачи о покрытии и анализ влияния конкретного алгоритма на показатели производительности работы гетерогенной распределённой системы – время выполнения всех заданий и коэффициент использования ресурсов [1].

Модель планирования на основе задачи о покрытии

Постановка задачи планирования формулируется следующим образом:

необходимо минимизировать

$$L_t = \sum_{j=1}^n x_j(t_k) \rightarrow \min \quad (1)$$

при ограничениях

$$\sum_{j=1}^n \beta_{ij} x_j(t_k) \geq 1, \quad i = \overline{1, m}; \quad (2)$$

$$\beta_{ij} \in \{0, 1\}; \quad x_j(t_k) \in \{0, 1\}.$$

где m – количество заданий, подлежащих планированию; n – количество ресурсов системы, доступных и свободных на момент планирования; $t_k \in [T_0, T_N]$.

Планирование осуществляется на интервале времени $[T_0, T_N]$, где T_0 – время начала планирования; T_N – время окончания планирования заданий очереди. Данный подход базируется на следующих положениях.

1. Система планирования организована в виде двухуровневой структуры, на первом уровне которой из глобальной очереди, например, по приоритету, выбирается множество заданий (пакет заданий, называемый в данной модели пулом), подлежащие планированию, к ним применяется метод решения задачи (1), (2). Далее выбранные, как результат ее решения, задания назначаются на доступные и свободные на момент планирования ресурсы и решаются на них под управлением локального планировщика.

2. Метод планирования на каждом шаге планирования максимально загружает минимальное количество свободных и доступных на момент планирования ресурсов. На следующем этапе распределения заданий очереди количество ресурсов для возможного назначения на них заданий будет максимальным.

3. Алгоритм решения задачи (1), (2) должен иметь малую временную сложность его реализации для минимизации времени, отводимого на процесс планирования заданий для их выполнения на ресурсах.

4. Система планирования использует пакетную технологию: задания, организованные в форме пакета (пула) заданий, выбираются из глобальной очереди, и по мере их планирования на ресурсы помещаются в пакет заданий на назначенный ресурс (ресурсы) и передаются на решение на этот ресурс (ресурсы).

5. Система планирования предусматривает использование идеи балансировки загрузки ресурсов: в случае, если ресурс, вошедший в покрытие на момент распределения на него задания (заданий) занят, оно перепланируется на другой, но свободный на данный момент ресурс, также вошедший в покрытие или на который в настоящий момент имеется очередь с меньшим количеством заданий в пакете (без повторного решения задачи о покрытии).

Постановка компьютерных экспериментов и анализ результатов

В качестве единицы времени планирования и проводимых расчетов используется внутреннее время имитационной модели системы – 1 такт, которое соответствует времени решения одного задания, имеющего сложность 100 000 MI (million instructions), на ресурсе, производительность которого 100 000 MIPS (million instructions per second).

Для динамического планирования используются случайные значения параметров модели с различными законами распределения. Это дает возможность учесть динамику изменений в инфраструктуре распределённой системы и входных потоков заданий (табл. 1).

При проведении вычислительных экспериментов использованы следующие параметры:

сложность задания: задания малой сложности (трудоемкости) определялись в интервале от 100 до 300 тактов; средней сложности – от 300 до 10 000 тактов; высокой сложности – от 10 000 и выше тактов. Сложность заданий определялась нормальным законом с среднеквадратическим отклонением (СКО) – 10 % от математического ожидания;

производительность ресурсов: низкая производительность – 10, средняя производительность – 50, высокая производительность – более 100. Для генерации данных по производительности использовался нормальный закон с СКО = 1.

Размер пула выбирался в диапазоне от 10 до 100, размер пакета заданий – в диапазоне от 1 до 50, периодичность выбиралась в диапазоне от 10 до 10 000.

Интенсивность поступления заданий задавалась по нормальному закону с математическим ожиданием равным 50, СКО = 5, для экспоненциального закона интенсивности математическое ожидание выбиралось равным 50.

Для анализа влияния алгоритмов были выбраны: жадный алгоритм [2], точный алгоритм на основе рангового подхода [3] и алгоритм с малой временной сложностью и низкой погрешностью, предложенный и исследованный в работе [4].

Исследовалось поведение системы – влияние изменения параметров настроек модели на время выполнения заданий и коэффициент использования ресурсов следующих параметров:

универсальности;

размера пула при фиксированной универсальности заданий (50 %) и периодичности планирова-

ния, равной среднему времени освобождения ресурса (отношению сложности заданий к производительности ресурса);

периодичности планирования для фиксированного размера пула (250) и пакета заданий равного 50. Все эксперименты проводились для количества ресурсов в диапазоне от 10 до 100.

Для повышения эффективности работы алгоритма МС использовались его модификации: МС2 (выбирается тот ресурс из нескольких возможных, на который на момент планирования количество заданий в очереди на ресурс меньше) и МС3 (выбирается тот ресурс из нескольких возможных, производительность которого выше).

Для анализа влияния работы различных алгоритмов на производительность системы использованы следующие сценарии:

1. Точный, жадный и МС.

2. Жадный и МС.

3. Модификации алгоритма МС и алгоритм FCFS.

Таблица 1

Параметризация настроек динамической модели планирования

№ п/п	Наименование характеристики для задания и ресурса	Параметр	Закон распределения
1	Задание	количество заданий, которое будут подано на вход модели для решения	нормальный, экспоненциальный, равномерный
2	Ресурс	количество всех ресурсов, которыми решаются задания	нормальный, экспоненциальный, равномерный
3	Сложность решения задания	максимальное количество тактов, за которое будет решено задание	нормальный, экспоненциальный, равномерный
4	Универсальность задания	количество ресурсов, которыми может быть решено задание, в процентах к общему количеству ресурсов	постоянная величина
5	Пул заданий	промежуточный временной стек системы определенного размера, в который загружаются входящие задания	экспоненциальный, нормальный, постоянная величина
6	Производительность ресурса	максимальное количество тактов решения задания, которое ресурс может выполнить за единицу времени	нормальный, экспоненциальный, равномерный
7	Пакет заданий	количество заданий, которое может «выстроиться» в очередь на выбранный ресурс после планирования	постоянная величина
8	Коммуникационная задержка	количество тактов, которое характеризует временную задержку передачи задания на выбранный для ее решения ресурс	равномерный, нормальный, постоянная величина
9	Периодичность планирования	количество тактов, через которое осуществляется процедура планирования для заданий, поступивших в пул	постоянная величина

Сценарий 1. Проведен сравнительный анализ точного, жадного и МС метода для определения влияния универсальности заданий на коэффициент использования для 25 ресурсов (рис. 1, 2). Результаты показали преимущество МС по отношению к жадному алгоритму на 40 %, а по отношению к точному – на 175 %. Следует отметить общее увеличение эффективности методов с увеличением универсальности заданий.

Сценарий 2. Проведен сравнительный анализ жадного алгоритма и МС на большом количестве

ресурсов (100) для времени выполнения (рис. 3, 4). Результаты показали, что МС имеет значительное преимущество по отношению к жадному алгоритму при размере пула больше 10, при пуле 100 преимущество составляет 300 %. Коэффициент использования у МС имеет максимальное значение по отношению к жадному алгоритму при размере пула равном 10. Отметим, что при экспоненциальном законе поступления заданий результаты по коэффициенту использования аналогичны результатам для нормального закона поступления заданий.

Сценарий 3. Проведенный сравнительный анализ методов МС (МС1, МС2, МС3) и метода FCFS для 10 и 100 ресурсов (рис. 5, 6) показал для заданий высокой сложности выигрыш МС3 по времени выполнения заданий по отношению к FCFS при изменении периодичности планирования. При этом коэффициент использования для 100 ресурсов больше у FCFS, так как интенсивность заданий меньше ко-

личества ресурсов. При количестве ресурсов 10 коэффициент использования максимальный у всех методов МС (МС1, МС2, МС3).

Полученные результаты показывают влияние периодичности планирования на время выполнения для метода МС, т.е. существует значение периодичности, которое определяет минимальное время выполнения заданий.

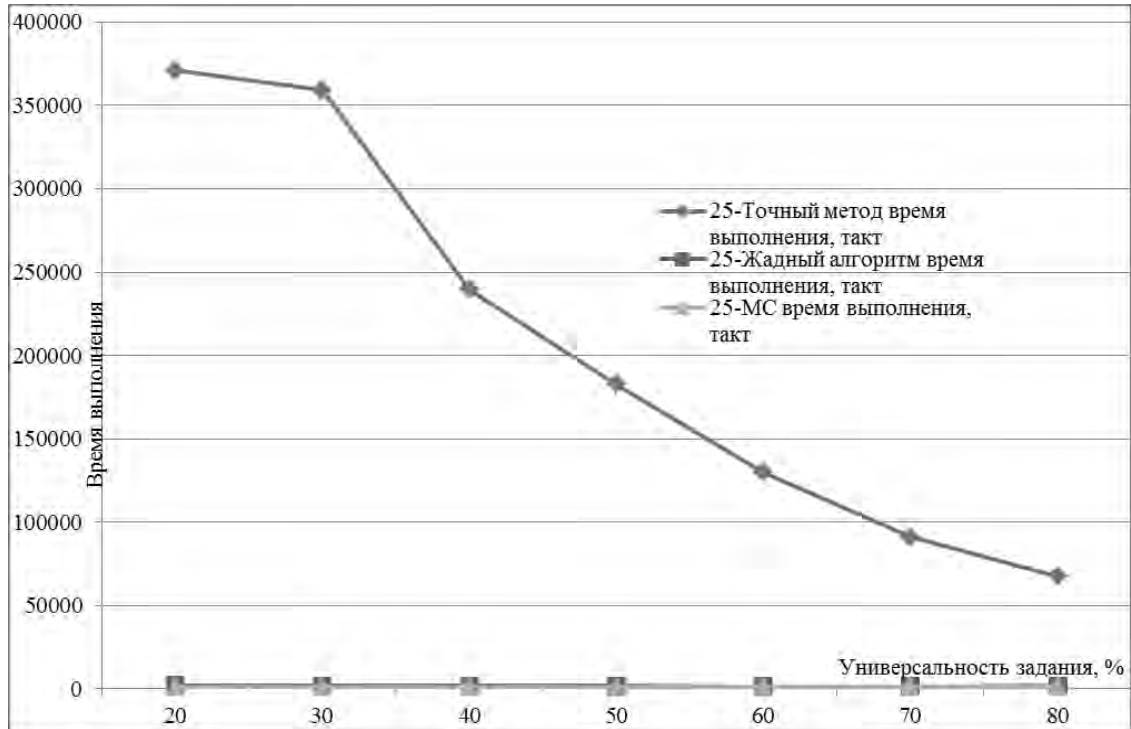


Рис. 1. Сценарий 1

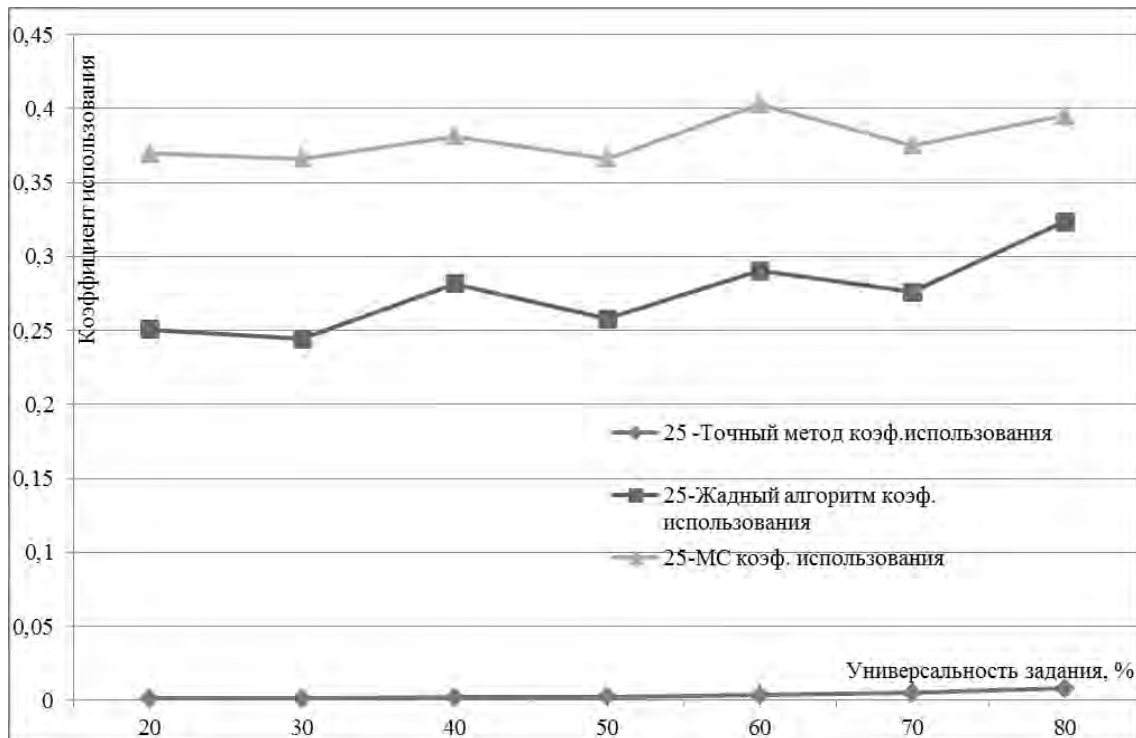


Рис. 2. Сценарий 1

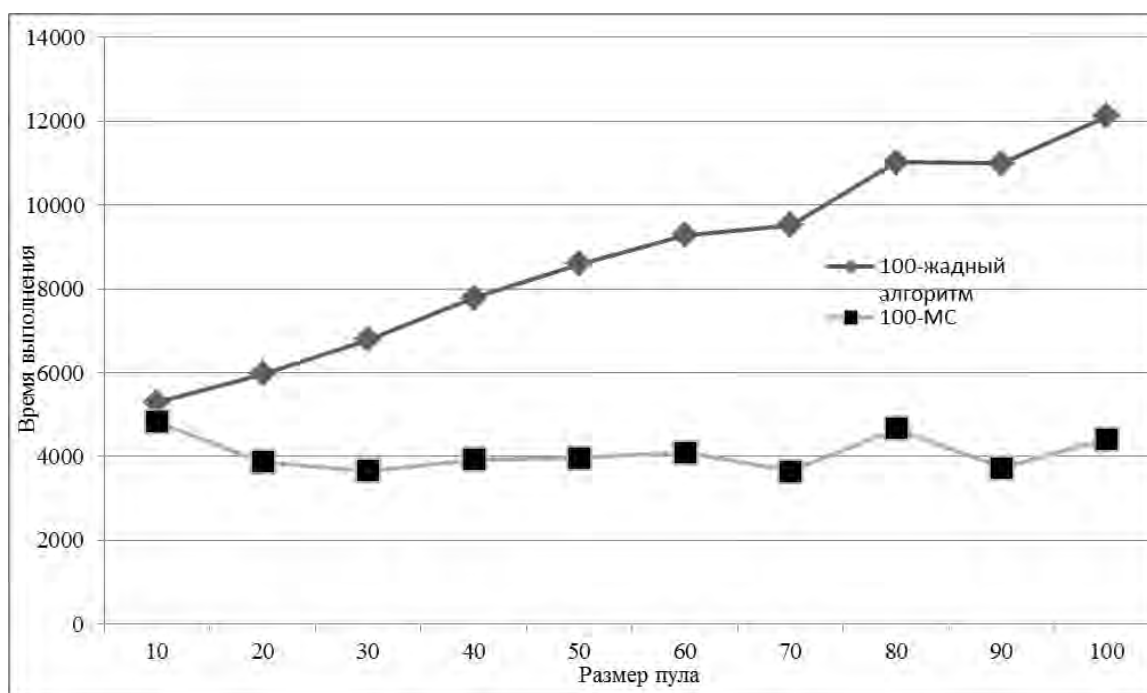


Рис. 3. Сценарий 2

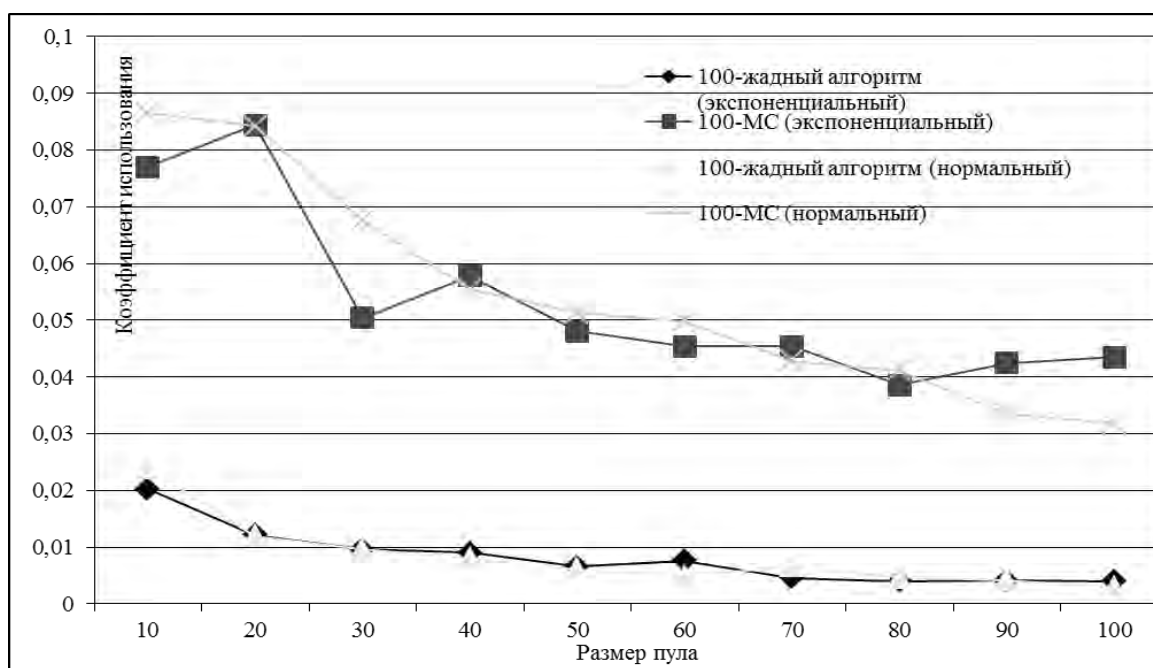


Рис. 4. Сценарий 2

Выводы

В результате проведенных экспериментов был осуществлен сравнительный анализ алгоритмов решения задачи о покрытии, который показал преимущества предложенного в работе алгоритма МС по отношению к жадному и точному алгоритмам. При этом, наибольшее влияние на производительность системы оказывают количество ресурсов, размер пула (оптимальный размер пула зависит от интенсивности заданий и количества ресурсов), перио-

дичности планирования (определяется средним временем освобождения ресурса). Особенностью работы МС является устойчивость по отношению к законам поступления заданий на вход системы и изменению их универсальности. Получила подтверждение гипотеза о том, что для заданий высокой сложности в случае выбора одного ресурса из нескольких свободных предпочтение необходимо отдать ресурсу с большей производительностью (алгоритм МС3).

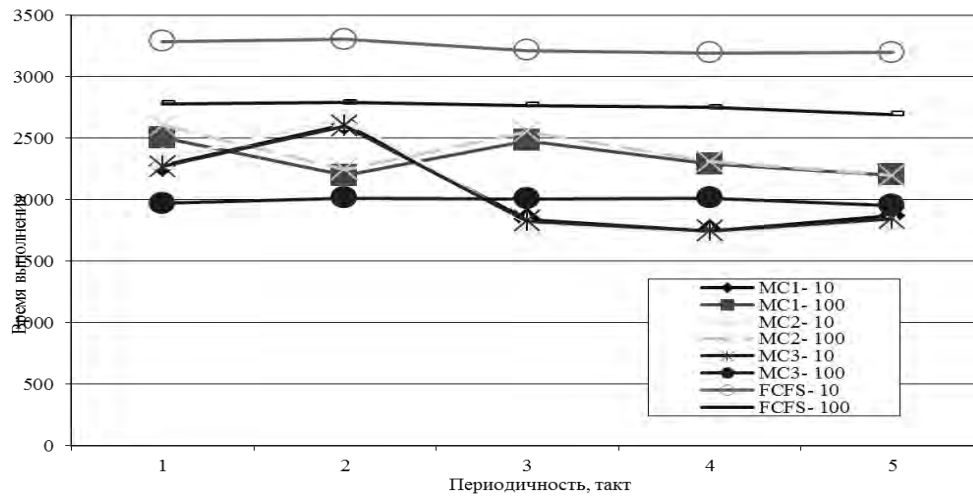


Рис. 5. Сценарий 3

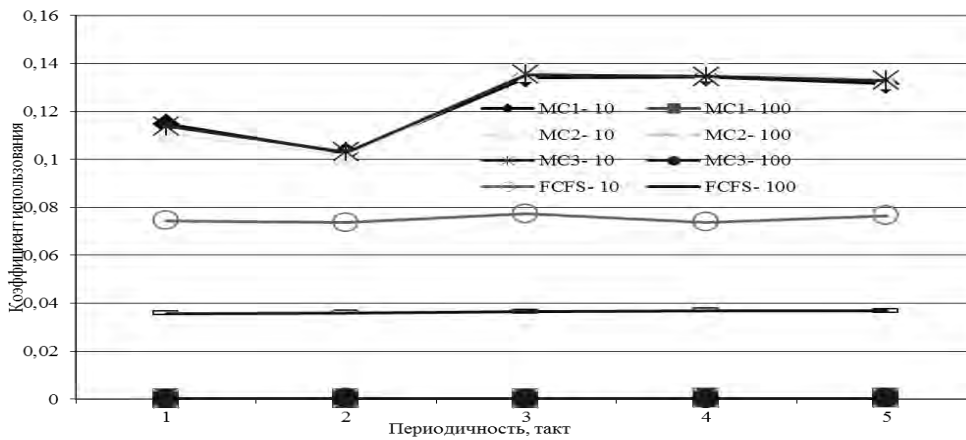


Рис. 6. Сценарий 3

Список литературы

1. Листровой С.В.. Модель и подход к планированию распределения ресурсов в гетерогенных Грид-системах / С.В. Листровой, С.В. Минухин Проблемы управления и информатики. – 2012. – № 5. – С. 120-133.
 2. Chvátal V. A greedy-heuristic for the set covering problem / V. Chvátal // Math. Oper. Res. – 1979. – № 4. – P. 233-235.
 3. Листровой С.В. Точный алгоритм решения задачи о минимальном покрытии. / С.В. Листровой, А.Ю. Гуль, Е.С. Листровая // Информатика. Вып. 5. – К.: 1998. – С. 32-36.

4. Листровой С.В. Метод решения задач о минимальном вершинном покрытии в произвольном графе и задачи о наименьшем покрытии / С.В. Листровой, С.В. Минухин // Электронное моделирование.– 2012. – Т. 34. – № 1. – С. 29 – 43.

Поступила в редколлегию 1.10.2012

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

МОДЕЛЮВАННЯ ТА АНАЛІЗ ВПЛИВУ ЗАСНОВАНОГО НА МЕТОДІ ПОКРИТТЯ АЛГОРИТМІВ ПЛАНУВАННЯ ЗАВДАНЬ ТА РОЗПОДІЛУ РЕСУРСІВ НА ПРОДУКТИВНІСТЬ ГЕТЕРОГЕННИХ РОЗПОДІЛЕНИХ СИСТЕМ

С.В. Мінухін

В роботі досліджується похід до планування завдань і ресурсів гетерогенної розподіленої системи на основі методу покриття. Розглянуто основні положення підходу, математична модель та проведено порівняльний аналіз ефективності евристичних і точного алгоритмів розв'язання задачі про покриття для вирішення завдань планування. Наведено результати комп'ютерних експериментів і аналіз відповідно до розроблених сценаріями моделювання, які показали переваги запропонованих алгоритмів за умови зміни параметрів моделі при фіксованій кількості та інтенсивності потоків завдань.

Ключові слова: алгоритм, найменше покриття, гетерогенна розподілена система, планування, продуктивність.

MODELING AND ANALYSIS OF IMPACT OF BASED ON MINIMAL COVER METHOD ALGORITHMS FOR TASKS SCHEDULING AND RESOURCE ALLOCATION ON PERFORMANCE OF HETEROGENEOUS DISTRIBUTED SYSTEMS

S.V. Minukhin

In this paper the approach to scheduling of tasks and resources in a heterogeneous distributed system based on minimal cover method. The main provisions of the approach, the mathematical model and the comparative analysis of the effectiveness of heuristic and exact algorithms for the covering problem to solve scheduling. The results of computer experiments and comparative analysis use the developed simulation scenarios, the advantages of the proposed algorithms provided changing the model parameters for a fixed amount of resources and intensity of the task stream.

Keywords: algorithm, the minimal cover, heterogeneous distributed systems, scheduling, performance.