

Алгоритмы оптимизации энергопотребления и повышения эффективности процессоров с масштабированием частоты и напряжения гетерогенного кластера

С.В. Минухин, М.И. Сухонос

Харьковский национальный экономический университет 9-а, пр. Ленина, Харьков, Украина

ms_vl@mail.ru, maksim.sukhonos@gmail.com

Аннотация. Рассмотрены задачи повышения эффективности работы процессоров гетерогенного кластера на основе оптимизации режимов переключения частоты и напряжения для заданий с директивными сроками выполнения для приоритетного режима EDF. Приведены примеры реализации предлагаемых алгоритмов при решении задачи выбора частоты процессора. Разработано программное обеспечение для моделирования режимов переключения частот для 4- и 12-уровневых типов процессоров. Приведены результаты компьютерного моделирования предлагаемых алгоритмов, подтверждающие их преимущества по сравнению с существующими *offline* алгоритмами, не требующих определения точного значения частоты и напряжения процессора. Полученные результаты обосновывают возможность использования алгоритмов на практике для оптимизации энергопотребления и повышения коэффициента загрузки процессоров гетерогенного кластера.

Ключевые слова

Кластер, работа, директивный срок, масштабирование, энергопотребление, частота, напряжение, процессор, алгоритм, коэффициент использования

1 Введение

Оптимизация работы вычислительных кластеров, серверов и отдельных ПК в настоящее время связано с решением целого ряда задач, которые позволяют повысить их экономическую эффективность. Одной из таких задач является оптимизация энергопотребления на основе динамического масштабирования частоты процессора (DVFS Scheduling Algorithms), позволяющая в зависимости от характеристик заданий (работ) – длительности решения и директивного срока – и диапазона изменений частот процессора управлять изменением частоты и напряжения так, чтобы повысить коэффициент его использования, минимизировать время завершения выполнения заданий, величину суммарного запаздывания и, таким образом, уменьшить величину штрафа за несвоевременное их выполнение [1–9].

В данной работе развивается подход, использующий режим EDF (earliest deadline first), который является приоритетным правилом выборки заданий из очереди. Суть его заключается в следующем: все поступившие на решение задания сортируются в порядке возрастания (неубывания) их директивных сроков, после этого определяется максимальная загрузка процессора, рассчитываемая как отношение суммы времени выполнения всех заданий к максимальному директивному сроку, для которой по соответствующей шкале определяется частота процессора и напряжение.

Основная идея подхода, использующего EDF – минимизировать возможные запаздывания за счет такого выбора частоты, которая гарантирует минимальное запаздывание (или отсутствие запаздывания) для всех поступивших на выполнение заданий.

Необходимо отметить, что такой подход в значительной степени зависит от соотношения между величиной процессорного времени выполнения работы (задания) и его директивным сроком: чем меньше эта разница, тем больше вероятность того, что при большом количестве заданий частоту в результате небольшого

количества переключений необходимо установить на ее максимальное значение, что приведет к тому, что процедура масштабирования частоты и напряжения не будет эффективной..

2 Постановка задачи и описание алгоритмов оптимального масштабирования частоты и напряжения процессора

Обозначим текущее множество выполняемых работ как $T_k = \{t_{k,j}(e_{k,j}, d_{k,j}) \mid j = 1, \dots, n_k\}$, где $e_{k,j}$ и $d_{k,j}$ оставшееся время выполнения и директивный срок j -й работы. Если длина оставшейся работы $l_{k,i}$, то оставшееся время выполнения определяется как $l_{k,i}/Q_m$. N_k – набор работ в T_k .

Поскольку схема назначения приоритетов базируется на принципе EDF, T_k сортируется по директивному сроку, из чего следует что $d_{k,i} \leq d_{k,i+1}$, где $i = 1, \dots, n_k-1$. Планировщик всегда выполняет работу с более ранним директивным сроком. Обозначим текущий режим работы процессора как v_k . Текущий коэффициент использования $u_{k,i}$ определяется по следующей формуле:

$$u_{k,i} = \frac{\sum_{j=1}^i e_{k,j}}{d_{k,i}}.$$

Коэффициент использования $u_{k,i}$ определяет необходимую загрузку процессора для работы $t_{k,i}$ при использовании правила EDF. Схема управления частотой основывается на работе [1], таким образом, что режим работы процессора для работы с наивысшим приоритетом (наименьшим директивным сроком) находится по следующей формуле:

$$\tilde{s}_k = \max_{i=1}^{n_k} \{u_{k,i}\}.$$

Поскольку режимы работы процессора дискретны в диапазоне от $V_1 - V_m$, напряжение v_k для работы $t_{k,i}$ определяется как наименьшее V_i при условии, что S_i меньше или равно s_k (см. формулу (2)). Когда PE_k выбирает работу с наиболее ранним директивным сроком в своей очереди, текущее напряжение выбирается как величина v_k по формуле:

$$v_k = \min_{i=1}^m \{V_i \mid S_i \geq \tilde{s}_k\}. \quad (2)$$

Для сравнительного анализа базового алгоритма [1] (далее называемого simple) и предлагаемых рассмотрим набор работ T_k и их характеристики, представленные в табл. 1. Режимы работы процессора (4 уровня) и соответствующие относительные скорости (частоты) процессора приведены в табл. 2.

Таб. 1. Характеристики работ

№ работы	L	D
1	6	10
2	3	14
3	7	39
4	8	57
5	1	58

Таб. 2. Режимы работы процессора

Режим	Относительная скорость
x1	0,25
x2	0,5
x3	0,75
x4	1

Следуя работе [1], значения $u_{k,1}$, $u_{k,2}$, $u_{k,3}$, $u_{k,4}$, $u_{k,5}$ для этих работ равны $6/10=0,6$, $9/14=0,64$, $16/39=0,41$, $24/57=0,42$ и $25/58=0,43$ соответственно. Рассчитаем необходимые для построения графиков режимов переключения для базового алгоритма: S_k первой работы равно 0,64, ближайший режим работы – x3; для второй работы $u_{k,2}$, $u_{k,3}$, $u_{k,4}$, $u_{k,5}$ равны $3/14=0,21$, $10/39=0,26$, $18/57=0,32$, и $19/58=0,33$, соответственно; S_k второй работы равно 0,33, ближайший режим работы – x2; S_k третьей работы равно 0,28, ближайший режим работы – x2; S_k четвертой работы равно 0,16, ближайший режим работы – x1; S_k пятой работы равно 0,02, ближайший режим работы – x1; для предлагаемого метода: $u_{k,1}$, $u_{k,2}$, $u_{k,3}$, $u_{k,4}$, $u_{k,5}$ для первой работы равны $6/10=0,6$, $9/14=0,64$, $16/39=0,41$, $24/57=0,42$ и $25/58=0,43$ соответственно, S_k первой работы равно 0,64, ближайший режим работы – x3; для второй работы $l_{k,1}=1/0,75$ (относительная скорость режима x3)*6=8.

Таким образом, $u_{k,2}$, $u_{k,3}$, $u_{k,4}$, $u_{k,5}$ равны $11/14=0,79$, $18/39=0,46$, $26/57=0,46$, и $27/58=0,47$, соответственно; S_k второй работы равно 0,79, ближайший режим работы – x4; S_k третьей работы равно 0,47, ближайший режим работы – x2; S_k четвертой работы равно 0,59, ближайший режим работы – x3; S_k пятой работы равно 0,63, ближайший режим работы – x3.

Полученные в результате приведенных расчетов режимы переключения для базового и модифицированного (Improved) алгоритмов приведены на рис. 2 и рис. 3, соответственно.

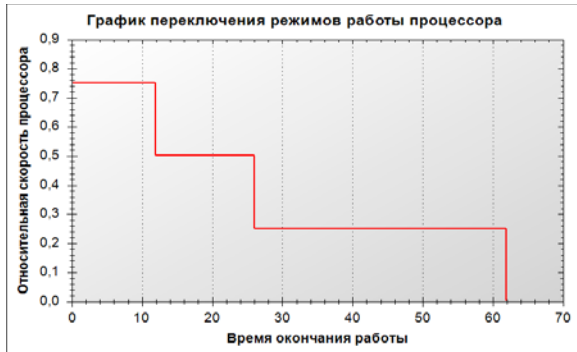


Рис. 2. Режимы переключения для базового алгоритма

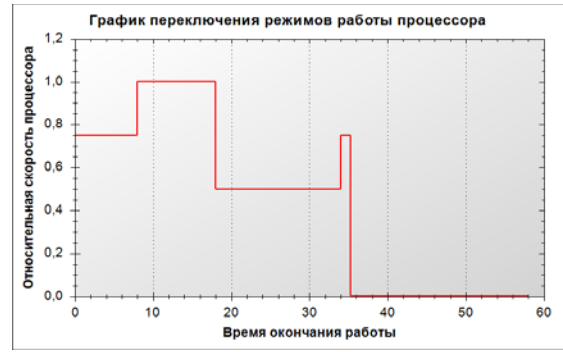


Рис. 3. Режимы переключения для нового алгоритма

Из приведенных графиков следует, что для базового алгоритма количество переключений составляет 3, а для предлагаемого 4, но при этом у базового алгоритма имеется существенное суммарное запаздывание выполняемых работ, что определяет большой уровень коэффициента использования, но вместе с тем имеет существенный недостаток – необходимость работы процессора в течении значительного промежутка времени на максимальной частоте.

Так, например, для количества работ 5 с длительностью в интервале от 1 до 10, с множителем директивных сроков в интервале от 1 до 3, распределенных по равномерному закону, использование рассматриваемых алгоритмов показывает следующие результаты: для базового алгоритма суммарное запаздывание составляет 10, для предлагаемого алгоритма 0; потребление электроэнергии для базового – 120, для предлагаемого алгоритма – 80. Это подтверждается графиками на ниже приведенных рис. 3, 4 для количества работ в диапазоне от 20 до 100.

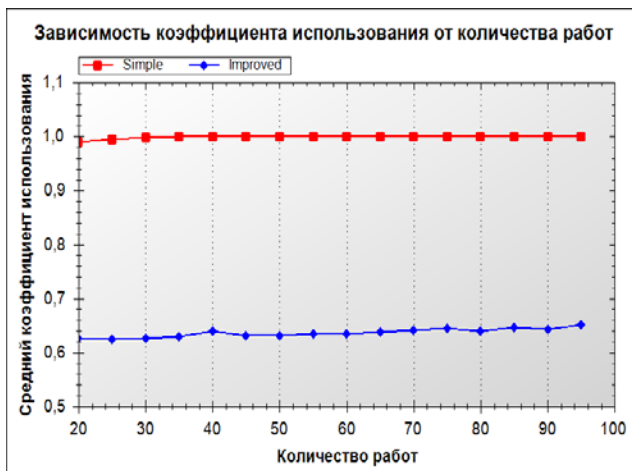


Рис. 3. Зависимость коэффициента использования от количества работ

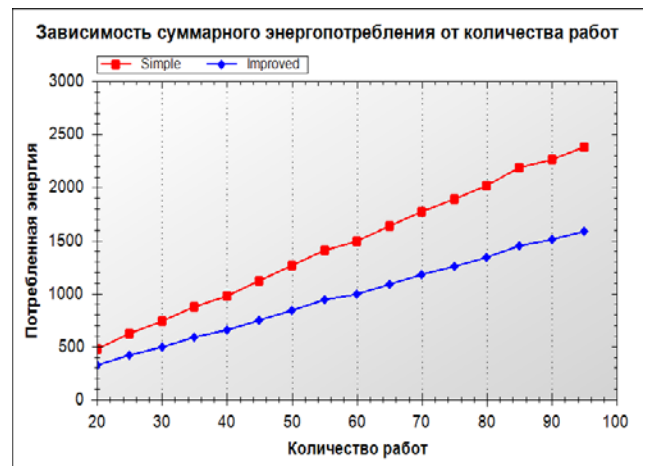


Рис. 4. Зависимость суммарного энергопотребления от количества работ

Для моделирования работы предлагаемых алгоритмов Improved I, Dual-mode II и Ternary-mode III был разработан программный продукт, реализованный на языке программирования C#, псевдокоды которых приведены на рис. 5.

```
// Improved I
for i ← 1 to n
do   contMode ← Max(c[] / d[])
     mode[i] ← Aproximate(contMode)
     l[i] ← 1 / mode[i] * l[i]

     for k ← 1 to n
     do   if k = 1
          then c[k] ← l[k]
          else c[k] ← c[k-1] + l[k]

// Dual-mode II
```

```
maxMode ← Max(mode[])
avgContMode ← Avg(mode[])
avgMode ← Approximate(avgContMode)

for i ← 1 to n
  do    if    mode[i] <= avgMode
        then mode[i] ← avgMode
        else mode[i] ← maxMode

    l[i] ← 1 / mode[i] * l[i]
// Ternary-mode III
maxMode ← Max(mode[])
avgContMode ← Avg(mode[])
avgMode ← Approximate(avgContMode)
for i ← 1 to n
  do    if    mode[i] < avgMode
        then modesUnderAvarage[].Add(mode[i])
if    modesUnderAvarage[].Count() = 0
then  underAvgMode = avgMode
else  underAvgMode = Approximate(Avg(modesUnderAvarage[]))
for i ← 1 to n
  do    if    mode[i] <= underAvgMode
        then mode[i] ← underAvgMode
        else if mode[i] <= avgMode
        then mode[i] ← avgMode
```

Рис. 5. Псевдокоды предлагаемых алгоритмов планирования расписания выполнения заданий

3 Проведение вычислительных экспериментов и анализ результатов

Для проведения экспериментов использованы следующие значения настроек:

- процессорное время работ выбиралось по равномерному закону в интервале [1, 10];
- множители директивных сроков выполнения работ выбирались по равномерному закону в интервале [1, 3];
- количество наблюдений для генерации исходных данных и определения математического ожидания расчетных параметров выбиралось в количестве 40;
- количество режимов процессора – 4;
- количество работ – 100.

В качестве метрик, характеризующих эффективность исследуемых алгоритмов, использованы следующие:

- минимальное, максимальное и среднее (суммарное) время запаздывания всех работ;
- максимальное, минимальное и среднее количество переключений процессора при выполнении всех работ;
- максимальный, минимальный и средний коэффициент использования процессора [1] для всех работ;
- минимальное, максимальное и среднее время работы процессора в холостом режиме;
- минимальное, максимальное и среднее время работы процессора в активном режиме;
- минимальное, максимальное и среднее энергопотребление процессора в холостом режиме [1];
- минимальное, максимальное и среднее энергопотребление процессора в активном режиме [1];
- минимальное, максимальное и среднее общее энергопотребление процессора.

Для проведения сравнительного анализа алгоритмов были рассчитаны наиболее важные с точки зрения эффективности использования масштабирования частоты и напряжения процессора зависимости коэффициента использования и общего энергопотребления от диапазонов изменения параметра D, приведенные на рис. 6–13.

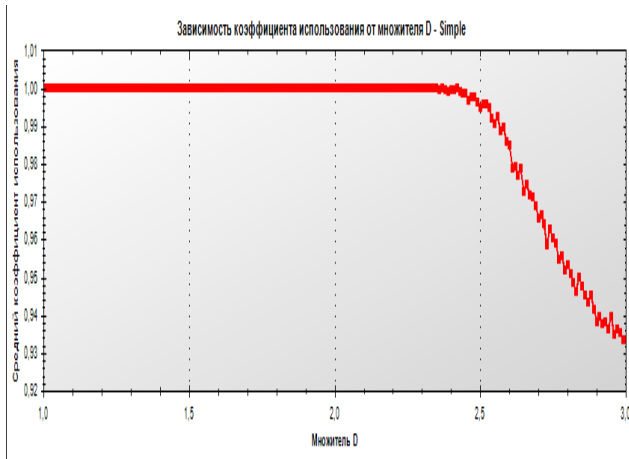


Рис. 6. Зависимость коэффициента использования от D для базового алгоритма

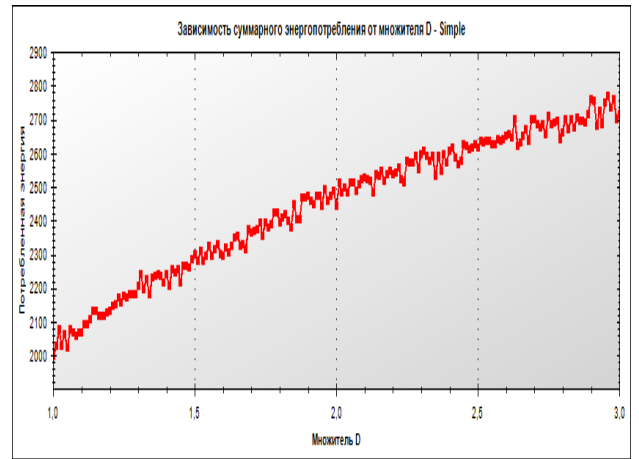


Рис. 7. Зависимость общего энергопотребления от D для базового алгоритма

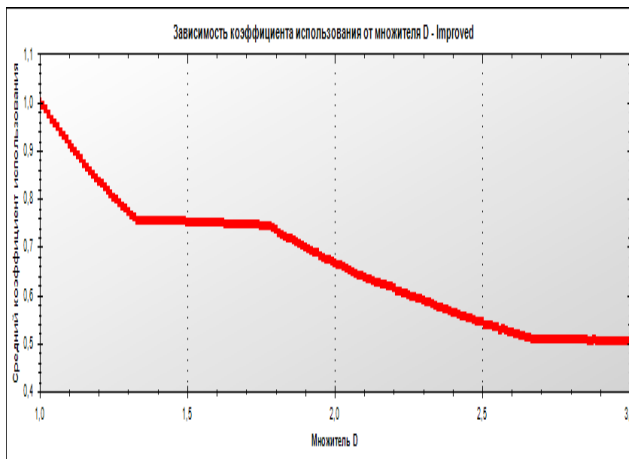


Рис. 8. Зависимость коэффициента использования от D для алгоритма Improved I

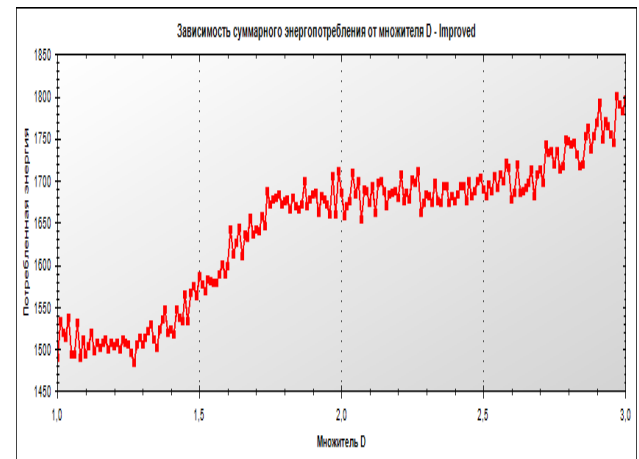


Рис. 9. Зависимость общего энергопотребления от D для алгоритма Improved I

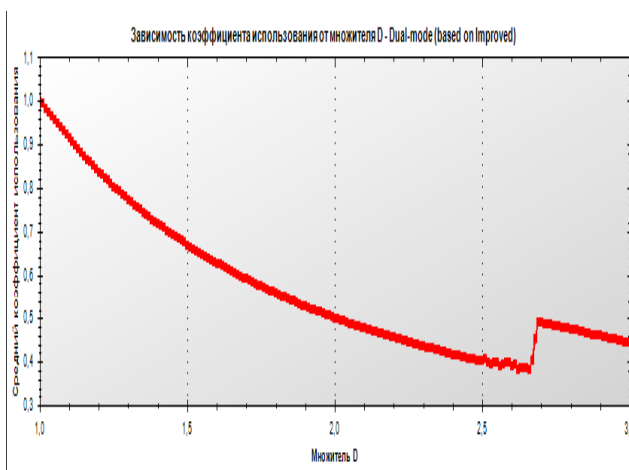


Рис. 10. Зависимость коэффициента использования D от D для алгоритма Dual-mode II

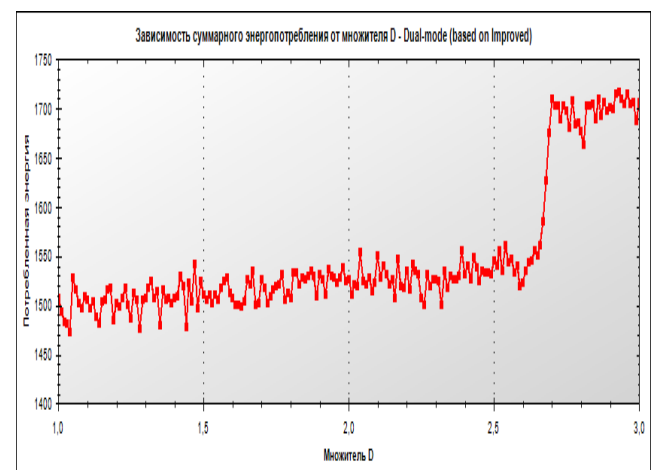


Рис. 11. Зависимость общего энергопотребления от D для алгоритма Dual-mode II

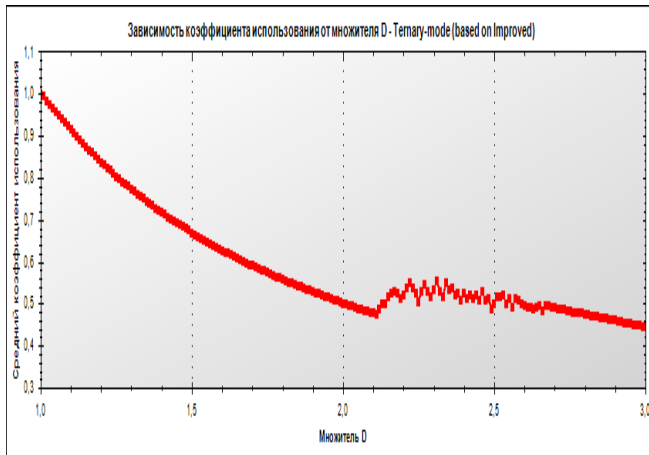


Рис. 12. Зависимость коэффициента использования D от D для алгоритма Ternary-mode III

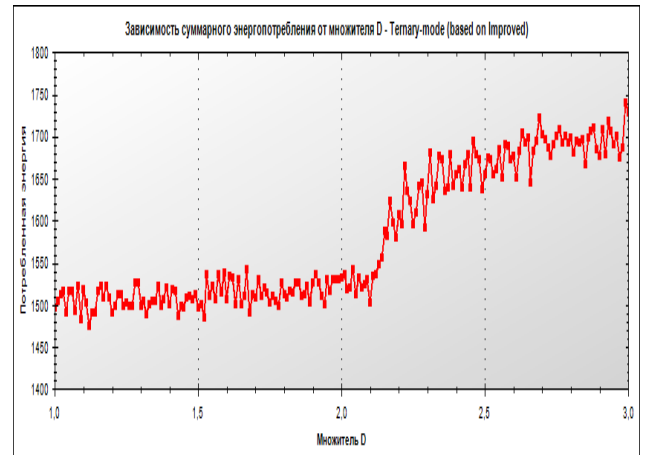


Рис. 13. Зависимость общего энергопотребления от D для алгоритма Ternary-mode III

Для проверки гипотезы о необходимости увеличения режимов настройки предлагаемых алгоритмов при увеличении количества уровней (P-states) масштабирования частот и напряжений процессора были проведены эксперименты с целью анализа влияния добавляемых уровней определения возможных частот и процедуры усреднения частот на эффективность процедуры переключения режимов.

Результаты для количества уровней равного 12 переключения частот в активном состоянии процессора приведены на рис. 14–17.



Рис. 14. Зависимость коэффициента использования от D для базового алгоритма

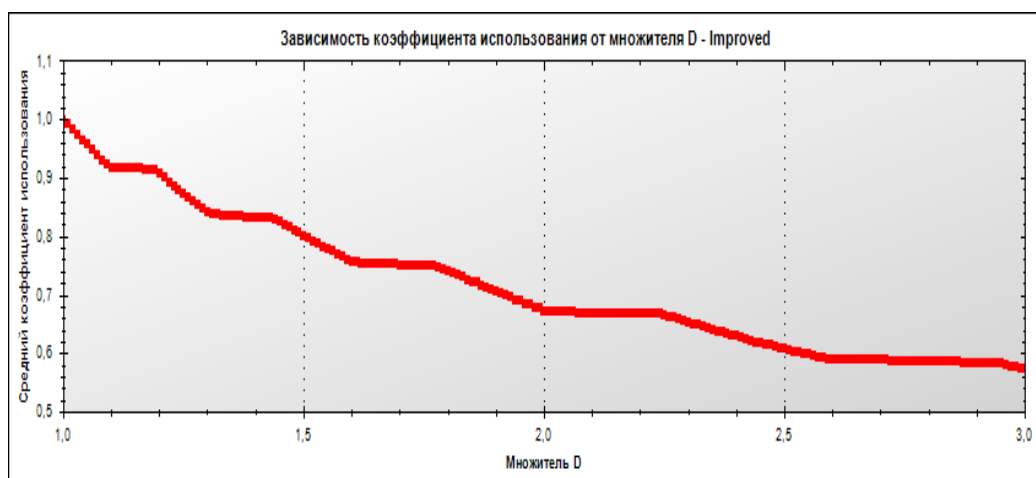


Рис. 15. Зависимость коэффициента использования от D для алгоритма Improved I

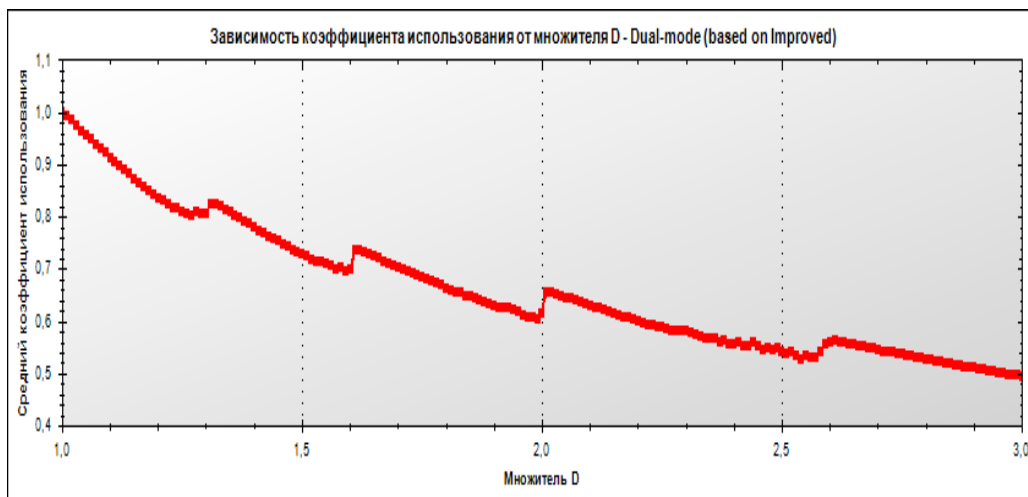


Рис. 16. Зависимость коэффициента использования от D для алгоритма Dual-mode II

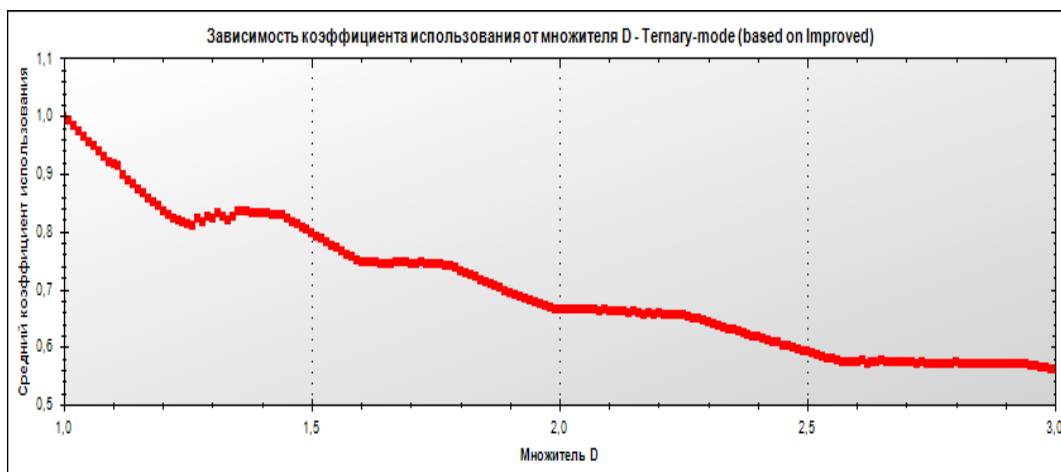


Рис. 17. Зависимость коэффициента использования от D для алгоритма Ternary-mode III

Сравнительный анализ результатов, приведенных на рис. 14–17, с результатами на рис. 6, 8, 10, 12 показывают, что все предложенные алгоритмы обеспечивают повышение коэффициента использования для процессора с 12 уровнями по сравнению с аналогичным показателем для процессора с 4 уровнями. Для базового алгоритма коэффициент использования равен 1 (рис. 14), но расчет энергопотребления показывает значительное увеличение ее величины, а также наличие другого негативного результата – большое суммарное запаздывание работ.

Это обосновывает гипотезу о том, что с увеличением количества уровней переключения частот процессора необходима адаптация используемой схемы предлагаемых алгоритмов, то есть перехода к алгоритмам, использующих большее количество уровней усреднения значений частот по сравнению с существующими, что было использовано в алгоритме Dual-mode II (2 уровня) и алгоритме Ternary-mode III (3 уровня).

Из полученных результатов следует 2 важных вывода:

теоретический: предлагаемые алгоритмы реализуют аппроксимацию точного значения частоты для выполнения множества заданий и адаптируют схему усреднения частот в зависимости от количества уровней (режимов) переключений процессора;

практический: для повышения эффективности работы гетерогенного кластера необходимо устанавливать схемы переключений процессоров, основываясь на моделях процессоров и встроенных механизмах реализации переключений их частот. Это позволит повысить эффективность работы вычислительного кластера по производительности на 7–10 % и уменьшить потребление электроэнергии на 10–20 % в зависимости от количества узлов кластера и используемых типов (моделей) процессоров.

На рис. 18, 19 приведены результаты расчетов потребления электроэнергии для процессора Intel Pentium M 1,6 ГГц для базового (рис. 18) и предлагаемого алгоритма Improved I (рис. 19).



Рис. 18. Зависимость общего энергопотребления от D для базового алгоритма для процессора Intel Pentium M



Рис. 19. Зависимость общего энергопотребления от D для алгоритма Improved I для процессора Intel Pentium M

4 Выводы

Разработаны алгоритмы для оптимизации переключения частоты и напряжения процессора с встроенными возможностями для управления и выбора необходимой частоты, позволяющие обеспечить выполнение заданий с отсутствием суммарного запаздывания, повысить коэффициент использования (загрузки) процессора и обеспечить уменьшение общего энергопотребления путем многократного усреднения частот, адаптированного к количеству уровней переключений процессора.

Проведённые компьютерные эксперименты подтверждают возможность использования предложенных алгоритмов в гетерогенных кластерах для выполнения заданий с директивными сроками в широком диапазоне (multiple deadlines) и для различных типов процессоров с различными соотношениями величин потребляемой электроэнергии в активном и холостом режимах.

Литература

- [1] K. H. Kim, R. Buyya, J. Kim: Power Aware Scheduling of Bag-of-Tasks Applications with Deadline Constraints on DVS-enabled Clusters. *Proceedings of the 7th IEEE International Symposium on Cluster Computing and the Grid*, 2007, 541–548.
- [2] Nikzad B. Rizvandi, Javid Taheri, Albert Y. Zomaya: Some Observations on Optimal Frequency Selection in DVFS-based Energy Consumption Minimization. *Journal of Parallel and Distributed Computing*, 71(8): 1154-1164, 2011.
- [3] R. Ge, X. Feng, Kirk W. Cameron: Improvement of Power-Performance Efficiency for High-End Computing. *Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS 05) – HPPAC '05 Workshop*, 2005, Vol. 12. Workshop - 25 - 11: the 1st IEEE Workshop on High-Performance, Power-Aware Computing (HPPAC'05). Denver, Colorado, USA. 2005;233.2.

- [4] C. Lefurgy, K. Rajamani, F. Rawson, W. Felter, M. Kistler, T. Keller: Energy Management for Commercial Servers. *Computer*, 2003; 36(12):39–48.
- [5] T. Horvath, K. Skadron: Multi-mode energy management for multi-tier server clusters. *Proceedings of the 17th International Conference on Parallel Architectures and Compilation Techniques (PACT'08)*. Toronto, Ontario, Canada. 25-29 Oct., 2008; 270–279.
- [6] S. Saha, B. Ravindran: An Experimental Evaluation of Real-Time DVFS Scheduling Algorithms. *SYSTOR '12 Proceedings of the 5th Annual International Systems and Storage Conference*, 2012.
- [7] S. Huang, W. Feng: Energy-Efficient Cluster Computing via Accurate Workload Characterization. *Proc. CCGRID '09 Proceedings of the 2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid*, 2009, 68–75.
- [8] Giorgio Luigi Valentini, Walter Lassonde, Samee Ullah Khan, Nasro Min-Allah, Sajjad A. Madani, Juan Li, Limin Zhang, Lizhe Wang, Nasir Ghani, Joanna Kolodziej, Hongxiang Li, Albert Y. Zomaya: An overview of energy efficiency techniques in cluster computing systems. *Cluster Computing*, 10:1–13, 2011.
- [9] J.-K. Kim, H. J. Siegel, A. A. Maciejewski, R. Eigenmann: Dynamic resource management in energy constrained heterogeneous computing systems using voltage scaling. *IEEE Trans. on Parallel and Distrib. Syst.*, 19 (11): 1445–1457, 2008.