Оригинальная статья/Original article

УДК 004.75

DOI: http://doi.org/10.20914/2310-1202-2017-1-95-98

Определение оптимального числа дополнительных слоёв передаваемых данных схемы сокрытия сетевой латентности

Алексей Б. Новиков [1] a.b.novikov@vniia.ru
Георгий И. Евтушенко [1] evtushenko.georgy@gmail.com

[1] Всероссийский научно-исследовательский институт автоматики имени Н.Л. Духова, ул. Сущевская, 22, Москва, 101000, Россия

Реферат. Ключевым компонентом эффективности параллельных вычислений является организация обмена данными между вычислительными узлами. Для повышения эффективности параллельных вычислений необходимо сокращать задержки на обмен данными. Для этого был разработан алгоритм перекрытия задержек обмена данными B+2R. В отечественных и зарубежных работах не рассматривается способ выбора числа слоёв дополнительно передаваемых данных R. Для возможности применения математического аппарата оптимизации в работе вводятся модели всех систем, влияющих на время выполнения параллельного расчёта. Вводится модель сети передачи данных и модель параллельного расчётного приложения. Время вычисления ячейки считается постоянной величиной, зависящей от конкретного расчёта. Вводится оценка времени счёта в зависимости от количества слоёв дополнительно передаваемых данных. Далее, вводится производная зависимости времени счёта параллельного приложения от количества слоёв дополнительно передаваемых данных. Наименьший действительный положительный корень получившегося кубического уравнения является минимумом времени счёта параллельного приложения. Может оказаться так, что уравнение не будет иметь действительных положительных корней, это соответствует существенно большему времени локальной сетки с приграничными слоями по отношению к задержкам обмена данными, что делает не целесообразным применение рассматриваемого алгоритма. Для проверки полученных зависимостей был проведён вычислительный эксперимент, результаты которого согласуются с прогнозируемыми величинами. Стоит отметить, что целью проведения вычислительного эксперимента является не столько совпадение полученного времени счёта параллельного приложения с данным числом слоёв данных и времени счёта вычисляемого по предлагаемой модели, сколько совпадение минимумов этих зависимостей. Это объясняется тем, что цель разработанной модели - достижение минимального времени счёта, а не его прогнозирование. Результатом работы служит зависимость, позволяющая по ряду параметров вычислительного комплекса и задачи определить оптимальное количество слоёв дополнительно передаваемых данных.

Ключевые слова: схемы сокрытия сетевой латентности, B+2R, структурированные сетки

Optimal additional data layers amount determining for interconnect latency hiding scheme

Aleksei B. Novikov [1] a.b.novikov@vniia.ru
Georgii I. Evtushenko [1] evtushenko.georgy@gmail.com

[1] VNIIA, Suschevskaya str. 22, Moscow, 101000, Russia

Summary. The key component of parallel computing efficiency is the structure of data exchange between computing nodes. It is necessary to reduce the latency of data exchange to improve the efficiency of parallel computing. A B+2R algorithm for overlapping delays in the data exchange was offered for this purpose. Existing works do not offer a method for selecting the additionally transmitted data layer count R. We introduce the models of all systems affecting the parallel calculation time. It makes possible to apply the analytical optimization. We introduce a data transmission network and a parallel computing application models. We consider the cell calculation time is a constant value. The cell calculation time depends on the specific calculation parameters. We introduce an estimate of the computation time. Computation time depends on the additionally transmitted data layers count. Further we introduce the derivative of computation time equation. We use lowest positive real roots of the cubic equation. It's possible that the final cubic equation hasn't real positive roots. It's mean that local structured grid calculation time is much bigger than network latency. In that case, it's not recommended to use latency hiding schemes. Otherwise we recommend to use R equal to 1. Purpose of out research is to find optimal R. Optimal R value should lead to a calculation time equation minimum. The method proposed in the paper correspond to experimental result. Designed analytical model for B+2R algorithm makes possible to select optimal R value, which leads to the best calculation speedup.

Keywords: interconnect latency hiding schemes, B+2R, structured grids

Введение

Одним из ключевых компонентов эффективности массивно-параллельных вычислений является организация обмена данными между вычислительными узлами. При передаче данных существенные временные задержки могут приходиться на сетевую латентность. Таким образом возникает необходимость в использовании алгоритмов перекрытия задержек передачи данных для повышения эффективности параллельных вычислений. На данный момент практика применения программных методов перекрытия задержек обмена данными показала существенное повышение масштабируемости и производительности параллельных приложений [1, 2]. Большое количество работ посвящено аппаратным реализациям методов перекрытия задержек обмена данными [3-9]. Стоит отметить, что разработка программных методик является предпочтительным направлением, так как позволяет, с одной стороны, не закупать

Для цитирования

Новиков А.Б., Евтушенко Г.И. Определение оптимального числа дополнительных слоёв передаваемых данных схемы сокрытия сетевой латентности // Вестник ВГУИТ. 2017. Т. 79. № 1. С. 95–98. doi:10.20914/2310-1202-2017-1-95-98

For citation

Novikov A. B., Evtushenko G.I.. Optimal additional data layers amount determining for interconnect latency hiding scheme. *Vestnik VGUIT* [Proceedings of VSUET]. 2017. Vol. 79. no. 1. pp. 95–98. (in Russian). doi:10.20914/2310-1202-2017-1-95-98

дополнительное оборудование, а с другой, может применяться совместно с аппаратными методиками. В отечественных и зарубежных работах не рассматривается способ выбора числа слоёв дополнительно передаваемых данных R. Наша гипотеза состоит в том, что оптимальное число дополнительно передаваемых данных может быть выражено через параметры среды передачи данных и параметров параллельного приложения.

1.1 Модель сети

Для оценки времени передачи данных нами использовалась модель, предложенная Хокни:

$$\tau_{пд} = \tau_л + \frac{m}{B} \quad (1)$$

где $\tau_{пд}$ – время передачи m бит данных с пропускной способностью сети B [бит/секунда].

1.2 Оценка размера сетки на процессе

В данной работе рассматривается простейший случай декомпозиции структурированной сетки по процессам:

$$N_d^{(i)} = \begin{cases} \frac{N_d}{N_d^p} + 1 & P_d^{(i)} < N_d \% N_d^p \\ \frac{N_d}{N_d^p} & P_d^{(i)} \geq N_d \% N_d^p \end{cases}, d \in \{x, y, z\}$$

где $N_d^{(i)}$ – количество ячеек по измерению d на i-ом процессе,. N_d – количество ячеек по заданному измерению в базовой сетке, $P_d^{(i)}$ – номер i-го процесса по заданному измерению в сетке процессов. Таким образом, количество ячеек на одном процессе выражается как:

$$N_{ячеек}^{(i)} = \prod_{d \in \{x,y,z\}} N_d^{(i)}$$

И должно выполняться соотношение:

$$N_{ячеек} = \prod_{d \in \{x,y,z\}} N_d = \sum_{i=0}^{N_p} N_{ячеек}^{(i)}$$

1.3 Оценка времени счёта

В отсутствие наложения счёта и передачи данных, время счёта можно представить суммой времени обмена данными и непосредственно вычислений.

$$\tau = N_{it}(\tau_c(N_{ячеек}^{(i)}) + 2\tau_{пд}(N_{гя}^{(i)})) \quad (2)$$

Где количество граничных ячеек на i-ом процессе для двумерного случая без использования стратегии B+2R выражается как:

$$N_{гя}^{(i)} = 4N_x^{(i)}N_y^{(i)}$$

Очевидно, что время счёта τ_c является функцией от числа ячеек, подвергающихся счёту на i-ом процессе $N_{гя}^{(i)}$, а время передачи данных – функцией от количества ячеек и объема памяти, занимаемой ими.

Рассмотрим случай использования стратегии B+2R. Под R будем понимать количество слоёв ячеек соседних процессов.

$$\tau = \frac{N_{it}}{R}\left[2N_n\tau_{пд}\left(N_R(R)\right) + \sum_{r=1}^{R-1}\tau_c\left(N_{ячеек}^{(i)} + N_R(r)\right)\right] \quad (3)$$

где N_n – максимальное число соседей в сетке процессов. Для двумерного случая с использованием стратегии B+2R $N_n = 8$, для трёхмерного – $N_n = 26$, при условии, что $R < N_d$. Функция, определяющая количество ячеек передаваемых как R слоёв ячеек от соседа, будет определена как показано в уравнении 4 для двумерного случая.

$$N_R = 2N_x^{(i)}R + 2N_y^{(i)}R + 4R^2 = 2R\left[N_x^{(i)} + 2R + N_y^{(i)}\right] \quad (4)$$

$$\tau = \frac{N_{it}}{R}\left[2N_n\tau_n + \tau_c\left(N_{ячеек}R + \frac{2R(2R-1)(R-1)}{3} + R(R-1)(N_x + N_y)\right) + \frac{4N_c^s N_c^{Base} R(N_x + N_y + 2R)}{B}\right] \quad (5)$$

$$\frac{\partial\tau}{\partial R} = \left(\frac{8}{3}N_{it}\tau_c\right)R^3 + \left(\frac{N_{it}}{3B}24N_c^s N_c^{Base} + 3N_{it}B\tau_c(N_x + N_y - 2)\right)R^2 - 2N_nN_{it}\tau_n \quad (6)$$

Взяв производную уравнения 5, мы получаем выражение 6, приравняв которое 0, получаем кубическое уравнение, решением которого является оптимальное для данного вычислительного комплекса и данной задачи R. Стоит отметить, что решение необходимо выбирать, следуя нескольким ограничениям. С одной стороны, R должно быть положительным вещественным числом. При определённых конфигурациях МВК может оказаться, что уравнение 6 не имеет решений в области вещественных чисел, это соответствует случаю, когда небольшая латентность сети позволяет пересылать данные существенно быстрее, чем считается ячейка.

$$N_R = 2R\left(4R^2 + N_xN_y + N_xN_z + N_yN_z + 2R\left(N_x + N_y + N_z\right)\right) \quad (7)$$

$$\tau = \frac{N_{it}}{R}\left[2N_n\tau_n + \tau_c\left(2R^2\left(R-1\right)^2 + N_cR + \frac{2}{3}R\left(2R-1\right)\left(R-1\right)\left(N_x + N_y + N_z\right) + \right.\right. \quad (8)$$

$$R\left(R-1\right)\left(N_xN_y + N_xN_z + N_yN_z\right)\Big) +$$

$$\left.\frac{4}{B}N_c^{Base}N_c^sR\left(4R^2 + N_xN_y + N_xN_z + N_yN_z + 2R\left(N_x + N_y + N_z\right)\right)\right]$$

$$\frac{\partial\tau}{\partial R} = \left(6N_{it}\tau_c\right)R^9 + \left(\frac{N_{it}}{3B}\left(-24B\tau_c + 96N_c^{Base}N_c^s + 8B\tau_c\left(N_x + N_y + N_z\right)\right)\right)R^8 + \quad (9)$$

$$\left(\frac{N_{it}}{3B}\left(6B\tau_c + 24N_c^{Base}N_c^s\left(N_x + N_y + N_z\right) - 6B\tau_c\left(N_x + N_y + N_z\right) + 3B\tau_c\left(N_xN_y + N_xN_z + N_yN_z\right)\right)\right)R^7 +$$

$$\left(-2N_{it}N_n\tau_n\right)R^5$$

В выражении 8 представлена зависимость времени счёта от числа слоёв передаваемых соседями данных. Приравнивая производную выражения 8 нулю, мы находим минимум искомой функции (рисунок 1). Данное выражение не учитывает времени, которое затрачивается на подготовку данных к пересылке, и обработки данных во время приёма. Тем не менее, выражение хорошо согласуется с результатами экспериментов. Для эксперимента использовалась конфигурация из 20 машин. Решатель содержал уравнение переноса для трёхмерного случая.

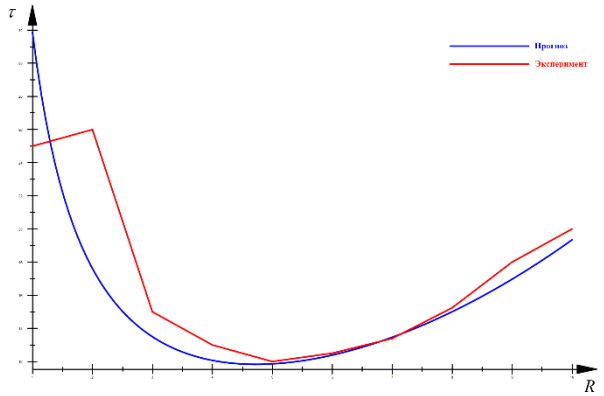

Рисунок 1. Зависимость времени счёта от R

Figure 1. Dependence of calculation time on R

График производной зависимости времени счёта от числа слоёв передаваемых данных (рисунок 2) показывает пересечение близко к 5 (половина сетки) слоям для данной конфигурации запуска, что соответствует полученному в экспериментах минимуму. Среднеквадратичное отклонение запусков эксперимента составляет $\sigma = 0.335$ секунды.

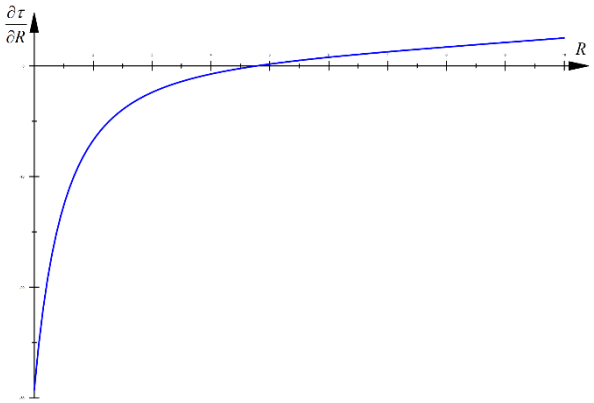

Рисунок 2. Производная зависимости времени счёта от R

Figure 2. Derivative of dependence of calculation time on R

Заключение

В работе представлен метод определения оптимального числа передаваемых слоёв данных для схемы сокрытия задержек. На наличие минимума функции времени счёта от числа слоёв существенно влияет латентность сети передачи данных, время счёта одной ячейки и количество соседей у данного процесса.

ЛИТЕРАТУРА

1 Brandon G.A., Kalyan S.P., Sudip K.S. Efficient Simulation of Agent-Based Modoels on Multi-GPU and Multi-Core Clusters. Proceedings of SIMUTools. 2010 March 15–19

2 Калмыков В.В., Ибраев, Р.А. Алгоритм с перекрытиями для решения системы уравнений мелкой воды на параллельных компьютерах с распределённой памятью // Вестник УГАТУ. 2013. № 5. С. 252-259.

3 Jaehyuk H. Hardware Techniques to Reduce Communication Costs in Multiprocessors. Doctoral dissertation, 2006.

4 Cicotti P. Tarragon: a programming model for latency-hiding scientific computations. Doctoral dissertation, 2011.

5 Alameldeen Alaa R. Using Compression to improve chip multiprocessor performance. Doctoral dissertation, 2006.

6 Afsahi A. Design and Evaluation of Communication Latency Hiding/Reduction Techniques for Message-Passing Environments. Doctoral dissertation, 2000.

7 Chen Li-li, Huang Jian-xin, Zhang Jing A Latency-Hiding Algorithm for ABMS on Parallel/Distributed Computing Environment. ACM/IEEE/SCS 26th Workshop on Principles of Advanced and Distributed Simulation, 2012.

8 Yong Chen, Surendra Byna, Xian-He Sun, Rajeev Thakur et al. Hiding I/O latency with pre-execution prefetching for parallel applications. In Proceedings of the 2008 ACM/IEEE conference on Supercomputing (SC '08). IEEE Press, Piscataway, NJ, USA, 2008, Article 40, pp. 10.

9 Hakan Grahn Comparative Evaluation of Latency-Tolerating and -Reducing Techniques for Hardware-Only and Software-Only Directory Protocols. Journal of Parallel and Distributed Computing 60, 2000, pp. 807-834.

REFERENCES

1 Brandon G.A. Kalyan S.P. Sudip K.S. Efficient Simulation of Agent-Based Modoels on Multi-GPU and Multi-Core Clusters. Proceedings of SIMUTools. 2010 March 15–19

2 Kalmicov, V.V., Ibraev, R.A. Latency hiding algorithm for shallow water equations solve on parallel computers. *Vestnik UGATU* [Ufa State Aviation Technical University] 2013, no 5, pp. 252-259. (in Russian)

3 Jaehyuk H. Hardware Techniques to Reduce Communication Costs in Multiprocessors. Doctoral dissertation, 2006.

4 Cicotti P. Tarragon: a programming model for latency-hiding scientific computations. Doctoral dissertation, 2011.

5 Alameldeen Alaa R. Using Compression to improve chip multiprocessor performance. Doctoral dissertation, 2006.

6 Afsahi A. Design and Evaluation of Communication Latency Hiding/Reduction Techniques for Message-Passing Environments. Doctoral dissertation, 2000.

7 Chen Li-li, Huang Jian-xin, Zhang Jing A Latency-Hiding Algorithm for ABMS on Paral-lel/Distributed Computing Environment. ACM/IEEE/SCS 26th Workshop on Principles of Ad-vanced and Distributed Simulation, 2012.

8 Yong Chen, Surendra Byna, Xian-He Sun, Rajeev Thakur et al. Hiding I/O latency with pre-execution prefetching for parallel applications. In Proceedings of the 2008 ACM/IEEE conference on Supercomputing (SC '08). IEEE Press, Piscataway, NJ, USA, 2008, Article 40, pp. 10.

9 Hakan Grahn, Comparative Evaluation of Latency-Tolerating and -Reducing Techniques for Hardware-Only and Software-Only Directory Protocols. Journal of Parallel and Distributed Computing 60, 2000, pp. 807-834.

СВЕДЕНИЯ ОБ АВТОРАХ

Георгий И. Евтушенко аспирант, младший научный сотрудник лаборатории высокопроизводительных вычислений, Всероссийский научно-исследовательский институт автоматики имени Н.Л. Духова, ул. Сущевская, 22, г. Москва, 101000, Россия, evtushenko.georgy@gmail.com

Алексей Б. Новиков заместитель начальника научно-исследовательского отдела-начальник научно-исследовательской лаборатории, Всероссийский научно-исследовательский институт автоматики имени Н.Л. Духова, ул. Сущевская, 22, г. Москва, 101000, Россия, a.b.novikov@vniia.ru

INFORMATION ABOUT AUTHORS

Georgii I. Evtushenko junior research fellow, VNIIA, Suschevskaya str. 22, Moscow, 101000, Russia, evtushenko.georgy@gmail.com

Aleksei B. Novikov deputy head of the research department, research laboratory chief, VNIIA, Suschevskaya str. 22, Moscow, 101000, Russia, a.b.novikov@vniia.ru

КРИТЕРИЙ АВТОРСТВА

Георгий И. Евтушенко предложил методику определения оптимального количества слоёв передаваемых данных, отвечал за написание исходных кодов

Алексей Б. Новиков предложил методику проведения эксперимента, отвечал за написание исходных кодов

CONTRIBUTION

Georgii I. Evtushenko proposed a method of the optimal amount of additional data layers, writes the source codes

Aleksei B. Novikov proposed a scheme of the experiment and writes the source codes

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

ПОСТУПИЛА 31.08.2016

ПРИНЯТА В ПЕЧАТЬ 17.02.2017

CONFLICT OF INTEREST

The authors declare no conflict of interest.

RECEIVED 8.31.2016

ACCEPTED 2.17.2017