10.7256/2305-6061.2013.4.10548

Гришенцев А.Ю., Коробейников А.Г.

ПОСТАНОВКА ЗАДАЧИ ОПТИМИЗАЦИИ РАСПРЕДЕЛЁННЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ

Аннотация: *Разработана модель и постановка задачи оптимизации распределённых вычислительных систем. Результаты работы хорошо согласуется с законом Амдала и позволяют при помощи методов теории игр и оптимизации отыскивать наиболее удачные, с точки зрения эффективности использования вычислительных ресурсов, решения при проектировании или модернизации распределённых вычислительных систем. Рассматривается поточная модель распределённой вычислительной системы непрерывного времени. Недостатком такой модели является возможность моделирования только поточной распределённой вычислительной системы, для рассмотрения случая передачи данных блоками необходимо ввести модель системы дискретного времени. Современные распределённые вычислительные системы (РВС) могут содержать множество отдельных вычислительных единиц связанных коммуникационной сетью и распределённых по разным частям Земли и околоземного пространства. Рассматривается блочная модель распределённой вычислительной системы дискретного времени. Такая модель позволяет рассматривать как поточную, так и блочную обработку данных, учитывать время задержки необходимое для синтеза и передачи данных. Решение задачи оптимизации возможно путём последовательного перебора, с применением методов теории игр и оптимизации, для вычислительных задач, ресурсов узлов*

Ключевые слова: *поточная модель, распределённая вычислительная система, оптимизация, закон Амдала, система дискретного времени, орграф, узел графа, блочная модель РВС, время вычислительного канала, теория игр*

Поточная модель распределённой вычислительной системы непрерывного времени

Рассмотрим некоторую абстрактную вычислительную систему которую можно задать как ориентированный граф (*орграф*) $\mathbf{G}(\mathbf{V},\mathbf{E})$. Множество вершин орграфа $\mathbf{V}=\{v_1,v_2,...,v_n\}$ соответствует некоторым вычислительным элементам (узлам) потребляющим и производящим данные. Рёбра орграфа $\mathbf{E}$ соответствуют каналам обмена информацией между элементами множества $\mathbf{V}$. Будем обозначать ребро e_{ij}, соответствующее информационному каналу, соединяющему вершину v_i с вершиной v_j. Положим, что каждый канал e_{ij} предназначен для осуществления основного потока данных от узла v_i к v_j, при этом канал может поддерживать некоторый, существенно меньший поток данных в обратном

10.7256/2305-6061.2013.4.10548

направлении (от v_i к v_j), что на практике соответствует запросам от v_i к v_j на: отправку, подтверждение получения, проверку целостности данных и т.д. Пусть каждый канал e_{ij} характеризуется весовой характеристикой определяющей скорость передачи данных в прямом направлении. С учётом, что объём данных x_{ij} передаваемых в прямом направлении существенно больше объёма данных передаваемых в обратном направлении, временем передачи данных в обратном направлении можно пренебречь. Каждый вычислительный узел, соответствующий узлу графа v_i, производит данные y_j. Каждый вычислительный узел v_i, характеризуется: вычислительной мощностью Q_j, единицы измерения [операций/с]; вычислительной сложностью алгоритма C_j (единицы измерения [операций]), которая является функцией $C_j = f(\int_{\vartheta_1} e_{1j}dt, \int_{\vartheta_2} e_{2j}dt,..., \int_{\vartheta_n} e_{nj}dt)$ от получаемых узлом v_j данных $x_{1j}, x_{2j},..., x_{nj}$ за время $\vartheta_1, \vartheta_2,..., \vartheta_n$. Таким образом, множество C_j соответствует элементам (локальным задачам) образующим общую (глобальную) вычислительную задачу решаемую РВС.

Во многих системах, вычисления происходят порциями (блоками), измеряемыми объёмами данных существенно больших размеров, чем бит или байт. Таким образом, каждый узел ожидает некоторого объёма данных достаточных для начала вычислений, если при этом предыдущие вычисления на узле были завершены, то вычислительная мощность данного узла будет простаивать, что может негативно повлиять на общее время вычислений. При поточной обработке вычислительному узлу необходим некоторый малый объём данных, передача которого не занимает существенного времени и вычислительный процесс происходит непрерывно. В реальных системах обычно присутствуют как поточные вычисления, так и блочные.

Для построения данной модели допустим, что вычисления происходят поточно, причём при изменении времени $\Delta t \to 0$ приращение данных в потоке так же стремиться к нулю, что возможно выразить следующей системой дифференциальных уравнений (1):

$$\begin{cases} e_{ij} = \dfrac{dx_{ij}}{dt} \\ m_j \dfrac{Q_j}{C_j} = \dfrac{dy_j}{dt} \\ C_j = f(\int_{\vartheta_1} e_{1j}dt, \int_{\vartheta_2} e_{2j}dt,..., \int_{\vartheta_n} e_{nj}dt) \end{cases}, \qquad (1)$$

где e_{ij} – скорость передачи данных от узла v_i к v_j m_j – коэффициент пропорциональности, таким образом, выражение $m_j Q_j / C_j$ определяет скорость производства данных y_j вычислительным узлом v_j. Недостатком такой модели является возможность моделирования только поточной распределённой

10.7256/2305-6061.2013.4.10548

вычислительной системы, для рассмотрения случая передачи данных блоками необходимо ввести модель системы дискретного времени.

Блочная модель распределённой вычислительной системы дискретного времени

В таблице (табл. 1.) рассмотрена послойная блочная модель РВС дискретного времени. В качестве примера выделено три слоя: передающий – содержит вычислительные узлы $v_p, v_{p+1},..., v_{p+k}$, где для соответствующих выражений индекс i принимает значения из множества $p, p+1,..., p+k$ $(i \in \{p, p+1,..., p+k\})$, узлы данного слоя осуществляют синтез и начало передачи данных;

принимающий – содержит вычислительные узлы $v_s, v_{s+1},..., v_{s+m}$, где для соответствующих выражений индекс j принимает значения из множества $j \in \{s, s+1,..., s+m\}$ $(j \in \{s, s+1,..., s+m\})$, узлы данного слоя осуществляют приём и синтез данных;

транспортный – содержит каналы передачи данных между передающим и принимающим слоями и осуществляет непосредственно передачу.

Собственно каждый вычислительный узел может одновременно быть принимающим и передающим, причём возможны петли, т.е. случаи, когда узел передаёт данные самому себе. Множества передающих и принимающих смежных узлов могут пересекаться.

Такая модель позволяет рассматривать как поточную, так и блочную обработку данных, учитывать время задержки необходимое для синтеза и передачи данных.

Рассмотрим пример: пусть некоторый узел v_s получает данные от узлов $v_p, v_{p+1},..., v_{p+k}$ начавших вычисления одновременно, причём v_s может начать вычисления только после получения всех блоков данных от узлов $v_p, v_{p+1},..., v_{p+k}$, тогда время задержки начала вычислений в узле v_s с начала вычислений в узлах $v_p, v_{p+1},..., v_{p+k}$ можно определить как: $\tau_s = \max(\Delta t_p + \Delta t_{p,s}, \Delta t_{p+1} + \Delta t_{p+1,s},..., \Delta t_{p+k} + \Delta t_{p+k,s})$. Заметим, что возможно определить время передачи и обработки данных по некоторому пути следования соответствующего всему циклу обработки (от ввода исходных данных до получения конечного результата), обозначим такое время T_w и назовём его *временем вычислительного канала*, где w – индекс, принадлежащий множеству $\mathbf{W}$ индексов конечных узлов в цепи вычисления РВС.

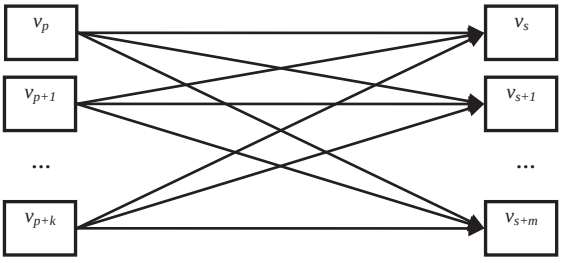
Таким образом, вычислительный процесс может быть задержан. Особенно существенной может быть задержка при многослойной модели РВС, возможен значительный простой вычислительных мощностей при значительной же вариабельности времени задержек $(\Delta t_p + \Delta t_{p,s}, \Delta t_{p+1} + \Delta t_{p+1,s},..., \Delta t_{p+k} + \Delta t_{p+k,s})$.

Частично данную проблему решает буферизация данных с предобработкой. Например: передающий узел v_p, вычислив новый блок данных может передать его принимающему узлу v_s. Принимающий узел

10.7256/2305-6061.2013.4.10548

v_s, произведя предварительную предобработку (не дожидаясь прочих данных необходимых для получения результата y_s) блока данных x_{sp} размещает их в буфере тем самым освобождая канал связи e_{sp} для передачи нового блока. Такой подход может несколько уменьшить время простоя, но не решает вопрос неэффективного использования вычислительных ресурсов полностью. Координальным решением проблемы будет оптимизация распределения вычислительной нагрузки по узлам РВС и планирования трафика каналов связи.

Таблица 1. Послойная модель РВС дискретного времени.

	Слои РВС		
характеристика	вычислительные узлы передающий слой	каналы передачи данных транспортный слой	вычислительные узлы принимающий слой
данные	$y_i=m_iQ_i\,\Delta t_i/C_i$	$x_{ij}=e_{ij}\,\Delta t_{ij}$	$y_j=m_jQ_j\,\Delta t_j/C_j$
скорость	m_iQ_i/C_i	e_{ij}	m_jQ_j/C_j
время	Δt_i	Δt_{ij}	Δt_j
			

Полученные уравнения (табл. 1) запишем в виде системы (2):

$$\begin{cases} e_{ij} = \dfrac{x_{ij}}{\Delta t_{ij}} \\ \dfrac{m_j Q_j}{C_j} = \dfrac{y_j}{\Delta t_j} \\ C_j = f(y_i, y_2, ..., y_n) \end{cases} . \qquad (2)$$

На базе полученной модели, возможно, решать задачу оптимизации использования РВС при проектировании или рефакторинге.

Постановка задачи оптимизации РВС

При постоянстве числа узлов **V**, транспортных каналов **E**, суммы вычислительных мощностей $\sum_{i=1}^{n} Q_i = const$ и маршрутов следования данных необходимо таким образом распределить вычислительные

10.7256/2305-6061.2013.4.10548

ресурсы Q_i между вычислительными задачами C_i, что бы минимизировать время получения конечного результата $\max_{\mathbf{w}}(T_w)$. Разделим T_w на компоненты образованные задержкой в каналах передачи данных T_w^E и вычислительных узлах T_w^V.

Задачу оптимизации распределения ресурсов Q_i между вычислительными задачами C_j, без учёта задержки при передаче данных, можно выразить в виде системы (3):

$$\begin{cases} \sum_{i=1}^{n} Q_i = const \\ \dfrac{m_j Q_j}{C_j} = \dfrac{y_j}{\Delta t_j} \\ C_j = f(y_i, y_2, ..., y_n) \\ T_w^V = \sum_{S} \max_{N}(\Delta t_j) \\ \max_{\mathbf{w}}(T_w^V) \to \min \end{cases}, \qquad (3)$$

для $i, j \in \{1,2,...,n\}$, время задержки вычислительного канала T_w^V рассчитывается как сумма максимальных значений $\max_{N}(\Delta t_j)$ (при N элементах в слое) из каждого слоя образующего цепочку вычислительного канала.

При оптимизации распределения ресурсов Q_i между вычислительными задачами C_j, с учётом задержки при передаче данных система (3) будет иметь вид (4):

$$\begin{cases} \sum_{i=1}^{n} Q_i = const \\ e_{ij} = \dfrac{x_{ij}}{\Delta t_{ij}} \\ y_j = \dfrac{m_j Q_j}{C_j} \Delta t_j \\ C_j = f(y_i, y_2, ..., y_n) \\ T_w = \varphi(\Delta t_{ij}, \Delta t_j) \\ \max_{\mathbf{w}}(T_w) \to \min \end{cases}. \qquad (4)$$

где $\varphi(\Delta t_{ij}, \Delta t_j)$ – функция расчёта максимального времени обработки (передачи и вычисления) данных в вычислительном канале.

10.7256/2305-6061.2013.4.10548

Заключение

Полученная модель хорошо согласуется с законом Амдала[1], который гласит: «В случае, когда задача разделяется на несколько частей, суммарное время её выполнения на параллельной системе не может быть меньше времени выполнения самого длинного фрагмента». Ускорение расчёта в системе параллельных вычислений, по сравнению с последовательными вычислениями, можно определить по выражению (5):

$$W \le \frac{1}{\chi + \frac{1-\chi}{p}}, \qquad (5)$$

где $\chi \in [0,1]$ – часть вычислений которая может быть вычислена последовательно, $(1-\chi) \in [0,1]$ – часть вычислений которая может быть распараллелена, p – число параллельных процессов.

В соответствии, с полученными моделями (3) и (4) оптимизация сводится к снижению времени выполнения T_w^V и T_w соответственно, затраченного на решение цепочки задач, образованной вычислительными задачами C_j, решаемыми за счёт ресурсов узлов Q_i, и передачи данных между узлами.

Решение задачи оптимизации возможно путём последовательного перебора, с применением методов теории игр и оптимизации, для вычислительных задач C_j, ресурсов узлов Q_i.

Библиография:

1. Таненбаум Э. Распределенные системы. Принципы и парадигмы / Э. Таненбаум, М. Ван Стен — СПб.: Питер, 2003. — 877 с: ил.
2. Гришенцев А. Ю., Муромцев Д. И. Система управления данными наблюдений солнечно-земной физики «MI» // Регистрация программы для ЭВМ от 21.07.2011 г. – № 2011615714.
3. Гришенцев А. Ю., Коробейников А. Г. Обратная задача радиочастотного зондирования ионосферы. Российская академия наук «Журнал радиоэлектроники» электронный журнал. Web: http://jre.cplire.ru/jre/oct10/6/text.html №10-октябрь 2010 г.
4. Антонов А. Под законом Амдала (рус.) Компьютерра. — 11.02.2002. — № 430. Web: http://old.computerra.ru/offline/2002/430/15838/

References:

1. Tanenbaum E. Raspredelennye sistemy. Printsipy i paradigmy / E. Tanenbaum, M. Van Sten — SPb.: Piter, 2003. — 877 s: il.
2. Grishentsev A. Yu., Muromtsev D. I. Sistema upravleniya dannymi nablyudenii solnechno-zemnoi fiziki «MI» // Registratsiya programmy dlya EVM ot 21.07.2011 g. – № 2011615714.
3. Grishentsev A. Yu., Korobeinikov A. G. Obratnaya zadacha radiochastotnogo zondirovaniya ionosfery. Rossiiskaya akademiya nauk «Zhurnal radioelektroniki» elektronnyi zhurnal. Web: http://jre.cplire.ru/jre/oct10/6/text.html №10-oktyabr' 2010 g.
4. Antonov A. Pod zakonom Amdala (rus.) Komp'yuterra. — 11.02.2002. — № 430. Web: http://old.computerra.ru/offline/2002/430/15838/

