

А.Г. Басыров¹
В.В. Ширококов²

A.G. Basyrov
V.V. Shirobokov

**ПОДХОД К РАСПРЕДЕЛЕННОЙ
ОБРАБОТКЕ ИНФОРМАЦИИ
В МОБИЛЬНОЙ НЕОДНОРОДНОЙ
ВЫЧИСЛИТЕЛЬНОЙ СЕТИ**

**APPROACH TO THE DISTRIBUTION
OF INFORMATION PROCESSING
IN HETEROGENEOUS COMPUTING
MOBILE NETWORK**

Рассмотрена проблема организации распределенных вычислений в мобильных неоднородных вычислительных сетях с изменяющейся пропускной способностью каналов связи при ограничениях на объем памяти и энергоресурс автономных источников питания.

The problem of distributed computing in heterogeneous mobile computing systems with variable bandwidth communication channels by the restrictions on the amount of memory and energy independent power supply is considered.

Ключевые слова: распределенная обработка, мобильная вычислительная система, неоднородная вычислительная сеть.

Keywords: distributed processing, mobile computer system, heterogeneous computer network.

На современном этапе развития сетевых технологий большое распространение получили мобильные вычислительные сети. Отличительные свойства данных сетей состоят в том, что они имеют большое число различающихся по производительности элементов (вычислительных устройств), обменивающихся данными по ненадежным каналам с переменной пропускной способностью, объединены в общий ресурс, характеризуются сложной зависимостью производительности от режимов энергопотребления и характеристик каналов связи, а также изменением структуры в процессе функционирования. В данных сетях имеет место зависимость взаимного положения узлов на скорость обработки данных.

Примером подобных сетей могут служить мобильная сеть переносных компьютеров, рас-

пределенная вычислительная система, состоящая из бортовых компьютеров группировки микроспутников и т.д.

Расширение масштабов использования мобильных неоднородных вычислительных сетей влечет необходимость повышения автономности их функционирования [1; 2]. Перенос решения ряда задач со стационарных на мобильные вычислительные средства приводит к необходимости увеличения производительности отдельного мобильного компьютера и совершенствование подходов к построению и организации функционирования мобильной вычислительной сети в целом. Нарастивание производительности вычислительного устройства мобильного компьютера вступает в противоречие с ограничениями по массе, энергозатратам, требованиям по надежности. При этом нарастаемые ресурсы мобильного компьютера не будут использованы в полной мере, а будут востребованы лишь на относительно коротких интервалах его функционирования.

Решением данного противоречия является подход, в основу которого положена технология предоставления информационно-вычисли-

¹ Доктор технических наук, профессор, начальник кафедры информационно-вычислительных систем и сетей Военно-космической академии им. А.Ф. Можайского.

² Адъюнкт кафедры информационно-вычислительных систем и сетей Военно-космической академии им. А.Ф. Можайского.

тельных ресурсов на основе модели «клиент-сервер» с конфигурацией «тонкий клиент». Данный подход позволяет минимизировать требования к аппаратно-программным ресурсам компьютера-клиента, перенося часть информационно-вычислительной нагрузки на центры обработки данных (серверы).

При организации распределенной обработки информации в мобильной неоднородной вычислительной сети возникает задача оптимального распределения вычислительной нагрузки между клиентом (α) и сервером (β) высокой производительности при ограниченном энергоресурсе бортовых источников питания. Это приводит к необходимости создания модели распределенной обработки информации в мобильной неоднородной вычислительной сети (рис. 1), учитывающей: состав и характеристики специального программного обеспечения (ПО) Ψ , входные данные $V_{вх}$, поступающие на обработку, динамически изменяющееся расстояние d между клиентом и сервером. Распределение вычислительной нагрузки между α и β заключается в разделении последовательности задач ПО Ψ на два непересекающихся подмножества: задач, выполняемых на α , и задач, выполняемых на β . Для этого выбирается целое число $1 \leq k \leq n$, такое, что задачи с номерами $1, 2, \dots, k$ будут выполняться на α , а задачи с номерами $k + 1, k + 2, \dots, N$ будут выполняться на β . При этом формируются целевые показатели обработки информации – время T выполнения целевой задачи и энергоресурсы E , затраченные на ее выполнение:

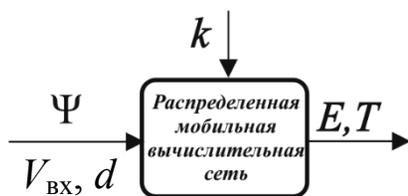


Рис. 1. Абстрактная модель распределенного выполнения целевой задачи

Для выполнения возложенных на клиента задач предназначается ПО $\Psi = \{\psi_1, \psi_2, \dots, \psi_n\}$, включающее n -подзадач (рис. 2). Каждая подзадача включает $\psi_i = \{S_i, P_i\}$ набор инструкций, состоящий из S_i инструкций данного кода, выполняющихся только последовательно друг за другом на одном ядре вычислительного устройства, и P_i программно-независимыми друг от друга инструкций, которые могут выполняться одновременно на нескольких ядрах процессора мобильной вычислительной системы (МВС).

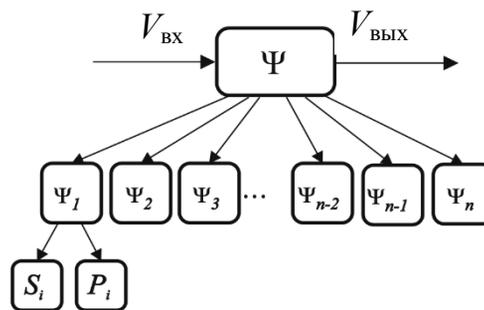


Рис. 2. Структура целевой задачи

Входными данными (рис. 3) для ПО Ψ является информация $V_{вх}$ объемом v_1 , поступающая с входных устройств, установленных на клиенте в соответствии с его предназначением, а выходными данными $V_{вых}$ объемом v_{n+1} является результат выполнения программы Ψ , который затем передается потребителем. При этом должно выполняться условие:

$$\left(\sum_{i=1}^k v_i + p_\alpha \right) \leq W_\alpha \cap \left(\sum_{i=k+1}^n v_i + p_\beta \right) \leq W_\beta, \quad (1)$$

где W_α, W_β – объем памяти на α или β , соответственно;

p_α, p_β – объем памяти, занимаемой программами на α или β , соответственно.

Объем памяти программ p_α, p_β определяется по формулам:

$$p_\alpha = \sum_{i=1}^k p_i, \quad (2)$$

$$p_\beta = \sum_{i=k+1}^n p_i, \quad (3)$$

где p_i – объем памяти, занимаемый программой для выполнения i -й подзадачи на α или β , который может быть найден по формуле:

$$p_i = (S_i + P_i)u, \quad (4)$$

где u – размер одной инструкции (байт);

S_i и P_i – количество инструкций программного кода, составляющих подзадачу ψ_i .

При расчете объема памяти, занимаемой программой для выполнения последовательности подзадач на МВС α или β , должно выполняться условие (1).

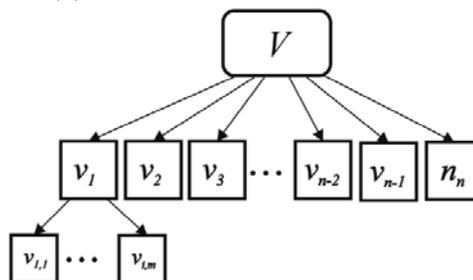


Рис. 3. Входные данные для выполнения подзадачи

Общая схема выполнения программы на МВС клиента представлена на рис. 4.

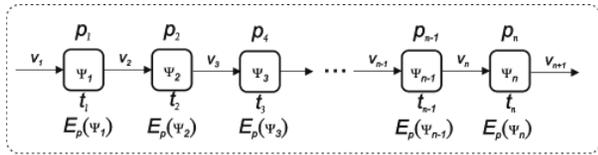


Рис. 4. Схема выполнения задачи на клиенте

Сложные алгоритмы обработки входных данных требуют наличия высокопроизводительной МВС и приемлемого бортового энергоресурса для обеспечения вычислительного процесса. Ограничение по масса-габаритным характеристикам клиента не позволяет устанавливать на него высокопроизводительную МВС и мощную систему бортового энергообеспечения. Кроме того, при выполнении сложных вычислительных алгоритмов на МВС клиента уменьшается ресурс системы бортового энергообеспечения, который в основном и определяет срок его активного существования. Для большинства клиентов мобильной вычислительной сети после истечения гарантийного срока активного существования характеристики системы энергопитания значительно ухудшаются, и клиент используется с ограничениями по целевому назначению.

Перенос обработки части подзадач с компьютера α на компьютер β позволит уменьшить нагрузку на систему электроснабжения компьютера α . Перенос осуществляется по технологии предоставления информационно-вычислительных ресурсов на основе модели «клиент-сервер» с конфигурацией «тонкий клиент» (рис. 5).



Рис. 5. Модель «клиент-сервер» с конфигурацией «тонкий клиент»

Перенос части вычислений с компьютера α на компьютер β разделяет выполнение программы на две части. Первая часть подзадач ψ_i , $i \in (1, \dots, k)$ выполняется на α до момента окончания выполнения k -й подзадачи, затем результат выполнения k -й подзадачи передается по каналу связи на β и продолжается обработка подзадач ψ_i , $i = k + 1, \dots, n$. После обработки данных в МВС β результат передается потребителю.

Схема распределенной обработки программы в компьютерах α и β представлена на рис. 6, где p_i – объем памяти, занимаемый программой для выполнения i -й подзадачи клиента, t_i – длительность монопольного решения i -й подзадачи при входных данных v_i , $E_p(\psi_i)$ – потребляемая процессором компьютера мощность при выполнении i -й подзадачи.

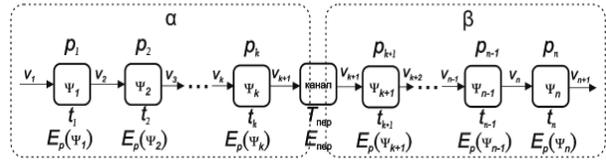


Рис. 6. Схема выполнения задачи в клиент-серверной системе

Цель организации распределенной обработки информации между клиентом и сервером заключается в минимизации времени выполнения целевых задач при ограничениях на объем памяти каждого компьютера и на энергоресурс автономных источников питания.

Формально постановку задачи с учетом (1) можно представить как необходимость поиска оптимального значения k^* при условии:

$$\left\{ \begin{array}{l} k^* = \arg \min T(k, d, V, \Psi, L) \\ \left(\sum_{i=1}^{k^*} v_i + p_\alpha \right) \leq W_\alpha \\ \left(\sum_{i=k^*+1}^n v_i + p_\beta \right) \leq W_\beta \\ E(k^*, d, T, P_{pow}, P_{ппа}) \leq E_{тр} \end{array} \right. , \quad (5)$$

где k – наибольший номер задачи, выполняемой на α ;

T – время решения целевой задачи;

d – расстояние между α и β ;

V – массив входных данных для выполнения каждой подзадачи;

Ψ – специальное программное обеспечение;

L – количество инструкций, выполняемых процессором в единицу времени;

W_α, W_β – объем памяти на α или β , соответственно;

P_{pow} – потребляемая мощность процессором;

$P_{ппа}$ – потребляемая мощность приемо-передающей аппаратурой.

Так как временные и энергетические ресурсы складываются из затрат на обработку информации и на ее передачу между клиентом и сервером, для решения поставленной задачи необходимо найти:

– время выполнения целевой задачи и затраченного суммарного количества энергии на ее выполнение;

– пропускную способность канала связи;

– количество потребляемой энергии бортовой вычислительной системой при выполнении набора инструкций;

– количество потребляемой энергии при передаче данных.

Общее время выполнения задачи в системе, основанной на модели «тонкий клиент», определяется по формуле:

$$T = T_{\alpha} + T_{\beta} + T_{\text{пер}}, \quad (6)$$

где T_{α} – время выполнения последовательности подзадач на α ;

T_{β} – время выполнения последовательности подзадач на β ;

$T_{\text{пер}}$ – время передачи данных по каналу связи между α и β .

Время выполнения одной последовательности подзадач на α определяется по формуле:

$$T_{\alpha} = \sum_{i=1}^k t_i(\psi_i), \quad (7)$$

где $t_i(\psi_i)$ – длительность монопольного решения i -й подзадачи ψ_i на α .

Время выполнения одной последовательности подзадач на β определяется по формуле:

$$T_{\beta} = \sum_{i=k+1}^n t_i(\psi_i), \quad (8)$$

где $t_i(\psi_i)$ – длительность монопольного решения i -й подзадачи ψ_i на β .

Зависимость меры пропускной способности канала от помех выражена в теореме Шеннона следующим образом [2]:

$$C = \frac{P_{\text{пер}} G S_{\text{эф}} \eta_{\text{пр}} \eta_{\text{пер}} L \log_2 \left(1 + \frac{P_c}{P_{\text{ш}}} \right)}{d^2 4\pi w \left(\frac{P_c}{P_{\text{ш}}} \right)}, \quad (9)$$

где d – расстояние между α и β ;

G – коэффициент направленности передающей антенны;

$P_{\text{пер}}$ – мощность передатчика;

$S_{\text{эф}}$ – эффективная площадь антенны;

w – коэффициент пропорциональности мощности шумов к ширине полосы пропускания;

$P_c/P_{\text{ш}}$ – отношение мощности сигнала к мощности помех;

$\eta_{\text{пр}}, \eta_{\text{пер}}$ – коэффициенты полезного действия приемного и передающего антенно-фидерных трактов;

L – потери, вызванные затуханием энергии сигнала в свободном пространстве, ионизацией

слоев атмосферы, шумом в приёмных трактах линий связи, неточностью наведения антенн, вращением плоскости поляризации и другими факторами.

Пропускная способность канала будет снижаться пропорционально информационной нагрузке абонентов. При этом возможны два варианта временного разделения абонентов:

– фиксированное разделение пропорционально количеству абонентов q , информационной нагрузке отдельных абонентов;

– случайный множественный доступ абонентов.

При расчете пропускной способности канала связи следует учитывать невысокую мощность установленных на клиенте приемо-передающего оборудования, обусловленную ограничениями, вносимыми системой электропитания и массогабаритными характеристиками клиентов.

При фиксированном разделении интенсивности информационного трафика в канале на q абонентов эффективная пропускная способность канала, выделяемая каждому абоненту, составит

$$C_q = \frac{C}{q}. \quad (10)$$

Суммарный энергоресурс, затраченный на выполнение ПО Ψ , с учетом энергии $E_{\text{пер}}(v_{k+1})$, затраченной на передачу выходных данных v_{k+1} k -й подзадачи по каналу связи, определяется следующим выражением:

$$E = \sum_{i=1}^k E_p(\psi_i) + E_{\text{пер}}(v_{k+1}) + \sum_{i=k+1}^n E_p(\psi_i). \quad (11)$$

Время выполнения одной i -й подзадачи на одноядерном процессоре ($z = 1$) в МВС при входных данных v_i и условии $\langle v_{i,j}, \psi_i \rangle$ составит

$$t_i(\psi_i) = \frac{S_i + P_i}{L}, \quad (12)$$

где $i \in \{1, \dots, n\}$, S_i и P_i – количество инструкций программного кода, составляющих подзадачу ψ_i , L – количество инструкций, выполняемых процессором в единицу времени.

Количество инструкций, выполняемых за такт, зависит от микроархитектуры процессора, напряжения питания, а также от технологии производства, определяющей минимальные размеры используемых транзисторов, их быстродействие и время задержки передачи сигнала в межуровневых соединениях.

В случае использования многоядерного процессора, число ядер которого равно z , время выполнения одной i -й подзадачи окажется меньше за счет параллельного выполнения ψ_i команд на разных ядрах процессора и составит

$$t_i(\psi_i) = \frac{S_i}{L} + \frac{P_i}{zL}, \quad (13)$$

где z – количество ядер в многоядерном процессоре БВС, S_i – число инструкций этого кода, выполняющихся только последовательно друг за другом, P_i – количество инструкций, являющихся программно независимыми друг от друга и способных выполняться одновременно на всех ядрах процессора.

Зависимость потребляемой процессором мощности от его тактовой частоты можно представить следующей формулой [5]:

$$P_{pow} = CU^2 f, \quad (14)$$

где f – тактовая частота процессора;

U – напряжение питания процессора;

C – динамическая емкость, определяемая микроархитектурой процессора, количеством транзисторов в микросхеме и их активностью переключения.

Учитывая, что тактовая частота обусловлена напряжением питания процессора, потребляемая мощность нелинейным образом зависит от частоты процессора. Соответственно получаем нелинейную связь между производительностью процессора и потребляемой им мощностью.

Следовательно, количество энергии, затраченной на выполнение набора инструкций ψ_i в i -й подзадаче, определяется по формуле (15):

$$E_p(\psi_i) = P_{pow} t_i(\psi_i), \quad (15)$$

где P_{pow} – потребляемая процессором мощность МВС.

Время передачи данных по каналу связи между α и β определяется по формуле:

$$T_{пер} = \frac{V_{\alpha\beta}}{C} + \frac{d}{c}, \quad (16)$$

где $V_{\alpha\beta}$ – объем трафика;

C – пропускная способность канала связи между α и β ;

d – расстояние между α и β ;

c – константа, $c = 299\,792\,458$ м/с.

Энергия, затраченная для передачи данных по каналу связи между α и β , определяется по формуле:

$$E_{пер} = T_{пер} P_{ппа}, \quad (17)$$

где $P_{ппа}$ – потребляемая мощность приемо-передающей аппаратуры.

Таким образом, с учетом выражений (6), (11), время распределенной обработки целевой задачи в мобильной неоднородной вычислительной сети может быть представлено в виде:

$$T = \frac{1}{L} \left(\sum_{i=1}^k \left(S_i + \frac{P_i}{z} \right) + \sum_{i=k+1}^n \left(S_i + \frac{P_i}{z} \right) \right) + \frac{V_{\alpha\beta}}{C} + \frac{d}{c}, \quad (18)$$

а затраты энергоресурса:

$$E = \frac{P_{pow}}{L} \left(\sum_{i=1}^k \left(S_i + \frac{P_i}{z} \right) + \sum_{i=k+1}^n \left(S_i + \frac{P_i}{z} \right) \right) + P_{пер} \left(\frac{V_{\alpha\beta}}{C} + \frac{d}{c} \right). \quad (19)$$

В предложенном подходе распределенной обработки информации в мобильной неоднородной вычислительной сети учитывается время выполнения набора инструкций, затраты энергоресурса источников питания компьютера-клиента и сервера, требуемый объем памяти МВС с учетом пропускной способности канала связи и расстояния между клиентом и сервером.

Современные сетевые технологии способны обеспечить высокоскоростной обмен информацией по каналам связи в мобильной вычислительной сети, а технологии программирования и построения распределенных вычислительных сервисов позволяют организовать управляемую, масштабируемую высокопроизводительную систему.

Применение предложенного подхода обеспечивает организацию вычислительного процесса с предоставлением вычислительных ресурсов сервера «по запросу» клиента.

Литература

1. Гладышев А.И. Вопросы создания единого информационного пространства в космотехносфере // Вестник Российского нового университета. – 2014. – Вып. 4. – С. 137–140.
2. Фатеев В.Ф. Инфраструктура малых космических аппаратов. – М.: Радиотехника, 2011. – 432 с.
3. Невзоров Ю.В., Козак О.И., Васильев О.В. Факторы, влияющие на скорость передачи информации по космической радиолинии: развитие систем СВЧ радиосвязи // Электросвязь: научно-технический журнал по проводной и радиосвязи, телевидению, радиовещанию. – 2012. – № 8. – С. 29–31. – ISSN 0013-5771.
4. Пахомов С. Эра многоядерных энергоэффективных процессоров. – URL: <http://compres.ru/Article.aspx?id=16962> (дата обращения 01.05.2015).