

*А.С. Хританков*

Московский физико-технический институт (государственный университет)

## Математическая модель характеристик производительности распределённых вычислительных систем

В данной работе предлагается модель производительности распределенной системы, в рамках которой определяются такие понятия, как эталонная производительность, эффективность и ускорение. Показано, что для параллельной системы, которую можно рассматривать как частный случай распределенной системы постоянного состава и содержащей только однородные компоненты, предложенная модель совпадает с традиционной моделью производительности параллельной системы. Также был обобщен закон Амдала на случай неоднородных систем.

**Ключевые слова:** эталонная производительность, параллельная система, распределенная система, вычислительная система, закон Амдала.

### I. Введение

Для решения многих задач физики, вычислительной химии и биологии требуются значительные вычислительные ресурсы, превышающие возможности обычных однопроцессорных рабочих станций. По этой причине широкое распространение получили различные варианты многопроцессорных архитектур [1]. Такие системы также принято называть параллельными.

В последнее время также активно развиваются технологии распределённых вычислений [2], в частности, — Грид-технологии [3]. Распределенная система представляет собой инфраструктуру, объединяющую совокупность географически распределённых неоднородных вычислительных ресурсов. Ресурсы, входящие в распределенную систему, будем называть узлами этой системы. Узлы могут быть как обычными рабочими станциями с одним процессором, так и многопроцессорными вычислительными комплексами (МВК).

Параллельные системы применяются достаточно давно, поэтому для них разработано большое число различных методов и моделей, и, в

частности, моделей производительности [4]. Такие понятия, как ускорение, эффективность, загрузка, пиковая производительность, широко используются при оценке качества параллельных систем и приложений. Распределённые системы стали применяться в вычислительных целях относительно недавно. Видимо, по этой причине аналогичных параметров для распределённых систем к настоящему моменту не разработано.

Заметим, что распределенные системы и параллельные системы имеют ряд общих черт, но при этом отличаются характеристиками, перечисленными в табл. 1. Даже такое поверхностное сравнение даёт основание говорить о том, что прямое перенесение характеристик производительности параллельных систем на случай распределённых в общем случае необоснованно. Например, понятие ускорения, естественным образом определяемое в случае однородной многопроцессорной системы как отношение времени решения задачи на одном процессоре ко времени решения задачи на всей системе, не может быть непосредственно перенесено на распределенную систему, так как непонятно, какой из узлов брать в качестве эталона для сравнения.

Т а б л и ц а 1

Сравнение параллельных и распределённых систем

Характеристика	Параллельная система	Распределенная система
Степень связности	Высокая степень связности: высокопроизводительная сеть (интерконнект) либо общая память.	Слабая связность: каналы с относительно низкими показателями пропускной способности и задержки.
Узлы	Как правило, однородные по производительности и архитектуре процессорные устройства.	Неоднородные по производительности и архитектуре вычислительные машины.
Готовность выполнять вычисления	Выделенные процессоры доступны в течение всего времени вычислений.	Узлы могут выходить из состава распределенной системы или, наоборот, добавляться к ней в процессе вычислений.

## II. Модель параллельной системы

Рассмотрим традиционную модель производительности параллельной системы, предложенную в [4]. Под параллельной системой будем понимать вычислительную систему, состоящую из  $n$  одинаковых функциональных устройств, объединённых вычислительной сетью [5].

Рассмотрим решение одной вычислительной задачи  $A$  на параллельной системе. Пусть время решения задачи на этой системе составляет  $T$ . Зафиксируем некоторый эталонный последовательный алгоритм решения задачи  $A$ . Пусть время решения задачи  $A$  с помощью этого алгоритма одним устройством составляет  $T_0$ . Ускорение  $S$  определяется как отношение  $S = T_0/T$ . Таким образом, ускорение показывает, во сколько раз можно уменьшить время решения задачи с помощью применения параллельной системы. Заметим, что в качестве эталонного алгоритма обычно выбирают наиболее быстрый известный последовательный алгоритм.

Принято определять эффективность  $E$  параллельной системы как  $E = S/n$  (см. также [4]). Эффективность можно переписать как

$$E = \frac{S}{n} = \frac{T_0}{nT}. \quad (1)$$

Представленная модель описывает характеристики системы, отражающие решение задачи в целом, и может быть применена после того, как известно время решения  $T$ . Кроме того, модель требует знания времени решения  $T_0$  задачи  $A$  с помощью эталонного алгоритма на одном устройстве системы.

## III. Модель системы с расписанием

В данном разделе мы обобщим модель параллельной системы на вычислительные системы с расписанием. Распределённой вычислительной системой будем называть вычислительную систему, состоящую из совокупности функциональных устройств, работающих во времени. При этом устройства могут быть различными по своим характеристикам, а также могут быть доступны только часть времени. Вычислительные системы с расписанием отличаются от распределённых систем тем, что для каждого устройства мы полагаем известным, когда устройство выделено для решения задачи. В данной работе расписание считается известным заранее.

Рассмотрим вычислительную систему с расписанием, состоящую из  $n$  функциональных устройств. Пусть дана задача  $A$  из некоторого класса задач  $\mathbf{A}$ . Предположим, что функциональные устройства рассматриваемой системы могут решать задачи этого класса, при этом время решения задачи  $A_i$ -м устройством составляет

$T_i(A) > 0$ . Также для каждого устройства задано расписание, которое мы опишем функцией времени  $h_i(t) : R^+ \rightarrow \{0,1\}$ . При этом  $h_i(t) = 1$ , если устройство в момент времени  $t$  выделено для решения, и  $h_i(t) = 0$  в противном случае. Рассматривать будем только расписания, задаваемые функцией  $\bar{h}(t) = (h_i(t))$ , имеющей не более конечного числа точек разрыва на любом ограниченном интервале времени.

**Определение.** Под доступностью  $\rho_i(t)$  устройства будем понимать долю временного интервала  $[0,t]$ , в течение которого устройство было выделено для решения задачи  $A$ . При этом значение доступности  $i$ -го устройства определяется формулой

$$\rho_i(T) = \frac{1}{T} \int_0^T h_i(t) dt.$$

Построенная модель позволяет описывать вычислительные системы с расписанием. Обозначим вычислительную систему с расписанием через  $\mathbf{R}$ . Формальное определение дадим позже.

Для параллельной системы, согласно выражению (1), эффективность определяется как  $E = T_0/(nT)$ . Данное выражение можно переписать в виде

$$E = \frac{T_0/n}{T},$$

где выражение  $T_0/n$  соответствует некоторому 'эталонному' времени решения задачи  $A$  системой из  $n$  устройств. Для системы с расписанием определим эффективность аналогичным образом.

**Определение.** Эффективностью  $E$  системы с расписанием будем называть величину

$$E = \frac{\bar{T}(A)}{T(A)},$$

где  $\bar{T}(A)$  — эталонное время решения задачи  $A$ , а  $T(A)$  — реальное время решения задачи  $A$ .

Для определения эталонного времени решения задачи нам понадобится величина, называемая эталонной производительностью  $\pi(A,t)$  системы  $\mathbf{R}$ . Зафиксируем некоторый эталонный последовательный алгоритм решения задачи  $A$ . Обозначим через  $\bar{T}_i(A) > 0$  время решения задачи  $A_i$ -м устройством с помощью данного алгоритма. Заметим, что в качестве эталонного алгоритма обычно выбирается наиболее быстрый известный последовательный алгоритм. Будем называть эталонной производительностью  $\pi_i(A)$  устройства  $i$  при решении задачи  $A$  величину

$$\pi_i(A) = \frac{L(A)}{\bar{T}_i(A)},$$

здесь  $L(A)$  — функция трудоёмкости задачи (см. также [4]). Функция трудоёмкости  $L(A) : \mathbf{A} \rightarrow R^+$  определена на некотором множестве задач  $\mathbf{A}$  и выражает наше априорное знание о сложности решения задачи — вычислительных ресурсах, которые

необходимо затратить на её решение. Примером такой функции может служить оценка числа элементарных операций, выполняемых алгоритмом. Заметим, что функцию  $L(A)$  можно в некоторых случаях принять тождественно равной единице  $L(A) \equiv 1$ .

Эффективность параллельной системы можно представить как

$$E = \frac{1/T}{n \cdot 1/T_0} = \frac{L(A)/T}{n \cdot \pi_0(A)},$$

где  $\pi_0(A)$  — эталонная производительность одного устройства.  $L(A)/T$  можно рассматривать как реальную производительность параллельной системы,  $n \cdot \pi_0(A)$  — как эталонную.

Введём эталонную производительность  $\pi(A, t)$  системы с расписанием  $\mathbf{R}$  как сумму эталонных производительностей устройств с учётом расписания. Эталонная производительность  $\pi(A, t)$  запишется в виде

$$\pi(A, t) = \sum_{i=1}^n \pi_i(A) h_i(t).$$

При полной доступности устройств  $h_i(t) \equiv 1$  ( $i = 1 \dots n$ ) в процессе решения задачи и при одинаковой эталонной производительности  $\pi_i(A) = \pi_0$  эталонная производительность системы будет совпадать с  $n \cdot \pi_0$ , то есть с эталонной производительностью параллельной системы.

Дадим формальное определение вычислительной системы с расписанием.

**Определение.** Вычислительной системой с расписанием  $\mathbf{R}$  для решения задач из множества  $\mathbf{A}$  будем называть совокупность

$$\mathbf{R} = \langle \vec{\pi}(A), \vec{h}(t) \rangle,$$

где  $\vec{\pi}(A) : \mathbf{A} \rightarrow \mathbf{R}_+^n$  — вектор эталонных производительностей устройств,  $\vec{h}(t) : \mathbf{R}^+ \rightarrow \{0, 1\}^n$  — функция расписания, при этом  $n$  соответствует числу устройств в системе.

**Определение.** Эталонным временем решения  $\bar{T}(A)$  задачи  $A$  системой с расписанием  $\mathbf{R}$  будем называть решение задачи оптимизации

$$\min_t t, \quad \int_{\tau=0}^t \pi(A, \tau) d\tau = L(A), \quad 0 \leq t < \infty. \quad (2)$$

Итак, эффективность вычислительной системы с расписанием определяется как

$$E = \frac{\bar{T}(A)}{T(A)},$$

где  $T(A)$  — время решения, определяемое экспериментально, а эталонное время  $\bar{T}(A)$  рассчитывается по формуле (2). Доступность  $i$ -го устройства при эталонном решении задачи будем обозначать через  $\bar{\rho}_i = \rho_i(\bar{T})$ .

Ускорение для параллельной системы определяется как отношение времени решения задачи на одном устройстве ко времени решения задачи на всей системе. В распределенной системе с расписанием устройства могут быть различными, поэтому вводить понятие ускорения таким же образом, как и для однородных параллельных систем, некорректно, так как непонятно, по отношению к какому устройству считать ускорение. Поэтому предпочтительнее ввести более общее понятие относительного ускорения.

**Определение.** Ускорением  $S$  системы с расписанием  $\mathbf{R}_1$  относительно системы  $\mathbf{R}_2$  будем называть отношение времён решения задачи  $A$  этими системами:

$$S(\mathbf{R}_1, \mathbf{R}_2) = \frac{T_1}{T_2}.$$

Определим коэффициент ускорения  $S_i$  как отношение эталонного времени решения задачи на устройстве  $i$  ко времени решения задачи на всей системе:

$$S_i = \frac{\bar{T}_i}{T}.$$

Под относительным ускорением системы с расписанием  $\mathbf{R}$  будем понимать вектор

$$\vec{S} = (S_1, \dots, S_n).$$

**Утверждение 1.** Для распределенной системы с расписанием  $\mathbf{R}$  справедливо следующее соотношение:

$$E = \left( \sum_{i=1}^n \frac{\bar{\rho}_i}{S_i} \right)^{-1}.$$

**Доказательство.** Действительно, по определению эффективность системы с расписанием составляет  $E = \bar{T}/T$ , из определения эталонного времени решения (2) следует:

$$\int_0^{\bar{T}} \sum_{i=1}^n \pi_i h_i(t) d\tau = L(A).$$

Преобразуем данное выражение. Заметим, что если подставить выражение для доступности, то получим

$$\int_0^{\bar{T}} \sum_{i=1}^n \pi_i h_i(t) d\tau \equiv \sum_{i=1}^n \pi_i \bar{T} \frac{1}{\bar{T}} \int_0^{\bar{T}} h_i(t) d\tau \equiv \sum_{i=1}^n \pi_i \bar{T} \bar{\rho}_i,$$

значит, эталонное время может быть вычислено согласно формуле

$$\bar{T} = \frac{L(A)}{\sum_{i=1}^n \pi_i \bar{\rho}_i}.$$

Отсюда следует, что, так как  $\pi_i = L(A)/\bar{T}_i$ , эффективность можно представить в виде

$$E = \frac{L(A)}{T \sum_{i=1}^n \bar{\rho}_i \pi_i} = \frac{L(A)}{T \cdot L(A) \sum_{i=1}^n \frac{\bar{\rho}_i}{\bar{T}_i}} = \frac{1}{\sum_{i=1}^n \frac{\bar{\rho}_i}{S_i}}.$$

Тем самым утверждение доказано.

Интересно заметить, что если в распределенной системе с расписанием положить, что все устройства доступны в течение всего времени расчёта, то выражение для эффективности приобретает вид

$$E_{hetero} = \left( \sum_{i=1}^n \frac{1}{S_i} \right)^{-1}. \quad (3)$$

Если же дополнительно положить, что все устройства решают задачу  $A$  за одинаковое время, то распределенная система будет эквивалентна однородной параллельной системе, и выражение для эффективности для неё будет иметь вид

$$E_{parallel} = \left( \sum_{i=1}^n \frac{1}{S_i} \right)^{-1} = \frac{S}{n},$$

что совпадает с выражением для эффективности для параллельных систем. Заметим, что аналогичный результат можно получить сразу из определения эффективности для распределенной системы, если предположить, что все устройства одинаковы и доступны на протяжении всего времени решения задачи  $A$ :

$$h_i(t) \equiv 1,$$

$$\pi_i(A) = \pi_0.$$

С учётом указанных соотношений эффективность системы с расписанием можно представить как

$$E_{parallel} = \frac{\bar{T}(A)}{T(A)} = \frac{L(A)/(n \cdot \pi_0)}{T(A)} = \frac{T_0}{nT}, \quad (4)$$

что также совпадает с определением эффективности для параллельной системы

#### IV. Закон Амдала для неоднородных систем

Напомним, что закон Амдала устанавливает максимально возможное ускорение при выполнении параллельного алгоритма с последовательной частью на однородной параллельной системе. В данном разделе предлагается обобщение закона Амдала для однородных параллельных систем на случай неоднородных параллельных систем. Также показано, что закон Амдала для неоднородных систем не может быть применен к вычислительным системам с расписанием напрямую.

Под неоднородной вычислительной системой будем понимать такую систему с расписанием, для которой каждое устройство доступно в течение всего времени расчёта, то есть для любого  $i$  расписание  $h_i(t) \equiv 1$ . В дальнейшем под последовательной частью алгоритма будем понимать тот его участок, который может быть выполнен только на одном устройстве, при этом никакие другие

устройства не могут выполнять другие части алгоритма в то же самое время.

Рассмотрим выражение для закона Амдала для параллельных систем [4]. Пусть  $\beta$  выражает долю последовательных вычислений. Обычно, закон Амдала записывают для ускорения параллельной системы в виде

$$S \leq \frac{1}{\beta + (1 - \beta)/n}. \quad (5)$$

Обобщим последнее неравенство на случай неоднородных систем. Для неоднородных вычислительных систем относительное ускорение есть вектор из  $S_i$ . Доля времени, затраченная системой на выполнение последовательной части алгоритма на  $j$ -м устройстве, составляет  $\beta \bar{T}_j$ . Параллельная часть алгоритма выполняется всеми устройствами системы за время  $(1 - \beta) \bar{T}$ . Поэтому, согласно определению относительного ускорения, максимальное относительное ускорение  $S_{ij}^*$  для  $i$ -го устройства при выполнении последовательной части алгоритма полностью на  $j$ -м функциональном устройстве:

$$S_{ij}^* = \frac{\bar{T}_i}{\beta \bar{T}_j + (1 - \beta) \bar{T}}. \quad (6)$$

При этом для однородной параллельной системы выражение для максимального ускорения будет совпадать с правой частью неравенства в законе Амдала (5):

$$S^* = \frac{1}{\beta + (1 - \beta)/n}.$$

Для однородной параллельной системы максимальная эффективность, согласно определению, составляет

$$E_p^* = \frac{S^*}{n} = \frac{1}{\beta n + (1 - \beta)}.$$

Для неоднородной системы из соотношений (3) и (6) можно получить выражение для максимальной эффективности при выполнении последовательной части алгоритма только на  $j$ -м устройстве:

$$E_j^* = \left( \sum_i 1/S_{ij}^* \right)^{-1},$$

$$E_j^* = \left( \beta \sum_i \frac{\bar{T}_j}{\bar{T}_i} + (1 - \beta) \right)^{-1}.$$

Учитывая, что  $\sum_i 1/\bar{T}_i = 1/\bar{T}$ , получим

$$E_j^* = \frac{\bar{T}}{\beta \bar{T}_j + (1 - \beta) \bar{T}}.$$

Подставив выражения для эталонной производительности устройства  $j$  и системы в целом, получим

$$E_j^* = \frac{1}{\beta \frac{\pi}{\pi_j} + 1 - \beta}, \quad (7)$$

здесь  $\pi = \sum_i \pi_i$  — производительность системы.

Поскольку различные значения  $j$  описывают различные варианты распределения нагрузки, то очевидно, что

$$E \leq \max_j E_j^*.$$

Покажем, что в неоднородной системе с помощью выбора устройства для выполнения последовательной части можно добиться большей эффективности, чем в параллельных системах, состоящих из такого же числа устройств. Примем трудоёмкость задачи за единицу:  $L(A) = 1$ . Перепишем выражение для максимальной эффективности  $E_j^*$ , учитывая, что  $\pi_i = 1/\bar{T}_i$ :

$$\begin{aligned} \frac{1}{E_j^*} &= \frac{\beta}{\pi_j} \sum_i \pi_i + (1 - \beta), \\ \sum_i \frac{\pi_j}{E_j^*} &= n\beta \sum_i \pi_i + (1 - \beta) \sum_j \pi_j, \\ \sum_j \pi_j \left( \frac{1}{E_j^*} - \frac{1}{E_p^*} \right) &= 0. \end{aligned} \quad (8)$$

Полученные соотношения можно интерпретировать следующим образом. Без ограничения общности можно считать, что  $\pi_1 \geq \pi_2 \geq \dots \geq \pi_n$ . Тогда значения максимальных эффективностей  $E_j^*$  будут удовлетворять неравенствам

$$E_1^* \geq E_2^* \geq \dots \geq E_n^*.$$

Кроме того, из равенства (8) следует соотношение, связывающее оценки эффективности в неоднородной и однородной системах:

$$E_1^* \geq E_p^* \geq E_n^*.$$

Заметим, что максимальная эффективность достигается в случае, когда последовательная часть алгоритма выполняется на вычислительном узле с наибольшей производительностью, что согласуется с интуитивным представлением об оптимальном размещении последовательной части. Поскольку  $\pi = \sum_i \pi_i \geq \pi_1$ , то из выражения для максимальной эффективности (7) следует, что  $E_1^* \leq 1$ , при этом  $\lim_{\pi_1/\pi \rightarrow 1} E_1^* = 1$ . Содержательно полученные соотношения означают, что эффективность неоднородной системы при решении задач с последовательной частью не превосходит единицы и будет близка к ней, если наиболее быстрое устройство системы сильно превосходит остальные устройства по производительности.

Приведённые рассуждения позволяют сформулировать закон Амдала для неоднородной системы в виде

$$E \leq \max_j E_j^* = \frac{\pi_1}{\beta\pi + (1 - \beta)\pi_1}.$$

Из соотношения (8) также следует, что обратная величина эффективности однородной системы равна обратному значению эффективности

неоднородной системы, усреднённой по  $j$  с весами, равными относительным эталонным производительностям.

$$\begin{aligned} \sum_j \frac{\pi_j}{\pi} \frac{1}{E_j^*} &= \frac{1}{E_p^*}, \\ E_p^* &= \left( \sum_{j=1}^n \frac{\pi_j}{\pi} (E_j^*)^{-1} \right)^{-1}. \end{aligned}$$

Приведём также выражение для относительного ускорения, подставив выражения для эталонных производительностей устройств:

$$S_{ij}^* = \frac{\pi_j \pi}{\beta\pi_i \pi + (1 - \beta)\pi_j \pi_i}.$$

Итак, справедливо следующее утверждение, выражающее закон Амдала для неоднородных вычислительных систем.

**Утверждение 2.** Пусть устройства неоднородной вычислительной системы  $\mathbf{R}$  упорядочены по убыванию эталонной производительности  $\pi_1 \geq \pi_2 \geq \dots \geq \pi_n$ , тогда справедливы соотношения:

$$\begin{aligned} E &\leq \frac{\pi_1}{\beta\pi + (1 - \beta)\pi_1}, \\ S_i &\leq \frac{\pi_1 \pi}{\beta\pi_i \pi + (1 - \beta)\pi_1 \pi_i}, \end{aligned}$$

где  $\beta$  — доля последовательных вычислений.

Покажем, что для систем с расписанием закон Амдала в виде утверждения 2 неприменим. Рассмотрим частный случай однородной параллельной системы с расписанием. Пусть параметр  $\beta = 1/2$ , система состоит из двух устройств одинаковой производительности  $\pi_i = \pi_0$  с расписанием:

$$\begin{aligned} h_1(t) &= \begin{cases} 0, t \in [0, 1/2], \\ 1, t \in [1/2, 1], \end{cases} \\ h_2(t) &\equiv 1. \end{aligned}$$

Тогда по формулам для максимального ускорения и эффективности для параллельных систем:

$$\begin{aligned} S^* &= \frac{1}{1/2 + 1/4} = \frac{4}{3}, \\ E^* &= \frac{S^*}{2} = \frac{2}{3}. \end{aligned}$$

Согласно определению эффективности и относительного ускорения для систем с расписанием, получим:

$$\begin{aligned} \bar{T} &= 1/2 + 1/4 = \frac{3}{4}, \\ E &= \frac{\bar{T}}{T} = 1, \\ S_1 &= \frac{\bar{T}_1}{T} = \frac{4}{3}, \\ S_2 &= \frac{\bar{T}_2}{T} = \frac{4}{3}. \end{aligned}$$

Видно, что закон Амдала выполняется для ускорения, но не для эффективности. Для неоднородных систем максимальное относительное ускорение, определяемое формулой

$$S_i^* = \frac{\pi_1 \pi}{\beta \pi_i \pi + (1 - \beta) \pi_1 \pi_i},$$

превосходит максимальное значение ускорения, даваемое законом Амдала для параллельных систем

$$S^* = \frac{\pi}{\beta \pi + (1 - \beta)},$$

так как  $\pi_i \leq \pi_1 (i = 1 \dots n)$ . Рассмотрим теперь другой пример.

Рассмотрим распределенную систему с расписанием, состоящую из трёх устройств. Пусть эталонные производительности устройств равны  $\pi_1 = 2$ ,  $\pi_2 = 3/2$ ,  $\pi_3 = 3/2$  и расписание задается соотношениями

$$h_1(t) = \begin{cases} 0, & t \in [0, 1/2), \\ 1, & t \in [1/2, 1], \end{cases} \quad h_2(t) \equiv 1, h_3(t) \equiv 1.$$

Пусть  $\beta = 1/2$ . Если принять трудоёмкость задачи за единицу  $L(A) = 1$ , то эталонное время решения задачи  $A$  данной системой составляет  $\bar{T} = 1/3$ . Относительные ускорения устройств:

$$S_1 = \frac{1/2}{1/3} = \frac{3}{2}, \quad S_2 = S_3 = \frac{2/3}{1/3} = 2.$$

Соответствующие максимальные ускорения, вычисленные при помощи соотношений для неоднородной системы:

$$S_1^* = \frac{2 \cdot 5}{1/2 \cdot 2 \cdot 5 + 1/2 \cdot 2 \cdot 2} = \frac{10}{7} < S_1,$$

$$S_2^* = \frac{10}{1/2 \cdot 3/2 \cdot 5 + 1/2 \cdot 2 \cdot 3/2} = \frac{40}{21} < S_2,$$

$$S_3^* = \frac{40}{21} < S_3.$$

Таким образом, соотношение  $S_i \leq S_i^*$ , выражающее закон Амдала для неоднородных систем, неприменимо к вычислительным системам с расписанием.

Итак, были представлены примеры вычислительных систем, которые показывают, что доступность устройств только часть времени в процессе решения задачи не позволяет использовать для оценки ускорения и эффективности систем с расписанием как закон Амдала в общепринятой форме (5), так и выражение для неоднородных систем, представленное в утверждении 2.

## V. Заключение

В работе была предложена модель производительности распределенной системы. В модели были введены основные характеристики производительности такие, как эталонная производительность, эффективность и ускорение. Различия между параллельной и распределенной системами учитываются в модели при помощи расписания устройств и их неоднородности. В работе также показано, что для параллельной системы, которую можно рассматривать как частный случай распределенной системы постоянного состава и содержащей только однородные устройства, предложенная модель совпадает с традиционной моделью производительности. Кроме того, получено соотношение, связывающее эффективность и ускорение для распределенной системы с расписанием, аналогичное соотношению для параллельных систем. Также был обобщен закон Амдала на случай неоднородных систем.

В работе не рассматривается вопрос измерения и оценки необходимых параметров, считается, что характеристики производительности рассчитываются после того, как система завершила решение задачи. Дальнейшие исследования будут посвящены вопросу о промежуточном состоянии системы, что позволит использовать предложенную модель также для анализа алгоритмов распределения нагрузки.

## Литература

1. *Хорошевский В.Г.* Архитектура вычислительных систем: учеб. пособие для вузов. — М.: Изд-во МГТУ им. Н.Э. Баумана, 2005.
2. Проблемы вычислений в распределенной среде: организация вычислений в глобальных сетях / под. ред. С.В. Емельянова, А.П. Афанасьева. — М.: РОХОС, 2004.
3. Интернет-портал по Грид-технологиям. — URL: [www.gridclub.ru](http://www.gridclub.ru).
4. *Grama A., Gupta A., Karypis G., Kumar V.* Introduction to Parallel Computing, Second Edition. — Addison Wesley, 2003.
5. *Воеводич В.В., Воеводич Вл. В.* Параллельные вычисления. — СПб.: БХВ-Петербург, 2002.

Поступила в редакцию 10.01.2008.