КОМПЬЮТЕРНЫЕ ИССЛЕДОВАНИЯ И МОДЕЛИРОВАНИЕ 2025 Т. 17 № 3 С. 423–436
DOI: 10.20537/2076-7633-2025-17-3-423-436

Ки&М

МОДЕЛИ В ФИЗИКЕ И ТЕХНОЛОГИИ

УДК: 004.75

Математические модели и методы организации вычислений в мультипроцессорных системах

П. А. Павлов

Полесский государственный университет,
Республика Беларусь, 225710, г. Пинск, ул. Днепровской флотилии, д. 23

E-mail: pavlov.p@polessu.by

Получено 23.12.2024, после доработки — 02.04.2025.
Принято к публикации 28.05.2025.

В работе предложена и исследована математическая модель распределенной вычислительной системы параллельных взаимодействующих процессов, конкурирующих за использование ограниченного числа копий структурированного программного ресурса. В случаях неограниченного и ограниченного параллелизма по числу процессоров мультипроцессорной системы решены задачи определения оперативных и точных значений времени выполнения неоднородных и одинаково распределенных конкурирующих процессов в синхронном режиме, при котором обеспечивается линейный порядок выполнения блоков структурированного программного ресурса внутри каждого из процессов без задержек. Полученные результаты можно использовать при сравнительном анализе математических соотношений для вычисления времени реализации множества параллельных распределенных взаимодействующих конкурирующих процессов, математическом исследовании эффективности и оптимальности организации распределенных вычислений, решении задач построения оптимальной компоновки блоков одинаково распределенной системы, нахождения оптимального числа процессоров, обеспечивающих директивное время выполнения заданных объемов вычислений. Предложенные модели и методы открывают новые перспективы при решении проблем оптимального распределения ограниченных вычислительных ресурсов, синхронизации множества взаимодействующих конкурирующих процессов, минимизации системных затрат при выполнении параллельных распределенных процессов.

Ключевые слова: распределенная вычислительная система, процесс, программный ресурс, структурирование, конвейеризация, неоднородная система, одинаково распределенная система, неограниченный параллелизм, ограниченный параллелизм

Работа выполнена по заданию «Развитие методов исследования операций, теории графов, комбинаторики и теории алгоритмов» Государственной программы научных исследований Республики Беларусь «Конвергенция-2025» (2021–2025 гг.) № 20212140.

COMPUTER RESEARCH AND MODELING
2025 VOL. 17 NO. 3 P. 423–436
DOI: 10.20537/2076-7633-2025-17-3-423-436

Ки&М

MODELS IN PHYSICS AND TECHNOLOGY

UDC: 004.75

Mathematical models and methods for organizing calculations in SMP systems

P. A. Pavlov

Polessky state university,
23 Dneprovskoy flotilii st., Pinsk, 225710, Belarus

E-mail: pavlov.p@polessu.by

Received 23.12.2024, after completion — 02.04.2025.
Accepted for publication 28.05.2025.

The paper proposes and investigates a mathematical model of a distributed computing system of parallel interacting processes competing for the use of a limited number of copies of a structured software resource. In cases of unlimited and limited parallelism by the number of processors of a multiprocessor system, the problems of determining operational and exact values of the execution time of heterogeneous and identically distributed competing processes in a synchronous mode are solved, which ensures a linear order of execution of blocks of a structured software resource within each of the processes without delays. The obtained results can be used in a comparative analysis of mathematical relationships for calculating the implementation time of a set of parallel distributed interacting competing processes, a mathematical study of the efficiency and optimality of the organization of distributed computing, solving problems of constructing an optimal layout of blocks of an identically distributed system, finding the optimal number of processors that provide the directive execution time of given volumes of computations. The proposed models and methods open up new prospects for solving problems of optimal distribution of limited computing resources, synchronization of a set of interacting competing processes, minimization of system costs when executing parallel distributed processes.

Keywords: distributed computing system, process, software resource, structuring, pipelining, heterogeneous system, identically distributed system, unlimited parallelism, limited parallelism

Citation: *Computer Research and Modeling*, 2025, vol. 17, no. 3, pp. 423–436 (Russian).

The work was completed according to the assignment “Development of methods of operations research, graph theory, combinatorics and theory of algorithms” of the State Scientific Research Program of the Republic of Belarus “Convergence-2025” (2021–2025) No. 20212140.

Введение

Быстрое развитие информационно-коммуникационных и сетевых технологий привело к интенсивному использованию географически распределенных вычислительных ресурсов и созданию на их основе динамически масштабируемых высокопроизводительных распределенных вычислительных систем (РВС). В литературе отсутствует каноническое определение того, что такое распределенная вычислительная система. Например, профессор Э. Таненбаум определяет распределенную систему как «набор соединенных каналами связи независимых компьютеров, которые, с точки зрения пользователя некоторого программного обеспечения, выглядят единым целым» [Tanenbaum, Steen, 2023]. В книге Ж. Теля «Введение в распределенные алгоритмы» сказано, что «под распределенной системой мы понимаем всякую вычислительную систему, в которой несколько компьютеров или процессоров так или иначе вступают во взаимодействие». Моделью распределенной системы может также быть набор программных средств, представляющий собой совокупность взаимосвязанных процессов, выполняемых на одном и том же вычислительном устройстве [Бабичев, Коньков, 2019]. На сегодняшний день существуют различные типы распределенных вычислительных систем: это вычислительные кластеры, симметричные мультипроцессоры (SMP), системы с распределенной разделяемой памятью (DSM), массово-параллельные системы (MPP) и мультикомпьютеры. Изучая распределенные вычисления нельзя не упомянуть облачные, параллельные и грид-вычисления [Антонов и др., 2021; Топорков, Емельянов, 2018]. Из приведенных примеров следует, что РВС — это сложные сетевые системы (системы систем, SoS), предназначенные для обработки больших объемов данных, одним из основных преимуществ которых является возможность параллельного выполнения процессов.

Проектирование распределенных вычислительных систем — это сложный процесс, требующий глубокого понимания проблем проектирования и нуждающийся в хорошо продуманном теоретическом обосновании, в том числе и в разработке математических моделей и методов, которые адекватно учитывали бы физические особенности функционирования РВС и в то же время позволяют получать количественные и качественные результаты их функционирования [Kshemkalyani, Singhal, 2008].

В данном исследовании *РВС рассматривается как некоторый агрегированный процесс, получаемый путем объединения множества параллельных распределенных взаимодействующих процессов, конкурирующих за использование структурированного программного ресурса* [Kovalenko et al., 2012; Павлов, Коваленко, 2024; Павлов, 2024]. *Программным ресурсом* (ПР) будем считать многократно выполняемую в мультипроцессорной системе (МС) программу или ее часть (блок). *Процесс* будем рассматривать как выполнение в вычислительной среде линейной последовательности блоков структурированного ПР. При этом процесс будем считать *распределенным*, если все блоки или их часть выполняются на разных процессорах. *Взаимодействующими* будем называть процессы, которые влияют на поведение друг друга путем обмена информацией. Множество процессов, одновременно использующих ПР, будем называть *конкурирующими*. Из вышесказанного следует, что при моделировании и создании РВС особую значимость приобретают задачи построения и исследования математических моделей организации выполнения множества параллельных распределенных взаимодействующих конкурирующих процессов.

Задачи математического моделирования функционирования систем взаимодействующих распределенных процессов, когда в общей памяти РВС находится одна копия ПР, с различных точек зрения были решены в работах [Kovalenko, Pavlov, 2012; Pavlov, 2012]. Несмотря на имеющиеся в этом направлении результаты, нерешенными остаются задачи оптимальной организации распределенных вычислений в различных режимах асинхронного и синхронного взаимодействия процессов, конкурирующих за использование ограниченного числа копий структурированного программного ресурса в условиях неограниченного и ограниченного параллелизма по числу процессоров мультипроцессорной системы. Ввиду дискретного и комбинаторного характера математических задач распределенного программирования определенный прогресс на пути их решения был достигнут за счет применения математического аппарата и методов дискретных систем и дискретной оптимизации, теории расписаний и сетевых графов, теории алгоритмов и множеств, алгебры матриц и др. Исследования в указанном направлении характерны для Института системного программирования Российской академии наук, Московского государственного университета имени М. В. Ломоносова, Объединенного института проблем информатики Национальной академии наук Беларуси, Национального исследовательского университета «Московский энергетический институт».

Хотелось бы также отметить, что принципы распределенной организации выполнения процессов являются не только универсальным способом достижения высокой производительности и надежности МС и комплексов, но носят и достаточно общий характер и присущи процессам различной природы, прежде всего они свойственны банковским системам, энергетическим системам, логистическим системам и др. [Zaiets, Shtepa, 2019].

1. Математическая модель распределенных вычислений при ограниченном числе копий программного ресурса

Математическая модель системы распределенной обработки взаимодействующих процессов, конкурирующих за использование ограниченного числа копий структурированного программного ресурса, включает в себя [Павлов, 2024]: $p \geqslant 2$ процессоров распределенной вычислительной системы, которые имеют доступ к общей разделяемой памяти; $n \geqslant 2$ распределенных взаимодействующих конкурирующих процессов; $s \geqslant 2$ блоков структурированного на блоки программного ресурса; матрицу $T = [t_{ij}]$, $i = \overline{1, n}$, $j = \overline{1, s}$, времен выполнения блоков программного ресурса распределенными взаимодействующими конкурирующими процессами; $2 \leqslant c \leqslant p$ копий структурированного на блоки программного ресурса, которые могут одновременно находиться в оперативной памяти, доступной для всех p процессоров; параметр $\theta > 0$, характеризующий время дополнительных системных расходов, связанных с параллельной обработкой блоков структурированного программного ресурса множеством взаимодействующих конкурирующих процессов при распределенной обработке. Будем также считать, что число процессов n кратно числу копий структурированного программного ресурса c, т. е. $n = mc$, где $m = \frac{n}{c} \geqslant 2$, и что взаимодействие процессов, процессоров и блоков программного ресурса подчинено следующим условиям:

1) ни один из процессоров не может обрабатывать одновременно более одного блока;

2) процессы выполняются в параллельно-конвейерном режиме группами, т. е. осуществляется одновременное выполнение c копий каждого блока в сочетании с конвейеризацией группы из c копий Q_j-го блока, $j = \overline{1, s}$, по процессам и процессорам;

3) обработка каждого блока программного ресурса осуществляется без прерываний;

4) в случае *неограниченного* параллелизма по числу процессоров многопроцессорной системы, т. е. когда число блоков программного ресурса $s \leqslant \left[\frac{p}{c}\right]$, где $[x]$ — целая часть числа, для каждого i-го процесса, где $i = c(l-1) + q$, $l = \overline{1, m}$, $q = \overline{1, c}$, распределение блоков Q_j, $j = \overline{1, s}$, программного ресурса по процессорам осуществляется по правилу: блок с номером j распределяется на процессор с номером $c(j-1) + q$.

2. Время реализации неоднородных распределенных систем конкурирующих процессов

Определение 1. Систему n распределенных конкурирующих процессов будем называть *неоднородной* (*гетерогенной*), если времена выполнения блоков программного ресурса $Q_1, Q_2, \ldots, Q_s$ зависят от объемов обрабатываемых данных и/или их структуры, т. е. разные для разных процессов.

В [Павлов, Коваленко, 2024] введен *синхронный* режим, обеспечивающий линейный порядок выполнения блоков программного ресурса внутри каждого из процессов без задержек, т. е. в случае, когда $2 \leqslant s \leqslant \left[\frac{p}{c}\right]$, момент завершения выполнения Q_j-го блока, $j = \overline{1, s-1}$, процессом с номером $i = (l-1)c + q$, $l = \overline{1, m}$, $q = \overline{1, c}$, на $((j-1)c + q)$-м процессоре совпадает с моментом начала выполнения следующего Q_{j+1}-го блока на процессоре с номером $(jA + q)$. На рис. 1 изображена диаграмма Ганта, иллюстрирующая функционирование распределенной системы, где $T^{\text{н}}_{1c}(p, n, s, c, \theta)$ — время выполнения $n = 6$ *неоднородных* конкурирующих процессов, использующих $c = 2$ копии программного ресурса, структурированного на $s = 3$ линейно-упорядоченных блоков Q_1, Q_2, Q_3 с матрицей времен выполнения с учетом дополнительных системных расходов $T^{\theta} = \left[t^{\theta}_{ij}\right]_{6\times 3} = [t_{ij} + \theta]_{6\times 3}$. Предполагается, что выполнение процессов осуществляется в вычислительной системе с $p = 7$ процессорами.

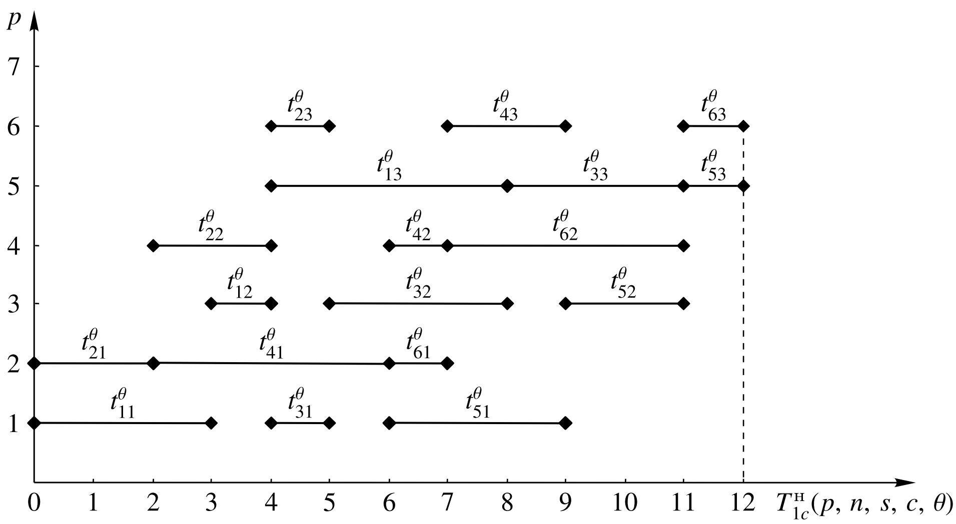

Рис. 1. Синхронный режим при неограниченном параллелизме

В случае *неограниченного* параллелизма все множество процессов разобьем на c подмножеств по m процессов в каждом, причем в каждое q-е подмножество, $q = \overline{1, c}$, будут включены процессы с номерами $i = c(l-1) + q$, $l = \overline{1, m}$, блоки которых будут выполняться на $(c(j-1) + q)$ процессорах, $j = \overline{1, s}$. Тогда для вычисления времени выполнения $T^{\text{н}}_{1c}(p, n, s, c, \theta)$ имеет место формула

$$T^{\text{н}}_{1c}(p, n, s, c, \theta) = \max_{1 \leqslant q \leqslant c}\left(\sum_{i=1}^{m-1} \max_{1 \leqslant u \leqslant s}\left[\sum_{j=1}^{u} t^{\theta}_{c(i-1)+q,j} - \sum_{j=1}^{u-1} t^{\theta}_{ci+q,j}\right] + \sum_{j=1}^{s} t^{\theta}_{c(m-1)+q,j}\right). \tag{1}$$

Здесь

$$\max_{1 \leqslant u \leqslant s}\left[\sum_{j=1}^{u} t^{\theta}_{c(i-1)+q,j} - \sum_{j=1}^{u-1} t^{\theta}_{ci+q,j}\right], \quad i = \overline{1, m-1},$$

— начало выполнения первого блока для каждого из последующих процессов, начиная с $(ci+q)$-го в q-м подмножестве процессов, а $\sum\limits_{j=1}^{s} t^{\theta}_{c(m-1)+q,j}$ — длительность выполнения последнего процесса q-го подмножества, где $q=\overline{1,\,c}$.

В случае *ограниченного* параллелизма, т. е. когда $s>\left[\frac{p}{c}\right]$ и $s=k\left[\frac{p}{c}\right]$, где $k>1$, выполним разбиение всех s блоков на k групп по $\left[\frac{p}{c}\right]$ блоков в каждой. Тогда матрица времен выполнения блоков структурированного ПР $T^{\theta}=\left[t^{\theta}_{ij}\right]$, $i=\overline{1,\,n}$, $j=\overline{1,\,s}$, разобьется на k подматриц T^{θ}_{φ} размерностью $n\times\left[\frac{p}{c}\right]$ каждая, где $\varphi=\overline{1,\,k}$ (рис. 2).

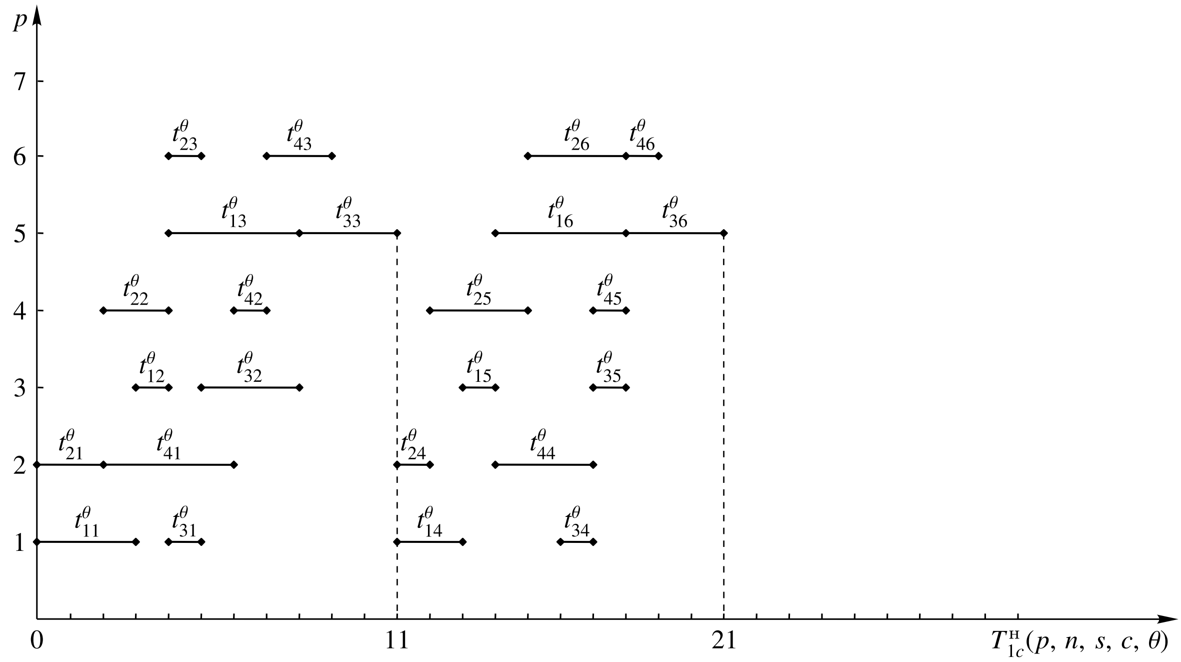

Рис. 2. Несовмещенные диаграммы Ганта при $s=k\left[\frac{p}{c}\right]$, $k>1$

Из анализа диаграмм Ганта видно, что каждая из диаграмм отображает во времени выполнение очередных $\left[\frac{p}{c}\right]$ блоков программного ресурса на $c\left[\frac{p}{c}\right]$ процессорах всеми n процессами. Введем следующие обозначения:

- $t^{\varphi,q}_{ij}$ — время выполнения в φ-й группе блоков i-м процессом из q-го подмножества процессов j-го блока с учетом параметра θ:

$$t^{\varphi,q}_{ij}=t^{\theta}_{c(i-1)+q,(\varphi-1)\left[\frac{p}{c}\right]+j},\quad \varphi=\overline{1,\,k},\quad q=\overline{1,\,c},\quad i=\overline{1,\,m},\quad j=\overline{1,\,\left[\frac{p}{c}\right]};$$

- T^{φ}_{q} — время выполнения в φ-й группе блоков q-го подмножества процессов:

$$T^{\varphi}_{q}=\sum_{i=1}^{m-1}\max_{1\leqslant u\leqslant\left[\frac{p}{c}\right]}\left[\sum_{j=1}^{u} t^{\theta}_{c(i-1)+q,(\varphi-1)\left[\frac{p}{c}\right]+j}-\sum_{j=1}^{u-1} t^{\theta}_{ci+q,(\varphi-1)\left[\frac{p}{c}\right]+j}\right]+ \\ +\sum_{j=1}^{[p/c]} t^{\theta}_{c(m-1)+q,(\varphi-1)\left[\frac{p}{c}\right]+j},\quad \varphi=\overline{1,\,k},\quad q=\overline{1,\,c};$$

- T_φ^θ — общее время выполнения φ-й группы блоков всеми n процессами на $c\left[\frac{p}{c}\right]$ процессорах с учетом параметра θ:

$$T_\varphi^\theta = \max_{1\leqslant q\leqslant c}\left(\sum_{i=1}^{m-1}\max_{1\leqslant u\leqslant\left[\frac{p}{c}\right]}\left[\sum_{j=1}^{u} t^\theta_{c(i-1)+q,(\varphi-1)\left[\frac{p}{c}\right]+j} - \sum_{j=1}^{u-1} t^\theta_{ci+q,(\varphi-1)\left[\frac{p}{c}\right]+j}\right] + \sum_{j=1}^{[p/c]} t^\theta_{c(m-1)+q,(\varphi-1)\left[\frac{p}{c}\right]+j}\right), \quad \varphi=\overline{1,\,k};$$

- $E_{ij}^{\varphi,q}$ — время завершения обработки в φ-й группе блоков i-м процессом из q-го подмножества j-го блока:

$$E_{ij}^{\varphi,q} = E^\varphi_{c(i-1)+q,j} = \sum_{\mu=1}^{i-1}\max_{1\leqslant u\leqslant\left[\frac{p}{c}\right]}\left[\sum_{w=1}^{u} t^\theta_{c(\mu-1)+q,(\varphi-1)\left[\frac{p}{c}\right]+w} - \sum_{w=1}^{u-1} t^\theta_{c\mu+q,(\varphi-1)\left[\frac{p}{c}\right]+w}\right] + \sum_{w=1}^{j} t^\theta_{c(i-1)+q,(\varphi-1)\left[\frac{p}{c}\right]+w}, \quad \varphi=\overline{1,\,k}, \quad q=\overline{1,\,c}, \quad i=\overline{1,\,m}, \quad j=\overline{1,\,\left[\frac{p}{c}\right]}.$$

Из анализа диаграмм на рис. 2 видно, что минимальное общее время выполнения *неоднородных* распределенных процессов, конкурирующих за использование c копий структурированного программного ресурса в случае $s = k\left[\frac{p}{c}\right]$, $k>1$, определяется как сумма длин составляющих диаграмм, т. е.

$$T^{\text{н}}_{1c}\left(p,\,n,\,k\left[\frac{p}{c}\right],\,c,\,\theta\right) = \sum_{\varphi=1}^{k} T_\varphi^\theta = \sum_{\varphi=1}^{k}\max_{1\leqslant q\leqslant c} T_q^\varphi.$$

Если воспользоваться совмещением последовательных диаграмм Ганта по оси времени справа налево, то время $T^{\text{н}}_{1c}\left(p,\,n,\,k\left[\frac{p}{c}\right],\,c,\,\theta\right)$ можно существенно сократить (рис. 3):

$$T^{\text{н}}_{1c}\left(p,\,n,\,k\left[\frac{p}{c}\right],\,c,\,\theta\right) = \sum_{\varphi=1}^{k} T_\varphi^\theta - \sum_{\varphi=1}^{k-1}\delta_\varphi,$$

где $\delta_\varphi = \min\left(\delta'_\varphi,\,\delta''_\varphi\right)$, $\varphi=\overline{1,\,k-1}$, — длина отрезка максимально возможного совмещения двух последовательных диаграмм Ганта по оси времени.

Здесь δ'_φ — отрезок возможного совмещения по оси времени, который представляет собой разность между моментом начала выполнения j-го блока первым процессом q-го подмножества процессов в $(\varphi+1)$-й группе блоков и моментом завершения выполнения j-го блока последним процессом q-го подмножества процессов в φ-й группе блоков, т. е.

$$\delta'_\varphi = \min_{1\leqslant q\leqslant c}\left(T_\varphi^\theta - T_q^\varphi + \min_{1\leqslant j\leqslant\left[\frac{p}{c}\right]}\left[\sum_{w=j+1}^{[p/c]} t^\theta_{c(m-1)+q,(\varphi-1)\left[\frac{p}{c}\right]+w} + \sum_{w=1}^{j-1} t^\theta_{q,\varphi\left[\frac{p}{c}\right]+w}\right]\right) = \min_{1\leqslant q\leqslant c}\left(\min_{1\leqslant j\leqslant\left[\frac{p}{c}\right]}\left[T_\varphi^\theta - E^\varphi_{c(m-1)+q,j} + \sum_{w=1}^{j-1} t^\theta_{q,\varphi\left[\frac{p}{c}\right]+w}\right]\right), \quad \varphi=\overline{1,\,k-1}.$$

Значение δ''_φ представляет собой разность между началом выполнения первого блока i-м процессом в $(\varphi+1)$-й группе блоков и моментом завершения выполнения последнего блока i-м

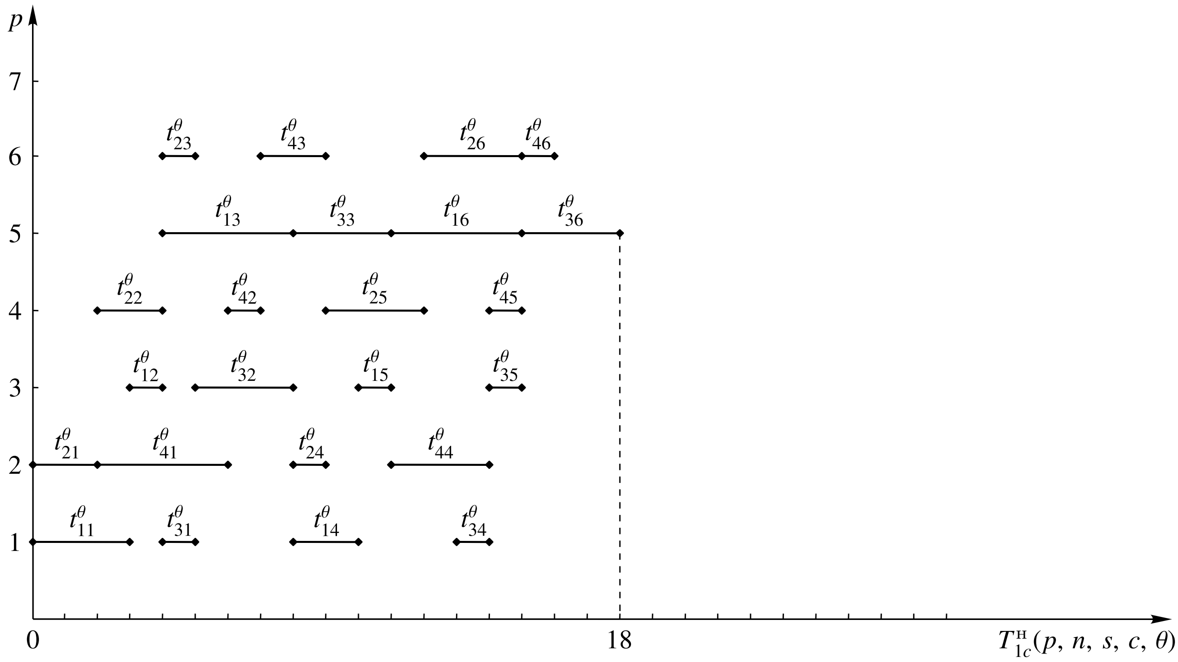

Рис. 3. Совмещенные диаграммы Ганта при $s = k\left[\frac{p}{c}\right]$, $k > 1$

процессом в φ-й группе блоков, т. е.

$$\delta''_{\varphi} = \min_{1\leqslant q\leqslant c}\left(\min_{1\leqslant i\leqslant m}\left[T^{\theta}_{\varphi} - E^{q,\varphi}_{i,\left[\frac{p}{c}\right]} + E^{q,\varphi+1}_{i1} - t^{q,\varphi+1}_{i1}\right]\right) =$$

$$= \min_{1\leqslant q\leqslant c}\left(\min_{1\leqslant i\leqslant m}\left[T^{\theta}_{\varphi} - E^{\varphi}_{c(i-1)+q,\left[\frac{p}{c}\right]} + E^{\varphi+1}_{c(i-1)+q,1} - t^{\theta}_{c(i-1)+q,\varphi\left[\frac{p}{c}\right]+1}\right]\right), \quad \varphi = \overline{1,\, k-1}.$$

В случае когда $s = k\left[\frac{p}{c}\right] + r$, $k \geqslant 1$, $1 \leqslant r < \left[\frac{p}{c}\right]$, все множество из s блоков разбивается на $k+1$ группу по $\left[\frac{p}{c}\right]$ блоков в каждой, за исключением последней, в которой будет только r блоков ПР. Тогда

$$T^{\text{н}}_{1c}\left(p,\, n,\, k\left[\frac{p}{c}\right] + r,\, c,\, \theta\right) = \sum_{\varphi=1}^{k} T^{\theta}_{\varphi} + T^{\theta}_{k+1} - \sum_{\varphi=1}^{k-1}\delta_{\varphi} - \delta_k = \sum_{\varphi=1}^{k}\max_{1\leqslant q\leqslant c} T^{\varphi}_{q} + \max_{1\leqslant q\leqslant c} T^{k+1}_{q} - \sum_{\varphi=1}^{k-1}\delta_{\varphi} - \delta_k.$$

Здесь

- T^{θ}_{k+1} — время выполнения $(k+1)$-й группы из r блоков всеми n процессами:

$$T^{\theta}_{k+1} = \max_{1\leqslant q\leqslant c}\left(\sum_{i=1}^{m-1}\max_{1\leqslant u\leqslant r}\left[\sum_{j=1}^{u} t^{\theta}_{c(i-1)+q,k\left[\frac{p}{c}\right]+j} - \sum_{j=1}^{u-1} t^{\theta}_{ci+q,k\left[\frac{p}{c}\right]+j}\right] + \sum_{j=1}^{r} t^{\theta}_{c(m-1)+q,k\left[\frac{p}{c}\right]+j}\right) = \max_{1\leqslant q\leqslant c} T^{k+1}_{q};$$

- T^{k+1}_{q} — время выполнения q-го подмножества из m процессов в $(k+1)$-й группе блоков:

$$T^{k+1}_{q} = \sum_{i=1}^{m-1}\max_{1\leqslant u\leqslant r}\left[\sum_{j=1}^{u} t^{\theta}_{c(i-1)+q,k\left[\frac{p}{c}\right]+j} - \sum_{j=1}^{u-1} t^{\theta}_{ci+q,k\left[\frac{p}{c}\right]+j}\right] + \sum_{j=1}^{r} t^{\theta}_{c(m-1)+q,k\left[\frac{p}{c}\right]+j}, \quad q = \overline{1,\, c};$$

- δ_k — величина максимально допустимого совмещения по оси времени k-й и $(k+1)$-й диаграмм:

$$\delta_k = \min\left(\delta'_k,\ \delta''_k\right).$$

Отрезки возможного совмещения по оси времени δ'_k и δ''_k будут определяться по формулам

$$\delta'_k = \min_{1\leqslant q\leqslant c}\left(T^\theta_k - T^k_q + \min_{1\leqslant j\leqslant r}\left[\sum_{w=j+1}^{[p/c]} t^\theta_{c(m-1)+q,(k-1)\left[\frac{p}{c}\right]+w} + \sum_{w=1}^{j-1} t^\theta_{q,k\left[\frac{p}{c}\right]+w}\right]\right) =$$

$$= \min_{1\leqslant q\leqslant c}\left(\min_{1\leqslant j\leqslant r}\left[T^\theta_k - E^k_{c(m-1)+q,j} + \sum_{w=1}^{j-1} t^\theta_{q,k\left[\frac{p}{c}\right]+w}\right]\right);$$

$$\delta''_k = \min_{1\leqslant q\leqslant c}\left(\min_{1\leqslant i\leqslant m}\left[T^\theta_k - E^{\theta,k}_{i,\left[\frac{p}{c}\right]} + E^{\theta,k+1}_{i1} - t^{q,k+1}_{i1}\right]\right) =$$

$$= \min_{1\leqslant q\leqslant c}\left(\min_{1\leqslant i\leqslant m}\left[T^\theta_k - E^k_{c(i-1)+q,\left[\frac{p}{c}\right]} + E^{k+1}_{c(i-1)+q,1} - t^\theta_{c(i-1)+q,k\left[\frac{p}{c}\right]+1}\right]\right),$$

где

$$E^{q,k+1}_{ij} = E^{k+1}_{c(i-1)+q,j} = \sum_{\mu=1}^{i-1}\max_{1\leqslant u\leqslant r}\left[\sum_{w=1}^{u} t^\theta_{c(\mu-1)+q,(k+1)r+w} - \sum_{w=1}^{u-1} t^\theta_{c\mu+q,(k+1)r+w}\right] + \sum_{w=1}^{j} t^\theta_{c(i-1)+q,(k+1)r+w},$$

$$q = \overline{1,\ c},\quad i = \overline{1,\ m},\quad j = \overline{1,\ r}.$$

3. Время выполнения одинаково распределенных конкурирующих процессов

Определение 2. Систему распределенных конкурирующих процессов будем называть *одинаково распределенной* (*равнораспределенной*), если времена выполнения всех блоков ПР каждым из процессов совпадают и равны t^θ_i, т. е. справедлива цепочка равенств $t^\theta_{i1} = t^\theta_{i2} = \ldots = t^\theta_{is} = t^\theta_i$, для всех $i = \overline{1,\ n}$.

Пусть $t^\theta_1, t^\theta_2, \ldots, t^\theta_n$ — времена выполнения Q_j-го блока всеми процессами, $j = \overline{1,\ s}$, а $T^\theta_n = \sum\limits_{i=1}^{n} t^\theta_i$ — суммарное время выполнения каждого из блоков Q_j n процессами.

Обозначим через $T^{\text{op}}_{1c}(p, n, s, c, \theta)$ минимальное общее время выполнения на p процессорах n *одинаково распределенных* конкурирующих процессов, использующих c копий структурированного на s линейно-упорядоченных блоков $Q_1, Q_2, \ldots, Q_s$ ПР с учетом дополнительных системных расходов θ в синхронном режиме.

На рис. 4 для случая $2 \leqslant s \leqslant \left[\frac{p}{c}\right]$ изображена диаграмма Ганта выполнения одинаково распределенных конкурирующих процессов в вычислительной системе с параметрами $p = 7$, $n = 6$, $s = 3$, $c = 2$,

$$T^\theta = \begin{bmatrix} 1 & 1 & 1 \\ 3 & 3 & 3 \\ 2 & 2 & 2 \\ 4 & 4 & 4 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix}.$$

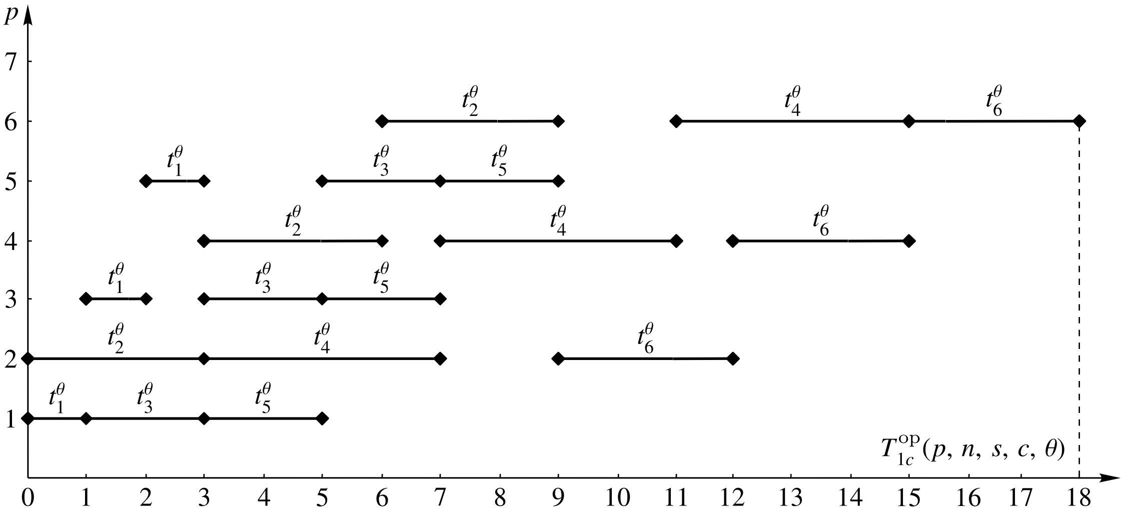

Рис. 4. Выполнение одинаково распределенных процессов при неограниченном параллелизме

Для нахождения $T_{1c}^{\mathrm{op}}(p, n, s, c, \theta)$ при $2 \leqslant s \leqslant \left[\frac{p}{c}\right]$, подставив в формулу (1) значения $t^{\theta}_{c(i-1)+q,j} = t^{\theta}_{c(i-1)+q}$, $i = \overline{1, m}$, $q = \overline{1, c}$, $j = \overline{1, s}$, и выполнив преобразования, получим

$$T_{1c}^{\mathrm{op}}(p, n, s, c, \theta) = \max_{1\leqslant q\leqslant c}\left(\sum_{i=1}^{m-1}\max_{1\leqslant u\leqslant s}\left[\sum_{j=1}^{u} t^{\theta}_{c(i-1)+q} - \sum_{j=1}^{u-1} t^{\theta}_{ci+q}\right] + \sum_{j=1}^{s} t^{\theta}_{c(m-1)+q}\right) =$$

$$= \max_{1\leqslant q\leqslant c}\left(\sum_{i=1}^{m} t^{\theta}_{c(i-1)+q} + (s-1)\left[t^{\theta}_{c(m-1)+q} + \sum_{i=2}^{m}\max\left(t^{\theta}_{c(i-2)+q} - t^{\theta}_{c(i-1)+q},\, 0\right)\right]\right). \quad (2)$$

В случае когда $s = k\left[\frac{p}{c}\right]$, где $k > 1$, времена выполнения каждых $\left[\frac{p}{c}\right]$ блоков программного ресурса всеми n процессами равны между собой (рис. 5), поэтому общее время $T_{1c}^{\mathrm{op}}\left(p, n, k\left[\frac{p}{c}\right], c, \theta\right)$ будет определяться по формуле

$$T_{1c}^{\mathrm{op}}\left(p, n, k\left[\frac{p}{c}\right], c, \theta\right) = kT_{1c}^{\mathrm{op}}\left(p, n, \left[\frac{p}{c}\right], c, \theta\right) - (k-1)\min\{\sigma_1, \sigma_2\},$$

где $T_{1c}^{\mathrm{op}}\left(p, n, \left[\frac{p}{c}\right], c, \theta\right)$ находится по формуле (2) заменой s на $\left[\frac{p}{c}\right]$:

$$T_{1c}^{\mathrm{op}}\left(p, n, \left[\frac{p}{c}\right], c, \theta\right) = \max_{1\leqslant q\leqslant c}\left(\sum_{i=1}^{m} t^{\theta}_{c(i-1)+q} + \left(\left[\frac{p}{c}\right] - 1\right)\left[t^{\theta}_{c(m-1)+q} + \sum_{i=2}^{m}\max\left(t^{\theta}_{c(i-2)+q} - t^{\theta}_{c(i-1)+q},\, 0\right)\right]\right).$$

Величины σ_1 и σ_2 представляют собой временные отрезки максимально возможного совмещения φ-й и $(\varphi+1)$-й диаграмм Ганта, $\varphi = \overline{1, k-1}$:

$$\sigma_1 = \left(\left[\frac{p}{c}\right] - 1\right)\min_{1\leqslant q\leqslant c}\left\{t^{\theta}_{q},\, t^{\theta}_{(m-1)c+q}\right\},$$

$$\sigma_2 = \min_{1\leqslant q\leqslant c}\left(\sum_{i=1}^{m} t^{\theta}_{c(i-1)+q} + \left(\left[\frac{p}{c}\right] - 1\right)\left[t^{\theta}_{c(m-1)+q} + \sum_{i=2}^{m}\max\left(t^{\theta}_{c(i-2)+q} - t^{\theta}_{c(i-1)+q},\, 0\right)\right] - \left[\frac{p}{c}\right]\max_{1\leqslant i\leqslant m} t^{\theta}_{c(i-1)+q}\right).$$

Пусть $p = 7$, $n = 4$, $s = 9$, $c = 2$,

$$T^{\theta} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\ 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

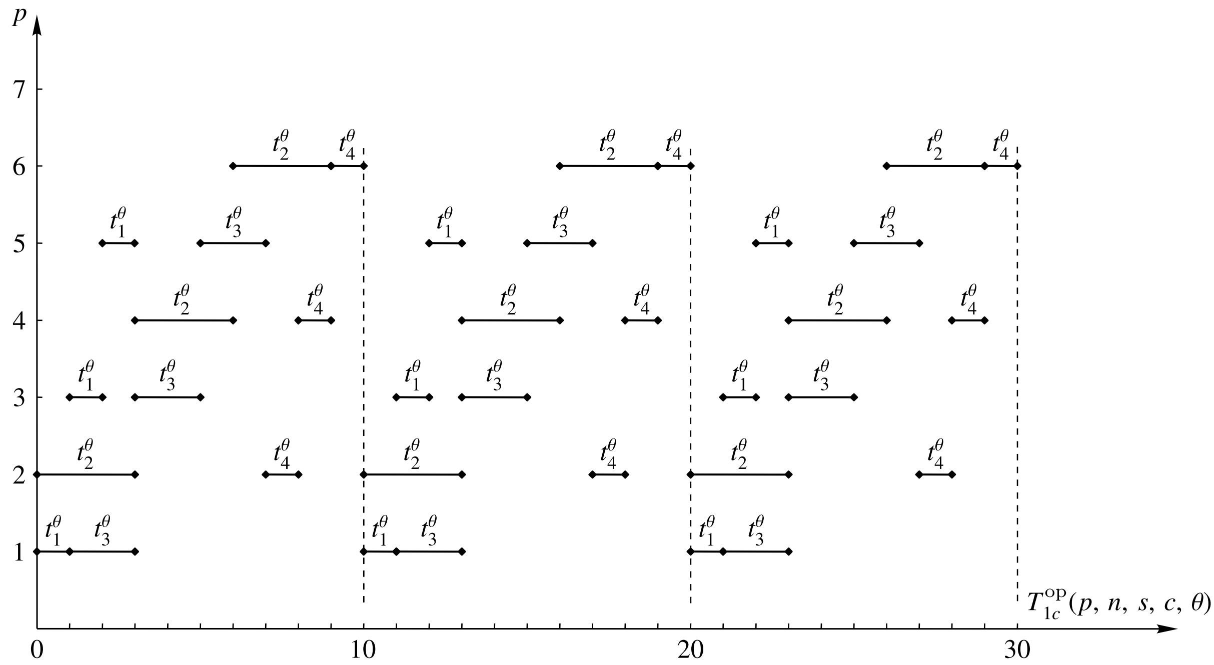

Рис. 5. Несовмещенные диаграммы Ганта при $s = k\left[\frac{p}{c}\right]$, $k > 1$

(рис. 5), тогда

$$T_{1c}^{\text{op}}(7,\ 4,\ 3,\ 2,\ \theta) = \max_{1\leqslant q\leqslant 2}\left(\sum_{i=1}^{2} t^{\theta}_{2(i-1)+q} + 2\left[t^{\theta}_{2+q} + \sum_{i=2}^{2}\max\left(t^{\theta}_{2(i-2)+q} - t^{\theta}_{2(i-1)+q},\ 0\right)\right]\right) = 10,$$

$$\sigma_1 = 2\min\{1,\ 2,\ 3,\ 1\} = 2, \quad \sigma_2 = \min\{7-6,\ 10-9\} = 1.$$

Получим, что $T_{1c}^{\text{op}}(7,\ 4,\ 9,\ 2,\ \theta) = 28$ (рис. 6).

В случае когда $s = k\left[\frac{p}{c}\right] + r$, $k \geqslant 1$, $1 \leqslant r < \left[\frac{p}{c}\right]$, минимальное общее время $T^1_{\text{op}}(p,\ n,\ kp+r,\ \varepsilon)$ определяется по формуле

$$T_{1c}^{\text{op}}\left(p,\ n,\ k\left[\frac{p}{c}\right] + r,\ c,\ \theta\right) = kT_{1c}^{\text{op}}\left(p,\ n,\ \left[\frac{p}{c}\right],\ c,\ \theta\right) - (k-1)\min\{\sigma_1,\ \sigma_2\} + T_{1c}^{\text{op}}(p,\ n,\ r,\ c,\ \theta) - \xi.$$

Здесь $T_{1c}^{\text{op}}(p,\ n,\ r,\ c,\ \theta)$ находится из (2) путем замены s на r, т. е.

$$T_{1c}^{\text{op}}(p,\ n,\ r,\ c,\ \theta) = \max_{1\leqslant q\leqslant c}\left(\sum_{i=1}^{m} t^{\theta}_{c(i-1)+q} + (r-1)\left[t^{\theta}_{c(m-1)+q} + \sum_{i=2}^{m}\max\left(t^{\theta}_{c(i-2)+q} - t^{\theta}_{c(i-1)+q},\ 0\right)\right]\right),$$

а $\xi = \min\{\xi_1,\ \xi_2\}$ — величина отрезка максимально возможного совмещения по оси времени φ-й и $(\varphi+1)$-й диаграмм, где

$$\xi_1 = \min_{1\leqslant q\leqslant c}\left((r-1)\min\left\{t^{\theta}_{q},\ t^{\theta}_{(m-1)c+q}\right\} + \left(\left[\frac{p}{c}\right] - r\right)t^{\theta}_{(m-1)c+q}\right),$$

$$\xi_2 = \min_{1\leqslant q\leqslant c}\left(T_{1c}^{\text{op}}\left(p,\ n,\ \left[\frac{p}{c}\right],\ c,\ \theta\right) - \max_{1\leqslant i\leqslant m}\left[T_{1c}^{\text{op}}\left(p,\ i,\ \left[\frac{p}{c}\right],\ c,\ \theta\right) - T_{1c}^{\text{op}}(p,\ i,\ r,\ c,\ \theta) + rt^{\theta}_{(i-1)c+q}\right]\right).$$

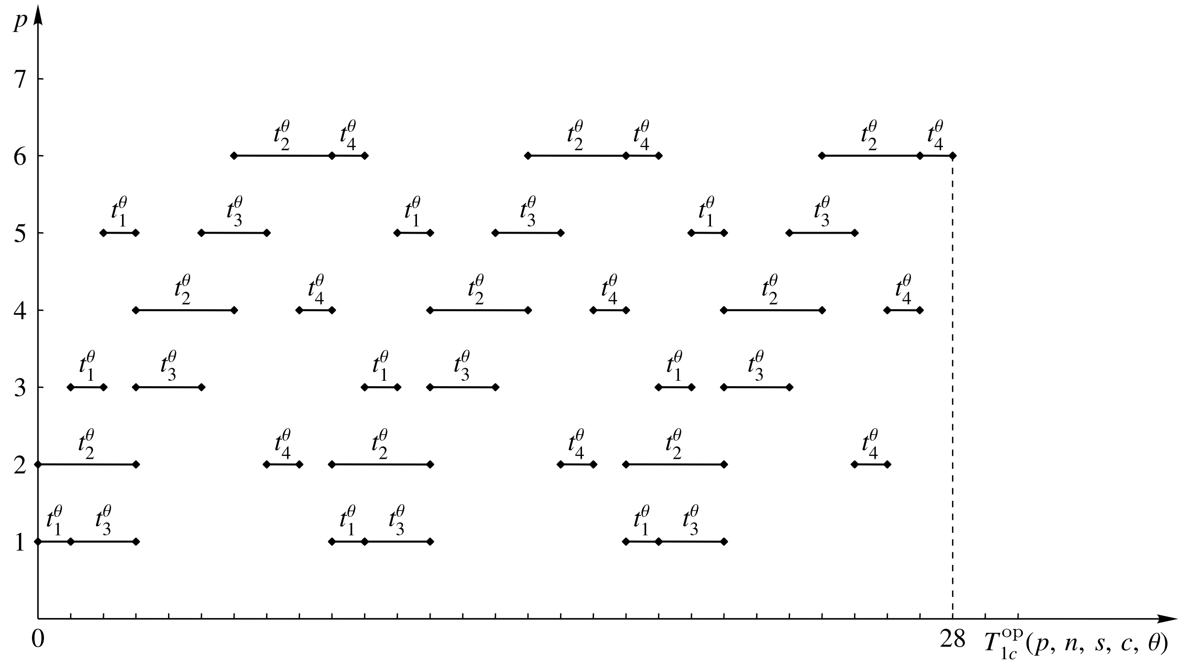

Рис. 6. Совмещенные диаграммы Ганта при $s = k\left[\frac{p}{c}\right]$, $k > 1$

В свою очередь, $T_{1c}^{\mathrm{op}}\left(p,\, i,\, \left[\frac{p}{c}\right],\, c,\, \theta\right)$ и $T_{1c}^{\mathrm{op}}(p,\, i,\, r,\, c,\, \theta)$ будут определяться по формулам

$$T_{1c}^{\mathrm{op}}\left(p,\, i,\, \left[\frac{p}{c}\right],\, c,\, \theta\right) = \sum_{\mu=1}^{i} t_{c(\mu-1)+q}^{\theta} + \left(\left[\frac{p}{c}\right] - 1\right)\left[t_{c(i-1)+q}^{\theta} + \sum_{\mu=2}^{i} \max\left(t_{c(\mu-2)+q}^{\theta} - t_{c(\mu-1)+q}^{\theta},\, 0\right)\right],$$

$$T_{1c}^{\mathrm{op}}(p,\, i,\, r,\, c,\, \theta) = \sum_{\mu=1}^{i} t_{c(\mu-1)+q}^{\theta} + (r-1)\left[t_{c(i-1)+q}^{\theta} + \sum_{\mu=2}^{i} \max\left(t_{c(\mu-2)+q}^{\theta} - t_{c(\mu-1)+q}^{\theta},\, 0\right)\right].$$

Таким образом, справедлива следующая

Теорема 1. *Если взаимодействие процессов, процессоров и блоков структурированного программного ресурса подчинено условиям синхронного режима, то для любых* $p \geqslant 2$, $n \geqslant 2$, $s \geqslant 2$, $2 \leqslant c \leqslant p$, $\theta > 0$ *минимальное общее время выполнения одинаково распределенных конкурирующих процессов определяется следующим образом*:

$$T_{1c}^{op}(p,\, n,\, s,\, c,\, \theta) = \begin{cases} \max\limits_{1\leqslant q\leqslant c}\left(\sum\limits_{i=1}^{m} t_{c(i-1)+q}^{\theta} + (s-1)\left[t_{c(m-1)+q}^{\theta} + \sum\limits_{i=2}^{m} \max\left(t_{c(i-2)+q}^{\theta} - t_{c(i-1)+q}^{\theta},\, 0\right)\right]\right) \\ \text{при } 2 \leqslant s \leqslant \left[\dfrac{p}{c}\right], \\ kT_{1c}^{op}\left(p,\, n,\, \left[\dfrac{p}{c}\right],\, c,\, \theta\right) - (k-1)\min\{\sigma_1,\, \sigma_2\} \text{ при } s = k\left[\dfrac{p}{c}\right],\ k > 1, \\ kT_{1c}^{op}\left(p,\, n,\, \left[\dfrac{p}{c}\right],\, c,\, \theta\right) - (k-1)\min\{\sigma_1,\, \sigma_2\} + T_{1c}^{op}(p,\, n,\, r,\, c,\, \theta) - \min\{\xi_1,\, \xi_2\} \\ \text{при } s = k\left[\dfrac{p}{c}\right] + r,\ k \geqslant 1,\ 1 \leqslant r < \left[\dfrac{p}{c}\right]. \end{cases}$$

Определение 3. Равнораспределенная система конкурирующих процессов называется *стационарной*, если $t_1^\theta = t_2^\theta = \ldots = t_n^\theta = t^\theta$.

Теорема 2. *Для вычисления минимального общего времени выполнения стационарных распределенных конкурирующих процессов в синхронном режиме при ограниченном числе копий программного продукта имеют место формулы*

$$T_{1c}^{cc}(p,\, n,\, s,\, c,\, \theta) = \begin{cases} (m+s-1)t^{\theta}, & \left[\dfrac{p}{c}\right] \geqslant \min\left(\dfrac{n}{c},\, s\right), \\ \left(km + \left[\dfrac{p}{c}\right] - 1\right)t^{\theta}, & \left[\dfrac{p}{c}\right] < \min\left(\dfrac{n}{c},\, s\right),\ s = k\left[\dfrac{p}{c}\right],\ k > 1, \\ ((k+1)m + r - 1)t^{\theta}, & \left[\dfrac{p}{c}\right] < \min\left(\dfrac{n}{c},\, s\right),\ s = k\left[\dfrac{p}{c}\right] + r,\ k \geqslant 1,\ 1 \leqslant r < \left[\dfrac{p}{c}\right]. \end{cases}$$

Доказательство. Для доказательства первой формулы теоремы 2 рассмотрим формулу для одинаково распределенных систем теоремы 1 в случае, когда $\left[\frac{p}{c}\right] \geqslant \min(m,\, s)$, где $m = \frac{n}{c}$. Так как $t_i^{\theta} = t^{\theta}$, $i = \overline{1,\, n}$, получим

$$T_{1c}^{cc}(p,\, n,\, s,\, c,\, \theta) = \sum_{i=1}^{m} t^{\theta} + (s-1)\left[t^{\theta} + \sum_{i=2}^{m} \max\left(t^{\theta} - t^{\theta},\, 0\right)\right] = (m+s-1)t^{\theta}.$$

Рассмотрим случай, когда $s = k\left[\frac{p}{c}\right]$, $k > 1$ и $\left[\frac{p}{c}\right] < \min(m,\, s)$. Для стационарных распределенных процессов получим

$$T_{1c}^{\text{op}}\left(p,\, n,\, \left[\frac{p}{c}\right],\, c,\, \theta\right) = \left(m + \left[\frac{p}{c}\right] - 1\right)t^{\theta}, \quad \sigma_1 = \left(\left[\frac{p}{c}\right] - 1\right)t^{\theta}, \quad \sigma_2 = (m-1)t^{\theta},$$

тогда

$$T_{1c}^{cc}\left(p,\, n,\, k\left[\frac{p}{c}\right],\, c,\, \theta\right) = k\left(m + \left[\frac{p}{c}\right] - 1\right)t^{\theta} - (k-1)\min\left\{\left(\left[\frac{p}{c}\right] - 1\right)t^{\theta},\, (m-1)t^{\theta}\right\} = \left(km + \left[\frac{p}{c}\right] - 1\right)t^{\theta}.$$

Для случая, когда $\left[\frac{p}{c}\right] < \min(m,\, s)$, $s = k\left[\frac{p}{c}\right] + r$, $k \geqslant 1$, $1 \leqslant r < \left[\frac{p}{c}\right]$, имеем

$$T_{1c}^{cc}(p,\, n,\, r,\, c,\, \theta) = (m+r-1)t^{\theta}, \quad \xi_1 = \left(\left[\frac{p}{c}\right] - 1\right)t^{\theta}, \quad \xi_2 = (m-1)t^{\theta},$$

тогда

$$T_{1c}^{cc}\left(p,\, n,\, k\left[\frac{p}{c}\right] + r,\, c,\, \theta\right) = \left(km + \left[\frac{p}{c}\right] - 1\right)t^{\theta} + (m+r-1)t^{\theta} - \left(\left[\frac{p}{c}\right] - 1\right)t^{\theta} = ((k+1)m + r - 1)t^{\theta}.$$

Теорема доказана. □

Заключение

В данной работе проведено концептуальное развитие математической модели с одним программным ресурсом на случай ограниченного числа программных ресурсов, что позволило установить взаимосвязь мультиконвейерной обработки с аналогичной обработкой при одном ПР, получить аналитические оценки общего времени выполнения неоднородных и одинаково распределенных конкурирующих процессов при неограниченном и ограниченном параллелизме по числу процессоров мультипроцессорной системы для синхронного режима. Полученные результаты можно использовать при сравнительном анализе математических соотношений для вычисления времени реализации множества параллельных распределенных взаимодействующих конкурирующих процессов, при математическом исследовании эффективности и оптимальности мультиконвейерной организации вычислений, при решении задач построения оптимальной компоновки блоков одинаково распределенной системы, нахождения оптимального числа процессоров, обеспечивающих директивное время выполнения заданных объемов вычислений и др.

Список литературы (References)

Антонов А. С., Афанасьев И. В., Воеводин В. В. Высокопроизводительные вычислительные платформы: текущий статус и тенденция развития // Вычислительные методы и программирование. — 2021. — Т. 22. — С. 135–177.

Antonov A. S., Afanasyev I. V., Voevodin V. V. Vysokoproizvoditel'nye vychislitel'nye platformy: tekushchii status i tendentsiya razvitiya [High-performance computing platforms: current status and development trends] // Computational methods and programming. — 2021. — Vol. 22. — P. 135–177 (in Russian).

Бабичев С. Л., Коньков К. А. Распределенные системы. — М.: Юрайт, 2019.

Babichev S. L., Konkov K. A. Raspredelennye sistemy [Distributed systems]. — Moscow: Yurayt, 2019 (in Russian).

Павлов П. А. О времени реализации распределенных вычислений в синхронном режиме при ограниченном числе копий программного ресурса // Труды Института системного программирования РАН. — 2024. — Т. 36, № 4. — С. 81–98.

Pavlov P. A. O vremeni realizatsii raspredelennykh vychislenii v sinkhronnom rezhime pri ogranichennom chisle kopii programmnogo resursa [On the implementation time of distributed computing in synchronous mode with a limited number of copies of a software resource] // Proceedings of the Institute for System Programming of the Russian Academy of Sciences. — 2024. — Vol. 36, No. 4. — P. 81–98 (in Russian).

Павлов П. А., Коваленко Н. С. Синхронный режим распределенных вычислений при непрерывном выполнении блоков ограниченного числа копий программного ресурса // Программные продукты и системы. — 2024. — № 1. — С. 43–53.

Pavlov P. A., Kovalenko N. S. Sinkhronnyi rezhim raspredelennykh vychislenii pri nepreryvnom vypolnenii blokov ogranichennogo chisla kopii programmnogo resursa [Synchronous mode of distributed computing with continuous execution of blocks of a limited number of copies of a software resource] // Software products and systems. — 2024. — No. 1. — P. 43–53 (in Russian).

Топорков В. В., Емельянов Д. М. Модели, методы и алгоритмы планирования в грид- и облачных вычислениях // Вестник Московского энергетического института. — 2018. — № 6. — С. 75–86.

Toporkov V. V., Emelyanov D. M. Modeli, metody i algoritmy planirovaniya v grid- i oblachnykh vychisleniyakh [Models, methods, and algorithms for planning in grid and cloud computing] // Bulletin of the Moscow Energy Institute. — 2018. — No. 6. — P. 75–86 (in Russian).

Kovalenko N. S., Pavlov P. A. Optimal grouping algorithm of identically distributed systems // Programming and Computer Software. — 2012. — No. 3. — P. 143–150.

Kovalenko N. S., Pavlov P. A., Ovseec M. I. Asynchronous distributed computations with a limited number of copies of a structured program resource // Cybernetics and systems analysis. — 2012. — No. 1. — P. 86–98.

Kshemkalyani A. D., Singhal M. Distributed computing: principles, algorithms, and systems. — Cambridge University Press, 2008.

Pavlov P. A. The optimality of software resources structuring through the pipeline distributed processing of competitive cooperative processes // International Journal of Multimedia Technology (IJMT). — 2012. — Vol. 2, No. 1. — P. 5–10.

Tanenbaum A., Steen M. Distributed systems. — Amazon Digital Services LLC, 2023.

Zaiets N., Shtepa V. (eds.) Development of a resource-process approach to increasing the efficiency of electrical equipment for food production // Eastern-European Journal of Enterprise Technologies. — 2019. — Vol. 5, No. 8. — P. 59–65.