УДК 681.3.06

П. А. Павлов*
**Павлов Павел Александрович, к.ф.-м.н., доцент*
Полесский государственный университет, г. Пинск, Республика Беларусь
pin2535@tut.by

ВРЕМЯ РЕАЛИЗАЦИИ АСИНХРОННЫХ ПРОЦЕССОВ В СИСТЕМАХ МАКРОКОНВЕЙЕРНОГО ТИПА

Ключевые слова: параллелизм, программный ресурс, многопроцессорная система, маситабируемость, макроконвейер, структурирование.

Предлагается эффективная модель организации вычислений неоднородных многостадийных процессов в многопроцессорных системах макроконвейерного типа.

Введение

Среди наиболее перспективных концепций параллельной обработки является концепция *макроконвейерной* организации вычислений над структурами данных. В настоящее время интерес к этой концепции постоянно растет в связи с развитием и широким применением локальных и глобальных сетей, созданием вычислительных многопроцессорных систем (МС) и комплексов, сетевого аппаратного и прикладного программного обеспечения. Основная идея заключается в том, что при распараллеливании и распределении вычислений между процессорами "каждому отдельному процессору на очередном шаге вычислений дается такое задание, которое позволяет ему длительное время работать автономно без взаимодействия с другими процессорами" [1].

В настоящей работе предлагается математическая модель эффективной организации вычислений неоднородными процессами в многопроцессорных системах макроконвейерного типа, а также решение задач определения временных характеристик процессов вычислений конкурирующих за использование каналов обмена.

1. Макроконвейерная обработка и метод структурирования программных ресурсов

Структурирование (декомпозиция) – это основной способ уменьшения сложности больших задач, программ, систем и т.д. Основная идея состоит в обеспечении специального способа структурирования программного ресурса (PR) на блоки $Q_1, Q_2, \ldots, Q_s$ и организации параллельного использования этих блоков множеством конкурирующих процессов [2]. Структурирование программного ресурса на блоки осуществляется, как правило, либо исходя из физического смысла задачи на этапах создания математической модели и алгоритмов её решения, либо путём анализа готовой, последовательной программы с целью её декомпозиции.

Макроконвейерная технология вычислений над структурами данных предполагает декомпозицию структуры данных на большие информационно—слабозависимые подструктуры. Причем должен соблюдаться основной принцип макроконвейерной технологии обработки данных, который предполагает работу с достаточно крупными подструктурами, способными занимать процессор длительное время. Работа процессоров при этом организуется таким образом, чтобы обмен данными между ними занимал небольшое время по сравнению с временем вычислений.

Рассмотрим основные положения макроконвейерного способа вычислений применительно к программным ресурсам, одновременно используемым множеством процессов. Пусть PR – программный ресурс, который может быть использован двумя и более конкурирующими процессами, т.е. $n \geq 2$, $p \geq 2$ – число процессоров МВС, причем память может быть как общей для всех процессоров, так и распределенной по процессорам.

Рассмотрим следующие возможные организации вычислений.

Каждому i-му процессу, $i = \overline{1,n}$, представляется отдельная копия программного ресурса PR. При такой стратегии, в случае $p \geq n$, все n процессов могут выполняться одновременно при условии, что в МС достаточно памяти для размещения n копий программного ресурса (в случае МС с общей памятью) или память каждого процессора МС вмещает отдельную копию программного ресурса (в случае МС с распределенной памятью). Если же $p < n$, то возможна организация циклического выполнения n процессов группами по p.

Программный ресурс PR может быть структурирован на блоки $Q_1, Q_2, ..., Q_s$, а вычисления в этом случае организуются в соответствии с методом структурирования.

Вторая стратегия может применяться при организации вычислений в МС всякий раз, если имеются ограничения на оперативную память, как общую, так и память каждого процессора. В качестве блоков $Q_1, Q_2, ..., Q_s$ может выступать также упорядоченный набор программ, выполнение которых составляет вычислительный процесс.

2. Математическая модель макроконвейерной организации асинхронных конкурирующих процессов

Пусть МС характеризуется следующими параметрами: p – число процессоров, каждый из которых имеет собственную локальную память, $p \geq 2$; k – число каналов, через которые каждый из процессоров имеет доступ к внешней памяти, общей для всех процессоров, $k \geq 1$.

Предполагается, что в МС выполняется n процессов, $n \geq 2$, каждый из которых состоит из s блоков обмена и s блоков счета, $s \geq 1$. Времена обмена и счета для каждого из процессов представлены в виде матриц размерности $n \times s$, в которых i-е строки соответствуют i-му процессу:

$$t = \begin{bmatrix} t_{11} & t_{12} & \dots & t_{1s} \\ t_{21} & t_{22} & \dots & t_{2s} \\ \dots & \dots & \dots & \dots \\ t_{n1} & t_{n2} & \dots & t_{ns} \end{bmatrix}, \quad T = \begin{bmatrix} T_{11} & T_{12} & \dots & T_{1s} \\ T_{21} & T_{22} & \dots & T_{2s} \\ \dots & \dots & \dots & \dots \\ T_{n1} & T_{n2} & \dots & T_{ns} \end{bmatrix}.$$

Взаимодействие процессов с каналами и процессорами характеризуется следующими условиями: 1) к выполнению одновременно готовы p процессов из n; 2) в каждый момент времени k процессов из n, одновременно протекающих в МС, выполняются синхронно, остальные в очереди ждут освобождения каналов; 3) во время обмена каждый процесс монополизирует один и тот же канал, во время счета – процессор; 4) очередной j–й блок счета на каждом процессоре выполняется только после завершения соответствующего j–го блока обмена, а каждый $(j+1)$–й блок обмена выполняется после завершения j–го блока счета; 5) процессы считаются равноприоритетными, а режим работы каналов является циклическим. Условия 1–5 определяют *асинхронный* режим взаимодействия процессов, каналов и процессоров, который допускает как простои каналов из–за занятости процессоров, так и простои процессоров из–за занятости каналов обмена.

В этих предположениях рассмотрим решение задач получения математических соотношений для вычисления минимального общего времени реализации множества конкурирующих процессов.

3. Время реализации асинхронных процессов в системах макроконвейерного типа

Обозначим через $T_n(k)$ общее время выполнения всех n процессов, которые используют k каналов. Заметим, что при $p \geq k \geq n$ в рамках принятой модели макроконвейерных вычислений $T_n(k)$ составит величину $T_n(k) = T_n(n) = \max\limits_{1 \leq i \leq n} \sum\limits_{j=1}^{s} (t_{ij} + T_{ij})$. Если окажется, что $p > k > n$, то $k - n$ каналов будут не задействованы, а $p - n$ процессоров будут простаивать.

Из физических соображений наибольший интерес в рамках концепции макроконвейерных вычислений представляет случай ограниченного числа каналов, т.е. когда $k << n$, $n = mk$, $m > 1$, что означает, что процессы конкурируют за использование каналов.

Пусть имеется только один канал, т.е. $k = 1$. Предположим, что $p = n$. На рис. 1 приведена диаграмма Ганта [3-5], отображающая взаимодействие n процессов (номер процесса изображен справа в прямоугольнике) с одним каналом и p процессорами.

Лемма 1. Времена вынужденных простоев каналов из–за ожидания ввода информации для i–го процесса на j–м шаге вычислений δ_{ij} определяются из соотношений:

$$\delta_{ij} = \max\left\{0; T_{ij} - \sum_{r=i+1}^{p} (\delta_{r,j-1} + t_{rj}) - \sum_{r=1}^{i-1} (\delta_{rj} + t_{r,j+1})\right\},$$

$$j = \overline{1, s-1},\ i = \overline{1, p},\ \delta_{i0} = 0, \qquad (1)$$

Справедливость формулы (1) проверяется методом математической индукции.

Лемма 2. Время δ от конца ввода каналом последней порции информации t_{ps} до окончания счета всеми процессорами, определяется из соотношения:

$$\delta = \max_{1 \le l \le p} \left\{ T_{ls} - \sum_{r=l+1}^{p} (\delta_{r,s-1} + t_{rs}) \right\}. \tag{2}$$

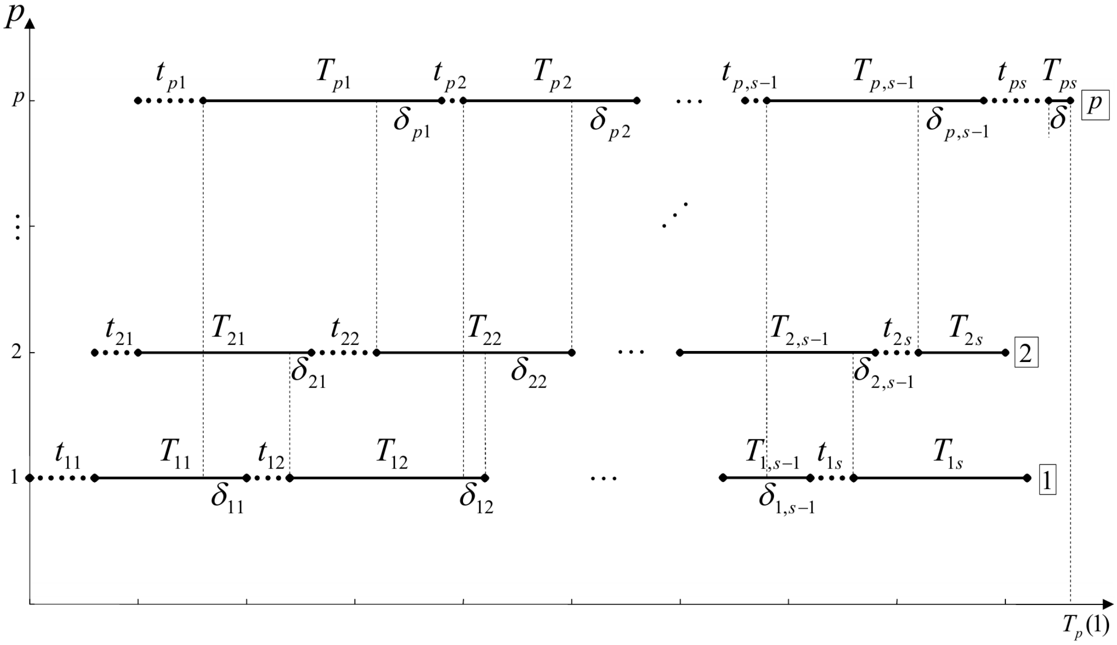

Рис. 1. Диаграмма Ганта с одним каналом обмена

Теорема 1. Общее время выполнения n $(n \ge 2)$ процессов p $(p \ge 2)$ процессорами, конкурирующими за использование одного канала, в случае $p = n$, определяется по формуле:

$$T_n(1) = T_p(1) = \sum_{i=1}^{p} \sum_{j=1}^{s} t_{ij} + \sum_{i=1}^{p} \sum_{j=1}^{s-1} \delta_{ij} + \delta . \tag{3}$$

Пример 1. Определить общее время $T_n(k)$ выполнения $n = 3$ процессов в МС с $p = 3$ процессорами и $k = 1$ каналом обмена. Каждый процесс состоит из $s = 4$ блоков обмена и $s = 4$ блоков счета, времена которых заданы следующими матрицами:

$$t_{3 \times 4} = \begin{bmatrix} 3 & 2 & 2 & 1 \\ 2 & 3 & 2 & 3 \\ 3 & 2 & 3 & 2 \end{bmatrix}, \quad T_{3 \times 4} = \begin{bmatrix} 4 & 9 & 6 & 3 \\ 6 & 9 & 9 & 5 \\ 8 & 10 & 2 & 2 \end{bmatrix}.$$

На рис. 2 приведена диаграмма Ганта, отображающая взаимодействие трех процессов с одним каналом обмена и тремя процессорами.

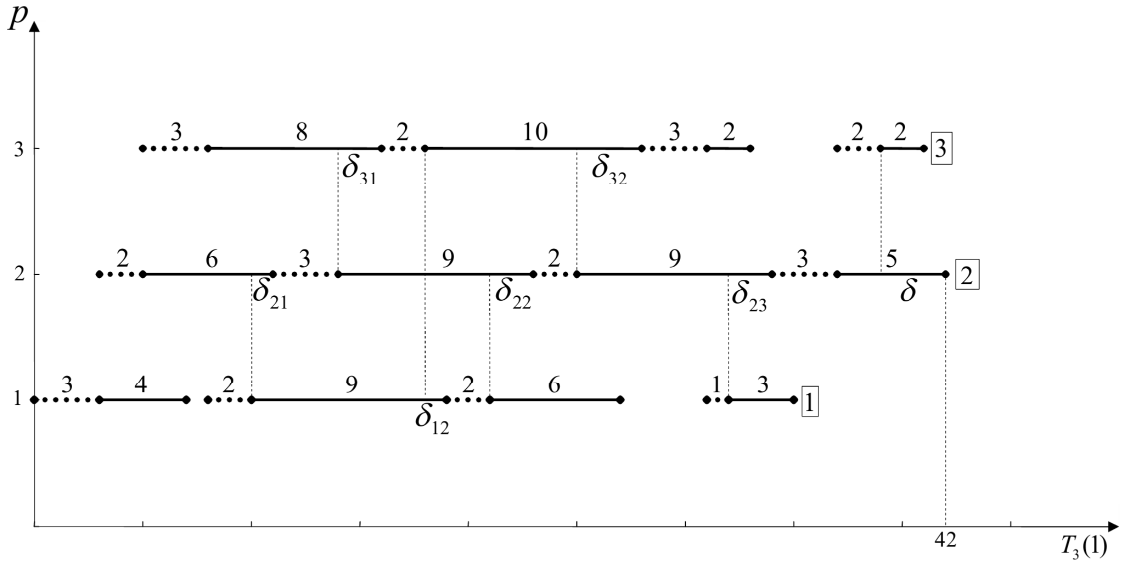

Рис. 2. Диаграмма Ганта (3 процесса, 1 канал)

Для нахождения $T_n(k)$ воспользуемся формулой (3). Предварительно найдем времена вынужденного простоя канала δ_{ij}, $j=\overline{1,s-1}$, $i=\overline{1,p}$, и время δ по формулам (1) и (2):

$\delta_{11}=0$, $\delta_{21}=1$, $\delta_{31}=2$, $\delta_{12}=1$, $\delta_{22}=2$, $\delta_{32}=3$, $\delta_{13}=0$, $\delta_{23}=2$, $\delta_{33}=0$, $\delta=3$.

Тогда $T_3(1)=\sum_{i=1}^{3}\sum_{j=1}^{4} t_{ij}+\sum_{i=1}^{3}\sum_{j=1}^{3}\delta_{ij}+\delta=28+11+3=42$. Таким образом, общее время выполнения 3 процессов на 3 процессорах, использующих 1 канал обмена, будет равно 42, что совпадает с временем выполнения на диаграмме Ганта.

Пусть далее МС характеризуется следующими параметрами: k – число каналов, $(k\geq 1)$; $n=mk$, $m>1$ – число процессов, которое кратно числу каналов $(n\geq 2)$. В этом случае возможен следующий способ взаимодействия процессов, каналов и процессоров: все множество из n процессов разбивается на m групп по k процессов в каждой, причем каждый канал будет обслуживать по одному процессу из каждой группы, т.е. g-й канал, $g=\overline{1,k}$, обслуживает группу из m процессов с номерами $(l-1)k+g$, где $l=\overline{1,m}$ (рис. 3).

Так как каждый g-й канал, $g=\overline{1,k}$, обслуживает группу из m процессов, то согласно формулам (1)-(3) время, затраченное на выполнение каждой группы из m процессов m процессорами каждым каналом составит:

$$T_m^1(1)=\sum_{i=1}^{m}\sum_{j=1}^{s} t_{[(i-1)k+1],j}+\sum_{i=1}^{m}\sum_{j=1}^{s-1}\delta_{[(i-1)k+1],j}+\delta_1,$$

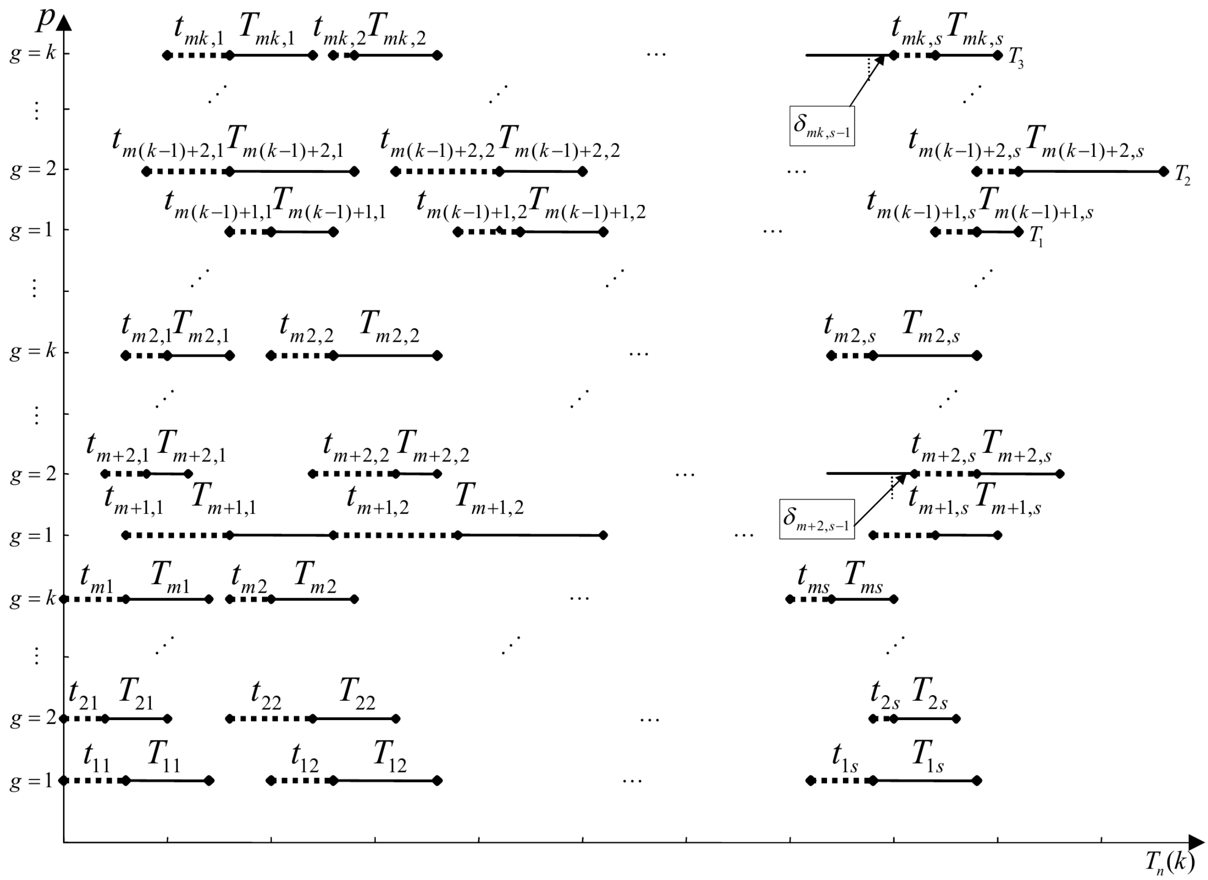

Рис. 3. Способ взаимодействия процессов, каналов и процессоров

где

$$\delta_{[(i-1)k+1],j} = \max\left\{0; T_{[(i-1)k+1],j} - \sum_{r=i+1}^{m}(\delta_{[(r-1)k+1],j-1} + t_{[(r-1)k+1],j}) - \right.$$

$$\left. - \sum_{r=1}^{i-1}(\delta_{[(r-1)k+1],j} + t_{[(r-1)k+1],j+1})\right\};$$

$$\delta_{[(i-1)k+1],0} = 0,\ i = \overline{1,m},\ j = \overline{1,s-1},$$

$$\delta_1 = \max_{1 \le l \le m}\left\{T_{[(l-1)k+1],s} - \sum_{r=l+1}^{m}(\delta_{[(r-1)k+1],s-1} + t_{[(r-1)k+1],s})\right\};$$

$$T_m^2(1) = \sum_{i=1}^{m}\sum_{j=1}^{s} t_{[(i-1)k+2],j} + \sum_{i=1}^{m}\sum_{j=1}^{s-1}\delta_{[(i-1)k+2],j} + \delta_2,$$

где

$$\delta_{[(i-1)k+2],j} = \max\left\{0; T_{[(i-1)k+2],j} - \sum_{r=i+1}^{m}(\delta_{[(r-1)k+2],j-1} + t_{[(r-1)k+2],j}) - \right.$$

$$-\sum_{r=1}^{i-1}(\delta_{[(r-1)k+2],j}+t_{[(r-1)k+2],j+1})\Bigg\},$$

$$\delta_{[(i-1)k+2],0}=0\,,\ i=\overline{1,m}\,,\ j=\overline{1,s-1}\,,$$

$$\delta_2=\max_{1\le l\le m}\left\{T_{[(l-1)k+2],s}-\sum_{r=l+1}^{m}(\delta_{[(r-1)k+2],s-1}+t_{[(r-1)k+2],s})\right\};\ldots;$$

$$T_m^k(1)=\sum_{i=1}^{m}\sum_{j=1}^{s}t_{ik,j}+\sum_{i=1}^{m}\sum_{j=1}^{s-1}\delta_{ik,j}+\delta_k\,,$$

где

$$\delta_{ik,j}=\max\left\{0;T_{ik,j}-\sum_{r=i+1}^{m}(\delta_{ik,j-1}+t_{ik,j})-\sum_{r=1}^{i-1}(\delta_{ik,j}+t_{ik,j+1})\right\},$$

$$\delta_{ik,0}=0\,,\ i=\overline{1,m}\,,\ j=\overline{1,s-1}\,,\ \delta_k=\max_{1\le l\le m}\left\{T_{lk,s}-\sum_{r=l+1}^{m}(\delta_{ik,s-1}+t_{ik,s})\right\}.$$

Теорема 2. Общее время выполнения p процессорами $(p\ge 2)$ $n=km$ $(m>1)$ процессов, которые конкурируют за использование k каналов $(k\ge 1)$, каждый из которых обслуживает m процессоров, в случае $p=n$ определяется из соотношения:

$$T_n(k)=\max_{1\le g\le k}T_m^g(1)=\max_{1\le g\le k}\left\{\sum_{i=1}^{m}\sum_{j=1}^{s}t_{[(i-1)k+g],j}+\sum_{i=1}^{m}\sum_{j=1}^{s-1}\delta_{[(i-1)k+g],j}+\delta_g\right\};\qquad(4)$$

где

$$\delta_{[(i-1)k+g],j}=\max\left\{0;T_{[(i-1)k+g],j}-\sum_{r=i+1}^{m}(\delta_{[(r-1)k+g],j-1}+t_{[(r-1)k+g],j})-\right.$$

$$\left.-\sum_{r=1}^{i-1}(\delta_{[(r-1)k+g],j}+t_{[(r-1)k+g],j+1})\right\},$$

$$\delta_{[(i-1)k+g],0}=0\,;\ j=\overline{1,s-1}\,,\ i=\overline{1,m}\,;$$

$$\delta_g=\max_{1\le l\le m}\left\{T_{[(l-1)k+g],s}-\sum_{r=l+1}^{m}(\delta_{[(r-1)k+g],s-1}+t_{[(r-1)k+g],s})\right\}.$$

Пример 2. Пусть $n=6$, $k=3$, $s=5$, а времена обмена и счета заданы матрицами:

$$t_{6x5} = \begin{bmatrix} 3 & 3 & 2 & 1 & 3 \\ 2 & 4 & 5 & 3 & 1 \\ 3 & 2 & 1 & 4 & 2 \\ 5 & 6 & 3 & 2 & 3 \\ 2 & 4 & 3 & 2 & 3 \\ 2 & 3 & 4 & 3 & 2 \end{bmatrix}, \quad T_{6x5} = \begin{bmatrix} 4 & 5 & 3 & 3 & 5 \\ 3 & 4 & 6 & 4 & 3 \\ 4 & 4 & 4 & 6 & 3 \\ 5 & 7 & 4 & 2 & 3 \\ 2 & 2 & 3 & 4 & 4 \\ 3 & 5 & 4 & 4 & 5 \end{bmatrix}.$$

Тогда каждый из трех каналов будет обслуживать группу из трех процессов. На рис.4 приведена диаграмма Ганта, отображающая взаимодействие в МС $n = 6$ процессов с $k = 3$ каналами и заданными временами выполнения блоков обмена и счета.

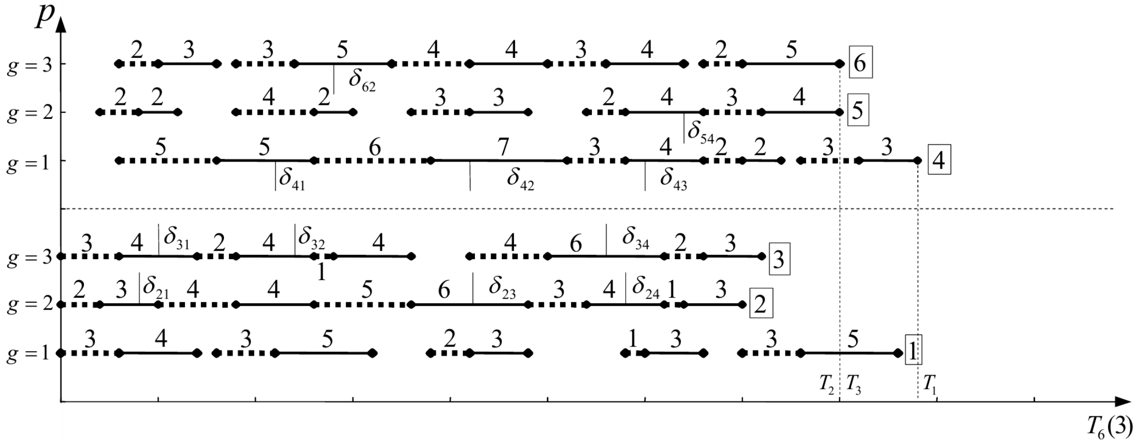

Рис. 4. Диаграмма Ганта (6 процессов, 3 канала)

Полагая в формуле (4) $k = 1,2,3$, $n = p = 6$, $s = 5$, находим $T_3^1(1) = 44$, $T_3^2(1) = 40$, $T_3^3(1) = 40$. Таким образом, общее время выполнения 6 процессов на 6 процессорах, использующих 3 канала, составит: $T_6(3) = \max(44,40,40) = 40$.

Заключение

Построенная модель организации макроконвейерных вычислений над структурами данных при ограниченном числе каналов обмена и разработанные аналитические методы расчета общего времени выполнения множества неоднородных конкурирующих процессов являются основой для постановки и решения ряда важных практических задач по расчету оптимальной балансировки числа процессоров и каналов, оптимизации числа блоков счета и обмена, минимизации общего времени выполнения процессов и др.

Библиографический список

1. Глушков В.М., Капитонова Ю.В., Летичевский А.А. Об одном подходе к реализации параллельных вычислений в многопроцессорных вычислительных

системах // В кн.: Параллельное программирование и высокопроизводительные системы. – Новосибирск, 1980. – ч.1. – С. 3–25.

2. Павлов П.А., Коваленко Н.С. Математическое моделирование параллельных процессов. – Germany: Lambert Academic Publishing, 2011. – 246 с.

3. Павлов П.А. Организация однородных конкурирующих процессов при распределенной конвейерной обработке / П.А. Павлов // Проблемы управления. – 2010. – №3. – С. 66–75.

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

5. Kovalenko N.S., Pavlov P.A., Ovseec M.I. Asynchronous distributed computations with a limited number of copies of a structured program resource / N.S. Kovalenko, P.A. Pavlov, M.I. Ovseec // Cybernetics and systems analysis. – 2012. – Vol.48, №1. – PP. 86–98.

6. Kovalenko N.S., Pavlov P.A. Optimal Grouping Algorithm of Identically Distributed Systems / N.S. Kovalenko, P.A. Pavlov // Programming and Computer Software. – 2012. – Vol.38, №3. – PP. 143–150.

СОДЕРЖАНИЕ