DOI: 10.25728/mlsd.2022.0952

РАСПРЕДЕЛЕНИЕ НЕОДНОРОДНОГО КОМПЛЕКСА РЕСУРСОВ ПРИ РЕГИОНАЛЬНОМ ПЛАНИРОВАНИИ В УСЛОВИЯХ НЕОПРЕДЕЛЕННОСТИ

Кононов Д.А.

Российский государственный гуманитарный университет
Институт проблем управления им. В.А. Трапезникова РАН,
Россия, г. Москва, ул. Профсоюзная, д.65
dmitrykon52@gmail.com

Фуругян М.Г.

Федеральный исследовательский центр «Информатика и управление» РАН,
Россия, г. Москва, ул. Вавилова, д. 40
rtsccas@yandex.ru

Аннотация. Рассматривается задача распределения неоднородного множества ресурсов при планировании сложного комплекса работ, состоящего из основных и срочных заданий (с неопределенными моментами запроса). Разработан метод построения допустимого расписания основных работ, которое минимально подвержено изменению вследствие поступления запросов на выполнение срочных заданий.

Ключевые слова: возобновляемые и невозобновляемые ресурсы, исполнительные механизмы, допустимое расписание, сетевая модель, поток минимальной стоимости, антагонистическая игра, оптимальные смешанные стратегии.

Введение

Задачи распределения ресурсов и календарного планирования возникают в различных областях деятельности человека. Например, при проектировании и работе производственных комплексов, транспортных и энергетических систем, в конвейерных производствах, при работе вычислительных комплексов и сетей, при проектировании, испытаниях и функционировании сложных технических объектов (атомные реакторы, летательные аппараты), в других областях. Задачам распределения возобновляемых ресурсов (машины, исполнительные механизмы, процессоры, приборы) посвящены многие работы по теории расписаний. Отметим, например, такие фундаментальные работы, как [1, 2], в которых исследованы различные задачи теории расписаний (составление однопроцессорных и многопроцессорных расписаний, расписаний с прерываниями и без прерываний, с отношениями предшествования и независимыми заданиями, с критериями быстродействия, минимального запаздывания, директивными сроками, другие NP-трудные и полиномиально разрешимые задачи). В [3] исследуются NP-трудные задачи быстродействия и минимизации максимального временного смещения для одного и нескольких приборов. Предлагается новый подход к поиску приближенных решений. В [4] предлагается методика построения оптимальных расписаний в задачах с нефиксированными параметрами (длительности, потребляемые ресурсы). Методика основана на использовании метода «ветвей и границ" и построении многогранников устойчивости решений. Задачи распределения невозобновляемых ресурсов (финансы, электроэнергия, горючие материалы) рассмотрены в [5, 6]. В предположении, что известна зависимость длительности выполнения работ от количества выделенных им ресурсов, исследованы задачи минимизации времени выполнения комплекса заданий, а также минимизации объема потребляемых ресурсов. В [5] данные задачи исследованы на примере системы ПЕРТ. Приведено большое количество практических задач, решение которых основано на использовании системы ПЕРТ. В [7] решена задача распределения неоднородного комплекса ресурсов, содержащего как возобновляемые, так и невозобновляемые ресурсы. Разработан полиномиальный алгоритм построения допустимого расписания с прерываниями для совокупности работ, использующих данный комплекс ресурсов. В [8] исследовано обобщение этой задачи на случай, когда доступные объемы возобновляемых (механизмов) и невозобновляемых ресурсов могут различаться в различные моменты времени. Кроме того, при определенных ограничениях допускается параллельное выполнение одной работы несколькими механизмами.

В настоящей статье рассматривается задача распределения неоднородного набора ресурсов (возобновляемых и невозобновляемых) при составлении допустимого расписания для сложного комплекса работ, состоящего из основных и дополнительных (срочных) заданий. Этот комплекс включает в себя как работы, допускающие прерывания, так и непрерываемые работы. Кроме того, в некоторые неопределенные моменты времени могут поступать запросы на дополнительные (срочные) работы, которые требуют немедленного их выполнения и выделения им требуемых ресурсов.

Например, в вычислительных системах реального времени при проведении летных испытаний требуется немедленное выполнение определенных программных модулей в случае, когда некоторые параметры выходят за границы допустимых значений. При этом возможно возникновение коллизии, т.е. ситуации, когда ресурс, требуемый для выполнения срочной работы, занят выполнением основного задания, которое будет отложено на более позднее время. Тем самым нарушается построенное ранее расписание выполнения основных работ и требуется его корректировка. Разработан алгоритм распределения ресурсов и построения допустимого расписания основных работ, при котором вероятность возникновения коллизии минимальна. Алгоритм основан на использовании сетевого моделирования, теории потоков в сетях и теории антагонистических игр.

1 Математическая постановка задачи

Имеется комплекс работ (заданий) W, который в горизонте планирования $[0;\ T]$ должен быть выполнен с использованием двух видов ресурсов – возобновляемых и невозобновляемых. Возобновляемые ресурсы (машины, приборы, процессоры) представлены набором идентичных исполнительных механизмов $M = M_1 \cup M_2$, где M_1 и M_2 – это множества, занумерованные соответственно от 1 до m_1 и от m_1+1 до m_2: $M_1 = \{\overline{1, m_1}\}$, $M_2 = \{\overline{m_1+1, m_2}\}$. Каждый исполнительный механизм может использоваться многократно. Невозобновляемый ресурс (финансы, электроэнергия, горючие материалы) выделяется работам в начальный момент горизонта планирования и не может использоваться повторно. Его объем составляет R.

Комплекс работ W состоит из трех множеств заданий: W=W_1∪W_2∪W_3; W_1 и W_2 – это основные работы, W_3 – дополнительные (или срочные). Множество $W_1 = \{\overline{1, n_1}\}$ – это работы, занумерованные от 1 до n_1 и выполняемые исполнительными механизмами M_1. При выполнении заданий $i \in W_1$ допускаются прерывания и переключения с одного исполнительного механизма на другой. Предполагается, что временными затратами на прерывания и переключения можно пренебречь. Множество W_2 представляется в виде $W_2 = \bigcup_{j=m_1+1}^{m_2} W_{2j}$, где работы из $W_{2j} = \{i_{j1}, i_{j2}, \ldots, i_{p_j}\}$ должны выполняться исполнительным механизмом $j \in M_2$ в определенном порядке: сначала i_{j1}, затем i_{j2} и т.д. Работы W_{2j} не допускают прерываний и переключений. Будем предполагать, что все работы W_2 занумерованы от n_1+1 до n_1+n_2. Множество дополнительных (срочных) работ W_3 представляется в виде $W_3 = \{i_j^3 : j = \overline{m_1+1, m_2}\}$, где работа i_j^3 должна выполняться исполнительным механизмом $j \in M_2$ без прерываний и переключений. Не допускается одновременное выполнение нескольких работ из W одним исполнительным механизмом и параллельное выполнение одной работы несколькими исполнительными механизмами.

Каждой работе $i \in W_1 \cup W_2$ может быть выделен невозобновляемый ресурс в объеме r_i. При этом должны выполняться следующие ограничения:

$$r_i \in [0;\ r_i^0],\ i = \overline{1, n_1+n_2}, \tag{1}$$

$$\sum_{i \in W_1 \cup W_2} r_i \le R\ . \tag{2}$$

При выделении работе $i \in W_1 \cup W_2$ невозобновляемого ресурса в объеме r_i ее длительность составляет

$$t_i = t_i^0 - a_i r_i. \tag{3}$$

Здесь r_i^0, t_i^0 и a_i – заданные положительные величины; t_i^0 – это длительность выполнения задания i в случае, когда невозобновляемый ресурс ему не выделяется; $t_i^0 - a_i r_i^0 > 0$. Предполагается, что

$$\sum_{i \in W_{2j}} t_i^0 \le T\ . \tag{4}$$

при всех $j = \overline{m_1+1, m_2}$.

Для работ $i \in W_1$ заданы директивные интервалы $[b_i, f_i]$, $0 \le b_i < f_i \le T$. У работ $i \in W_2 \cup W_3$ директивный интервал совпадает с горизонтом планирования $[0;\ T]$.

Работам $i_j^3 \in W_3$ невозобновляемый ресурс не выделяется и их длительности $t_{i_j^3}$ фиксированы. Запрос на выполнение каждой работы $i_j^3 \in W_3$ может поступить в любой момент времени $t \in [0;\ T]$. Исполнительный механизм $j \in M_2$, к которому приписана работа i_j^3, сразу же передается ей до ее завершения. Если при этом исполнительный механизм $j \in M_2$ занят некоторой работой $i \in W_2$, то выполнение последней прекращается и будет заново начато после завершения работы i_j^3. Такую ситуацию будем называть коллизией. При возникновении коллизии следует произвести корректировку построенного ранее расписания выполнения работ W_{2j}.

Задача заключается в распределении невозобновляемого ресурса, удовлетворяющего ограничениям (1), (2), и построении расписания работ $W_1 \cup W_2$, при котором:

– все работы $i \in W_1 \cup W_2$ выполняются в своих директивных интервалах ($[b_i, f_i]$ для $i \in W_1$ и $[0;T]$ для $i \in W_2$; такое расписание будем называть допустимым;

– вероятность возникновения коллизии минимальна.

Решение поставленной задачи разобьем на ряд этапов.

На первом этапе невозобновляемый ресурс распределяется между работами W_1, для которых также строится допустимое расписание их выполнения исполнительными механизмами M_1, или устанавливается, что такого расписания не существует. При этом потребление невозобновляемого ресурса работами W_1 минимизируется. Тем самым для работ W_2 остается максимально возможный объем невозобновляемого ресурса. Решение этой задачи основано на использовании теории потоков в сетях.

На втором этапе оставшийся свободным невозобновляемый ресурс распределяется между работами W_2 с тем, чтобы минимизировать длительности их выполнения.

На третьем этапе строится допустимое расписание выполнение работ W_2, при котором вероятность возникновения коллизий минимальна. При решении этой задачи используется теория антагонистических игр.

2 Распределение невозобновляемого ресурса и построение допустимого расписания для работ W_1

Для распределения невозобновляемого ресурса между работами W_1 и построения для них допустимого расписания воспользуемся методом, предложенным в [8].

Сначала по аналогии с тем, как это сделано в [1], построим потоковую сеть $G = (V, A)$ следующим образом (рис. 1). Пусть $y_0 < y_1 < \ldots < y_q$ – все различные значения b_i, f_i, $i \in W_1$. Будем рассматривать интервалы $I_j = [y_{j-1};\ y_j]$ и пусть $d_j = y_j - y_{j-1}$ – длина интервала I_j, $j = \overline{1,q}$. В сети G определим множество узлов $V = \{s, t, I_j, j = \overline{1,q}, w_i, i \in W_1\}$ (s, t – соответственно источник и сток) и ориентированное множество дуг $A = \{(s, I_j),\ j = \overline{1,q}, (I_j,\ w_i), j = \overline{1,q}, i \in W_1, (w_i,\ t),\ i \in W_1\}$. Узлы w_i соответствуют работам $i \in W_1$. Дуги $(I_j,\ w_i)$ присутствуют в сети G только в случае, когда $I_j \subseteq [b_i, f_i]$. Каждая дуга $(a, b) \in A$ имеет три параметра: нижнюю и верхнюю границы потока, соответственно $L(a, b)$ и $U(a, b)$, а также стоимость $C(a, b)$ единицы потока по дуге (a, b). Значения этих параметров приведены в таблице 1.

В сети G рассмотрим поток $g(a, b)$, $(a, b) \in A$.

Пусть

$$c(g) = \sum_{(a,b)\in A} C(a, b) g(a, b) - \tag{5}$$

стоимость потока. Как следует из [8], допустимое расписание для работ W_1 существует в том и только том случае, когда в сети G существует поток g, для которого выполнено неравенство

$$c(g) \leq R - \sum_{i \in W_1} t_i^0 / a_i. \tag{6}$$

Таким образом, для построения допустимого расписания работ W_1 следует найти поток g_0 минимальной стоимости (5) в сети G. Если неравенство (6) для $g = g_0$ не выполняется, то в поставленной задаче решения не существует. При выполнении неравенства (6) для $g = g_0$, как следует из [8], каждой работе $i \in W_1$ следует выделить невозобновляемый ресурс в количестве

$$r_i = \frac{t_i^0 - g_0(w_i, t)}{a_i}. \tag{7}$$

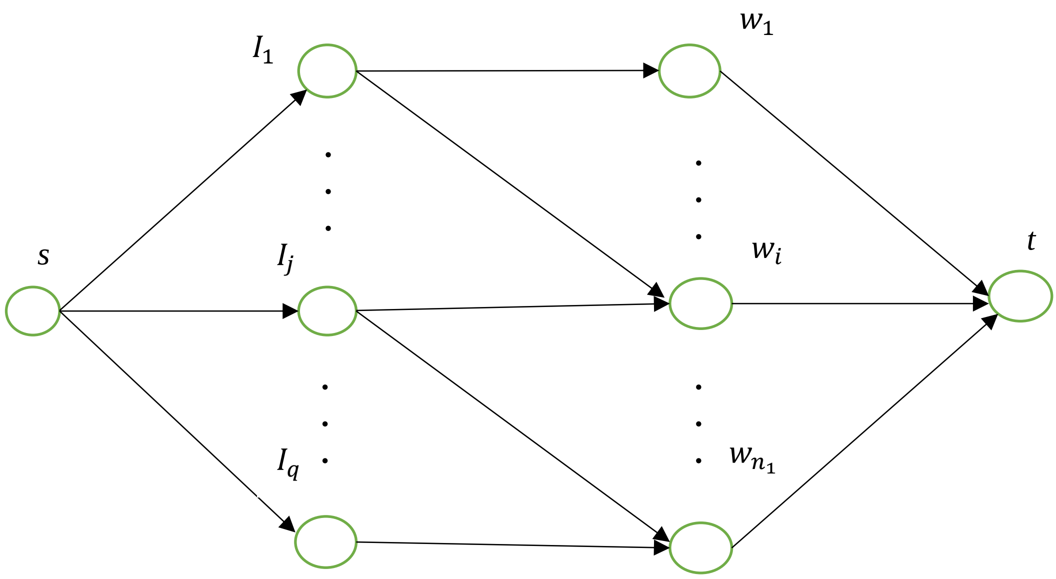

Рис. 1. Потоковая сеть G

Таблица 1. Значения параметров дуг сети G

Дуга	***L***	***U***	***C***
(s,I_j)	0	m_1d_j	0
(I_j,w_i)	0	d_j	0
(w_i,t)	$t^0_i - a_ir^0_i$	t^0_i	$-1/a_i$

При этом длительность работы $i \in W_1$ вычисляется согласно (3). Величина $g_0(I_j,\ w_i)$ равна суммарной длительности выполнения работы $i \in W_1$ исполнительными механизмами в интервале I_j. Для построения расписания выполнения работ W_1 исполнительными механизмами в интервале I_j следует применить алгоритм упаковки [1]. Для построения допустимого расписания выполнения работ W_1 в горизонте планирования [0; T] следует совместить расписания для всех интервалов $(I_j,\ w_i)$, $j = \overline{1,q}$.

Отметим, что искомый поток g_0, определяющий по формуле (7) величины r_i, $i \in W_1$, с одной стороны, позволяет построить допустимое расписание для работ W_1, а с другой стороны, минимизирует потребление невозобновляемого ресурса работами W_1. В этом смысле, вектор $(g_0(w_1,t), g_0(w_2,t), \ldots, g_0(w_{n_1},t))$ является парето-оптимальным. Действительно, если существует поток g_1, для которого $g_1(w_i,t) \geq g_0(w_i,t)$ при всех $i = \overline{1,n_1}$, а при некотором $i_0, 1 \leq i_0 \leq n_1$, $g_1(w_i,t) > g_0(w_i,t)$, то $c(g_1) < c(g_0)$, чего быть не может, поскольку g_0 – это поток минимальной стоимости.

3 Распределение невозобновляемого ресурса и построение допустимого расписания для работ W_2

Если после распределения невозобновляемого ресурса между работами W_1 выполнено соотношение $R > \sum_{i \in W_1} r_i$, где величины r_i, $i \in W_1$, вычислены согласно (7), то оставшуюся свободной часть невозобновляемого ресурса объемом $R_1 = R - \sum_{i \in W_1} r_i$ следует распределить между работами W_2. В силу (3) работам $i \in W_2$ с наибольшими значениями коэффициентов a_i будем выделять максимально возможный объем невозобновляемого ресурса с тем, чтобы минимизировать длительности их выполнения. Упорядочим работы W_2 по не возрастанию величин a_i, $i \in W_2$, и пусть $a_{n_1+1} \geq a_{n_1+2} \geq \ldots \geq a_{n_1+n_2}$. Распределим невозобновляемый ресурс между работами W_2 по следующему простому алгоритму.

Положить $k = n_1$.

Положить $k = k + 1$.

Положить $r_k = min(r^0_k, R_1)$.

Положить $R_1 = R_1 - r_k$.

Если $R_1 > 0$ и $k < n_1 + n_2$, то перейти на шаг 2.

Если $R_1 = 0$ или $k = n_1 + n_2$, алгоритм завершает работу.

В результате работы этого алгоритма будут определены длительности t_i всех работ $i \in W_2$, вычисленные согласно (3).

Будем строить допустимое расписание отдельно для каждого множества $W_{2j} = \{i_{j1},\ i_{j2}, \dots, i_{p_j}\}$, $j = \overline{m_1 + 1, m_2}$. Пусть $x_1,\ x_2, \dots, x_{p_j}$ – искомые моменты начала выполнения исполнительным механизмом j работ $i_{j1}, i_{j2}, \dots, i_{p_j}$ соответственно. Определим множество

$X = \{x = (x_1, x_2, \dots, x_{p_j})\colon 0 \le x_1;\ x_k + t_k \le x_{k+1}, k = \overline{1, p_j - 1}, x_{p_j} + t_{p_j} \le T\}$.

В силу (4) $X \neq \emptyset$. Пусть, далее, y – это момент поступления запроса на выполнение срочной работы $i_j^3 \in W_3$. Задача заключается в выборе такого $x \in X$, при котором вероятность возникновения коллизии минимальна. Определим на множествах чистых стратегий X и Y= [0;T] антагонистическую игру с платежной функцией $K(x, y)$ следующим образом:

$$K(x,y) = \begin{cases} 0, \text{если } (y + t_{i_j^3}) \cap [x_k, x_k + t_k] \neq \emptyset \text{ при некотором } k, 1 \le k \le p_j - 1; \\ \qquad\qquad 1 \text{ в противном случае.} \end{cases} \tag{8}$$

Из определения функции $K(x, y)$ видно, что если $K(x,y) = 1$, то коллизии не возникает, а при $K(x,y) = 0$ возникает коллизия.

Пусть $\alpha(x)$ – вероятностная мера на множестве X, т.е. некоторая смешанная стратегия первого игрока в антагонистической игре с платежной функцией (8). Тогда величина

$$\int_X K(x,y)\, d\alpha(x) \tag{9}$$

равна вероятности не возникновения коллизии в случае, если запрос на выполнение срочной работы i_j^3 поступает в момент времени y, а моменты начала выполнения работ W_{2j} выбираются согласно распределению $\alpha(x)$. Пусть, далее, $\alpha^*(x)$ – это оптимальная смешанная стратегия первого игрока в антагонистической игре с платежной функцией (8), а **Ω** – множество всех вероятностных мер на X.

Тогда справедливо равенство

$$\max_{\alpha(x)\in\Omega} \min_{y\in Y} \int_X K(x,y)\, d\alpha(x) = \min_{y\in Y} \int_X K(x,y) d\alpha^*(x) \tag{10}$$

Это означает, что моменты начала выполнения работ W_2, т.е. вектор $x \in X$, следует выбирать согласно распределению $\alpha^*(x)$, что обеспечит максимальную вероятность не возникновения коллизии между работами W_{2j} и i_j^3, т.е. минимальную вероятность ее возникновения.

Таким образом, задача свелась к нахождению оптимальной смешанной стратегии первого игрока в антагонистической игре с платежной функцией (8). Для этого может быть использован метод дискретной аппроксимации решения антагонистических игр с полунепрерывной платежной функцией, изложенный в [10].

Рассмотрим решение антагонистической игры с платежной функцией вида (8) на следующем модельном примере. Пусть $W_{2j} = \{i_{j1}\}$, t_{j1}=1, T = 3, $t_{i_j^3} = 1$, X = Y = [0; 3]. Тогда

$$K(x,y) = \begin{cases} 0 \text{ при } x - 1 \le y \le x + 1, \\ 1 \text{ при } y < x - 1 \text{ и } y > x + 1. \end{cases} \tag{11}$$

На рис. 2 дано графическое изображение платежной функции $K(x, y)$.

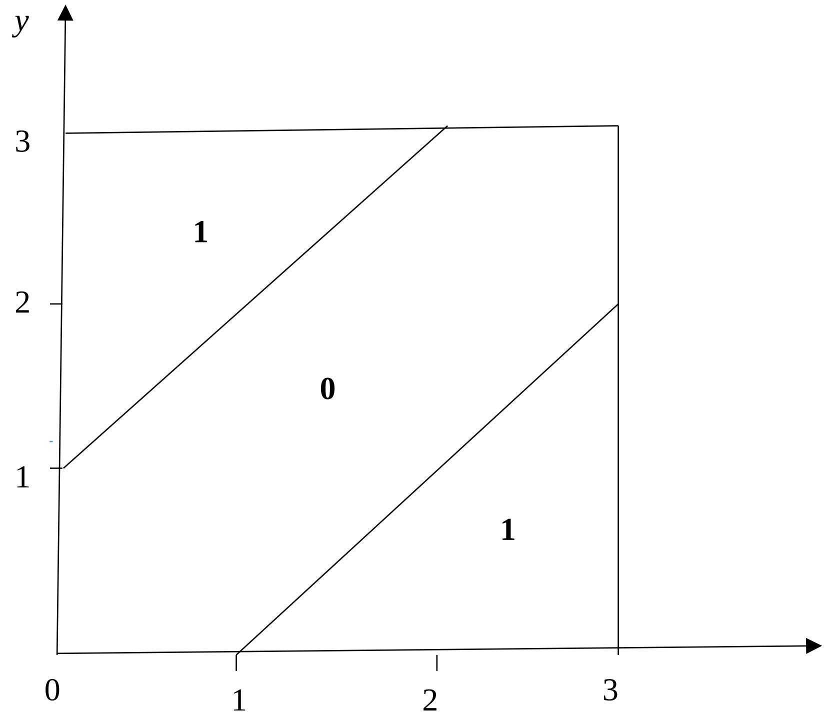

Рис. 2. Графическое изображение платежной функции K(x, y)

Пусть $\alpha^*(x)$ – вероятностная мера на X, сосредоточенная в точках x = 0 и x = 3, в которых имеет скачки, равные 0,5. Пусть $\beta^*(y)$ – вероятностная мера на Y, сосредоточенная в точках y = 1 и y = 2, в которых имеет скачки, равные 0,5. В этом случае в силу (9), (10)

$$\int_X K(x,y)\,d\alpha^*(x) \geq 0{,}5 \text{ при всех } y \epsilon Y, \tag{12}$$

$$\int_Y K(x,y)\,d\beta^*(y) \leq 0{,}5 \text{ при всех } x \epsilon X. \tag{13}$$

Из (12), (13) следует, что значение игры с платежной функцией (11) в смешанных стратегиях равно 0,5, а $\alpha^*(x)$ – оптимальная смешанная стратегия первого игрока. Таким образом, для минимизации вероятности коллизии с исполнительным механизмом j работу i_{j1} следует начинать в момент времени x = 0 с вероятностью 0,5 и в момент времени x = 3 также с вероятностью 0,5.

Заключение

В настоящей работе исследована задача распределения ресурсов двух типов, возобновляемых и невозобновляемых, при планировании сложного комплекса работ. Данный комплекс состоит из основных и дополнительных (срочных) работ. Основные работы используют оба типа ресурсов и делятся на допускающие прерывания при их выполнении и непрерываемые. Дополнительные работы являются непрерываемыми, используют только возобновляемые ресурсы, а запросы на их выполнение поступают в неопределенные моменты времени. Эти работы требуют немедленного выделения необходимых им ресурсов, вследствие чего некоторые основные работы могут быть отложены на более позднее время. Такая ситуация называется коллизией. Все работы характеризуются длительностями, которые линейно зависят от выделенного им невозобновляемого ресурса, и директивными интервалами. Разработан метод поиска распределения ресурсов и построения допустимого расписания основных работ, которое минимально подвержено изменению вследствие поступления запросов на выполнение дополнительных работ, т.е. при котором вероятность возникновения коллизии минимальна. Метод основан на использовании сетевого моделирования, теории потоков в сетях и антагонистических игр.

Дальнейшее развитие рассмотренных исследований предполагается в направлении изучения допустимых условий управления комплексами работ в условиях неопределенности, в том числе на основе применения вероятностной, игровой и конкурентной моделей выполнения контрактов, заключаемых региональными и муниципальными заказчиками.

Литература

1. *Танаев В.С., Гордон В.С., Шафранский Я.М.* Теория расписаний. Одностадийные системы. – М.: Наука, 1984. – 383 с.
2. *Brucker P.* Scheduling Algorithms. – Heidelberg: Springer, 2007. – 371 p.
3. *Лазарев А.А.* Теория расписаний. Оценка абсолютной погрешности и схема приближенного решения задач теории расписаний. – М.: МФТИ, 2008. – 222 с.
4. *Мищенко А.В., Кошелев П.С.* Оптимизация управления работами логистического проекта в условиях неопределенности // Известия РАН. Теория и системы управления. 2021, № 4. – С.123-134.
5. *Филлипс Д., Гарсиа-Диас А.* Методы анализа сетей. – М.: Мир, 1984. – 496 с.
6. *Давыдов Э.Г.* Исследование операций. – М.: Высшая школа, 1990. – 384 с.
7. *Косоруков Е.О., Фуругян М.Г.* Некоторые алгоритмы распределения ресурсов в многопроцессорных системах // Вестник МГУ. Сер. 15. 2009, № 4. – С. 34-37.
8. *Кононов Д.А., Фуругян М.Г.* Оптимизация использования неоднородного комплекса ресурсов при региональном планировании. Труды 14-й международной конференции «Управление развитием крупномасштабных систем. MLSD'2021. (27 – 29 сентября 2021 г., Москва)». – М.: ИПУ РАН, 2021. – С. 1231–1237.
9. *Корте Б., Фиген Й.* Комбинаторная оптимизация. Теория и алгоритмы. – М.: МЦНМО, 2015. – 720 с.
10. *Фуругян М.Г.* Приближенное решение одного класса бесконечных антагонистических игр с полунепрерывной платежной функцией // Вестник МГУ. Сер. 15. 1980, № 2. – С. 66-69.