



CONTROL SCIENCES

**Научно-технический
журнал**

6 номеров в год

ISSN 1819-3161

Издается с 2003 года

УЧРЕДИТЕЛЬ

Федеральное государственное
бюджетное учреждение науки
Институт проблем управления
им. В.А. Трапезникова РАН

Главный редактор
чл.-корр. РАН

Д.А. Новиков

Заместители главного
редактора

Л.П. Боровских, Ф.Ф. Пашенко

Редактор

Т.А. Гладкова

Выпускающий редактор

Л.В. Петракова

Издатель

ООО «Сенсидат-Плюс»

Адрес редакции

117997, ГСП-7, Москва,

ул. Профсоюзная, д. 65, к. 410.

Тел./факс (495) 334-92-00

E-mail: pu@ipu.ru

Интернет: http://pu.mtas.ru

Оригинал-макет и электронная версия
подготовлены

ИП Прохоров О. В.

Фото на четвертой странице обложки
В.М. Бабинова

Отпечатано в ООО «Авансд солишнз»

Заказ № PB418

Подписано в печать

16.07.2018 г.

Журнал зарегистрирован
в Министерстве Российской
Федерации по делам печати,
телерадиовещания и средств
массовых коммуникаций

Свидетельство о регистрации
ПИ № ФС 77-49203 от 30 марта 2012 г.

Журнал входит в RSCI на платформе
Web of Science и Перечень
рецензируемых научных изданий ВАК

Подписные индексы:

80508 и 81708 в каталоге Роспечати;

38006 в объединенном каталоге
«Пресса России»

Цена свободная

© Федеральное государственное
бюджетное учреждение науки
Институт проблем управления
им. В.А. Трапезникова РАН

ПРОБЛЕМЫ УПРАВЛЕНИЯ

4.2018

СОДЕРЖАНИЕ

Обзоры

- Вересников Г.С., Егоров Н.А., Кулида Е.Л., Лебедев В.Г.** Методы построения оптимальных очередей воздушных судов на посадку.
Ч. 1. Методы точного решения 2

Анализ и синтез систем управления

- Галеев А.А., Лысенко П.В.** Задача оптимального быстрогодействия при упругом и вязкоупругом взаимодействиях тела с поверхностью . . . 14
- Цирлин А.М., Овсепян В.С.** Робастная устойчивость и выбор настройки регуляторов для технологических объектов с запаздыванием 21
- Русев В.Н., Скориков А.В.** Аппроксимации функции восстановления и стратегия управления эксплуатационными затратами 28

Управление в социально-экономических системах

- Белов М.В., Новиков Д.А.** Основы теории комплексной деятельности.
Ч. 1. Структуры комплексной деятельности. Неопределенность и порождение комплексной деятельности 36

Информационные технологии в управлении

- Сомов С.К.** Задача оптимального размещения информационно-технологического резерва в распределенных системах обработки данных 46
- Подлазов В.С.** Бесконфликтная самомаршрутизация для трехмерного полного мультикольца 54

Управление в медико-биологических системах

- Бабушкина Н.А., Кузина Е.А., Лоос А.А., Беляева Е.В.** Результаты исследования режимов применения противоопухолевых вирусных вакцин на основе математического моделирования 61

Управление подвижными объектами и навигация

- Завадский В.К., Иванов В.П., Каблова Е.Б., Кленовая Л.Г.** Комплексирование расходомерной и уровнемерной информации в системах управления расходом топлива 71

* * *

- Contents and abstracts 80

МЕТОДЫ ПОСТРОЕНИЯ ОПТИМАЛЬНЫХ ОЧЕРЕДЕЙ ВОЗДУШНЫХ СУДОВ НА ПОСАДКУ¹.

Ч. 1. Методы точного решения

Г.С. Вересников, Н.А. Егоров, Е.Л. Кулида, В.Г. Лебедев

Отмечено, что задача формирования оптимальных очередей воздушных судов на посадку одна из основных задач при планировании и регулировании потоков воздушного движения, для решения которой требуется использование достаточно серьезного математического аппарата. Представлен краткий обзор основных подходов к постановке и решению этой задачи.

Ключевые слова: воздушное движение, критерий оптимизации, оптимальная очередь на посадку, линейное программирование, метод ветвей и границ.

ВВЕДЕНИЕ

Системы планирования и регулирования потоков воздушного движения (ВД) предназначены для обеспечения безопасного, упорядоченного и ускоренного ВД, в том числе в районах аэропортов и аэроузлов, уменьшения издержек авиакомпаний и пассажиров, уменьшения негативного влияния на окружающую среду с учетом неопределенности прогнозирования ВД.

Сложность планирования и регулирования потоков ВД в районе аэродрома и загрузка диспетчеров возрастают с увеличением интенсивности ВД. При этом падает эффективность управления, возрастают задержки воздушных судов (ВС), увеличивается расход топлива. Для предотвращения подобных ситуаций разрабатываются специальные средства поддержки работы диспетчера при планировании и регулировании потоков ВД, которые выполняют часть функций диспетчера и позволяют более эффективно использовать пропускную способность взлетно-посадочной полосы (ВПП) и воздушного пространства. В Европе такие средства получили название Arrival Manager (AMAN) и Departure Manager (DMAN) — менеджеры прилета и вылета.

Системы AMAN и их аналоги применяются в Европе и США более 25 лет, постоянно обновляясь и совершенствуясь. Современные системы активно внедряются по всему миру, а программы развития системы организации ВД SESAR и NextGen ведут обширные проекты по дальнейшему их совершенствованию.

В Российской Федерации есть несколько разработчиков систем планирования и регулирования прибывающих потоков ВД, относительно недавно начавших работы в данном направлении. Большинство из отечественных систем AMAN находятся на этапе разработки и тестирования, их внедрение пока только планируется. Однако в рамках модернизации Единой системы организации ВД Российской Федерации ведутся исследования ведущими научно-исследовательскими институтами и производителями систем автоматизации управления ВД, среди которых ГосНИИ гражданской авиации и филиал «Аэронавигация» [1], ГосНИИАС [2], Московский государственный технический университет гражданской авиации [3], Государственный университет гражданской авиации в Санкт-Петербурге [4], Московский авиационный институт [5], Ульяновское высшее авиационное училище [6, 7] и др.

«Построение оптимальных очередей на прилет является одной из основных задач при планировании и регулировании потоков ВД в районе аэродрома и аэроузла, для решения которой требуется

¹ Работа выполнена при частичной финансовой поддержке РФФИ (проект № 18-08-00822) и Программы I.30 Президиума РАН.



использование достаточно серьезного математического аппарата.

Самое простое решение такой задачи, по сути, не требующее математических вычислений, — это стратегия «первый пришел — первый обслужился» (first come — first served (FCFS)). При данном подходе не требуются перестановки в естественной очереди, и, казалось бы, посадка в первую очередь раньше прибывших ВС должна обеспечить минимальные задержки. Но разные летно-технические характеристики ВС, особенности структуры STAR (Standard Terminal Arrival Route) и интенсивное воздушное движение делают такой подход несостоятельным, в то время как некоторые перестановки в естественной последовательности (например, перестановка медленного ВС за быстрым) могут существенно улучшить результат. Поэтому в большинстве случаев при формировании оптимальной очереди на посадку приходится отступать от стратегии FCFS и каким-либо образом выбирать новый порядок в очереди, применяя при этом более сложные алгоритмы» [8].

Задача формирования оптимальной очереди на посадку (ФООП) стоит на повестке дня уже несколько десятилетий, но до сих пор не утратила своей актуальности. Первые публикации, посвященные задаче ФООП, появились в 1970-х гг. Однако сложность проблемы, новые требования к организации ВД, новые методы решения оптимизационных задач делают задачу актуальной и в настоящее время, о чем свидетельствуют многочисленные новые публикации на эту тему. В силу сложности и важности ФООП все еще остается одной из основных проблем в исследовании проблем управления потоками ВД [9].

В литературе рассматривается большое число различных постановок задачи ФООП [10]. Различаются статическая задача [11], которая решается для фиксированного набора ВС, и динамическая задача [12] для переменного набора ВС, когда часть ВС появляется в зоне регулирования, а другая их часть эту зону покидает. Различаются постановки задачи для посадки на одну или несколько ВПП. Рассматриваются различные критерии оптимизации, среди которых можно выделить два основных. Один из них применяется, когда плотность воздушного движения высока и необходимо эффективно использовать пропускную способность ВПП. Другой критерий применяется для того, чтобы все ВС приземлились как можно ближе к своему оптимальному времени посадки. Он обеспечивает наиболее эффективное и безопасное ВД, уменьшение издержек авиакомпаний и пассажиров, уменьшение негативного влияния на окружающую среду и может применяться при разумной плотности трафика.

В литературе представлены точные и приближенные методы решения задачи ФООП, Основными из них — это методы линейного программирования, метод ветвей и границ, эвристические и метаэвристические методы.

Точные алгоритмы основаны на анализе всех возможных перестановок ВС в очереди посадок, поэтому время работы экспоненциально растет с размером проблемы (числом ВС). Поэтому в точных алгоритмах решения задачи применяются различные приемы для сокращения перебора. Тем не менее, не удается получить оптимальное решение задач большой размерности в течение приемлемого времени. В литературе представлены примеры оптимальных решений задачи ФООП не более чем для 50 ВС.

Для получения приближенных решений задачи ФООП в литературе представлено большое число эвристических, метаэвристических, гибридных и прочих алгоритмов, в том числе с приведенными вычислительными результатами для 500 ВС и 5 ВПП, полученными за приемлемое время. Особенно успешными для приближенного решения задачи ФООП оказались меметические алгоритмы.

Обзор состоит из двух частей.

В первой части подробно рассматривается представленная в работе [11] классическая постановка статической задачи ФООП на основе линейного целочисленного программирования с булевыми переменными, которая лежит в основе множества других работ. Приводится обзор различных модификаций этой постановки.

Приводится краткий обзор в историческом разрезе методов решения задачи ФООП, начиная с самых ранних работ. Затем подробно рассматривается подход к решению поставленной задачи линейного программирования с помощью линейной релаксации [11] и решение задачи на основе метода ветвей и границ, дополненного методом, называемым «генерацией столбцов» [13].

Вторая часть обзора посвящена предлагаемому в мировой научной литературе перспективным алгоритмам оптимизации для приближенного решения задачи построения оптимальных очередей посадок ВС на одну или несколько ВПП.

1. ПОСТАНОВКА СТАТИЧЕСКОЙ ЗАДАЧИ ФОРМИРОВАНИЯ ОПТИМАЛЬНОЙ ОЧЕРЕДИ ВОЗДУШНЫХ СУДОВ НА ПОСАДКУ

В настоящее время классической постановкой для задачи формирования оптимальной очереди на посадку, которая лежит в основе множества других работ, считается постановка, представленная в работе [11]. В ее основе лежит задача линейного це-

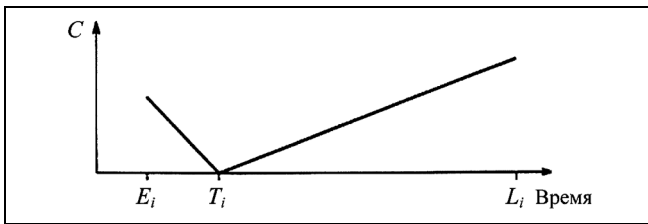


Рис. 1. Изменение затрат внутри временного окна посадки воздушных судов

лочисленного программирования с булевыми переменными.

1.1. Основные обозначения

Введем обозначения: P — число ВС, ожидающих посадки; E_i — самое раннее возможное время прибытия i -го ВС, $i = 1, \dots, P$; L_i — самое позднее возможное время прибытия i -го ВС, $i = 1, \dots, P$; T_i — оптимальное время прибытия i -го ВС, $i = 1, \dots, P$.

На рис. 1 представлена функция изменения затрат C в зависимости от времени посадки ВС в течение временного окна. Она нелинейная, однако, она состоит из двух линейных участков, что позволяет сформулировать задачу с линейной целевой функцией. Далее обозначим: S_{ij} — минимальный интервал между посадкой ВС j за ВС i , $j, i = 1, \dots, P$; g_i — штраф за единицу времени приземления i -го ВС, $i = 1, \dots, P$, до времени T_i ; h_i — штраф за единицу времени приземления i -го ВС, $i = 1, \dots, P$, после времени T_i .

Переменные: x_i — время приземления i -го ВС, $i = 1, \dots, P$; α_i — на сколько i -е ВС, $i = 1, \dots, P$, приземляется раньше времени T_i ; β_i — на сколько i -е ВС, $i = 1, \dots, P$, приземляется позднее времени T_i ;

$$\delta_{ij} = \begin{cases} 1, & \text{если ВС } i \text{ приземляется раньше ВС } j \\ i, j = 1, \dots, P, \\ 0 & \text{в противном случае.} \end{cases}$$

1.2. Ограничения, которым должно удовлетворять решение задачи

1. Практически во всех работах на время приземления налагаются ограничения, обусловленные летно-техническими характеристиками ВС и параметрами зоны приземления: ограничения на ускорения, конечные размеры зоны приземления, ограничения по топливу и др. Все эти факторы определяют фиксированный интервал допустимого времени приземления: ВС с номером i должно приземляться внутри временного окна $[E_i, L_i]$: $E_i \leq x_i \leq L_i$, $i = 1, \dots, P$, [11, 14–18].

2. Либо ВС i приземляется раньше ВС j ($\delta_{ij} = 1$), либо ВС j приземляется раньше ВС i ($\delta_{ji} = 1$): $\delta_{ij} + \delta_{ji} = 1$, $i, j = 1, \dots, P$.

На основе информации о временных окнах все пары (i, j) , $(i, j = 1, \dots, P)$ можно разделить на множества: U — множество пар (i, j) , для которых на основе анализа временных окон и ограничений задержки между последовательными посадками ВС (ограничения разделения) нельзя установить, какое ВС приземляется раньше; V — множество пар (i, j) , в которых ВС i должно приземлиться раньше ВС j , но при этом ограничения разделения не выполняются автоматически; W — множество пар (i, j) , в которых ВС i должно приземлиться раньше ВС j , при этом ограничения разделения автоматически выполняются: $\delta_{ij} = 1 \forall (i, j) \in W \cup V$.

3. Одно из наиболее распространенных ограничений в большинстве задач ФООП состоит в минимальной допустимой из соображений безопасности задержке между последовательной посадкой двух ВС [11, 14–16].

По длительности этой задержки ВС разделяются на ряд типов. Пример стандартной минимальной допустимой задержки между последовательными приземлениями ВС, приведенный в работе [14], представлен в табл. 1.

Эти ограничения разделения можно представить в виде:

$$x_j \geq x_i + S_{ij} \quad \forall (i, j) \in V,$$

$$x_j \geq x_i + S_{ij}\delta_{ij} - (L_i - E_j)\delta_{ji} \quad \forall (i, j) \in U.$$

Возможны два случая:

— если $\delta_{ij} = 1$, тогда из ограничения 2 следует $\delta_{ji} = 0$, и ограничение принимает вид: $x_j \geq x_i + S_{ij}$ для всех пар ВС, в которых ВС i приземляется раньше ВС j ;

— если $\delta_{ij} = 0$, тогда из ограничения 2 следует $\delta_{ji} = 1$, и ограничение принимает вид: $x_j \geq x_i + (L_i - E_j)$ или $x_j - E_j \geq x_i - L_i$, которое всегда выполняется, потому что левая часть больше или рав-

Таблица 1

Минимальная допустимая задержка между приземлениями, с

		Тип следующего приземляющегося ВС			
		1	2	3	4
Тип предыдущего приземляющегося ВС	1	96	200	181	228
	2	72	80	70	110
	3	72	100	70	130
	4	72	80	70	90
Примечание. 1 — Boeing 747, 2 — Boeing 727, 3 — Boeing 707, 4 — McDonnellDouglasDC 9.					



на 0, а правая часть меньше или равна 0. Таким образом, в этом случае ограничение эффективно деактивируется.

4. Ограничения для связи переменных α_i , β_i , x_i :

$$\begin{aligned}\alpha_i &\geq T_i - x_i, \quad i = 1, \dots, P; \\ 0 &\leq \alpha_i \leq T_i - E_i, \quad i = 1, \dots, P; \\ \beta_i &\geq x_i - T_i, \quad i = 1, \dots, P; \\ 0 &\leq \beta_i \leq L_i - T_i, \quad i = 1, \dots, P; \\ x_i &= T_i - \alpha_i + \beta_i, \quad i = 1, \dots, P.\end{aligned}$$

Введение дополнительных ограничений в модель позволяет учитывать ряд практически важных моментов, например, ограничения приоритета, балансировка рабочей нагрузки на ВПП, ограничение на число посадок в течение определенного периода времени. Последнее ограничение, при числе посадок, равном 0, позволяет обеспечить временные интервалы, в течение которых ВПП недоступна для посадок.

Предположим, например, что должно быть не более Q посадок за период времени.

Пусть $G = \{i | [t_1, t_2] \cap [E_i, L_i] \neq \emptyset, i = 1, \dots, P\}$, т. е. G — это множество ВС, которые могут приземлиться в течение периода времени $[t_1, t_2]$. Для простоты предположим, что все времена целочисленные. Воздушное судно i должно приземлиться в один из трех временных отрезков: $[E_i, \max(E_i, t_1) - 1]$, $[\max(E_i, t_1), \min(L_i, t_2)]$, $[\min(L_i, t_2) + 1, L_i]$. Заметим, что некоторые из этих временных отрезков могут не существовать.

Пусть

$$\gamma_{ij} = \begin{cases} 1, & \text{если ВС } i \text{ приземляется во время} \\ & \text{временного отрезка } j; \\ 0 & \text{в противном случае.} \end{cases}$$

Для несуществующих отрезков значение γ_{ij} устанавливается равным нулю.

После этого нужно добавить к модели уравнение

$$\sum_{i \in G} \gamma_{i2} = Q,$$

которое ограничивает число посадок за период времени $[t_1, t_2]$.

$$\gamma_{i1} + \gamma_{i2} + \gamma_{i3} = 1 \quad \forall i \in G.$$

Это уравнение гарантирует посадку ВС i только в один из трех временных отрезков:

$$x_i \leq [\max(E_i, t_1) - 1] \gamma_{i1} + L_i(1 - \gamma_{i1}) \quad \forall i \in G;$$

$$\max(E_i, t_1) \gamma_{i2} \leq x_i \leq \min(L_i, t_2) \gamma_{i2} + L_i(1 - \gamma_{i2}) \quad \forall i \in G;$$

$$x_i \leq [\min(L_i, t_2) + 1] \gamma_{i3} \quad \forall i \in G.$$

Последние три уравнения соотносят время посадки ВС i с соответствующим временным отрезком.

Отмечается, что эти ограничения можно применить к любому подмножеству ВС (например, можно ограничить число самолетов от определенной авиакомпании, которые приземляются в определенный период времени).

Условие $\sum_{i \in G} \gamma_{i2} = 0$ гарантирует отсутствие посадок в период времени $[t_1, t_2]$, и поэтому приведенные ограничения позволяют реализовать закрытие ВПП в указанный период времени [19].

На времена посадки могут налагаться и другие ограничения, например, недопустимость опережения ВС планового времени приземления. Подобные ограничения рассматриваются, в частности, в работах [9, 20–23].

В ряде работ делается попытка учета того, что ВС идут с нескольких различных маршрутов и происходит их слияние в некоторый единый маршрут. При этом порядок приземления ВС из одного маршрута должен совпадать с исходным порядком ВС в этом маршруте. Последующие ВС маршрута не допускаются к посадке до впереди идущих ВС. Сторонники подобного подхода обосновывают его вопросами полетной безопасности, а также принципами справедливости [24].

1.3. Постановка статической задачи формирования оптимальной очереди воздушных судов на посадку на несколько взлетно-посадочных полос

В большинстве оживленных международных аэропортов несколько ВПП. Постановка задачи формирования оптимальной очереди на посадку может быть расширена для нескольких ВПП введением дополнительных переменных и ограничений. Рассмотрим их.

Число ВПП равно R ,

$$z_{ij} = \begin{cases} 1, & \text{если ВС } i \text{ и ВС } j \text{ приземляются} \\ & \text{на одну ВПП,} \\ 0 & \text{в противном случае, } i, j = 1, \dots, P, i \neq j, \end{cases}$$

$$y_{jr} = \begin{cases} 1, & \text{если ВС } j \text{ приземляется на ВПП } r, \\ 0 & \text{в противном случае,} \\ j = 1, \dots, P, \quad r = 1, \dots, R, \end{cases}$$

$$\sum_{r=1}^R y_{ir} = 1, \quad i = 1, \dots, P.$$

Это уравнение означает, что каждое ВС приземляется точно на одну ВПП.

$$z_{ij} = z_{ji}, \quad i, j = 1, \dots, P; \quad j > i.$$

Это симметричное ограничение (если ВС i и ВС j приземляются на одну ВПП, то это верно и для ВС j и ВС i). Далее

$$z_{ij} \geq y_{ir} + y_{jr} - 1, \quad i, j = 1, \dots, P; \quad j > i; \\ r = 1, \dots, R.$$

Последнее уравнение гарантирует, что, если ВС i и ВС j приземляются на ВПП r , то $z_{ij} = 1$. Если $z_{ij} = 0$, то уравнение гарантирует, что ВС i и ВС j не могут приземлиться на одну ВПП.

Кроме того, должны соблюдаться ограничения, о которых говорилось в п. 1.2. В некоторых аэропортах ВПП могут располагаться таким образом, что необходимы дополнительные ограничения разделения между ВС, приземляющимися на разные ВПП. В этом случае вводятся дополнительные переменные s_{ij} — минимальный интервал между посадкой ВС j за ВС i на разные ВПП, $0 \leq s_{ij} \leq S_{ij}$, $i, j = 1, \dots, P$; $i \neq j$. Ограничения разделения в этом случае можно представить в виде:

$$x_j \geq x_i + S_{ij}z_{ij} + s_{ij}(1 - z_{ij}) \quad \forall (i, j) \in V, \\ x_j \geq x_i + S_{ij}\delta_{ij} + s_{ij}(1 - z_{ij}) - (L_i + S_{ij} - E_j)\delta_{ji} \\ \forall (i, j) \in U.$$

1.4. Критерии оптимизации

В настоящее время ведутся многочисленные исследования задач ФООП для различных целей оптимизации. В частности, модели ФООП могут разрабатываться с целью либо минимизации общей задержки в воздухе (общего времени задержки всех прибывающих ВС) [20, 22], либо дополнительных затрат (издержек, вызванных ранней или поздней посадкой) [11, 14, 15], либо времени обработки процесса приземления (времени, необходимого для всех посадок) [20, 21] и др.

В работе [14] проведен анализ некоторых наиболее распространенных целевых функций, используемых в различных моделях ФООП. Среди них отметим:

$$\min \sum_{i=1}^P [\max(x_i - T_i, 0)]$$

— минимум суммарной задержки всех ВС в условиях запрета опережения планового времени приземления, где x_i и T_i — фактическое и плановое времена приземления i -го ВС соответственно.

Большинство исследователей полагают, однако, что условие запрета опережения планового времени приземления является искусственным требованием, и рассматривают в силу этого более общую функцию минимизации суммарных затрат всех ВС.

В работе [11] приводятся два основных критерия оптимизации. Первый из них служит для ми-

нимизации стоимости отклонения от целевого времени посадки:

$$\min \sum_{i=1}^P (g_i \alpha_i + h_i \beta_i),$$

т. е. минимум суммарных затрат всех ВС, вызванных отклонениями фактического времени посадки от планового, где g_i и h_i — штрафы i -го ВС за опережение и задержку фактического времени приземления относительно планового. Этот критерий обеспечивает наиболее эффективное и безопасное ВД, уменьшение издержек авиакомпаний и пассажиров, уменьшение негативного влияния на окружающую среду и может применяться при разумной плотности трафика.

Второй критерий оптимизации предназначен для максимизации пропускной способности ВПП. Такая цель возникает при высокой плотности ВД и необходимости эффективного использования пропускной способности ВПП.

Для достижения этой цели рассматривается целевая функция

$$\min \max[x_i | i = 1, \dots, P],$$

т. е. последнее ВС должно приземлиться как можно быстрее. Эта целевая функция легко линеаризуется введением дополнительных переменных.

Рассматриваются и некоторые другие виды целевых функций:

$$\min(\max[x_1, \dots, x_P] - \min[x_1, \dots, x_P])$$

— минимум времени приземления всей группы ВС и др.

2. МЕТОДЫ ПОСТРОЕНИЯ ОПТИМАЛЬНЫХ ОЧЕРЕДЕЙ ПОСАДОК ВОЗДУШНЫХ СУДОВ НА ОДНУ ИЛИ НЕСКОЛЬКО ВЗЛЕТНО-ПОСАДОЧНЫХ ПОЛОС

Приведем краткий обзор в историческом разрезе исследования задачи ФООП, начиная с ранних работ [25], в которых исследована сильно упрощенная версия задачи ФООП. В этих работах предполагалось отсутствие задержек между последовательными приземлениями ВС.

В работе [26] представлена дискретно-событийная имитационная модель для оценки различных стратегий упорядочения ВС. Приведены вычислительные результаты по ряду моделируемых сценариев.

В работах [27, 28] обсуждаются статическая и динамическая задачи посадки ВС и представлен эвристический алгоритм ФООП на одну ВПП, основанный на технике, называемой скоординированным сдвигом позиции. Он включает в себя поиск для небольшого набора ВС наилучших возможных позиций для них в очереди посадки с



ограничением сдвига, так что ВС не может быть перемещено больше, чем на заданное число позиций от положения, которое оно имело бы при посадке на основе FCFS [29]. Представлены вычислительные результаты для трех моделируемых сценариев с участием 500 ВС. Ограничение разделения распространяется только на последовательные посадки ВС. Другими словами, если ВС i, j и k приземляются один за другим так, что ВС i приземляется перед ВС j , который приземляется перед ВС k , то неравенства

$$x_j \geq x_i + S_{ij} \text{ и } x_k \geq x_j + S_{jk}$$

гарантированно выполняются, но может быть, что $x_k < x_i + S_{ik}$.

Такая ситуация называется последовательным разделением. Другой случай — когда разделение гарантируется между всеми парами ВС (такая ситуация называется полным разделением). Отметим, что, если неравенство треугольника выполняется:

$$S_{ik} \leq S_{ij} + S_{jk} \quad (\forall i, \forall k \neq i, \forall j \neq i, k),$$

то последовательное разделение достаточно для обеспечения полного разделения.

Применяется также ограничение сдвига в статической задаче ФООП на одну ВПП [25]. Рассмотрена задача посадки группы одинаковых ВС. Вычислительные результаты представлены для трех групп ВС (всего 15 ВС). В дальнейшем рассматривались постановки, где все ВС (потенциально) разные.

Алгоритм поиска по дереву в глубину, основанный на перечислении всех возможных последовательностей ВС представлен в работе [30]. Ветви дерева отбрасываются, когда стоимость частично построенной последовательности превышает известное возможное решение. По вычислительным причинам предложено применять переборный алгоритм в эвристическом варианте со скользящим окном. Последовательность ВС упорядочивается, подпоследовательность ВС длины F (длина скользящего окна) размораживается, и для нее решается подзадача, после этого первое ВС из подпоследовательности замораживается, затем процедура повторяется.

В работе [31] представлена целочисленная постановка с булевыми переменными задачи ФООП на одну ВПП и эвристика на основе генетического алгоритма. Приведены вычислительные результаты по ряду задач, включающих до 20 ВС.

В работе [32] на основе теории массового обслуживания рассмотрена задача назначения приоритетов для ВС, ожидающих захода на посадку.

Специализированный симплекс-алгоритм, который с высокой вычислительной производи-

тельностью оценивает время посадки, основываясь на некоторой частичной информации, представлен в работе [33]. Используется предварительная обработка, предусматривающая сжатие временных окон и частичное упорядочение на основе данных задачи или верхней границы решения. Алгоритм тестировался на данных из OR-библиотеки [34], включающих до 50 ВС и 4 ВПП.

В дальнейших исследованиях учитывалась окружающая обстановка и допускалось определенное ограниченное изменение позиции приземления в изначальной последовательности ВС. Предлагался алгоритм оптимизации последовательности приземления ВС, основанный на методах динамического программирования.

Задача ФООП также рассматривалась как задача календарного планирования работ в цеху. Для ее решения были предприняты значительные усилия по применению математического программирования [18, 35–37]. В статье [38] представлен подход со скользящим горизонтом, который разбивает последовательность ВС на части и решает задачу ФООП отдельно для каждой из этих частей. Предложены способы разбиения заданной последовательности ВС и исследовано их влияние на качество решения.

В результате анализа существующих методов к наиболее гибким к постановке задачи и критериям оптимизации и дающим хорошие результаты следует отнести метод ветвей и границ, в котором для вычисления оценок применяется метод линейного программирования [11, 13]. Метод ветвей и границ сочетает в себе преимущества полного перебора, при этом, на основе различных техник сокращения перебора, позволяет отбрасывать большие ветви вычислений. Далее этот подход рассматривается более подробно.

2.1. Линейная релаксация задачи

Представленная выше постановка показывает, что целочисленное программирование с булевыми переменными достаточно гибко для того, чтобы учесть большое число ограничений, например, допустимое окно времени посадки и минимальное допустимое время между последовательными приземлениями. Для численного решения в работе [11] предлагается линейная релаксация постановки задачи путем замены булевых переменных непрерывными переменными, т. е. вместо булевых переменных, принимающих значения 0 и 1, рассматриваются непрерывные переменные в диапазоне $[0, 1]$. Проблема в том, что в этом случае решение также может оказаться нецелочисленным. Для улучшения решения в непрерывном пространстве вводятся дополнительные ограничения, которые в целочисленном случае избыточны.

2.1.1. Дополнительные ограничения

- $\delta_{ij} \geq (x_j - x_i)/(L_j - E_i)$

Если $(x_j - x_i) > 0$, то ВС j приземляется после ВС i и, следовательно, необходимо обеспечить условие $\delta_{ij} = 1$. В приведенном ограничении член $(L_j - E_i)$ является коэффициентом масштабирования и представляет максимальное значение, которое может принять разность $(x_j - x_i)$.

- Рассмотрим переменные δ_{ij} . Первое ВС приземляется перед $(P - 1)$ другими ВС, второе ВС приземляется перед $(P - 2)$ другими ВС и т. д. Таким образом, сумма всех переменных δ_{ij} должна быть равна $(P - 1) + (P - 2) + \dots + 1$, т. е. $P(P - 1)/2$.

Следовательно,

$$\sum_{i=1}^P \sum_{j=1, j \neq i}^P \delta_{ij} = P(P - 1)/2.$$

- Предположим, ВС j приземляется после ВС i , и $T_i < T_j$. Тогда промежуток между целевыми временами посадки $(T_j - T_i)$ может быть закрыт суммой отклонений от оптимального времени посадки $(\beta_i + \alpha_j)$ (рис. 2). Следовательно, мы имеем, что

$$\delta_{ij} \geq 1 - \frac{\beta_i + \alpha_j}{T_j - T_i}$$

поскольку, если $(\beta_i + \alpha_j)$ недостаточно для перекрытия промежутка, то ВС i должно приземлиться раньше ВС j .

- Предположим, что есть два ВС, таких что $T_i < T_j$ и $(T_j - T_i) < S_{ij}$, т. е. ВС i предпочтительнее приземлиться перед ВС j , но их оптимальные времена прибытия не удовлетворяют ограничению разделения.

Как видно из рис. 2, если ВС i приземляется перед ВС j ($\delta_{ij} = 1$) на одну ВПП, то отклонение должно быть не меньше, чем $S_{ij} - (T_j - T_i)$, а если ВС j приземляется перед ВС i ($\delta_{ji} = 1$), то отклонение должно быть не меньше, чем $(T_j - T_i) + S_{ji}$. Следовательно,

$$\begin{aligned} (\alpha_i + \beta_j) + (\alpha_j + \beta_i) &\geq [S_{ij} - (T_j - T_i)]\delta_{ij} + \\ &+ [(T_j - T_i) + S_{ji}]\delta_{ji} - \max\{[S_{ij} - (T_j - T_i)], \\ &[(T_j - T_i) + S_{ji}]\}(1 - z_{ij}) \quad \forall (i, j) \in U, \\ &\text{если } T_i < T_j \text{ \& } (T_j - T_i) < S_{ij}. \end{aligned}$$

Слагаемое, содержащее $(1 - z_{ij})$, становится неактивным, если ВС i и ВС j приземляются на одну ВПП и $z_{ij} = 1$, т. е. для случая с одной ВПП это слагаемое можно игнорировать.

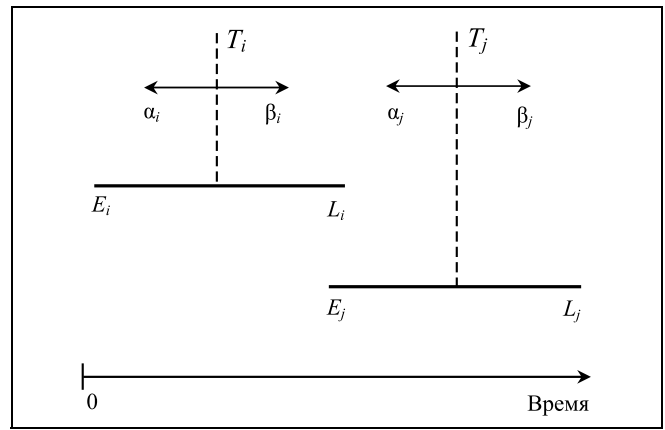


Рис. 2. Пример двух перекрывающихся временных окон посадки

Аналогичное ограничение может быть получено для ВС i и ВС j , если $T_i < T_j$ и $(T_j - T_i) < s_{ij}$.

2.2. Эвристический алгоритм решения задачи

Для решения задачи предложен эвристический алгоритм [11], основанный на нескольких перезапусках по мере улучшения найденного решения. Он заключается в следующем.

Шаг 1. Применить эвристику для создания верхней границы Z_{UB} .

Шаг 2. Воспользоваться значением Z_{UB} для сжатия временных окон.

Шаг 3. Решить получившуюся линейную релаксацию задачи:

— если будет найдено улучшенное возможное решение в течение P секунд, то прекратить его поиск, обновить значение Z_{UB} и перейти к шагу 2;

— в противном случае продолжить поиск решения до завершения.

Решение может перезапускаться несколько раз, каждый раз с улучшенным возможным решением.

Для решения смешанной целочисленно-бинарной задачи с помощью линейной релаксации авторы [11] воспользовались программным пакетом CPLEX.

2.2.1. Поиск верхней границы

Если найдена верхняя граница Z_{UB} оптимального решения задачи (значение целевой функции для некоторого известного допустимого решения), то можно сжать временные окна для каждого ВС. Кроме того, как будет описано ниже, верхней границей Z_{UB} можно воспользоваться для сокращения перебора в методе ветвей и границ.

Для нахождения верхней границы решения задачи применяется следующая эвристика.



Шаг 1. Пусть A_r ($r = 1, \dots, R$) — упорядоченное множество ВС, приземляющихся на ВПП r , где первоначально $A_r = \emptyset \forall r$.

Шаг 2. Рассмотрим ВС в порядке возрастания целевого времени T_j , и для каждого такого ВС j :

— вычислим для каждой ВПП r значение

$$B_r = \max[T_j, \max[x_k + S_{kj} | \forall k \in A_r], \\ \max[x_k + s_{kj} | u = 1, \dots, R, u \neq r \forall k \in A_u]];$$

это выражение дает наилучшее (наименьшее) время B_r , в которое ВС j может приземлиться на ВПП r , учитывая более ранние посадки на эту ВПП и более ранние посадки на все другие ВПП;

— пусть γ — ВПП, для которой соответствующее значение $B_\gamma = \min[B_r | r = 1, \dots, R]$; добавим j в множество A_γ и положим $x_j = B_\gamma$.

Шаг 3. Одна из особенностей шагов 1 и 2 заключается в том, что ВС никогда не приземляется до целевого времени, всегда вовремя или после него.

Чтобы улучшить полученное решение, на данном шаге пересчитываются времена посадок с помощью решения задачи линейного программирования. Это может быть сделано потому, что на шагах 1 и 2 сделан выбор всех целочисленных переменных задачи линейного программирования — определены номера ВПП для каждого ВС и порядок посадок ВС на каждую ВПП.

2.2.2. Сжатие временных окон

Если известна любая верхняя граница оптимального решения Z_{UB} , то исходные временные окна можно сжать. Поясним это для случая целевой функции вида:

$$\min \sum_{i=1}^P (g_i \alpha_i + h_i \beta_i).$$

Для ВС i можно ослабить ограничение E_i (см. п. 1.1), воспользовавшись формулой

$$E_i = \max \left[E_i, T_i - \frac{Z_{UB}}{g_i} \right], \quad i = 1, \dots, P.$$

Это можно сделать на основе предположения, что когда все остальные ВС вносят нулевой вклад в целевую функцию, то если ВС i отклонится от целевого времени сильнее, то мы получим решение хуже, чем Z_{UB} .

Аналогично можно ослабить ограничение L_i :

$$L_i = \min \left[L_i, T_i - \frac{Z_{UB}}{h_i} \right], \quad i = 1, \dots, P.$$

2.3. Модель генерации столбцов для задачи построения оптимальной очереди на посадку

Автор работы [13] считает, что рассмотренная линейная релаксация [11] обеспечивает плохую оценку оптимального значения для исходной смешанной целочисленной модели для задачи ФООП на несколько ВПП. Для протестированных примеров наихудшее решение составило около 77 % от оптимального. Для улучшения модели [11] предлагается переформулировать задачу как задачу генерации столбцов [13].

Обозначения: P — число ВС; S — множество допустимых последовательностей посадки, предполагается, что ВС из одной последовательности приземляются на одну ВПП;

$$c_s = \sum_{i=1}^P (g_i \alpha_i^s + h_i \beta_i^s),$$

c_s — стоимость последовательности.

Будем называть столбцом такой вектор a^s — вектор размерности P , в котором

$$a_i^s = \begin{cases} 1, & \text{если ВС } i \text{ принадлежит} \\ & \text{последовательности } s, s \in S, \\ 0 & \text{в противном случае.} \end{cases}$$

Результирующая модель задачи генерации столбцов:

минимизируется целевая функция, равная суммарной стоимости для всех последовательностей из S :

$$\min \sum_{s \in S} c_s z_s,$$

$$\text{где } z_s = \begin{cases} 1, & \text{если последовательность } s \\ & \text{включается в решение,} \\ 0 & \text{в противном случае.} \end{cases}$$

Ограничения: каждое ВС приземляется строго на одну ВПП:

$$\sum_{s \in S} a_i^s z_s = 1 \quad \forall i \in P \quad (1)$$

предел численности ВПП

$$\sum_{s \in S} z_s = R. \quad (2)$$

Поясним постановку задачи на примере, представленном в табл. 2, для трех ВС, которые приземляются на две ВПП. Время разделения (минимальный безопасный интервал) между любыми двумя ВС предполагается равным 10 [13].

Допустимые последовательности должны удовлетворять заданным временным окнам и ограничениям разделения. Например, последовательность $\{2 \rightarrow 1\}$ не является допустимой, поскольку $88 + 10 > 95$ ($E_2 + S_{21} > L_1$), это означает, что второе ВС не может приземлиться перед первым ВС на одну ВПП, и поэтому последовательность $\{2 \rightarrow 1\}$ не включена во множество допустимых последовательностей посадки S .

В табл. 3 представлены допустимые последовательности для примера, соответствующие времена приземления и соответствующие стоимости.

Далее рассматривается задача минимизации стоимости и нахождения оптимальных времен посадок для допустимой последовательности из S . Она может рассматриваться как задача ФООП для одной ВПП. Обозначим P_S множество ВС последовательности S , а U_S — множество упорядоченных пар последовательности множества S . Например, для последовательности 10 (см. табл. 3): $\{1 \rightarrow 2 \rightarrow 3\}$ $P_{10} = \{1, 2, 3\}$ и $U_{10} = \{(1, 3), (1, 2), (3, 2)\}$.

Формулировка задачи: найти

$$\min \sum_{i \in P_S} (g_i \alpha_i + h_i \beta_i)$$

Таблица 2

Пример задачи для трех ВС

ВС	E_i	T_i	L_i	g_i	h_i
1	50	88	95	3	1
2	88	95	105	3	1
3	75	100	120	3	1

Таблица 3

Допустимые последовательности для примера (см. табл. 2)

Последовательность	Допустимые последовательности	Времена приземления	Общая стоимость
1	{1}	{88}	0
2	{2}	{95}	0
3	{3}	{100}	0
4	{1 → 2}	{88 98}	3
5	{1 → 3}	{88 100}	0
6	{3 → 1}	{85 95}	52
7	{2 → 3}	{95 105}	5
8	{3 → 2}	{95 105}	25
9	{1 → 2 → 3}	{88 98 108}	11
10	{1 → 3 → 2}	{85 95 105}	34

при ограничениях:

$$\begin{aligned} E_i &\leq x_i \leq L_i \quad \forall i \in P_S; \\ x_j &\geq x_i + S_{ij} \quad \forall (i, j) \in U_S; \\ \alpha_i &\geq T_i - x_i \quad \forall i \in P_S; \\ 0 &\leq \alpha_i \leq T_i - E_i \quad \forall i \in P_S; \\ \beta_i &\geq x_i - T_i \quad \forall i \in P_S; \\ 0 &\leq \beta_i \leq L_i - T_i \quad \forall i \in P_S; \\ x_i, \alpha_i, \beta_i &\geq 0 \quad \forall i \in P_S. \end{aligned}$$

Это линейная задача, она позволяет находить оптимальное решение для любой заданной последовательности посадки.

Столбец в постановке не содержит информации о порядке посадок. Например, для столбца [1, 1, 1] существуют две допустимые последовательности посадок: последовательность 9: $\{1 \rightarrow 2 \rightarrow 3\}$ со стоимостью 11 и последовательность 10: $\{1 \rightarrow 3 \rightarrow 2\}$ со стоимостью 34 (см. табл. 3). Поскольку нужно минимизировать общую стоимость, то для столбца [1, 1, 1] коэффициент в целевой функции равен минимальной стоимости из всех стоимостей, соответствующих этому столбцу допустимых последовательностей.

Формулируется математическая модель определения минимальной стоимости для столбца. Эта задача оптимизации отличается от задачи минимизации стоимости для заданной последовательности посадок, поскольку порядок посадок не известен.

Вводятся дополнительные ограничения:

$$\delta_{ij} + \delta_{ji} = 1 \quad \forall i, j \in P_a, \quad i \neq j;$$

$$x_j \geq x_i + S_{ij} \delta_{ij} - (L_i - E_j) \delta_{ji} \quad \forall i, j \in P_a, \quad i \neq j,$$

где P_a — множество ВС, имеющих в столбце a .

В результате получается смешанная целочисленная постановка для определения минимальной стоимости, последовательности посадок и времени посадок для столбцов в задаче генерации столбцов (табл. 4).

Имея информацию о допустимых последовательностях посадок и соответствующих стоимостях, задачу ФООП можно сформулировать как задачу генерации столбцов. Соответствующая модель для рассматриваемого примера:

$$\min(0z_1 + 0z_2 + 0z_3 + 3z_4 + 0z_5 + 5z_6 + 11z_7)$$

при ограничениях (1):

$$z_1 + z_4 + z_5 + z_7 = 1;$$

$$z_2 + z_4 + z_6 + z_7 = 1;$$

$$z_3 + z_6 + z_7 = 1,$$



Таблица 4

Полная информация для задачи генерации столбцов

Переменные	c_s	a'_s	Допустимые последовательности	Времена приземления
z_1	0	[1 0 0]	{1}	{88}
z_2	0	[0 1 0]	{2}	{95}
z_3	0	[0 0 1]	{3}	{100}
z_4	3	[1 1 0]	{1 → 2}	{88 98}
z_5	0	[1 0 1]	{1 → 3}	{88 100}
z_6	5	[0 1 1]	{2 → 3}	{95 105}
z_7	11	[1 1 1]	{1 → 2 → 3}	{88 98 108}

и ограничении (2):

$$z_1 + z_2 + z_3 + z_4 + z_5 + z_6 + z_7 = 2.$$

Для рассматриваемого примера оптимальное решение имеет вид: $Z^* = [0, 1, 0, 0, 1, 0, 0]$;

ВВП	Столбец	Последовательность	Стоимость
1	[0 1 0]	{2}	0
2	[1 0 1]	{1 → 3}	0

2.4. Метод branch-and-price

В работе [13] для решения задачи ФООП предлагается применять метод branch-and-price — метод ветвей и границ, дополненный алгоритмом генерации столбцов. Авторы утверждают, что метод branch-and-price известен как эффективный метод при решении задач большой размерности в различных областях, но для решения задачи ФООП ранее не применялся.

2.4.1. Генерация столбцов

Для небольшого примера, включающего три ВС (см. табл. 2) можно перечислить все возможности последовательностей посадок, как показано в табл. 3. Однако для задачи с 50 ВС существует уже

$$\sum_{k=1}^{50} \binom{50}{k} \approx 1,1259 \times 10^{15}$$

допустимых столбцов и перечислять все эти столбцы неэффективно. Решение начинается с ограниченной задачи генерации столбцов (которая называется master problem — МР) с небольшим подмножеством переменных, полученных в процессе инициализации. При этом значения остальных переменных фиксируются равными нулю. Затем проверяется, может ли добавление одной или нескольких переменных улучшить решение. Согласно симплекс алгоритму, это может быть достигнуто путем добавления переменной с отрицательной

приведенной стоимостью. Задача нахождения переменной с отрицательной приведенной стоимостью называется подзадачей (subproblem — SUB). Приведенная стоимость определяется так:

$$\tilde{c}_j = c_j - \pi a \quad \forall j \in S,$$

где c_j — коэффициент стоимости для столбца a , π — это двойственная переменная, соответствующая каждому ограничению линейной системы, представленному в оптимальном базовом решении МР.

Математическая постановка задачи генерации столбцов SUB: найти

$$\min \left(\sum_{i=1}^P (g_i \alpha_i + h_i \beta_i) - \sum_{i=1}^P \pi_i a_i - \lambda \right)$$

при ограничениях:

$$\begin{aligned} a_i E_i &\leq x_i \leq a_i L_i \quad \forall i \in P; \\ \delta_{ij} &\leq a_i \quad \forall i, j \in P, \quad i \neq j, \\ \delta_{ij} &\leq a_j \quad \forall i, j \in P, \quad i \neq j, \\ 1 &\geq \delta_{ij} + \delta_{ji} \geq a_i + a_j - 1 \quad \forall i, j \in P, \quad i \neq j, \\ x_j &\geq x_i + S_{ij} \delta_{ij} - (L_i - E_j) \delta_{ji} - (a_i L_i - a_j E_j) + \\ &\quad + (L_i - E_j)(\delta_{ij} + \delta_{ji}) \quad \forall i, j \in P, \quad i \neq j, \\ \alpha_i &\geq a_i T_i - x_i \quad \forall i \in P; \\ 0 &\leq \alpha_i \leq T_i - E_i \quad \forall i \in P; \\ \beta_i &\geq x_i - a_i T_i \quad \forall i \in P; \\ 0 &\leq \beta_i \leq L_i - T_i \quad \forall i \in P; \\ x_i &= a_i T_i - \alpha_i + \beta_i \quad \forall i \in P, \end{aligned}$$

δ_{ij}, a_i — бинарные переменные $\forall i, j \in P$,

где π_i — двойственная переменная, соответствующая i -му ограничению (1), $i = 1, \dots, P$, λ — двойственная переменная, соответствующая ограничению (2).

$$a_i = \begin{cases} 1, & \text{если ВС } i \text{ присутствует} \\ & \text{в последовательности посадки,} \\ 0 & \text{в противном случае.} \end{cases}$$

Эту постановку можно рассматривать как задачу ФООП с одной ВПП. Если ВС не присутствует в последовательности посадки, то оно не влияет на целевую функцию.

Если в этой постановке значение целевой функции неотрицательно, то задача решена, в противном случае соответствующий столбец добавляется в задачу МР, и задача решается снова.

Общая структура метода генерации столбцов: сгенерировать начальное допустимое решение, например, с помощью эвристического алгоритма, предложенного в работе [11];

повторять:

- решить задачу МР с помощью линейного программирования;
 - найти столбцы с отрицательной приведенной стоимостью;
 - добавить найденные столбцы в МР;
- завершить, когда не существует столбцов с отрицательной приведенной стоимостью.

2.4.2. Метод ветвей и границ

В большинстве случаев решение задачи генерации столбцов с помощью метода линейной релаксации является дробным. Чтобы получить целочисленное решение, применяется известный метод ветвей и границ, в котором нижняя граница для каждой вершины оценивается с помощью метода генерации столбцов (так называемый метод branch-and-price).

Решение начинается с рассмотрения исходной задачи на множестве допустимых значений, которая называется корневой задачей. К корневой задаче применяются процедуры поиска нижней и верхней границ решения. Если эти границы совпадают, то найдено оптимальное решение, и процедура прекращается. Иначе множество допустимых значений делится на два или более подмножеств. Эти подзадачи (для подмножеств) становятся дочерними вершинами корневой задачи. Алгоритм применяется рекурсивно к подзадачам, генерируя дерево подзадач.

Верхняя граница Z_{UB} является значением целевой функции, соответствующим лучшему текущему целочисленному решению. Первое такое решение получается в процессе инициализации с помощью эвристического алгоритма. На каждой итерации для одной подзадачи (вершины) ищется решение Z_{SV} с помощью описанного выше метода генерации столбцов. Возможно три случая.

Случай 1. Если получено целочисленное решение Z_{SV} , оно сравнивается с текущей верхней границей решения Z_{UB} . Если $Z_{SV} < Z_{UB}$, то эта вершина становится лучшей целочисленной вершиной, и верхняя граница решения задачи обновляется: $Z_{UB} = Z_{SV}$, а нижняя граница каждой активной вершины дерева сравнивается с этой новой верхней границей. Те вершины, нижняя граница Z_{SV} которых больше, чем Z_{UB} , удаляются из дерева.

Случай 2. Если решение дробное и значение $Z_{SV} \geq Z_{UB}$, то вершина удаляется из дерева, так как целочисленные решения ее потомков не будут лучше полученного дробного решения.

Случай 3. Если решение дробное и значение $Z_{SV} < Z_{UB}$, то принимается решение о создании двух дочерних узлов.

Решение продолжается до тех пор, пока для всех порожденных вершин либо получено решение, либо вершины исключены из рассмотрения. Реализация операции ветвления подробно описана в работе [13].

Алгоритм протестирован на открытых данных из OR-библиотеки, включающих до 50 ВС и 4 ВПП. Авторы утверждают, что все задачи были решены с использованием менее 450 столбцов. Это указывает на потенциальную возможность применения предложенного алгоритма для решения задачи ФООП с фиксированным числом ВПП.

Процедура построения оптимальных очередей на посадку на основе метода ветвей и границ была реализована в составе комплексного исследовательского стенда полунатурного имитационного моделирования интегрированных систем управления воздушным движением, разрабатываемом в ГосНИИАС [2]. Особенность предложенного метода состоит в учете ограничений не только на минимальные интервалы между посадками ВС, но и минимальные интервалы между пролетом некоторых точек воздушного пространства.

ЗАКЛЮЧЕНИЕ

В точных алгоритмах применяется метод ветвей и границ для поиска оптимального решения для задачи ФООП. Однако время работы экспоненциально растет с размером задачи (числом ВС). Следовательно, его нельзя применить для оптимального решения задач большой размерности в течение приемлемого времени. Из литературы известно, что задача ФООП решается оптимально до 50 ВС.

Для получения приближенных решений задачи ФООП в литературе представлено большое число эвристических, метаэвристических, гибридных и прочих алгоритмов, в том числе с приведенными вычислительными результатами для 500 ВС и 5 ВПП, полученными за приемлемое время.

Перспективным методом построения приближенных решений задачи ФООП на одну или несколько ВПП посвящена вторая часть обзора.

ЛИТЕРАТУРА

1. Кумков С.И., Пятко С.Г. Задача обнаружения и разрешения конфликтных ситуаций в автоматизированной системе управления воздушным движением // Научный вестник ГосНИИ «Аэронавигация». — 2013. — № 12. — С. 35–46.
2. Дегтярев О.В., Сикачев В.Ю., Лященко А.В. Оптимальное управление прибывающим на аэродром потоком воздушных судов // Тр. девятого междунар. симпозиума / ГосНИИАС. — М., 2010. — С. 430–435.
3. Алешин А.В., Алешин В.И., Бабаев Н.В., Крыжановский Г.А. Методика имитационного моделирования процессов управления потоками прилетающих и вылетающих воздушных судов при УВД в районе аэродрома // Научный вестник МГТУ ГА. — 2011. — Т. 171, № 9. — С. 121–128.



4. Алешин А.В. Алгоритмы формирования очередности захода на посадку при маневрировании в зоне ожидания // Вестник Санкт-Петербургского гос. ун-та гражданской авиации. — 2013. — Т. 1 (5). — С. 73–84.
5. Лебедев Г.Н., Малыгин В.Б., Нечаев Е.Е., Тин Пхон Чжо. Использование системы приоритетного обслуживания при внедрении автоматизированного управления прилетом-вылетом в воздушном пространстве московского аэроузла // Научный вестник МГТУ ГА. — 2012. — Т. 180. — № 6. — С. 90–98.
6. Казаков В.А. Планирование и организация потоков воздушного движения: учеб. пособ. — Ульяновск: УВАУ ГА, 2008. — 89 с.
7. Стиюнов М.В. Управление воздушным движением в районе аэродрома: учеб. пособ. — Ульяновск: УВАУ ГА(И), 2011. — 103 с.
8. Методы и алгоритмы планирования и регулирования потоков воздушного движения с учетом требований по безопасности и эффективности выполнения полетов и различных систем регулирования / Аналитический обзор. — М.: ГосНИИАС, 2016.
9. Harikopoulou D., Neogi N. Polynomial-Time Feasibility Condition for Multiclass Aircraft Sequencing on a Single-Runway Airport // IEEE Trans. on Intelligent Transportation Systems. — March 2011. — Vol. 12, N 1. — P. 2–14.
10. Bennel J.A., Mesgarpour M., and Potts C.N. Airport runway scheduling, 4OR-Quarterly // Journal of Operations Research. — 2011. — Vol. 9. — P. 115–138.
11. Beasley J.E., Krishnamoorthy M., Sharaiha Y.M., and Abramson D. Scheduling aircraft landings — the static case // Transportation Science. — 2000. — Vol. 34, N 2. — P. 180–197.
12. Beasley J.E., Krishnamoorthy M., Sharaiha Y.M., and Abramson D. Displacement problem and dynamically scheduling aircraft landings // Journal of the Operational Research Society. — 2004. — Vol. 55, N 1. — P. 54–64.
13. Wen M. Algorithms of Scheduling Aircraft Landing Problem. — Technical University of Denmark, 2005.
14. Xiao-Peng Ji, Xian-Bin Cao, and Ke Tang. Sequence searching and evaluation: a unified approach for aircraft arrival sequencing and scheduling problems // Memetic Computing. — 2016. — Vol. 8, N 2. — P. 109–123.
15. Beasley J.E., Sonander J., and Havelock P. Scheduling aircraft landings at London Heathrow using a population heuristic // Journal of the Operational Research Society. — 2001. — Vol. 52, N 5. — P. 483–493.
16. Yu S.P., Cao X.B., and Zhang J. A real-time schedule method for Aircraft Landing Scheduling problem based on Cellular Automation // Applied Soft Computing. — 2011. — Vol. 11, N 4. — P. 3485–3493.
17. Soomer M.J., Franx G.J. Scheduling aircraft landings using airlines' preferences // European Journal of Operational Research. — 2008. — Vol. 190, N 1. — P. 277–291.
18. Faye A. Solving the Aircraft Landing Problem with time discretization approach // European Journal of Operational Research. — 2015. — Vol. 242, N 3. — P. 1028–1038.
19. Erzberger H. Design principles and algorithms for automated air traffic management. / In: Knowledge-based Functions in Aerospace Systems. — 1995. — P. 7:1–7:31.
20. Hu X.B., Chen W.H. Genetic Algorithm Based on Receding Horizon Control for Arrival Sequencing and Scheduling // Engineering Applications of Artificial Intelligence. — 2005. — Vol. 18, N 5. — P. 633–642.
21. Hu X.B., Chen W.H. Receding Horizon Control for Aircraft Arrival Sequencing // IEEE Trans. on Intelligent Transportation Systems. — 2005. — Vol. 6, N 2. — P. 189–197.
22. Hu X.B., Di Paolo E. Binary-representation-based genetic algorithm for aircraft arrival sequencing and scheduling // IEEE Trans. on Intelligent Transportation Systems. — 2008. — Vol. 9, N 2. — P. 301–310.
23. Zhan Z.H., Zhang J., Li Y., et al. An efficient ant colony system based on receding horizon control for the aircraft arrival sequencing and scheduling problem // IEEE Trans. on Intelligent Transportation Systems. — 2010. — Vol. 11, N 2. — P. 399–412.
24. Tang K., Wang Z., Cao X., and Zhang J. A multi-objective evolutionary approach to aircraft landing scheduling problems // IEEE World Congress on Computational Intelligence, 2008. — P. 3650–3656.
25. Psaraftis H.N. A Dynamic Programming Approach for Sequencing Groups of Identical Jobs // Operations Research. — 1980. — Vol. 23, N 6. — P. 1347–1359.
26. Andreucci A., Bianco L., and Ricciardelli S. A simulation model for aircraft sequencing in the near terminal area // European Journal of Operational Research. — 1981. — Vol. 8, N 4. — P. 345–354.
27. Dear R.G., Sherif Y.S. The Dynamic Scheduling of Aircraft in High Density Terminal Areas // Microelectronics Reliability. — 1989. — Vol. 29, N 5. — P. 743749.
28. Dear R.G., Sherif Y.S. An Algorithm for Computer Assisted Sequencing and Scheduling of Terminal Area Operations // Transportation Research Part A: General. — 1991. — Vol. 25, N 2-3. — P. 129–139.
29. Dear R.G. The Dynamic Scheduling of Aircraft in the Near Terminal Area / M.I.T., Flight Transportation Laboratory report R76-9, 1976.
30. Brinton C.R. An Implicit Enumeration Algorithm for Arrival Aircraft Scheduling // Proc. of the IEEE/AIAA 11th Digital Avionics Systems Conference. Seattle, 1992. — P. 268–274.
31. Abela J., Abramson D., Krishnamoorthy M., et al. Computing Optimal Schedules for Landing Aircraft // Proc. of the 12th national conference of the Australian Society for Operations Research. Adelaide, 1993. — P. 71–90.
32. Milan J. The Flow Management Problem in Air Traffic Control: A Model of Assigning Priorities for Landings at a Congested Airport // Transportation Planning and Technology. — 1997. — Vol. 20, N 2. — P. 131–162.
33. Ernst A.T., Krishnamoorthy M., and Storer T.H. Heuristic and Exact Algorithms for Scheduling Aircraft Landings // Networks. — 1999. — Vol. 34, N 3. — P. 229–241.
34. Beasley J.E. OR-Library: distributing test problems by electronic mail // Journal of the Operational Research Society. — 1990. — Vol. 41, N 11. — P. 1069–1072.
35. Tavakkoli-Moghaddam R. Scheduling the sequence of aircraft landings for a single runway using a fuzzy programming approach // Journal of Air Transport Management. — 2012. — Vol. 25. — P. 15–18.
36. Briskorn D., Stolletz R. Aircraft landing problems with aircraft classes // Journal of Scheduling. — 2014. — Vol. 17, N 1. — P. 31–45.
37. Lieder A., Briskorn D., and Stolletz R. A dynamic programming approach for the aircraft landing problem with aircraft classes // European Journal of Operational Research. — 2015. — Vol. 243, N 1. — P. 61–69.
38. Furini F., Kidd M.P., Persiani C.A., and Toth P. Improved rolling horizon approaches to the aircraft sequencing // Journal of Scheduling. — 2015. — Vol. 18, N 5. — P. 435–447.

Статья представлена к публикации членом редколлегии А.А. Лазаревым.

Вересников Георгий Сергеевич — канд. техн. наук, ст. науч. сотрудник, ✉ veresnikov@mail.ru,

Кулида Елена Львовна — канд. техн. наук, ст. науч. сотрудник, ✉ lenak@ipu.ru,

Егоров Николай Александрович — канд. техн. наук, ст. науч. сотрудник, ✉ egorov@ipu.ru,

Лебедев Валентин Григорьевич — д-р техн. наук, уч. секретарь, ✉ lebedev@ipu.ru,

Институт проблем управления им. В.А. Трапезникова РАН, г. Москва.

ЗАДАЧА ОПТИМАЛЬНОГО БЫСТРОДЕЙСТВИЯ ПРИ УПРУГОМ И ВЯЗКОУПРУГОМ ВЗАИМОДЕЙСТВИЯХ ТЕЛА С ПОВЕРХНОСТЬЮ¹

А.А. Галяев, П.В. Лысенко

Рассмотрены две задачи оптимального по быстродействию управления механической системой, состоящей из материальной точки и препятствия, взаимодействующих между собой посредством пружины, обладающей упругими или вязкоупругими свойствами. Временной интервал взаимодействия определяется физическими условиями контакта. Исследованы зависимости времени окончания взаимодействия и коэффициента восстановления от вязкоупругих свойств пружины. Разработан программный модуль на языке Python и проведено моделирование поведения системы «материальная точка — препятствие».

Ключевые слова: механическая система с ударами, среда Кельвина — Фойгта, управление по быстродействию.

ВВЕДЕНИЕ

Моделирование ударов в механической системе основано на представлении, в соответствии с которым траектория системы по достижении в фазовом пространстве ограничения в общем случае мгновенно «перескакивает» в другую точку с теми же пространственными координатами, но с другими значениями проекций скоростей. Эти новые значения проекций скоростей задаются при помощи ньютоновского коэффициента восстановления, в то время как сам процесс удара (взаимодействия между телами) оказывается скрыт в этом коэффициенте, поскольку длительность фазы удара полагается равной нулю. Управление в таких системах входит только в безударную фазу движения. В настоящей работе рассматривается конечная (ненулевая) длительность фазы удара (взаимодействия с поверхностью препятствия). Решение уравнений движения системы в фазе взаимодействия позволяет связать послеударные макроскопические характеристики системы (скорость, полную энергию и др.) с доударными [1–8], причем взаимодействие может описываться как линейными по фазо-

вым координатам уравнениями (например, для среды Кельвина — Фойгта [6]), так и нелинейными (например, теория удара Герца [7]). Другая отличительная особенность настоящей работы заключается во введении в фазу взаимодействия управляющего воздействия [4, 5]. При решении задачи применяется метод пространственно-временной замены координат, описанной в работе [8].

Решаются две задачи быстродействия при одностороннем упругом и вязкоупругом взаимодействии материальной точки с управляемой поверхностью. Условия окончания взаимодействия определяются через силу, действующую со стороны поверхности препятствия на тело. В литературе такие условия окончания взаимодействия называются физическими [2].

1. ЗАДАЧА БЫСТРОДЕЙСТВИЯ ПРИ УПРУГОМ ВЗАИМОДЕЙСТВИИ ТЕЛА С ПОВЕРХНОСТЬЮ

Постановка задачи оптимального управления. Рассмотрим динамику физической системы, состоящей из поверхности, управляемой по скорости, и элемента, включающего в себя пружину и тело конечной массы. Элемент, состоящий из тела и пружины, является единым целым. Поверхность и элемент испытывают взаимодействие при наличии упругой силы, которая пропорциональна сжа-

¹ Работа выполнена при частичной финансовой поддержке РФФИ (проект № 16-08-01285-а).

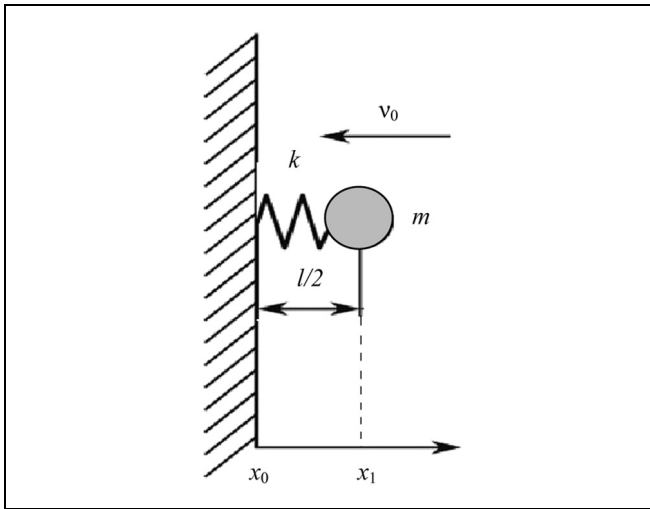


Рис. 1. Состояние системы в начальный момент времени

тию изначально недеформированной пружины. Взаимодействие тела с поверхностью происходит посредством одностороннего контакта безмассовой пружины. Данная модель описывает фазу удара и дает закон восстановления при стремлении жесткости пружины к бесконечности. Динамика системы (рис. 1) описывается уравнениями движения [1]

$$\begin{cases} \dot{x}_0(t) = u(t), \\ \ddot{x}_1(t) = -\omega_0^2(x_1 - x_0 - l/2). \end{cases} \quad (1)$$

Здесь x_0 — координата поверхности; x_1 — координата элемента, состоящего из тела массой m с прикрепленной к нему пружины длиной $l/2$ и жесткости k , $\omega_0 = \sqrt{k/m}$ — круговая частота. Без потери общности считаем, что $\omega_0 = 1$. Ось x направлена от препятствия к телу. Далее предположим, что на управление наложено ограничение $u \leq |u_0|$. Краевые условия, описывающие начальный и финальный моменты взаимодействия тела с препятствием, имеют вид:

$$\begin{cases} x_0(0) = 0, \\ x_1(0) = l/2, \\ \dot{x}_1(0) = -v_0, \\ x_1(\tau) - x_0(\tau) = l/2, \end{cases} \quad (2)$$

где τ — длительность взаимодействия стенки с пружиной, которую выберем в качестве критерия рассматриваемой задачи оптимального управления. По окончании взаимодействия потребуем, чтобы

$$\tau \rightarrow \min_u. \quad (3)$$

Получили классическую задачу оптимального управления, когда нужно найти минимум критерия (3) на траекториях системы (1) при краевых ограничениях вида (2) и ограничении на управление $u \leq |u_0|$.

Решение задачи. Приведем систему (1) к нормальной форме Коши, совершив замену координат $\dot{x}_1 = x_2$:

$$\begin{cases} \dot{x}_0 = u, \\ \dot{x}_1 = x_2, \\ \dot{x}_2 = -x_1 + x_0 + l/2. \end{cases} \quad (4)$$

Воспользуемся для решения задачи (1)–(4) принципом максимума Понтрягина [9]. Функция Понтрягина для задачи быстрогодействия

$$\begin{aligned} H(x_0, x_1, x_2, \psi_1, \psi_2, \psi_3, u) = \\ = \psi_1 u + \psi_2 x_2 + \psi_3(-x_1 + x_0 + l/2) - 1. \end{aligned}$$

Сопряженная система принимает вид:

$$\begin{cases} \dot{\psi}_1 = -\psi_3, \\ \dot{\psi}_2 = \psi_3, \\ \dot{\psi}_3 = -\psi_2. \end{cases} \quad (5)$$

Из системы (5) получаем выражения для сопряженных функций

$$\begin{aligned} \psi_3 &= A \sin(t - \alpha_0), \\ \psi_1 &= A \cos(t - \alpha_0) + B, \end{aligned}$$

где A, B, α_0 — некоторые константы. Условие максимума функции H по u означает, что на оптимальных траекториях (x_0^*, x_1^*, x_2^*) выполнено

$$\begin{aligned} H(x_0^*, x_1^*, x_2^*, \psi_1, \psi_2, \psi_3, u^*) = \\ = \max_{u \leq |u_0|} H(x_0^*, x_1^*, x_2^*, \psi_1, \psi_2, \psi_3, u), \end{aligned}$$

т. е. оптимальное управление u^* максимизирует функцию Понтрягина. Из вида функции H в нашем случае следует, что

$$u^* = u_0 \text{sign}(\psi_1) = u_0 \text{sign}(A \cos(t - \alpha_0) + B).$$

Откуда получаем управление релейного типа, которое переключается в моменты времени t_i , удовлетворяющие уравнению

$$A \cos(t_i - \alpha_0) + B = 0.$$

Теперь возникает вопрос выбора знака управления в начальный момент времени. Поскольку решается задача минимизации времени взаимодействия тела с препятствием, то выберем $u(0) = u_0 > 0$. Выпишем уравнения динамики системы до и после

ле первого момента переключения управления. Первый отрезок времени длительностью τ_1 с управлением $u = u_0$ определяет движение препятствия и динамику взаимодействия тела с препятствием:

$$\begin{cases} x_0(t) = u_0 t, \\ \ddot{x}_1(t) = -(x_1 - x_0 - l/2). \end{cases} \quad (6)$$

Проинтегрировав второе из уравнений (6), получим его общее решение в виде

$$x_1 = C \sin(t - \beta_0) + u_0 t + l/2.$$

Подставляем сюда краевые условия (2) и получаем, что

$$x_1(0) = \frac{l}{2} = C \sin(-\beta_0) + \frac{l}{2} \beta_0 = 0,$$

$$\dot{x}_1(0) = -v_0 = C \cos(0) + u_0 C = -(u_0 + v_0).$$

Таким образом, координаты препятствия и тела в момент первого переключения управления $t_1 = \tau_1$:

$$\begin{cases} x_0(t_1) = u_0 t_1, \\ x_1(t_1) = -(u_0 + v_0) \sin(t_1) + u_0 t_1 + l/2. \end{cases} \quad (7)$$

Второй отрезок времени длительностью τ_2 с управлением $u = -u_0$ определяет динамику системы вплоть до момента окончания взаимодействия, а именно:

$$\begin{cases} x_0(s) = u_0 \tau_1 - u_0 s, \\ \ddot{x}_1(s) = -(x_1 - x_0 - l/2) \end{cases} \quad (8)$$

во времени $s = t - t_1$, которое отсчитывается от момента $t_1 = \tau_1$. Проинтегрировав второе из уравнений (8), получим общий вид решения с неизвестными константами C_1 и β_1 :

$$x_1(s) = C_1 \sin(s - \beta_1) + u_0 \tau_1 - u_0 s + l/2.$$

Из условия неразрывности координат $x(0) = x(t_1)$, $\dot{x}(0) = \dot{x}(t_1)$ в выражении (7) находим соотношения между константами в виде:

$$\begin{aligned} C_1 \sin(\beta_1) &= (u_0 + v_0) \sin \tau_1, \\ C_1 \cos(\beta_1) &= 2u_0 - (u_0 + v_0) \cos \tau_1. \end{aligned} \quad (9)$$

Конечный момент взаимодействия из краевых условий (2) задает соотношение между интервалом τ_2 и константой β_1 :

$$\begin{aligned} C_1 \sin(\tau_2 - \beta_1) + u_0 \tau_1 - u_0 \tau_2 + \frac{l}{2} - (u_0 \tau_1 - u_0 \tau_2) &= \\ = \frac{l}{2} \tau_2 = \beta_1. \end{aligned} \quad (10)$$

Из соотношений (9) и (10) следует, что

$$\operatorname{tg} \tau_2 = \frac{(u_0 + v_0) \sin \tau_1}{2u_0 - (u_0 + v_0) \cos \tau_1}.$$

Общее время взаимодействия

$$\tau = \tau_1 + \tau_2 = \tau_1 + \operatorname{arctg} \frac{(u_0 + v_0) \sin \tau_1}{2u_0 - (u_0 + v_0) \cos \tau_1}. \quad (11)$$

Аналитическое исследование функции $\tau(\tau_1)$ (11) в пакете Maple показывает, что она достигает минимума в первой точке перелома — нуле знаменателя тангенса, а именно, при

$$2u_0 - (u_0 + v_0) \cos \tau_1 = 0.$$

Следовательно, длительности интервалов управления

$$\tau_1 = \arccos \frac{2u_0}{u_0 + v_0}, \quad \tau_2 = \frac{\pi}{2}. \quad (12)$$

Выражение (12) означает, что оптимальная длительность второго интервала управления при любых, имеющих физический смысл, u_0 и v_0 равна четверти периода.

Продemonстрируем полученный результат несколькими примерами.

Приведем соответствующие (рис. 2) графику оптимальные значения интервалов τ_1 и τ , получен-

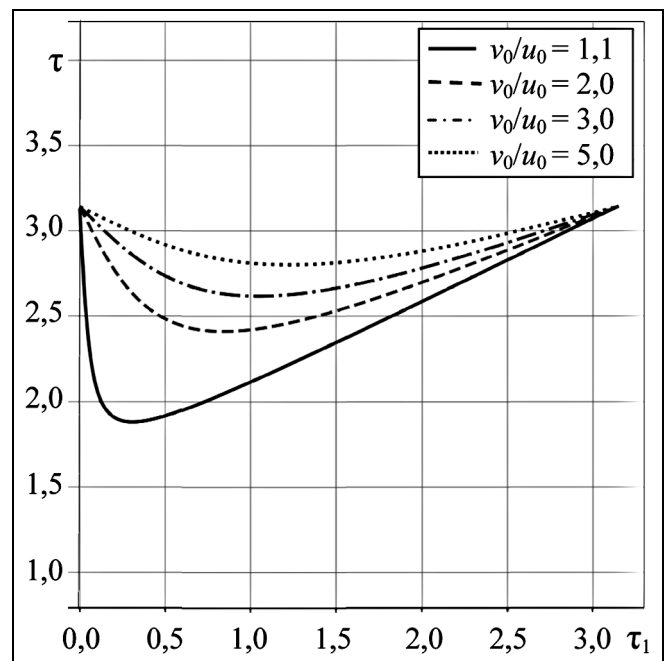


Рис. 2. Зависимость времени взаимодействия поверхности с телом от длительности первого интервала управления для разных соотношений начальной скорости тела к уровню управления



ные численным анализом функций $\tau(\tau_1)$ при различных отношениях v_0/u_0 :

$$\begin{aligned}\tau_1 &= 0,31, \quad \tau_2 = 1,57, \quad \tau = 1,88, \\ \tau_1 &= 0,64, \quad \tau_2 = 1,57, \quad \tau = 2,21, \\ \tau_1 &= 0,84, \quad \tau_2 = 1,57, \quad \tau = 2,41, \\ \tau_1 &= 1,05, \quad \tau_2 = 1,57, \quad \tau = 2,62.\end{aligned}$$

Численный расчет показывает, что интервал τ_2 с любой точностью всегда равен $\pi/2$. Также численно проверено, что функция $\tau(\tau_1)$ имеет минимум в точке, удовлетворяющей условию (12). Все численные расчеты выполнены средствами языка Python.

2. ЗАДАЧА БЫСТРОДЕЙСТВИЯ ПРИ ВЯЗКОУПРУГОМ ВЗАИМОДЕЙСТВИИ ТЕЛА С ПОВЕРХНОСТЬЮ

Постановка задачи оптимального управления. Положим теперь, что в рассмотренной выше системе взаимодействие препятствия с телом определяются не только упругой силой пружины, но еще и вязкой силой Кельвина — Фойгта, пропорциональной относительной скорости тела. Уравнения динамики системы (1) принимают вид

$$\begin{cases} \dot{x}_0(t) = u(t), \\ \ddot{x}_1(t) = -(x_1 - x_0 - l/2) - 2\alpha(\dot{x}_1 - \dot{x}_0) \end{cases} \quad (13)$$

Снова на управление наложено ограничение $u \leq |u_0|$. Коэффициент вязкости находится в интервале $\alpha \in [0, 1]$. Краевые условия имеют вид:

$$\begin{cases} x_0(0) = 0, \\ x_1(0) = l/2, \\ \dot{x}_1(0) = -v_0, \\ \ddot{x}_1(\tau) = 0, \end{cases} \quad (14)$$

где τ — длительность взаимодействия препятствия с пружиной, которую выберем в качестве критерия задачи. Итак, необходимо найти

$$\tau \rightarrow \min_u,$$

на решениях системы (13) и краевых условиях (14).

Решение задачи проводим аналогично решению задачи быстрогодействия в § 1. Приводим систему (13) к нормальной форме Коши, совершив замену координат $\dot{x}_1 = x_2$, получаем

$$\begin{cases} \dot{x}_0 = u, \\ \dot{x}_1 = x_2, \\ \dot{x}_2 = -x_1 + x_0 + l/2 - 2\alpha(x_2 - u). \end{cases} \quad (15)$$

Функция Понтрягина для задачи быстрогодействия (13)—(15)

$$H = \psi_1 u + \psi_2 x_2 + \psi_3 (-x_1 + x_0 + l/2 - 2\alpha(x_2 - u)) - 1.$$

Тогда сопряженная система имеет вид

$$\begin{cases} \dot{\psi}_1 = -\psi_3, \\ \dot{\psi}_2 = \psi_3, \\ \dot{\psi}_3 = -\psi_2 + 2\alpha\psi_3. \end{cases} \quad (16)$$

Подставив второе уравнение системы (16) в третье, получим

$$\ddot{\psi}_2 + 2\alpha\dot{\psi}_2 + \psi_2 = 0.$$

Характеристическое уравнение $\lambda^2 - 2\alpha\lambda + 1 = 0$

имеет корни $\lambda_{1,2} = \alpha \pm i\sqrt{1-\alpha^2}$, поэтому сопряженные функции представимы в явном виде:

$$\begin{cases} \psi_1 = -\psi_2 + C, \\ \psi_2 = Ae^{\alpha t} \sin(\omega t + \phi_0), \\ \psi_3 = A\alpha e^{\alpha t} \sin(\omega t + \phi_0) + A\omega e^{\alpha t} \cos(\omega t + \phi_0), \end{cases}$$

где $\omega = \sqrt{1-\alpha^2}$, а A , C и ϕ_0 — некоторые константы.

Согласно рассуждениям, аналогичным рассуждениям в § 1, из вида функции Понтрягина следует, что

$$\begin{aligned} u^*(t) &= u_0 \text{sign}(\psi_1 + 2\alpha\psi_3) = \\ &= u_0 \text{sign}(-Ae^{\alpha t} \sin(\omega t + \phi_0) + 2\alpha(A\alpha e^{\alpha t} \sin(\omega t + \phi_0) + \\ &\quad + A\omega e^{\alpha t} \cos(\omega t + \phi_0)) + C). \end{aligned}$$

Далее, выпишем уравнения, характеризующие поведение системы, до и после первого переключения. На первом отрезке времени длительностью τ_1 с управлением $u = u_0$ уравнения (13) принимают вид:

$$\begin{cases} \dot{x}_0(t) = u_0, \\ \ddot{x}_1(t) + 2\alpha\dot{x}_1 + x_1 = x_0 + l/2 + 2\alpha\dot{x}_0. \end{cases} \quad (17)$$

Подставив первое из уравнений (17) во второе, получим

$$\ddot{x}_1(t) + 2\alpha\dot{x}_1 + x_1 = u_0 t + l/2 + 2\alpha u_0.$$

Решение уравнения имеет вид:

$$x_1(t) = Ae^{-\alpha t} \sin(\omega t + \phi_0) + u_0 t + l/2.$$

Подставив начальные условия из краевых условий (14), получим

$$x_1(0) = l/2 = A \sin(\phi_0) + (l/2)\phi_0 = 0,$$

$$\dot{x}_1(0) = -v_0 = A\omega + u_0 A = -\frac{v_0 + u_0}{\omega}.$$

Таким образом, координаты и скорость тела в момент первого переключения управления $t_1 = \tau_1$

$$\begin{cases} x_1(t_1) = -\frac{v_0 + u_0}{\omega} e^{-\alpha t_1} \sin(\omega t_1) + u_0 t_1 + l/2, \\ \dot{x}_1(t_1) = -\frac{v_0 + u_0}{\omega} (-\alpha e^{-\alpha t_1} \sin(\omega t_1) + \\ + \omega e^{-\alpha t_1} \cos(\omega t_1)) + u_0. \end{cases} \quad (18)$$

На втором отрезке времени длительностью τ_2 с управлением $u = -u_0$ уравнения (13) принимают вид:

$$\begin{cases} \dot{x}_0(s) = u_0 \tau_1 + u_0 s, \\ \ddot{x}_1(s) + 2\alpha \dot{x}_1 + x_1 = x_0 + l/2 + 2\alpha \dot{x}_0. \end{cases} \quad (19)$$

Время $s = t - t_1$ здесь отсчитывается от момента времени $t = \tau_1$. Для уравнения

$$\ddot{x}_1(s) + 2\alpha \dot{x}_1 + x_1 = u_0 \tau_1 - u_0 s + l/2 - 2\alpha u_0,$$

аналогично рассуждениям выше, находим решение системы уравнений (19)

$$x_1(s) = B e^{-\alpha s} \sin(\omega s - \phi_1) + u_0 \tau_1 - u_0 s + l/2,$$

где B и ϕ_1 — некоторые константы, которые могут быть найдены из краевых условий. Из условия непрерывности координат $x(0) = x(t_1)$, $\dot{x}(0) = \dot{x}(t_1)$ в выражениях (18), получаем

$$\begin{cases} -B \sin \phi_1 = -\frac{v_0 + u_0}{\omega} e^{-\alpha \tau_1} \sin(\omega \tau_1), \\ B(\omega \cos \phi_1 + \alpha \sin \phi_1) - u_0 = \\ = -\frac{v_0 + u_0}{\omega} (-\alpha e^{-\alpha \tau_1} \sin(\omega \tau_1) + \omega e^{-\alpha \tau_1} \cos(\omega \tau_1)) + u_0. \end{cases} \quad (20)$$

Условие окончания взаимодействия из краевых условий (14) принимают вид:

$$(2\alpha^2 - 1) \sin(\omega \tau_2 - \phi_1) - 2\alpha \omega \cos(\omega \tau_2 - \phi_1) = 0. \quad (21)$$

Объединив выражения (20) и (21), запишем систему, позволяющую найти $\tau_2(\tau_1)$:

$$\begin{cases} B \sin \phi_1 = -A e^{-\alpha \tau_1} \sin(\omega \tau_1), \\ B \cos \phi_1 = \frac{2u_0}{\omega} + A e^{-\alpha \tau_1} \cos(\omega \tau_1), \\ (2\alpha^2 - 1) \sin(\omega \tau_2 + \phi_1) - 2\alpha \omega \cos(\omega \tau_2 - \phi_1) = 0. \end{cases} \quad (22)$$

Разделив первое из уравнений (22) на второе и сделав преобразования в третьем уравнении, можно перейти к более компактному виду:

$$\begin{cases} \operatorname{tg} \phi_1 = \frac{A \omega e^{-\alpha \tau_1} \sin(\omega \tau_1)}{2u_0 + A \omega e^{-\alpha \tau_1} \cos(\omega \tau_1)}, \\ \operatorname{tg}(\omega \tau_2 - \phi_1) = \frac{2\alpha \omega}{2\alpha^2 - 1}. \end{cases} \quad (23)$$

Легко удостовериться в том, что при подстановке $\alpha = 0$ уравнения (23) переходят в выражения (12) в предыдущей задаче.

Продemonстрируем данный результат несколькими примерами. Приведем численные результаты для первого случая ($v_0/u_0 = 1,1$):

$$\alpha = 0,0: \quad \tau_1 = 0,31, \quad \tau_2 = 1,57, \quad \tau = 1,88,$$

$$\alpha = 0,1: \quad \tau_1 = 0,31, \quad \tau_2 = 1,29, \quad \tau = 1,6,$$

$$\alpha = 0,5: \quad \tau_1 = 0,33, \quad \tau_2 = 0,49, \quad \tau = 0,82,$$

$$\alpha = 0,8: \quad \tau_1 = 0,34, \quad \tau_2 = 0,08, \quad \tau = 0,42.$$

Откуда и из рис. 3 следует, что с увеличением α оптимальное время уменьшается, также как и длительность второго интервала взаимодействия, при этом длительность первого интервала управления меняется незначительно в широком диапазоне изменения α . Это связано с тем, что с увеличением α все больше начальной кинетической энергии тела рассеивается в течение первого интервала. Построим зависимость оптимального времени от α при разных отношениях v_0/u_0 .

Рис. 4 показывает, что длительность интервала взаимодействия тела с поверхностью возрастает при увеличении отношения начальной скорости тела к уровню управления, и убывает с увеличением коэффициента вязкости α .

Особый интерес представляет коэффициент восстановления v_f/v_0 (отношение послеударной скорости тела к доударной) как характеристика взаимодействия тела с поверхностью. Послеударная скорость получается вследствие решения уравнений динамики взаимодействия тела с поверхностью в момент их контакта. Видим, что при различных отношениях начальной скорости тела к уровню управления возможно получить одинаковые коэффициенты восстановления при одних α . На рис. 5 видно, что для фиксированных α и u_0 есть максимум коэффициента восстановления по начальной скорости тела v_0 .

В полученные выражения оптимальных значений τ жесткость k и круговая частота ω_0 не входят,

поскольку изначально было высказано предположение, что $\omega_0 = 1$. В случае когда $\omega_0 = \sqrt{k/m}$, то замена времени $t \rightarrow t/\omega_0$ и ограничения $u_0 \rightarrow u_0/\omega_0$ приводит к тому, что оптимальные значения τ , τ_1 и τ_2 , полученные при решении задач быстрогодействия, следует разделить на ω_0 . Тогда при $k \rightarrow \infty$ эти временные интервалы стремятся к нулю. Однако структура решений задач быстрогодействия сохраня-

ется, временной интервал, на котором действует оптимальное управление, состоит из двух интервалов: на первом управление действует навстречу начальной скорости тела, на втором — в противоположную. Терминальное значение скорости не зависит от жесткости, а значит и от предельного перехода при $k \rightarrow \infty$. Тем самым в зависимости от вязкоупругих свойств пружины, начальной скорости тела и ограничения управления однозначным образом по решению системы (1) или (13) опреде-

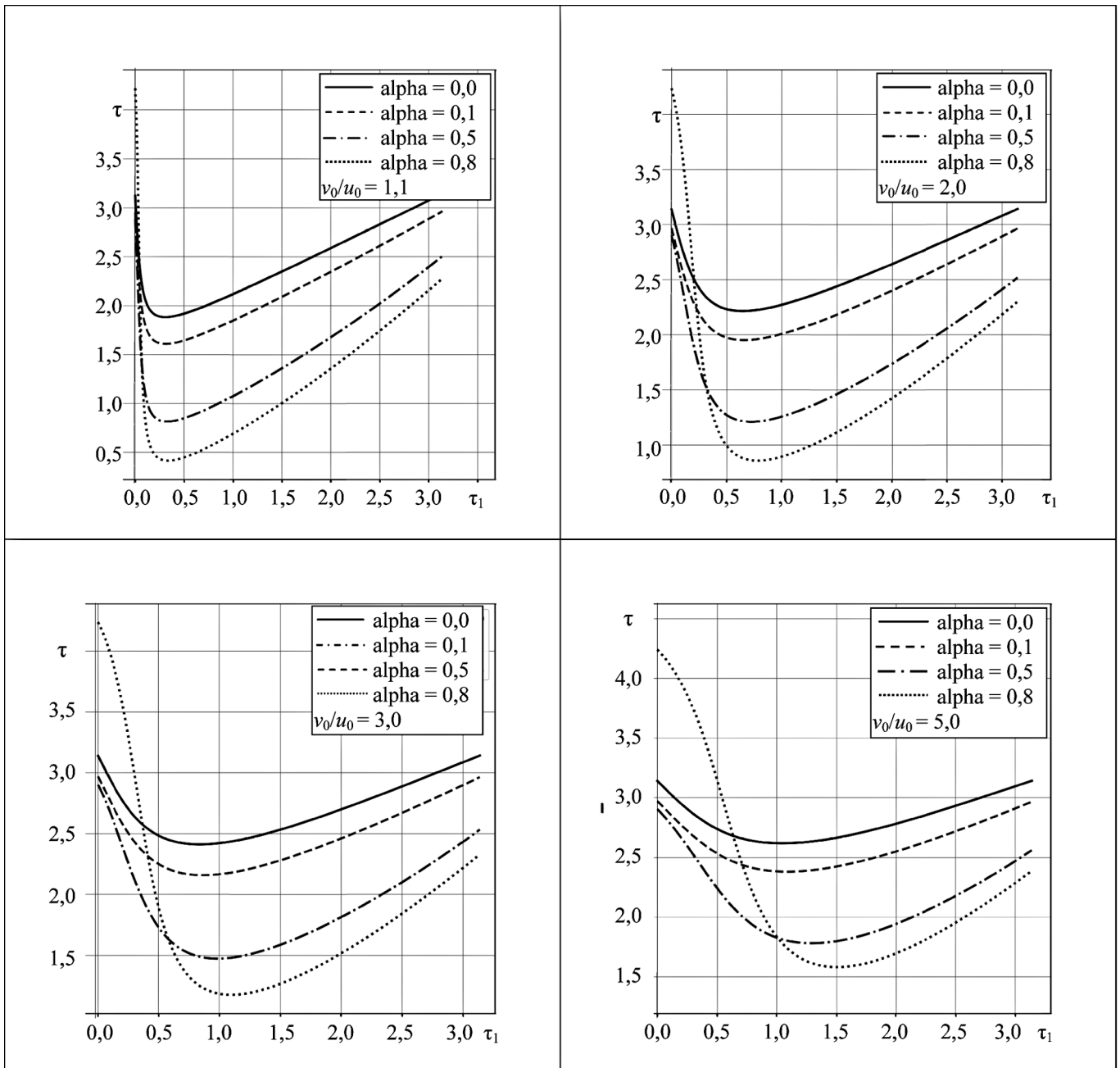


Рис. 3. Зависимость времени взаимодействия поверхности с телом от длительности первого этапа управления при разных отношениях начальной скорости элемента к уровню управления и разных коэффициентах вязкости α

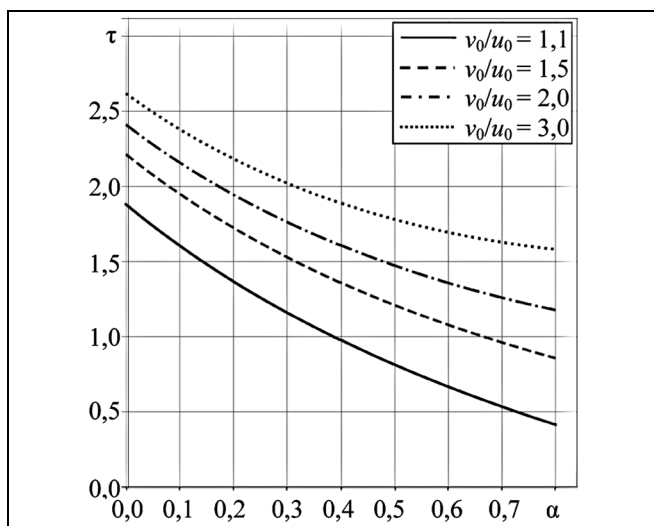


Рис. 4. Зависимость оптимального времени взаимодействия поверхности с телом от коэффициента вязкости α при разных отношениях начальной скорости элемента к уровню управления

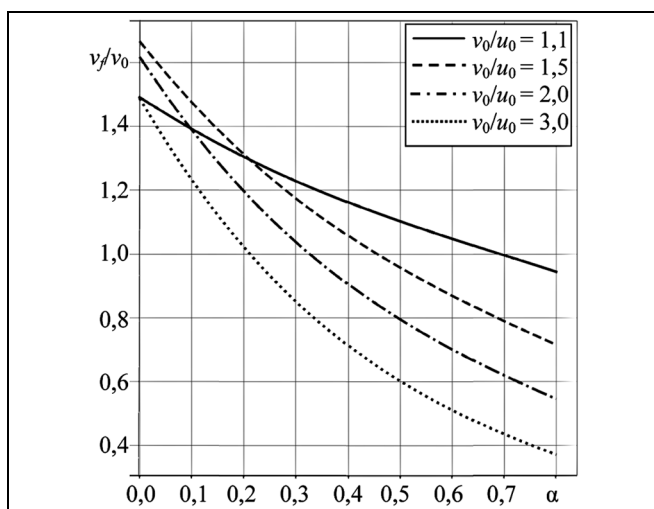


Рис. 5. Зависимость коэффициента восстановления от коэффициента вязкости α при разных отношениях начальной скорости тела к уровню управления

ляется терминальное послеударное значение скорости тела.

ЗАКЛЮЧЕНИЕ

Допредельная управляемая система является одномерной моделью взаимодействия в фазе удара. Чтобы дополнительно снабдить препятствие упругими свойствами, нужно знать пространственные распределения напряжений в среде либо пользоваться готовыми моделями, например, нелинейной моделью Герца. Показано, что даже в случае простой линейной модели появляется нетривиальный закон управления, с неоднозначной связью

закона восстановления с параметрами задачи. Закон оптимального управления, а также коэффициент восстановления, полученные в работе, служат основой численного моделирования поведения системы и дают начальные условия для описания движения точки после взаимодействия с препятствием. Предложенный подход к нахождению послеударных макроскопических характеристик системы позволит решать и другие задачи оптимального управления механической системой в фазе удара.

Решена задача оптимального быстрогодействия при одностороннем вязкоупругом взаимодействии материальной точки с управляемой поверхностью. Исследованы зависимости времени окончания взаимодействия и коэффициента восстановления от вязких свойств. Разработан программный модуль на языке Python и проведено моделирование поведения системы «материальная точка — препятствие».

ЛИТЕРАТУРА

1. Миллер Б.М., Рубинович Е.Я. Разрывные решения в задачах оптимального управления и их представление с помощью сингулярных пространственно-временных преобразований // Автоматика и телемеханика. — 2013. — № 12. — С. 56—103.
2. Козлов В.В., Трещев Д.В. Биллиарды. Генетическое введение в динамику систем с ударами. — М.: Изд-во МГУ, 1991. — 168 с.
3. Галаев А.А. О математической модели импульсного воздействия, вызванного ударом системы материальных точек об абсолютно жесткое препятствие // Автоматика и телемеханика. — 2006. — № 6. — С. 27—40. [Galyaev A.A. Impact of a system of material points against an absolutely rigid obstacle: a model for its impulsive action // Automation and Remote Control. — 2006. — Vol. 67, N 6. — P. 856—867.]
4. Галаев А.А. Оптимальное импульсное управление динамической системой в фазе удара // Автоматика и телемеханика. — 2006. — № 1. — С. 75—88. [Galyaev A.A. Optimal pulse control of dynamic systems in the shock phase // Automation and Remote Control. — 2006. — Vol. 67, N 1. — P. 65—77.]
5. Галаев А.А. Об одной задаче оптимального управления в фазе удара и унификации моментов окончания взаимодействия // Автоматика и телемеханика. — 2010. — № 12. — С. 10—21.
6. Галаев А.А. Об одномерном ударе цепочки тел, обладающей вязкоупругими свойствами // Автоматика и телемеханика. — 2015. — № 10. — С. 40—49.
7. Боровин Г.К., Лапшин В.В., Юрин Е.А. Нелинейная модель коллинеарного удара с сухим трением // Препринт ИПМ им. М.В. Келдыша. — 2014. — № 46. — С. 1—14.
8. Miller B.M., Rubanovich E.Ya., Bentsman J.G. Singular Space-Time Transformations. Method for Solving the Painleve Problem // Journal of Mathematical Sciences. — 2016. — Vol. 219, N 2. — С. 208—219.
9. Гирсанов И.В. Лекции по математической теории экстремальных задач. — Ижевск: Регулярная и хаотическая динамика, 2003. — 119 с.

Статья представлена к публикации членом редколлегии Е.Я. Рубиновичем.

Галаев Андрей Алексеевич — чл.-корр. РАН, гл. науч. сотрудник, galyaev@ipu.ru,

Лысенко Павел Владимирович — математик, Институт проблем управления им. В.А. Трапезникова РАН, г. Москва.

РОБАСТНАЯ УСТОЙЧИВОСТЬ И ВЫБОР НАСТРОЙКИ РЕГУЛЯТОРОВ ДЛЯ ТЕХНОЛОГИЧЕСКИХ ОБЪЕКТОВ С ЗАПАЗДЫВАНИЕМ

А.М. Цирлин, В.С. Овсепян

Рассмотрена задача о робастной устойчивости и выборе параметров типовых регуляторов одноконтурных линейных систем с обратной связью для технологических объектов с запаздыванием. Множество возможных значений параметров передаточной функции объекта предполагается замкнутым и ограниченным, при этом модуль и фаза амплитудно-фазовой характеристики разомкнутой системы монотонно уменьшаются с частотой.

Ключевые слова: системы с запаздыванием, робастная устойчивость, технологические объекты, робастная настройка.

Памяти В.Я. Ротача

ВВЕДЕНИЕ. ОСОБЕННОСТИ ТЕХНОЛОГИЧЕСКИХ ОБЪЕКТОВ РЕГУЛИРОВАНИЯ

В последние годы значительное внимание исследователей привлекает задача управления объектами с изменяющимися динамическими характеристиками и синтеза робастных регуляторов, которые без перенастройки способны управлять целым классом объектов или одним объектом в широком диапазоне изменения его параметров, нагрузок и пр. Методы синтеза таких систем с оценкой допустимой области возможных параметров объекта развиты в работах [1–4] и др. Большая часть этих работ посвящена линеаризованным системам, левая часть характеристического уравнения которых

$$P_n(a, p) = a_0 + a_1 p + \dots + a_n p^n = 0 \quad (1)$$

представляет собой полином, причем коэффициенты полинома могут принимать значения, принадлежащие некоторому множеству V_a . Полином $P_n(a, p)$ робастно устойчив, если он устойчив для любого $a \in V_a$. Для этого необходимо, но недостаточно, чтобы его коэффициенты были положительны $\forall a \in V_a$.

В работах [1, 2] задача робастной устойчивости полинома решена для случая, когда множество V_a

представляет собой параллелепипед, выделяемый интервальными ограничениями

$$\underline{a}_i \leq a_i \leq \bar{a}_i, \quad i = 0, \dots, n. \quad (2)$$

В работе [2] построены четыре полинома, для которых значения a выбраны так, что их устойчивость гарантирует устойчивость полинома (1). В работе [3] с использованием критерия устойчивости Михайлова по ограничениям (2) получена система, характеристики которой определены вектором $a_0 \in V_a$ и отклонениями $|a_i - a_0| \leq \gamma a_i$. Там показано, что требуется построить только один годограф Михайлова этой системы, чтобы узнать, будет ли исходный полином робастно устойчив для заданного γ и при каком максимальном γ эта устойчивость сохраняется.

Между тем, существует важный класс объектов регулирования, для которых полученные результаты неприменимы. Это технологические объекты в химической, металлургической, пищевой промышленности, энергетике и др. Динамика этих объектов определяется главным образом процессами тепло- и массопереноса. Они устойчивы, характеризуются распределенными параметрами, и апериодическим видом переходных процессов. На практике при расчете систем регулирования распределенность аппроксимируют введением чистого запаздывания.

Обозначим через $W_0(p)$ и $W_r(p)$ передаточные функции объекта и регулятора соответственно.

Передаточные функции технологических объектов содержат звено чистого запаздывания, а их характеристическое уравнение не имеет комплексных корней. Модуль и фаза амплитудно-фазовой характеристики (АФХ) таких объектов с ростом частоты монотонно уменьшаются, причем модуль стремится к нулю, а фаза к минус бесконечности (свойство «монотонности»). Свойство монотонности модуля и фазы технологических объектов в большинстве случаев справедливо и для разомкнутых систем регулирования, что, как показано ниже, позволяет упростить решение задачи о робастной устойчивости и о выборе робастных настроек регуляторов.

Знаменатель передаточной функции замкнутой системы регулирования технологических объектов не является полиномом и все результаты теории автоматического управления, основанные на свойствах полиномов, к таким системам неприменимы. Это относится к критериям устойчивости Рауса — Гурвица, Михайлова, к методам логарифмических частотных характеристик, методам пространства состояний и пр. Неприменимы к технологическим объектам и упомянутые выше результаты по робастной устойчивости динамических систем с полиномиальными характеристическими уравнениями. Особенности динамических свойств технологических объектов и систем управления ими неоднократно подчеркивал В.Я. Ротач [5].

Цель настоящей работы заключается в получении условий существования и алгоритма выбора робастных настроек типовых промышленных регуляторов в системах регулирования технологических объектов для ограниченного и замкнутого множества V_a возможных значений параметров передаточной функции объекта; уравнений, определяющих границу робастного D -разбиения в предположении, что модуль АФХ разомкнутой системы монотонно падает с ростом частоты, а также конкретизация этих результатов для типового технологического объекта.

1. ЗАДАЧА О МАКСИМАЛЬНОЙ СТЕПЕНИ УСТОЙЧИВОСТИ И ГИПОТЕЗА КРАТНОСТИ ДЕЙСТВИТЕЛЬНЫХ КОРНЕЙ

Характеристическое уравнение замкнутой системы регулирования имеет вид

$$1 + W(a, C, p) = 0.$$

Здесь $W(a, C, p)$ — передаточная функция разомкнутой системы, a и C — параметры объекта и регулятора, p — переменная преобразования Лапласа. Эта передаточная функция равна произведению передаточной функции объекта регулирования $W_0(a, p)$ на передаточную функцию регулятора $W_r(C, p)$, $C \in V_C$, а обратная связь отрицательная.

Регуляторы имеют типовую структуру, представляющую собой параллельное соединение пропорционального, дифференциального и интегрального блоков. Так, что передаточные функции объекта и регулятора имеют вид

$$W_0(p) = Ke^{-p\tau} \prod_{K=1}^N (1 + T_K p)^{-1}, \quad (3)$$

$$W_r(p) = -(C_0/p + C_1 + C_2 p). \quad (4)$$

Характеристическое уравнение замкнутой системы

$$W_0(a, p) W_r(C, p) = -1. \quad (5)$$

В общем случае левая часть уравнения (5) для технологических объектов может быть записана в форме квазиполинома

$$\begin{aligned} Z(C, p) &= \sum_{j=1}^r Q_j(a, p) e^{\tau_j p} - \sum_{k=1}^m C_k p^{k-1} = \\ &= f(p, a) - \sum_{k=1}^m C_k p^{k-1}. \end{aligned} \quad (6)$$

Здесь первое слагаемое не зависит от параметров регулятора. Обозначим его как $f(p, a)$.

В том случае, когда передаточные функции объекта и регулятора имеют вид (3) и (4), уравнение (5) принимает форму

$$\frac{K}{pe^{p\tau} \prod_{K=1}^N (1 + T_K p)} (C_0 + C_1 p + C_2 p^2) = 1$$

и может быть приведено к форме (6) как

$$pe^{p\tau} \prod_{K=1}^N (1 + T_K p) - K(C_0 + C_1 p + C_2 p^2) = 0.$$

Первое слагаемое в левой части этого уравнения обозначено выше через $f(p, a)$.

Решение задачи о выборе параметров регулятора, обеспечивающих максимум степени устойчивости системы, в работах [6—9] основывалось на предположении, что максимум степени устойчивости для системы, имеющей m варьируемых коэффициентов, достигается в случае, когда ближайшим к мнимой оси является действительный корень кратности $(m + 1)$. В этой точке характеристическое уравнение (для определенности рассмотрим ПИД-регулятор, для которого $m = 3$)

$$M(p, a, C) = f(p, a) - K(C_0 + C_1 p + C_2 p^2) = 0,$$

в котором комплексная переменная p может быть заменена на действительную переменную x , имеет корень кратности $(m + 1)$, т. е. не только значение



функции, но и значение всех ее производных до порядка m равно нулю, что приводит к системе уравнений:

$$\begin{aligned} M^{(3)} = f^{(3)}(x, a) = 0, \quad M^{(2)} = f^{(2)}(x, a) - 2KC_2 = 0, \\ M^{(1)} = f^{(1)}(x, a) - KC_1 - 2KC_2x = 0, \\ M(x, a, C) = 0. \end{aligned}$$

Только первое из этих уравнений нелинейно. Оно определяет зависимость максимальной степени устойчивости от параметров объекта $\eta^*(a)$. Для простых случаев оно имеет аналитическое решение. В общем случае приходится точно или приближенно находить эту зависимость численно. Остальные уравнения линейны по искомым параметрам регулятора, решаются поочередно, так что каждое из них содержит лишь одну переменную.

Гипотеза кратности действительных корней приводит к такой последовательности расчета параметров регулятора.

1. Находят максимальный корень $\zeta_m = \eta^*(a)$ уравнения

$$f^{(m)}(x, a) = 0, \quad m = 1, 2, 3. \quad (7)$$

Отметим, что в уравнение (7) входят только параметры объекта.

2. Для $m = 1$ (И-регулятор)

$$C_0^* = f(\zeta_1)/K. \quad (8)$$

Для $m = 2$ (ПИ-регулятор)

$$C_1^* = f'(\zeta_2)/K, \quad C_0^* = f(\zeta_2)/K - C_1^* \zeta_2. \quad (9)$$

Для $m = 3$ (ПИД-регулятор)

$$\begin{aligned} C_2^* = 0,5f''(\zeta_3)/K, \quad C_1^* = f'(\zeta_3)/K - C_2^* \zeta_3, \\ C_0^* = f(\zeta_3)/K - C_1^* \zeta_3 - C_2^* \zeta_3^2. \end{aligned} \quad (10)$$

Для пропорционального регулятора, как и для интегрального, $m = 1$, но

$$f(p, a) = \prod_{K=1}^N (1 + T_K p) e^{p\tau}.$$

Является ли найденное значение ζ_m максимальной степенью устойчивости системы, вообще говоря, не было доказано. В ряде случаев можно воспользоваться [9] достаточным условием того, что расстояние η корня ζ характеристического уравнения (6) от мнимой оси совпадает со степенью устойчивости: *для того чтобы действительный корень ζ характеристического уравнения системы с обратной связью был ближайшим к мнимой оси, достаточно выполнения неравенства*

$$|W_0(i\omega - \zeta) W_r(i\omega - \zeta)| < 1 \quad \forall \omega \neq 0.$$

При $\omega = 0$ левая часть этого неравенства равна единице, т. е. достаточно, чтобы модуль расширенной АФХ разомкнутой системы монотонно уменьшался с ростом частоты.

В работе [10] доказана теорема Попова — Пухова (ПП-теорема), приведем ее формулировку.

1. Степень устойчивости системы для квазиполинома $Z(C, p)$ вида (6) меньше либо равна модулю ζ_m , ближайшего к мнимой оси действительного корня уравнения (7). Так что, в случае, когда для фиксированных параметров объекта система уравнений (7)—(10) имеет действительное решение, найденная степень устойчивости является максимально-возможной, а соответствующие ей настройки регулятора оптимальными по этому показателю.

2. Эта система всегда имеет решение для технологических объектов, если $m < 3$.

3. Для системы с ПИД-регулятором условие кратности действительных корней выполнено (ближайший к мнимой оси корень — действительный кратности три), если отношение максимальной постоянной времени к времени запаздывания $T_{\max}/\tau > 0,5$.

Задача о выборе робастных параметров регуляторов в этих работах не рассматривалась.

2. РОБАСТНАЯ УСТОЙЧИВОСТЬ ТЕХНОЛОГИЧЕСКИХ СИСТЕМ РЕГУЛИРОВАНИЯ

Рассмотрим одноконтурные системы, у которых передаточная функция объекта зависит от коэффициентов $a \in V_a$, с типовыми регуляторами.

Для пропорционального регулятора $W_r(C, p) = -C_1$, для пропорционально-интегрального $W_r(C, p) = -(C_0/p + C_1)$. Наконец, для ПИД регулятора $W_r(C, p) = -(C_0/p + C_1 + C_2 p)$. Параметры регуляторов неотрицательны.

Будем предполагать разомкнутые системы устойчивыми или нейтральными, модуль и фаза которых удовлетворяют условиям монотонности:

$$M(a, C, \omega_k) \geq M(a, C, \omega_{k+1}), \quad (11)$$

где ω_k — решение уравнения

$$\begin{aligned} \varphi(a, C, \omega_k) = -\pi(1 + 2k), \\ k = 0, 1, \dots, \quad \frac{\partial \varphi}{\partial \omega} < 0 \quad \forall a \in V_a. \end{aligned} \quad (12)$$

Для технологического объекта и любого регулятора, у которого модуль и фаза АФХ не растут с частотой, эти условия выполнены. Если же регулятор содержит дифференцирующее звено, условия монотонности можно проверить по неравенству (11).

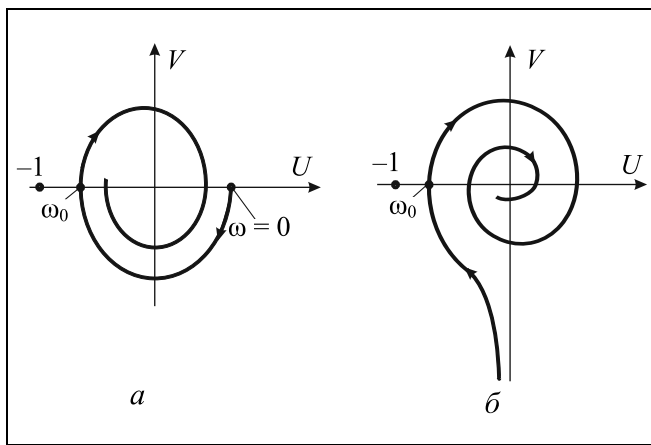


Рис. 1. Вид АФХ разомкнутой системы регулирования технологического объекта: *a* — регулятор не имеет интегральной составляющей; *б* — регулятор содержит интегральную составляющую

Согласно критерию Найквиста динамическая система с обратной связью устойчива, если разомкнутая система устойчива и при изменении ω от нуля до бесконечности годограф АФХ разомкнутой системы $W(a, C, i\omega) = W_r(C, i\omega)W_0(a, i\omega)$ не охватывает точку $(-1, i0)$.

Для систем с модулем АФХ, монотонно уменьшающимся с ростом ω , это значит, что при первом пересечении действительной оси на частоте $\omega_0 > 0$ выполнены условия

$$\varphi(a, C, \omega_0) = -\pi, \quad M(a, C, \omega_0) < 1.$$

Для краткости будем обозначать модуль АФХ разомкнутой системы в точке ее пересечения с отрицательной действительной полуосью как $M_\pi(a, C, \omega_0)$.

Величину $r(a, C, \omega_0) = 1 - M_\pi(a, C, \omega_0)$ называют *запасом устойчивости*. Системы, удовлетворяющие условиям (11) и (12), имеют АФХ вида, показанного на рис. 1.

Приведем несколько определений, касающихся робастных систем.

- Система робастно устойчива в замкнутом состоянии, если найдется такой допустимый вектор C , что $r(a, C) > 0 \quad \forall a \in V_a$.

Для систем, удовлетворяющих условиям (11) и (12), это означает, что существует вектор параметров регулятора $C^* \in V_C$ такой, что минимум по C максимального по a значения $M_\pi(a, C, \omega)$ меньше единицы:

$$\min_{C \in V_C} \left(\max_{a \in V_a} M_\pi(a, C, \omega) \right) = M_\pi^* < 1. \quad (13)$$

Разницу $r^* = 1 - M_\pi^*$ естественно назвать робастным запасом устойчивости.

- Настройки регулятора, выбранные по некоторому показателю качества, робастные, если для любых возможных параметров объекта $a \in V_a$ значение этого показателя не хуже его значения, соответствующего выбранным настройкам.
- Робастным D -разбиением в пространстве параметров регулятора называют выделение в этом пространстве множества всех значений параметров, для которых система робастно устойчива. Так что система робастно устойчива, если множество, определенное D -разбиением, не пусто.

Граница области робастной устойчивости в пространстве параметров регулятора C определена условиями

$$\left\{ \max_{a \in V_a, \omega} M(a, C, \omega) / \varphi(a, C, \omega) = -\pi \right\} = 1$$

для всех значений ω , соответствующих неотрицательным значениям C .

- Пусть система робастно устойчива при некотором значении C . Назовем робастной степенью устойчивости η такое неотрицательное число, что система, имеющая расширенную АФХ разомкнутой системы $W(a, C, i\omega - \eta)$, устойчива при всех $a \in V_a$ за исключением множества значений $a = a^*$, для которых она находится на границе устойчивости.

Для такой системы

$$\max_{a \in V_a, \omega} M(a, C, \eta, \omega) = 1 / \varphi(a, C, \eta, \omega) = -\pi. \quad (14)$$

Здесь $M(a, C, \eta, \omega)$ и $\varphi(a, C, \eta, \omega)$ — модуль и фаза расширенной АФХ разомкнутой системы.

Задача о выборе параметров регулятора по условиям максимума робастной степени устойчивости может быть с учетом введенных определенных сформулирована как

$$\eta^* = \max_C \min_{a \in V_a} \eta(a, C). \quad (15)$$

Эта задача очень сложна, так как степень устойчивости зависит от a и C алгоритмически и минимум ее по a нужно искать при всех возможных значениях параметров регулятора. Однако, как показано ниже, ее решение существенно упрощается, если порядок операций вычисления минимума и максимума можно изменить на обратный.

Покажем, что для систем регулирования с монотонной АФХ разомкнутой системы справедлива

Теорема. Задача (15) о максимуме робастной степени устойчивости имеет седловую точку (величина η^* не зависит от последовательности операций вычисления максимума и минимума).



Доказательство. Первоначально докажем, что для объектов с монотонной АФХ разомкнутой системы запас устойчивости и степень устойчивости монотонно связаны друг с другом.

Действительно, область, ограниченная расширенной АФХ, является отображением всех точек плоскости корней характеристического уравнения замкнутой системы, лежащих правее линии, параллельной мнимой оси, с абсциссой $-\eta$. В силу свойств конформного отображения эта область с ростом η расширяется. Так что модуль расширенной АФХ разомкнутой системы функцией для любого значения фазы ω , в частности, $M_\pi(a, C, \eta, \omega)$ с ростом η монотонно возрастает:

$$\frac{\partial M_\pi(a, C, \eta, \omega)}{\partial \eta} > 0. \quad (16)$$

Из неравенства (16) следует, что для того, чтобы система была робастно устойчивой, необходимо и достаточно существования такого $\eta > 0$, для которого выполнены условия (14).

Чем больше запас устойчивости системы (чем меньше $M_\pi(a, C, \omega)$), тем больше значение степени устойчивости, при котором модуль расширенной АФХ замкнутой системы станет равным единице в точке пересечения с отрицательной действительной полуосью.

Таким образом, решения задач (13) и (15) совпадают. Если первая из них имеет седловую точку, то это же справедливо и для задачи (15).

Задача о робастном значении запаса устойчивости для технологических объектов с монотонными АФХ разомкнутой системы имеет седловую точку (результат ее решения не зависит от порядка вычисления операций максимума и минимума), так как функция $M_\pi(a, C, \omega)$ мультипликативна. Модуль АФХ разомкнутой системы равен произведению модуля объекта на модуль регулятора, один из этих сомножителей зависит только от a , а другой — от C :

$$M_\pi(a, C, \omega) = F_a(a, \omega)F_C(C, \omega).$$

Обозначим максимум F_a по $a \in V_a$ и минимум F_C по C как $F_a^*(\omega)$ и $F_C^*(\omega)$. Изменение последовательности операций максимума и минимума не скажется на виде функций F_a^* и F_C^* . Как следствие эквивалентности, то же относится и к задаче о максимуме робастной степени устойчивости. Теорема доказана. ♦

С ростом коэффициента усиления разомкнутой системы величина M_π возрастает (запас устойчивости уменьшается), поэтому, если диапазон изменения коэффициента усиления K объекта не зависит от других его параметров, то минимуму по K степени устойчивости для любых фиксированных настроек регулятора соответствует максимум K .

Величина η^* заведомо не превосходит значения (см. работу [11])

$$\eta^0 = \min_{a \in V_a} \max_C \eta(a, C). \quad (17)$$

А в том случае, когда задача имеет седловую точку, она равна η^0 .

Множество значений вектора C , для которых система робастно устойчива, может быть построено как D -разбиение по условию устойчивости системы, имеющий расширенную АФХ разомкнутой системы $W(a, C, i\omega - \eta)$, где $a = a^*$, а $\eta < \eta^*$.

Если задачи (15) и (17) эквивалентны и $\eta^* = \eta^0$, то максимум робастной степени устойчивости можно найти, выбрав первоначально настройки регулятора по условию максимума η по C для любых допустимых параметров объекта, а затем вычислить минимум $\eta^*(a)$ по $a \in V_a$. Но первая (внутренняя) задача для рассматриваемого класса систем решена в работах [6–9] с учетом условия кратности действительных корней. Для любых значений коэффициентов a получены параметры регулятора, соответствующие максимальной степени устойчивости.

Внешняя задача является задачей нелинейного программирования о выборе на множестве V_a такого вектора параметров, для которого максимальная по C степень устойчивости (она зависит только от a) минимальна.

3. АЛГОРИТМ ВЫБОРА РОБАСТНЫХ НАСТРОЕК РЕГУЛЯТОРОВ

Использование результатов [6–9] позволяет предельно упростить задачу выбора робастных настроек регуляторов. Ее решение может быть проведено в такой последовательности.

1. Выбрав тип регулятора, записывают уравнение (7) и находят значение его максимального действительного корня $-\zeta = \eta^*(a)$ как функцию параметров объекта. Соответствующие настройки регулятора определены условиями (8)–(10), но считать их для всех $a \in V_a$ не требуется.

2. Решают задачу нелинейного программирования

$$\eta^*(a) \rightarrow \min_{a \in V_a}.$$

Если найденная минимальная степень устойчивости положительна, то систему можно сделать робастно устойчивой с помощью выбранного типа регулятора. Соответствующий вектор параметров a^* назовем «критическим». Критическим значением K^* коэффициента усиления объекта является максимальное из его возможных значений.

3. Для найденного критического вектора параметров объекта (только для него) по формулам (8)–(10) вычисляют настройки регулятора, обеспечивающие найденную робастную степень устойчивости системы.

4. Если $m > 2$, то проверяют для рассчитанной системы выполнение неравенства, связывающего

максимальную постоянную времени $T_{\max}$ и величину τ : $2T_{\max} \geq \tau$.

Отметим, что выбор параметров регулятора по условию максимума робастной степени устойчивости более естественен, чем по условию максимума запаса устойчивости, так как η^* прямо связана с продолжительностью переходного процесса в системе [12]. Кроме того, этот выбор существенно облегчается условием кратности ближайшего к мнимой оси действительного корня характеристического уравнения.

4. РОБАСТНЫЕ НАСТРОЙКИ СИСТЕМЫ С ТИПОВЫМ ТЕХНОЛОГИЧЕСКИМ ОБЪЕКТОМ

На практике при проектировании систем регулирования технологических объектов их характеристики представляют в форме звена чистого запаздывания или его последовательного соединения с интегрирующим или аperiodическим звеном. Поэтому для иллюстрации возможностей применения приведенных выше соотношений найдем зависимость достижимой робастной степени устойчивости для объекта, который представляет собой звено чистого запаздывания, последовательно соединенное с аperiodическим звеном первого порядка:

$$W_0(p) = Ke^{-p\tau} \frac{1}{Tp + 1}.$$

Такой объект часто называют *типовым технологическим объектом*, так как эта передаточная функция представляет собой простейшую аппроксимацию в окрестности равновесия распределенных процессов, включающих тепло- и массоперенос, транспорт продуктов, накопление вещества и пр.

Множество возможных параметров объекта зададим неравенствами:

$$K_{\min} \leq K \leq K_{\max}, \quad a = (T, \tau) \in V_0.$$

Множество V_0 лежит в первом квадранте и для любой его точки $2T \geq \tau$.

Условия максимума степени устойчивости приводят к системе уравнений (7)–(10), в которой

$$f(p, a) = pe^{p\tau}(Tp + 1).$$

И-регулятор. Решение уравнения (7) для $m = 1$ приводит к значению предельной степени устойчивости в функции параметров объекта

$$\eta^*(T, \tau) = \frac{1}{\tau} + \frac{1}{2T} - \sqrt{\frac{1}{\tau^2} + \frac{1}{4T^2}}. \quad (18)$$

Решение задачи о минимуме η^* по $(T, \tau) \in V_0$ определяет значения критических параметров объ-

екта T^*, τ^* . Соответствующее значение робастной настройки регулятора (8)

$$C_0 = \frac{1}{K_{\max}} \eta^* \exp^{-\eta^* \tau^*} (1 - T^* \eta^*).$$

ПИ-регулятор. В этом случае $m = 2$. Решение уравнения (7) для $m = 2$ приводит к значению предельной степени устойчивости:

$$\eta^* = \frac{2}{\tau} + \frac{1}{2T} - \sqrt{\frac{2}{\tau^2} + \frac{1}{4T^2}}. \quad (19)$$

Соответствующие робастные параметры регулятора для критических параметров объекта, минимизирующих ее на V_0 (см. выражения (8), (9)):

$$C_0 = \frac{\tau^*}{K_{\max}} (\eta^*)^2 e^{-\eta^* \tau^*} \left(1 + \frac{T^*}{\tau^*} - T^* \eta^* \right),$$

$$C_1 = \frac{1}{K_{\max}} (e^{-\eta^* \tau^*} [(2T^* + \tau^*) \eta^* - T^* \tau^* (\eta^*)^2 - 1]).$$

ПИД-регулятор. Аналогично двум рассмотренным выше случаям для $m = 3$ получим максимальную по C степень устойчивости

$$\eta^* = \frac{1}{2T} + \frac{3}{\tau} - \sqrt{\frac{3}{\tau^2} + \frac{1}{4T^2}}. \quad (20)$$

Соответствующие настроечные параметры для минимизирующих степень устойчивости T и τ :

$$C_0 = \frac{(\eta^*)^3}{K_{\max}} \left(T^* (\tau^*)^2 - T^* \eta^* (\tau^*)^3 + \frac{(\tau^*)^3}{2} \right) e^{-\eta^* \tau^*},$$

$$C_1 = \frac{e^{-\eta^* \tau^*}}{K_{\max}} (3T^* \tau^* (\eta^*)^2 + (\eta^*)^2 (\tau^*)^2 - \eta^* \tau^* - T^* (\eta^*)^3 (\tau^*)^2 - 1),$$

$$C_2 = \frac{e^{-\eta^* \tau^*}}{K_{\max}} \left(\frac{\eta^* \tau^*}{2} + 2T^* \eta^* - \frac{T^*}{\tau^*} - \frac{T^* \tau^* (\eta^*)^2}{2} - 1 \right).$$

П-регулятор. В этом случае $m = 1$, а функция $f(p, a) = e^{p\tau}(Tp + 1)$. Достижимая степень устойчивости (см. работу [9])

$$\eta^*(T, \tau) = \frac{1}{\tau} + \frac{1}{T}. \quad (21)$$

Робастная настройка пропорционального регулятора

$$C_1 = \frac{0.37 T^*}{K_{\max} \tau^*} e^{-\frac{\tau^*}{T^*}}.$$

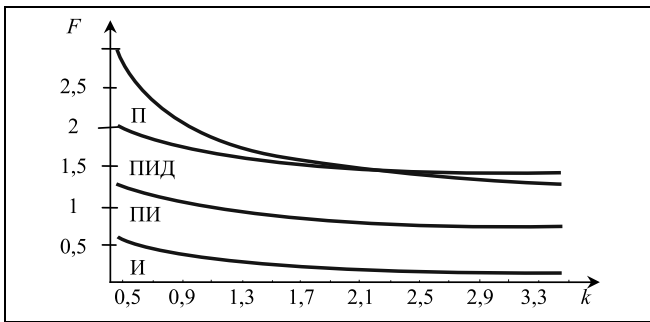


Рис. 2. Зависимость $F = \eta^* \tau$ от отношения $k = T/\tau$ для типового технологического объекта и регуляторов разного типа

Все выражения для предельной степени устойчивости могут быть записаны в форме:

$$\eta^*(T^*, \tau^*) = \frac{F(k)}{\tau^*},$$

где $k = T/\tau$.

Для каждого типа регулятора получим:

$$F_I(k) = 1 + \frac{1}{2k} - \sqrt{1 + \frac{1}{4k^2}},$$

$$F_{PI}(k) = 2 + \frac{1}{2k} - \sqrt{2 + \frac{1}{4k^2}},$$

$$F_{PID}(k) = 3 + \frac{1}{2k} - \sqrt{3 + \frac{1}{4k^2}},$$

$$F_P(k) = 1 + \frac{1}{k}.$$

Эти функции показаны на рис. 2. Они монотонны по k , так что для фиксированного значения τ минимуму степени устойчивости соответствует максимум k , а значит T .

Для заданного множества V_0 на плоскости с координатами T, τ для любого значения k минимум по τ достигается в точке пересечения, соответствующего луча с правой границей множества V_0 , соответствующей большим значениям τ . Множество таких точек задает зависимость $\tau^*(k)$. Если эта зависимость дифференцируема, то координаты критической точки, в которой отношение $\frac{F(k)}{\tau(k)}$ минимально, определяются условием

$$\frac{dF}{dk} \tau(k) = \frac{d\tau}{dk} F(k) = -\frac{T(k)}{k^2} F(k).$$

ЗАКЛЮЧЕНИЕ

Для систем регулирования, технологических объектов, содержащих запаздывание, доказана возможность выбора робастных настроек типо-

вых промышленных регуляторов минимизацией на множестве возможных параметров объекта (не обязательно характеризующегося интервальными ограничениями) максимально достижимой степени устойчивости, зависящей от этих параметров.

Для положительных ограниченных значений τ^* , T^* степень устойчивости η^* , найденная по формулам (18), (19), (20), (21), положительна, а значит, найдутся параметры типовых регуляторов, обеспечивающие робастную устойчивость.

Полученные условия позволяют решить и обратную задачу: найти тот диапазон изменения параметров объекта, для которого робастная устойчивость системы регулирования будет не ниже заданной.

ЛИТЕРАТУРА

1. Поляк Б.Т., Щербаков П.С. Робастная устойчивость и управление. — М.: Наука, 2002.
2. Харитонов В.Л. Асимптотическая устойчивость семейства линейных дифференциальных уравнений // Дифференциальные уравнения. — 1978. — Т. 1, вып. 11. — С. 2086—2088.
3. Поляк Б.Т., Цыпкин Я.З. Робастный критерий Найквиста // Автоматика и телемеханика. — 1992. — № 7. — С. 25—31.
4. Poljak B.T., Tsipkin J.Z. Robustness Nyquist-test // Automation and Remote Control. — 1992. — N. 7. — P. 25—31.
5. Ротач В.Я. Теория автоматического управления. — М.: МЭИ, 2008. — 396 с.
6. Татаринов А.В., Цирлин А.М. Предельная степень аperiodической устойчивости линейных систем и выбор параметров промышленных регуляторов // Моделирование и анализ информационных систем. — 2012. — Т. 19, № 2. — С. 87—96.
7. Гурецкий Х. Анализ и синтез систем управления с запаздыванием: пер. с польск. — М.: Машиностроение, 1974. — 326 с.
8. Шубладзе А.М. Способы синтеза систем управления максимальной степени устойчивости // Автоматика и телемеханика. — 1980. — № 1. — С. 28—37.
9. Загарий Г.И., Шубладзе А.М. Синтез систем управления на основе критерия максимальной степени устойчивости. — М.: Энергоатомиздат, 1988. — 170 с.
10. Попов А.Ю., Пухов С.С., Цирлин А.М. Минимизация точной верхней грани действительных частей корней квазимоночлена и предельная степень устойчивости линейных динамических систем с обратной связью // Автоматика и телемеханика. — 2015. — № 7. — С. 3—25.
11. Давыдов Э.Г. Исследование операций. — М.: Высшая школа, 1990. — 384 с.
12. Цыпкин Я.З. Основы теории автоматических систем. — М.: Наука, 1977. — 560 с.

Статья представлена к публикации членом редколлегии М.В. Хлебниковым.

Цирлин Анатолий Михайлович — д-р техн. наук, гл. науч. сотрудник, Институт программных систем им. А.К. Айламазяна РАН, г. Переславль-Залесский, ✉ tsirlin@sarc.botik.ru,

Овсепян Вардгес Сидракович — канд. техн. наук, доцент, Ванадзорский гос. ун-т им. Ов. Туманяна, Республика Армения, ✉ hovar@inbox.ru.

АППРОКСИМАЦИИ ФУНКЦИИ ВОССТАНОВЛЕНИЯ И СТРАТЕГИЯ УПРАВЛЕНИЯ ЭКСПЛУАТАЦИОННЫМИ ЗАТРАТАМИ

В.Н. Русев, А.В. Скориков

Рассмотрена аппроксимация функции восстановления для рекуррентных потоков восстановлений. Предполагается, что время безотказной работы объектов системы описывается с помощью двухпараметрического распределения Вейбулла — Гнеденко. Предложены аналитический и численный подходы (на основе методов дискретизации) решения уравнения Вольтерра. Полученные соотношения и алгоритмы проверены на контрольных примерах с помощью профессионального математического пакета Wolfram Mathematica. Изложенные результаты применены в стратегии «блоковой (групповой) политики замен» для обеспечения надежного и эффективного функционирования технологических объектов.

Ключевые слова: функция восстановления, распределение Вейбулла — Гнеденко, интегральное уравнение Вольтерра, производящая функция моментов, методы дискретизации интегральных уравнений, мониторинг надежности, блоковая политика замен, компьютерное моделирование в среде Wolfram Mathematica.

ВВЕДЕНИЕ

Функция восстановления $H(t)$ определяется как ожидаемое число отказов системы (или элемента) в течение времени $(0, t)$. Этот параметр используется для построения системы менеджмента надежности, определения оптимального плана профилактического обслуживания или замен [1–3]. Функция восстановления имеет важное значение для разработки политики гарантий [4], а также применяется в принятии решений планирования схемы поставок [5], построения планов страхования [6].

Для рекуррентных потоков восстановлений интегральное уравнение Вольтерра устанавливает соотношение между функцией восстановления $H(t)$ и функцией распределения $F(t)$ времени работы между отказами [7], когда плотность распределения $f(t) = F'(t)$ определена:

$$H(t) = F(t) + \int_0^t H(\tau)f(t - \tau)d\tau. \quad (1)$$

Решение в замкнутой форме этого уравнения невозможно, кроме некоторых случаев, когда поток восстановлений управляется экспоненциальным или эрланговским распределениями. Хорошо

известна [7] асимптотическая формула для функции восстановления (асимптота Смита)

$$H(t) \sim \frac{1}{\mu}t + \frac{1}{2}\left(\frac{\sigma^2 - \mu^2}{\mu^2}\right), \quad t \rightarrow \infty, \quad (2)$$

где μ и σ^2 — математическое ожидание и дисперсия времени работы между отказами.

Отметим также публикации [8, 9], посвященные прямому вычислению функции надежности для систем в терминах сверток.

В настоящей статье предполагается, что весь жизненный цикл функционирования объектов системы описывается с помощью двухпараметрического закона распределения Вейбулла — Гнеденко (см. например, работу [10]). Распределение Вейбулла — Гнеденко занимает важное место среди распределений времени безотказной работы систем, состоящих из групп большого числа элементов, отказы которых происходят взаимно независимо, так что отказ любого из элементов приводит к отказу всей системы (принцип «наислабейшего звена»).

Согласно государственному стандарту [11], «Планово-профилактические ремонты или замены полезны в случаях, когда отказы одной или нескольких ключевых составных частей изделия имеют четко выраженный износый и/или усталостный характер, что соответствует описанию



вероятности подобных отказов двухпараметрическим распределением Вейбулла. Зная параметры формы и масштаба этого распределения, можно установить рациональные значения периодичности профилактического обслуживания или замен этих составных частей». Поэтому, с позиций разработки мониторинговых систем оценивания факторов надежности элементов, для достижения оптимального плана технического обслуживания и ремонта изучение и анализ функции восстановления является весьма актуальной проблемой. Полученные аппроксимации функции восстановления применены в стратегии «групповой политики замен» для обеспечения надежного и эффективного функционирования технологических объектов.

Функция распределения и плотность распределения Вейбулла — Гнеденко имеют, соответственно, вид:

$$F(t) = \begin{cases} 1 - e^{-(\alpha t)^\beta}, & t > 0, \\ 0, & t \leq 0 \end{cases}$$

$$f(t) = \begin{cases} \alpha^\beta \beta t^{\beta-1} e^{-(\alpha t)^\beta}, & t > 0 \\ 0, & t \leq 0 \end{cases}$$

с параметрами формы $\beta > 0$, масштаба $\alpha > 0$.

Авторы работы [12] W.L. Smith and M.R. Leadbetter предложили метод для вычисления функции восстановления для распределения Вейбулла — Гнеденко с помощью разложения в степенной ряд для $\beta > 1$, численное вычисление этого ряда ограничено малым диапазоном t ; A.G. Constantine, N.I. Robinson [13] представили метод вычисления функции восстановления $H(t)$ с помощью вычетов, позволяющих получить представление для нее равномерно сходящимся рядом затухающих экспоненциальных членов. В статье [14] была предложена оценка функции восстановления, основанная на первых трех моментах распределения, уточняющая асимптотическую формулу (2). Получен также ряд других приближений, таких как полученных авторами работ [15–17], изучавших функцию восстановления для нормального, гамма-, равномерного распределений времени жизни.

1. АНАЛИТИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ ФУНКЦИИ ВОССТАНОВЛЕНИЯ. МЕТОД ПРОИЗВОДЯЩИХ ФУНКЦИЙ МОМЕНТОВ

Применим методы операционного исчисления. Из уравнения (1) получаем

$$\tilde{H}(s) = \tilde{F}(s) + \tilde{H}(s)\tilde{f}(s),$$

откуда следует

$$\tilde{H}(s) = \frac{\tilde{F}(s)}{1 - \tilde{f}(s)} = \frac{\tilde{f}(s)}{s(1 - \tilde{f}(s))}.$$

Для нахождения оригинала преобразования Лапласа $\tilde{H}(s)$ воспользуемся производящей функции моментов плотности распределения Вейбулла — Гнеденко.

Соотношение $h(t) = H'(t)$ в терминах преобразования Лапласа можно записать как $\tilde{H}(s) = \tilde{h}(s)/s$.

В работах [18, 19] получено представление преобразования Лапласа плотности восстановления

$$\tilde{h}(s) = \frac{1}{v_1 s} + \left(\frac{v_2}{2v_1^2} - 1 \right) + \frac{3v_2^2 - 2v_1 v_3}{12v_1^3} s +$$

$$+ \frac{3v_2^3 - 4v_1 v_2 v_3 + v_1^2 v_4}{24v_1^4} s^2 + \dots$$

Отсюда

$$\tilde{H}(s) = \frac{\tilde{h}(s)}{s} = \frac{1}{v_1 s^2} + \left(\frac{v_2}{2v_1^2} - 1 \right) \frac{1}{s} + \frac{3v_2^2 - 2v_1 v_3}{12v_1^3} +$$

$$+ \frac{3v_2^3 - 4v_1 v_2 v_3 + v_1^2 v_4}{24v_1^4} s + \dots$$

Используя обозначения

$$c_0 = \frac{1}{v_1}, \quad c_1 = \frac{v_2}{2v_1^2}, \quad c_2 = \frac{3v_2^2 - 2v_1 v_3}{12v_1^3},$$

$$c_3 = \frac{3v_2^3 - 4v_1 v_2 v_3 + v_1^2 v_4}{24v_1^4},$$

$$c_k =$$

$$= \begin{vmatrix} 1 & 0 & 0 & \dots & 0 & -m_0 \\ -m_0 & 1 & 0 & \dots & 0 & m_1 \\ m_1 & -m_0 & 1 & \dots & 0 & -m_2 \\ -m_1 & m_1 & -m_0 & \dots & 0 & m_3 \\ m_3 & -m_2 & m_1 & \dots & 0 & -m_4 \\ \dots & \dots & \dots & \dots & 1 & \dots \\ (-1)^{k+1} m_{k-2} & (-1)^k m_{k-3} & (-1)^{k-1} m_{k-1} & \dots & -m_0 & (-1)^k m_{k-1} \end{vmatrix},$$

где v_k — начальный момент случайной величины ξ с плотностью $f(t)$ порядка k :

$$v_k = \int_0^{+\infty} t^k f(t) dt = \frac{1}{\alpha^k} \Gamma\left(1 + \frac{k}{\beta}\right),$$

$$\Gamma(s) = \int_0^{+\infty} x^{s-1} e^{-x} dx \text{ — гамма-функция Эйлера, } s \in \mathbb{C},$$

имеем

$$\tilde{H}(s) = \frac{c_0}{s^2} + (c_1 - 1) \frac{1}{s} + c_2 + c_3 s + c_4 s^2 + \dots$$

Тогда

$$H(t) = F(t) + L^{-1} \left[\frac{c_0}{s^2} \tilde{f}(s) + (c_1 - 1) \frac{1}{s} \tilde{f}(s) + c_2 \tilde{f}(s) + c_3 s \tilde{f}(s) + c_4 s^2 \tilde{f}(s) + \dots \right],$$

где $L^{-1}[G(s)]$ обозначает обратное преобразование Лапласа для функции $G(s)$.

Таким образом,

$$\begin{aligned} H(t) &= F(t) + L^{-1} \left[\frac{c_0}{s^2} \tilde{f}(s) \right] + (c_1 - 1) \int_0^t f(\tau) d\tau + \\ &+ c_2 f(t) + c_3 f'(t) + c_4 f''(t) + \dots = \\ &= F(t) + L^{-1} \left[\frac{c_0}{s^2} \tilde{f}(s) \right] + (c_1 - 1) F(t) + c_2 F'(t) + \\ &+ c_3 F''(t) + c_4 F'''(t) + \dots \end{aligned}$$

В силу того, что

$$\begin{aligned} L^{-1} \left[\frac{c_0}{s^2} \tilde{f}(s) \right] &= c_0 \int_0^t (t - \tau) \alpha^\beta \beta \tau^{\beta-1} e^{-(\alpha\tau)^\beta} d\tau = \\ &= c_0 \alpha^\beta \beta \left[t \int_0^t \tau^{\beta-1} e^{-(\alpha\tau)^\beta} d\tau - \int_0^t \tau^\beta e^{-(\alpha\tau)^\beta} d\tau \right] = \\ &= c_0 \alpha^\beta \beta \left[t \frac{1}{\alpha^\beta \beta} \gamma(1, (\alpha t)^\beta) - \frac{1}{\alpha^{\beta+1} \beta} \gamma\left(1 + \frac{1}{\beta}, (\alpha t)^\beta\right) \right] = \\ &= c_0 \left[t F(t) - \frac{1}{\alpha} \gamma\left(1 + \frac{1}{\beta}, (\alpha t)^\beta\right) \right], \end{aligned}$$

где $\gamma(a, x) = \int_0^x t^{a-1} e^{-t} dt$ — неполная гамма-функция Эйлера и ввиду равенств

$$\begin{aligned} \int_0^t \tau^\mu e^{-(\alpha\tau)^\beta} d\tau &= \frac{1}{\alpha^{\mu+1} \beta} \gamma\left(\frac{\mu+1}{\beta}, (\alpha t)^\beta\right), \\ \gamma(1, (\alpha t)^\beta) &= \int_0^{(\alpha t)^\beta} e^{-z} dz = F(t), \end{aligned}$$

окончательно получаем представление:

$$\begin{aligned} H(t) &= H(t) + c_0 \left[t F(t) - \frac{1}{\alpha} \gamma\left(1 + \frac{1}{\beta}, (\alpha t)^\beta\right) \right] + \\ &+ (c_1 - 1) F(t) + c_2 F'(t) + c_3 F''(t) + c_4 F'''(t) + \dots = \\ &= F(t) [1 + c_0 t] - \frac{c_0}{\alpha} \gamma\left(1 + \frac{1}{\beta}, (\alpha t)^\beta\right) + (c_1 - 1) F(t) + \\ &+ c_2 F'(t) + c_3 F''(t) + c_4 F'''(t) + \dots = \\ &= c_0 t F(t) - \frac{c_0}{\alpha} \gamma\left(1 + \frac{1}{\beta}, (\alpha t)^\beta\right) + c_1 F(t) + c_2 F'(t) + \\ &+ c_3 F''(t) + c_4 F'''(t) + \dots, \end{aligned}$$

где коэффициенты c_k определены выше.

Другая запись полученного соотношения:

$$H(t) = F(t)(c_0 t + c_1) - \frac{c_0}{\alpha} \gamma\left(1 + \frac{1}{\beta}, (\alpha t)^\beta\right) + c_2 F'(t) + c_3 F''(t) + c_4 F'''(t) + \dots$$

Осталось только еще удостовериться в том, что полученное соотношение согласуется с асимптотической формой функции восстановления (2), которая в наших обозначениях имеет вид (см. далее рис. 4):

$$H(t) = c_0 t + c_1 - 1 + o(1), \quad t \rightarrow \infty.$$

2. ДИСКРЕТИЗАЦИЯ ИНТЕГРАЛЬНОГО УРАВНЕНИЯ

Для приближенного метода решения уравнения (1) относительно $H(t)$ применим методы дискретизации интегрального уравнения, которые в вариационных задачах называются прямыми методами. К таким методам относятся конечно-разностный метод Эйлера, метод Рунге, а также известный в задачах механики метод конечных элементов.

Дискретизация по методу Галеркина с применением кубических сплайнов была рассмотрена в работе [20]. Однако счет для большого числа точек разбиений, по предложенной в этой работе схеме, затруднителен из-за большого массива данных в оперативной памяти компьютера. Численное решение уравнения восстановления (прямой RS-метод) рассмотрено в статьях [21–23], где получены алгоритмы решения уравнения методами типа метода трапеций и метода Симпсона.

Используемая далее в работе дискретизация состоит в применении соответствующего алгоритма только к неизвестной функции, стоящей под знаком интеграла, а интегралы от ядер операторов вычисляются в пакете Mathematica. Таким образом, решение ищется, как в методе Рунге — Галеркина, в виде линейной комбинации соответствующих базисных функций, но для определения коэффициентов, как в методе коллокаций Канторовича [24], невязки в узлах полагаются равными нулю. Точность аппроксимации решения достигается не благодаря более сложным базисным функциям, как в методе Галеркина, а благодаря более мелкому разбиению области.

Рассмотрим три способа дискретизации уравнения (1). Разделим рассматриваемый отрезок времени $[0, t_{\max}]$ на n равных частей точками $t_0 = 0$, $t_1 = t_0 + \Delta$, ..., $t_n = t_0 + n\Delta$, где Δ — шаг разбиения.

Метод правых узлов. Будем искать решение $H(t)$ уравнения (1) в виде суммы кусочно-постоянных функций, равных на каждом отрезке разбиения $t_{k-1} \leq t \leq t_k$ значению искомого решения в правой

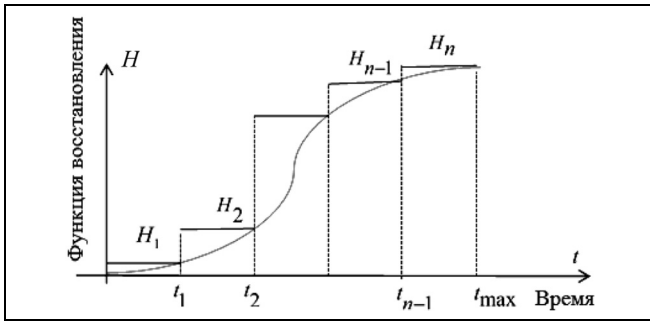


Рис. 1. Метод правых узлов

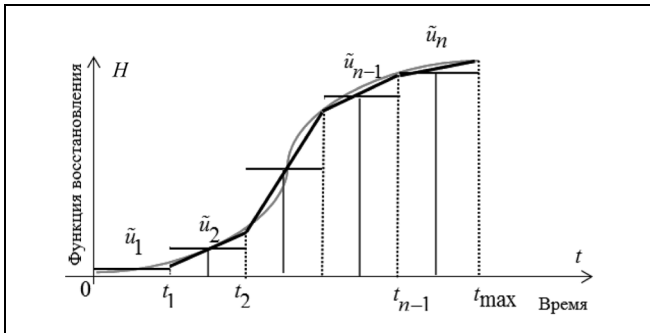


Рис. 2. Метод средних

точке t_k разбиения $H_k = H(t_k)$, также как в квадратурном методе прямоугольников (рис. 1). Если ввести координатные функции $I_k(t)$, равные единице при $t_{k-1} < t \leq t_k$ и нулю в остальных точках, то приближенное решение записывается в виде

$$H(t) \approx \sum_{k=1}^n H_k I_k(t).$$

Метод средних. Положим значение приближенного решения $u(t)$ на k -м отрезке, равным среднему значению $\bar{H}_k = \frac{H(t_{k-1}) + H(t_k)}{2}$, как в методе трапеций квадратурных формул (рис. 2). Приближенное решение, соответственно, имеет вид:

$$\tilde{H}(t) = \sum_{k=1}^n \bar{H}_k I_k(t).$$

Линейные сплайны. В качестве приближенного решения на k -м отрезке $\bar{H}_k(t)$ выберем интерполяционные многочлены Лагранжа первой степени, которые в узлах $t_{k-1}, t_k, k = 1, \dots, n$, равны значениям решения $H_{k-1} = H(t_{k-1}), H_k = H(t_k)$. Таким образом, на каждом отрезке разбиения искомое решение заменяется линейной функцией, т. е. рассматривается приближение решения ломаной (рис. 3).

Найдем решение по первому способу. Подставим точки разбиения в уравнение (1), заменим $H(t)$ на $H(t_k) = H_k$. Получим:

$$H_0 = F(0) = F_0,$$

$$\begin{aligned} H_1 &= F(t_1) + H_1 \int_0^{t_1} f(t_1 - s) ds = |t_1 - s = r| = \\ &= F(t_1) + H_1 \int_0^{t_1} f(r) dr, \\ H_1 &= F_1 + H_1 K_1, \end{aligned}$$

$$\text{где } F_1 = F(t_1), K_1 = \int_0^{t_1} f(r) dr,$$

$$\begin{aligned} H_2 &= F(t_2) + H_1 \int_0^{t_1} f(t_2 - s) ds + H_2 \int_{t_1}^{t_2} f(t_2 - s) ds = \\ &= |t_2 - s = r| = F(t_2) + H_1 \int_{t_2-t_1}^{t_2} f(r) dr + H_2 \int_0^{t_2-t_1} f(r) dr = \\ &= F(t_2) + H_1 \int_{t_1}^{t_2} f(r) dr + H_2 \int_0^{t_1} f(r) dr, \\ H_2 &= F_2 + H_1 K_2 + H_2 K_1, \end{aligned}$$

$$\text{где } F_2 = F(t_2), K_2 = \int_{t_1}^{t_2} f(r) dr,$$

$$\begin{aligned} H_n &= F(t_n) + H_1 \int_0^{t_1} f(t_n - s) ds + H_2 \int_{t_1}^{t_2} f(t_n - s) ds + \dots + \\ &+ H_n \int_{t_{n-1}}^{t_n} f(t_n - s) ds = |t_n - s = r| = \\ &= F(t_n) + H_1 \int_{t_{n-1}}^{t_n} f(r) dr + H_2 \int_{t_{n-2}}^{t_{n-1}} f(r) dr + \dots + H_n \int_0^{t_1} f(r) dr, \\ H_n &= F_n + H_1 K_n + H_2 K_{n-1} + \dots + H_n K_1, \end{aligned}$$

$$\text{где } F_n = F(t_n), K_n = \int_{t_{n-1}}^{t_n} f(r) dr.$$

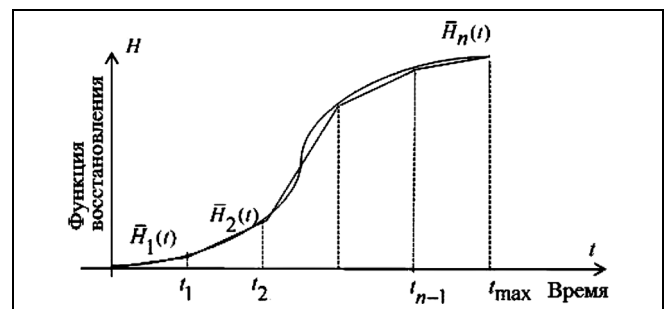


Рис. 3. Метод линейных сплайнов

Таким образом, для n точек разбиения получаем систему линейных уравнений вида:

$$\begin{cases} H_0 = F_0, \\ H_1 = F_1 + H_1 K_1, \\ H_2 = F_2 + H_1 K_2 + H_2 K_1, \\ \dots \\ H_n = F_n + H_1 K_n + H_2 K_{n-1} + \dots + H_n K_1. \end{cases} \quad (3)$$

Решение системы можно записать рекуррентными формулами:

$$\begin{cases} H_0 = F_0, \\ H_1 = F_1 / (1 - K_1) \\ H_2 = (F_2 + H_1 K_2) / (1 - K_1), \\ \dots \\ H_n = (F_n + H_1 K_n + H_2 K_{n-1} + \dots + H_{n-1} K_2) / (1 - K_1). \end{cases}$$

При втором способе дискретизации заменим $H(t)$ на соответствующем отрезке средним $\tilde{u}_k = \frac{u_{k-1} + u_k}{2}$. Соответственно, модифицируется система (3):

$$\begin{cases} H_0 = F_0, \\ H_1 = F_1 + \frac{H_1 + H_0}{2} K_1, \\ H_2 = F_2 + \frac{H_1 + H_0}{2} K_2 + \frac{H_2 + H_1}{2} K_1, \\ \dots \\ H_n = H_n + \frac{H_1 + H_0}{2} K_n + \frac{H_2 + H_1}{2} K_{n-1} + \dots + \frac{H_n + H_{n-1}}{2} K_1. \end{cases}$$

Отсюда получаем рекуррентные формулы для решения:

$$\begin{cases} H_0 = F_0, \\ H_1 = \left(F_1 + \frac{F_0}{2} K_1 \right) / \left(1 - \frac{K_1}{2} \right), \\ H_2 = \left(F_2 + \frac{H_1}{2} K_2 + \frac{H_0}{2} K_2 + \frac{H_1}{2} K_1 \right) / \left(1 - \frac{K_1}{2} \right), \\ \dots \\ H_n = \left(F_n + \frac{H_1}{2} K_n + \frac{H_2}{2} K_{n-1} + \dots + \frac{H_{n-1}}{2} K_2 + \frac{H_0}{2} K_n + \frac{H_1}{2} K_{n-1} + \dots + \frac{H_{n-1}}{2} K_1 \right) / \left(1 - \frac{K_1}{2} \right). \end{cases}$$

По третьему способу дискретизации приближенное решение на каждом отрезке разбиения ищется в виде интерполяционного многочлена Лагранжа первой степени, т. е.

$$H_k(t) = H_{k-1} \frac{t_k - t}{t_k - t_{k-1}} + H_k \frac{t - t_{k-1}}{t_k - t_{k-1}}.$$

Обозначим

$$l_k(t) = \frac{t_k - t}{t_k - t_{k-1}} = \frac{t_k - t}{\Delta}.$$

Тогда

$$H_k(t) = H_{k-1} l_k(t) - H_k l_{k-1}(t).$$

Для линейных функций l_k справедливо при $m \geq k$ равенство

$$\begin{aligned} l_k(t_m - t) &= \frac{t_k - t_m + t}{\Delta} = \frac{-(t_m - t_k) + t}{\Delta} = \\ &= -\frac{t_m - t_k}{\Delta} + \frac{t}{\Delta} = -l_{m-k}(t). \end{aligned} \quad (4)$$

Из уравнения (1) $u_0 = f(0) = f_0$, далее при $t = t_1$, учитывая равенство (4), имеем

$$\begin{aligned} H_1 &= F(t_1) + \int_0^{t_1} f(t_1 - s) (H_0 l_1(s) - H_1 l_0(s)) ds = \\ &= |t_1 - s = r| = F(t_1) + \int_0^{t_1} f(r) (H_0 l_1(t_1 - r) - \\ &\quad - H_1 l_0(t_1 - r)) dr = F_1 - H_0 \int_0^{t_1} f(r) l_0(r) dr + \\ &\quad + H_1 \int_0^{t_1} f(r) l_1(r) dr = F_1 - H_0 G_{10} + H_1 G_{11}. \end{aligned}$$

Отсюда

$$H_1(1 - G_{11}) = f_1 - H_0 G_{10}.$$

Подставим в уравнение (1) $t = t_n$:

$$\begin{aligned} H_n &= F(t_n) + \int_0^{t_1} f(t_n - s) (H_0 l_1(s) - H_1 l_0(s)) ds + \\ &\quad + \int_{t_1}^{t_2} f(t_n - s) (H_1 l_2(s) - H_2 l_1(s)) ds + \dots + \\ &\quad + \int_{t_{n-1}}^{t_n} f(t_n - s) (H_{n-1} l_n(s) - H_n l_{n-1}(s)) ds = |t_1 - s = r| = \\ &= f_1 + \int_{t_{n-1}}^{t_n} f(r) (H_0 l_1(t_n - r) - H_1 l_0(t_n - r)) dr + \\ &\quad + \int_{t_{n-2}}^{t_{n-1}} f(r) (H_1 l_2(t_n - r) - H_2 l_1(t_n - r)) dr + \dots + \\ &\quad + \int_0^{t_1} f(r) (H_{n-1} l_n(t_n - r) - H_n l_{n-1}(t_n - r)) dr = \end{aligned}$$

$$\begin{aligned}
&= F_n - H_0 \int_{t_{n-1}}^{t_n} f(r) l_{n-1}(r) dr + H_1 \int_{t_{n-1}}^{t_n} f(r) l_n(r) dr - \\
&- H_1 \int_{t_{n-2}}^{t_{n-1}} f(r) l_{n-2}(r) dr + H_2 \int_{t_{n-2}}^{t_{n-1}} f(r) l_{n-1}(r) dr - \dots - \\
&- H_{n-1} \int_0^{t_1} f(r) l_0(r) dr + H_n \int_0^{t_1} f(r) l_1(r) dr = \\
&= F_1 - H_0 G_{nn-1} + H_1 (G_{nn} - G_{n-1n-2}) + \dots + \\
&+ H_{n-1} (G_{22} - G_{10}) + H_n G_{11},
\end{aligned}$$

где

$$G_{kj} = \int_{t_{k-1}}^{t_k} f(r) l_j(r) dr.$$

Таким образом,

$$\begin{aligned}
H_n(1 - G_{11}) &= F_1 - H_0 G_{nn-1} + \\
&+ H_1 (G_{nn} - G_{n-1n-2}) + \dots + H_{n-1} (G_{22} - G_{10}).
\end{aligned}$$

Итак, система уравнений имеет вид

$$\begin{cases}
H_0 = F_0, \\
H_1(1 - G_{11}) = F_1 - H_0 G_{10}, \\
H_2(1 - G_{11}) = F_2 - H_0 G_{21} + H_1 (G_{22} - G_{10}), \\
\dots \\
H_n(1 - G_{11}) = F_n - H_0 G_{nn-1} + \\
+ H_1 (G_{nn} - G_{n-1n-2}) + \dots + H_{n-1} (G_{22} - G_{10}).
\end{cases} \quad (5)$$

Рекуррентные формулы для решения системы (5) имеют вид:

$$\begin{cases}
u_0 = f_0, \\
H_1 = (F_1 - H_0 G_{10}) / (1 - G_{11}), \\
H_2 = (F_2 - H_0 G_{21} + H_1 (G_{22} - G_{10})) / (1 - G_{11}), \\
\dots \\
H_n = (F_n - H_0 G_{nn-1} + H_1 (G_{nn} - G_{n-1n-2}) + \dots + \\
+ H_{n-1} (G_{22} - G_{10})) / (1 - G_{11}).
\end{cases}$$

3. КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ И СРАВНЕНИЕ РЕЗУЛЬТАТОВ

Полученные алгоритмы были проверены на контрольных примерах с помощью профессионального математического пакета Wolfram Mathematica. Погрешность приближенного решения определяется не только построенным алгоритмом, но также погрешностью численного интегрирования в пакете Wolfram Mathematica, равной 10^{-15} . Отметим, что погрешность решения контролируется применением различных методов и известным асимптотическим приближением (2). Для распре-

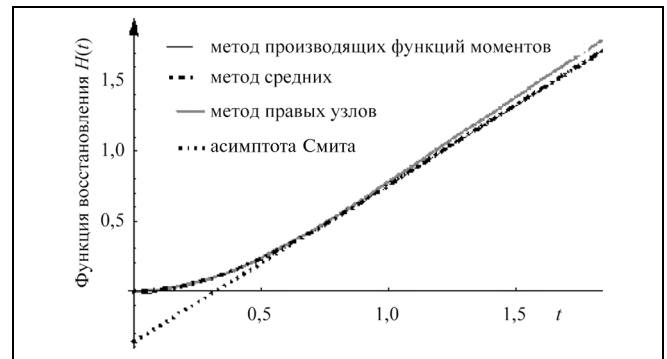


Рис. 4. Аппроксимация функции восстановления для распределения Рэлея

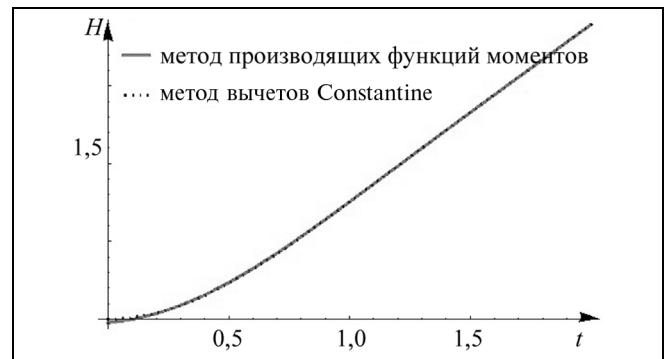


Рис. 5. Сравнение аппроксимаций функций восстановления для распределения Рэлея, полученных методом производящих функций моментов и методом вычетов Constantine, Robinson

деления Рэлея, важного частного случая распределения Вейбулла — Гнеденко ($\beta = 2, \alpha = 1$), функции восстановления представлены на рис. 4. Из графиков видно, что приближение линейными сплайнами и метод средних дает одно и то же решение при малом числе точек разбиения (20 точек), которое близко к приближенному решению, полученному методом производящих функций моментов. Рассмотренные приближенные решения быстро сходятся к асимптотическому значению функции восстановления (2). Метод правых узлов для рассмотренного числа точек разбиения (20 точек) дает большую погрешность при $t > 1$. Однако, увеличение числа точек разбиения до 200 дает практически совпадение приближенных решений, найденных рассмотренными выше методами правых узлов, средних и производящих функций. Время счета при таком разбиении составляет приблизительно 2–3 с.

Сравнение графиков функций восстановления, полученных для распределения Рэлея методом вычетов [13] и методом производящих функций моментов (рис. 5) показывает совпадение результатов.

4. ОПТИМАЛЬНАЯ СТРАТЕГИЯ ПРОФИЛАКТИЧЕСКОГО ОБСЛУЖИВАНИЯ

По ГОСТ Р 27.606—2013 [11]: «Если же отказ не несет угроз безопасности, но ведет к утрате изделием готовности к применению по назначению, то периодичность замен устанавливают, исходя из заданного уровня готовности, обеспечиваемого при оптимальных затратах, включающих в себя, в том числе, стоимость заменяемых изделий и экономический ущерб от отказов». Таким образом, необходим баланс между суммой, потраченной на профилактическое обслуживание, и суммой на замены при внезапном отказе.

Применим стратегию технического обслуживания, известную как «групповую или блоковую политику замен» («block replacement policy — BRP») [1, 3]. Предполагается, что объект заменяется новым изделием при постоянной длине интервала замен t_p , независимо от возраста объекта, а также замены объекта происходят столько раз, сколько требуется на интервале $(0, t_p)$ при внезапных отказах объекта. В работе [25] рассмотрены стратегии замен по эксплуатационным затратам, в частности, уточняющие определение оптимального срока

службы при восстановлении работоспособности объекта без выполнения его ресурса.

Пусть C_p — средняя стоимость профилактического обслуживания, C_f — средняя стоимость восстановления при отказе ($C_p < C_f$). Средняя стоимость на интервале $(0, t_p)$ профилактического обслуживания и восстановления после отказа равна $C_p + C_f H(t_p)$, где $H(t_p)$ — среднее число восстановлений на интервале $(0, t_p)$. В качестве критерия оптимальности рассматривается средняя стоимость эксплуатационных затрат (удельная стоимость затрат) единицу времени:

$$R(t_p) = \frac{C_p(1 + c_o H(t_p))}{t_p}, \quad (6)$$

где $c_o = C_f/C_p$ — коэффициент затрат. Точка минимума функции (6) дает соответствующее значение времени оптимального профилактического обслуживания t_p . Известно, что в точке t_0 экстремума $R(t_0) = C_f H'(t_0) = C_f h(t_0)$, т. е. значение функции $R(t)$ определяется плотностью восстановления $h(t_0)$. Более того, можно доказать равенство

$$R''(t_0) = t_0^2 C_f H''(t_0). \quad (7)$$

Из равенства (7) следует, что характер выпуклости функции $R(t)$ в точках экстремума совпадает с характером выпуклости функции $H(t)$ и осцилляция функции восстановления согласуется с осцилляцией средней стоимости эксплуатационных затрат.

Достаточным условием существования минимума функции $R(t)$ является условие на коэффициент затрат c_o [26, 27], имеющее в наших обозначениях вид:

$$c_o > \frac{2}{1 - CV^2}, \quad (8)$$

где $CV^2 = \sigma^2/\mu^2$ — квадрат коэффициента вариации. Отметим, что условие (8) не является необходимым. Приведем иллюстрацию данного утверждения, служащую контрпримером условию (8). Рассмотрим распределение Рэлея, для которого ограничение (8) означает, что $c_o > 2,8$. Однако, как видно из графика (рис. 6) при значении коэффициента $c_o = 2,6$, не удовлетворяющему условию (8), функция $R(t)$ имеет вполне конкретный минимум. Более того, при этом же коэффициенте затрат функция стоимости имеет еще и максимум. Иными словами, при большем значении интервала профилактического обслуживания можно попасть в точку максимума затрат, а из формулы (2) следует, что $R(\infty) = C_f/\mu$ является наименьшим значением функции стоимости. Поэтому в этом случае оптимальной стратегией будет замена лишь после наступления отказа.

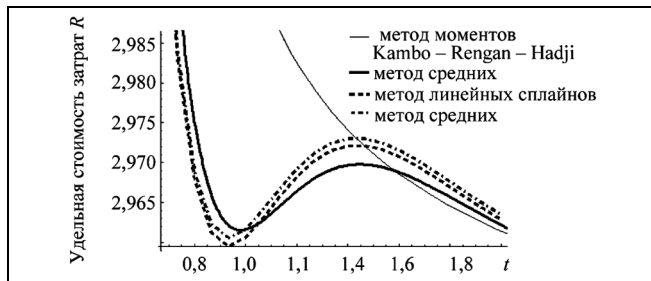


Рис. 6. Графики приближений функции стоимости $R(t)$ для распределения Рэлея при коэффициенте затрат $c_o = 2,6$

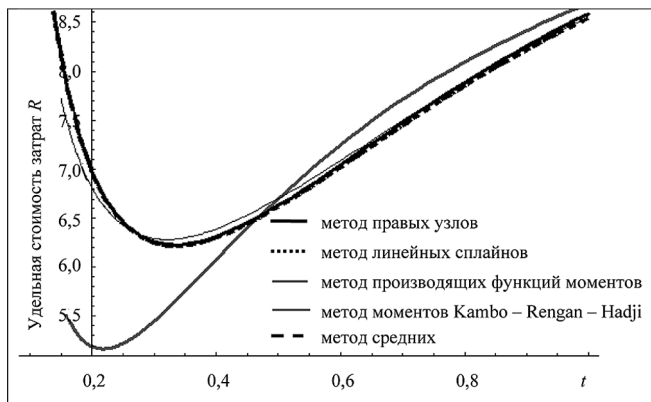


Рис. 7. Графики приближений функции $R(t)$ для распределения Рэлея при коэффициенте затрат $c_o = 10$



Заметим, что применение критерия (6) для определения времени оптимального профилактического обслуживания требует достаточно точного вычисления функции восстановления. Сравнение графиков (рис. 7) критерия (6) показывает значительное расхождение графиков $R(t)$ при вычислении функции восстановления по методу производящих моментов или дискретизацией с графиком $R(t)$ по методу [14] с использованием трех моментов.

Соответствующие координаты точек минимума: {0,32; 6,28}; {0,32; 6,22}; {0,22; 5,17}. Сравнение результатов, полученных в данной работе с результатами {0,34; 6,17} работы [16] показывают небольшое расхождение 5 и 2 % соответственно.

Таким образом, стратегия «групповой политики восстановлений» дает возможность обеспечения надежного и эффективного функционирования технологических объектов в рамках системного подхода к мониторингу эксплуатационных показателей.

ЗАКЛЮЧЕНИЕ

Проблеме вычисления функции восстановления, одной из популярных и важных задач теории надежности, посвящены многочисленные исследования, что отражено в достаточно подробном обзоре, выполненном в начале статьи. В работе предложено и исследовано несколько методов аппроксимации функции восстановления на основе ее аналитического представления: метод производящих функций моментов с последующим обращением преобразования Лапласа, методы решения интегрального уравнения восстановления путем его дискретизации. Приведены результаты компьютерного моделирования и сравнение на их основе различных методов аппроксимации. Полученные формулы для нахождения функции восстановления применены в стратегии «групповой политики замен» в задаче обеспечения надежного и эффективного функционирования технологических объектов с позиций мониторинга показателей надежности элементов.

ЛИТЕРАТУРА

1. Барлоу Р., Прошан Ф. Математическая теория надежности. — М.: Советское радио, 1969. — 488 с.
2. Северцев Н.А. Надежность сложных систем в эксплуатации и отработке. — М.: Высшая школа, 1989. — 432 с.
3. Jardine A.K.S., Tsang A.H.C. Maintenance, Replacement, and Reliability: Theory and Applications. — London, N.-Y.: Boca Raton, CRC/Taylor & Francis, 2006. — 330 p.
4. Blischke W., Murthy D.N.P. Warranty Cost Analysis. — N.-Y.: CRC Press, 1994.
5. Parsa H., Jin M. An improved approximation for the renewal function and its integral with an application in twoechelon // International Journal of Production Economics. — 2013. — Vol. 146, N 1. — P. 142–152.
6. Frees E.W. Warranty analysis and renewal function estimation // Naval Research Logistics Quarterly. — 1986. — Vol. 33. — P. 361–372.
7. Кокс Д.Р., Смит В.Л. Теория восстановлений. — М.: Советское радио, 1967. — 292 с.

8. Andronov A., Vishnevsky V., Rykov V. Reliability of a system with two parallel renewable channels // The 10-th Intern. Conf. on Mathematical Methods in Reliability (MMR-2017) July 3 — July 5, 2017, Grenoble, France. — P. 58.
9. Andronov A., Vishnevsky V., Rykov V. On Reliability Function of a Parallel System with Three Renewable Components // Материалы междунар. науч. конф. «Аналитические и вычислительные методы в теории вероятностей и ее приложениях (ABMTB-2017)», Москва, РУДН, 23–27 окт. 2017 г. / под общ. ред. А.В. Лебедева. — М., 2017. — С. 430–434.
10. Grigoriev L., Kucheryavy V., Rusev V., Sedyh I. Formation of estimates of reliability indicators for active elements in gas transport systems on the basis of refusals statistics // Journal of Polish Safety and Reliability Association. — 2014. — Vol. 5, N 2. — P. 41–47.
11. ГОСТ Р 27.606—2013. Надежность в технике. Управление надежностью. Техническое обслуживание, ориентированное на безотказность. — М.: Стандартинформ, 2014. — 38 с.
12. Smith W.L., Leadbetter M.R. On the Renewal Function for the Weibull Distribution // Technometrics. — 1963. — N 5. — P. 393–396.
13. Constantine A.G., Robinson N.I. The Weibull renewal function for moderate to large arguments. // Computational Statistics & Data Analysis. — 1997. — N 24. — P. 9–27.
14. Kambo N.S., Rangan A. & Ehsan Moghimi Hadji E. Moments based approximation to the renewal function // Communications in Statistics. — Theory and Methods. — 2012. — Vol. 41. — P. 851–868.
15. Cui L., Xie M. Some Normal Approximations for Renewal Function of Large Weibull Shape Parameter // Communications in Statistics. — 2003. — Vol. 32, N 1. — P. 1–16.
16. Smeitink E., Dekker R. Simple approximation to the renewal function // IEEE Trans. on Reliability. — 1990. — Vol. 39, N 1. — P. 71–75.
17. Maghsoodloo S., Helvaci D. Renewal and Renewal-Intensity Functions with Minimal Repair // Journal of Quality and Reliability Engineering. — 2014. — ID 857437. — 10 p.
18. Русев В.Н., Скориков А.В. Анализ элементов систем газоснабжения с помощью метода производящих функций моментов // Тр. РГУ нефти и газа им. И.М. Губкина. — 2016. — № 1 (282). — С. 68–79.
19. Rusev V., Skorikov A. On solution of renewal equation in the weibull model // Reliability: Theory & Applications. — 2017. — Vol. 12, N 4 (47). — P. 60–67.
20. Deligonul Z.S., Bilgen S. Solution of the Volterra equation of renewal theory with Galerkin technique using cubic splines // Journal of Statistical Computation and Simulation. — 1984. — N 20. — P. 37–45.
21. Xie M. On the solution of renewal-type integral equations // Communications in Statistics. — Simulation and Computation. — 1989. — Vol. 18 (1). — P. 281–293.
22. Xie M., Preuss W., Cui L. Error analysis of some integration procedures for renewal equation and convolution integrals // Journal of Statistical Computation and Simulation. — 2003. — Vol. 73 (1). — P. 59–70.
23. Tortorella M. Numerical solutions of renewal-type integral equation // Inform Journal of Computing. — 2005. — Vol. 17 (1). — P. 66–74.
24. Канторович Л.В., Акилов Г.П. Функциональный анализ: 3-е изд., перераб. — М.: Наука, 1984. — 752 с.
25. Лубков Н.В. Оптимизация срока службы оборудования по критерию эксплуатационных затрат // Автоматика и телемеханика. — 2008. — № 5. — С. 180–190.
26. Hanscom M.A., Cleroux R. The block replacement problem // J. Statist. Comput. Simul. — 1978. — Vol. 3. — P. 233–248.
27. Berg M. A marginal cost analysis for preventive maintenance policies // European Journal of Operational Research. — 1980. — N 4. — P. 136–142.

Статья представлена к публикации членом редколлегии В.М. Вишневым.

Русев Владимир Николаевич — ст. преподаватель,
✉ rusev.v@gubkin.ru,

Скориков Александр Васильевич — канд. физ.-мат. наук, доцент,
✉ skorikov.a@gubkin.ru,
Российский государственный университет нефти и газа (национальный исследовательский университет) имени И.М. Губкина.

ОСНОВЫ ТЕОРИИ КОМПЛЕКСНОЙ ДЕЯТЕЛЬНОСТИ.

Ч. 1. Структуры комплексной деятельности.

Неопределенность и порождение комплексной деятельности

М.В. Белов, Д.А. Новиков

Изложены основы теории комплексной деятельности, распространяющей методологию на случай любой сложной (имеющей нетривиальную многоуровневую внутреннюю структуру) человеческой деятельности. Рассмотрены структурные элементы комплексной деятельности, конструктивно описаны ее логическая, причинно-следственная и процессная структуры. Показано, что системы согласованных формальных моделей обеспечивает практическое применение результатов как специалистами-практиками, так и учеными, занимающимися исследованиями общих принципов организации деятельности (практической, научной и др.) и управления организационно-техническими системами. Впервые предложен формализм описания сложной деятельности вместе с субъектом — организационно-технической системой.

Ключевые слова: комплексная деятельность, неопределенность, организация, управление.

ВВЕДЕНИЕ

Подавляющее большинство комплексных систем¹ включают в себя людей. Наличие в составе организационно-технической системы² (ОТС) человека требует учитывать фактор его активности и, следовательно, активности системы в целом: целе-

полагание и намерения человека, его активный выбор, неопределенность его действий [1].

Отвечая этому требованию, далее в соответствии с *системным подходом* [2, 3] излагается согласованная система моделей комплексной деятельности (КД) — *теория комплексной деятельности* (ТКД) [4, 5] (гносеологический взгляд на эту теорию представлен в работе [6]). В ее рамках расширены понятия *деятельность*³ человека, с одной стороны, и *поведение*⁴ системы — с другой; введено понятие *комплексной деятельности*⁵, свойственной

¹ Система — совокупность элементов, находящихся в отношениях и связях друг с другом, которая образует определенную целостность, единство. «Система (искусственная) — совокупность взаимодействующих элементов, организованная для достижения одной или нескольких декларированных целей. ... Система (в прикладном смысле) часто рассматривается как продукт (деятельности) или как сервис, который система обеспечивает» [7, 8].

Комплексная система (Complex System) — система, обладающая свойством эмерджентности; открытая система с непрерывно взаимодействующими и конкурирующими элементами. Открытость понимается как свободное и неограниченное искусственными факторами участие и взаимодействие элементов друг с другом и окружающей средой [9].

Эмерджентность — свойство систем, состоящее в том, что свойства целого не сводятся к совокупности свойств частей, из которых оно состоит, и не выводятся из них.

² *Организационно-техническая система* — сложная система, включающая в себя людей, технические и природные элементы.

³ *Деятельность* [10, с. 4] — целенаправленная активность человека.

⁴ *Поведение* (системы, объекта) — последовательное во времени, хотя бы частично наблюдаемое, поддающееся измерению, объективной фиксации изменение его состояний. Для индивида-субъекта поведение — последовательность его действий, взаимодействие с окружающей средой, опосредованное его внешней (двигательной) или внутренней (психической) активностью.

⁵ Деятельность, обладающая нетривиальной внутренней структурой, с множественными и/или изменяющимися субъектом, технологией, ролью предмета деятельности в его целевом контексте.

Технология — система условий, критериев, форм, методов и средств последовательного достижения поставленной цели.

ОТС; исследована природа КД как комплексной системы; выявлены архитектура [11, 12] КД и аспекты развития КД во времени; разделены общесистемные и специфические аспекты КД, обобщены и исследованы первые из них.

Предметом ТКД служит диалектически связанная пара: комплексная деятельность и организационно-техническая система, являющаяся субъектом деятельности. Предметом и результатом комплексной деятельности, как правило, также служит комплексная система. Важно заметить, что применительно к создаваемым человеком системам ценность или полезность представляют не системы са-

ми по себе, а их функции и результаты деятельности. Поэтому и деятельность рассматривается как первичная в паре «КД — ОТС» [5].

Ключевые общесистемные категории КД: *цель, структура, неопределенность, жизненный цикл, организация и управление* — онтология основных категорий ТКД представлена на рис. 1.

Теория комплексной деятельности рассматривает элементы комплексной деятельности и конструктивно описывает ее структуру — базовый формализм описания КД, модели логической и причинно-следственной структур рассмотрены в § 1.

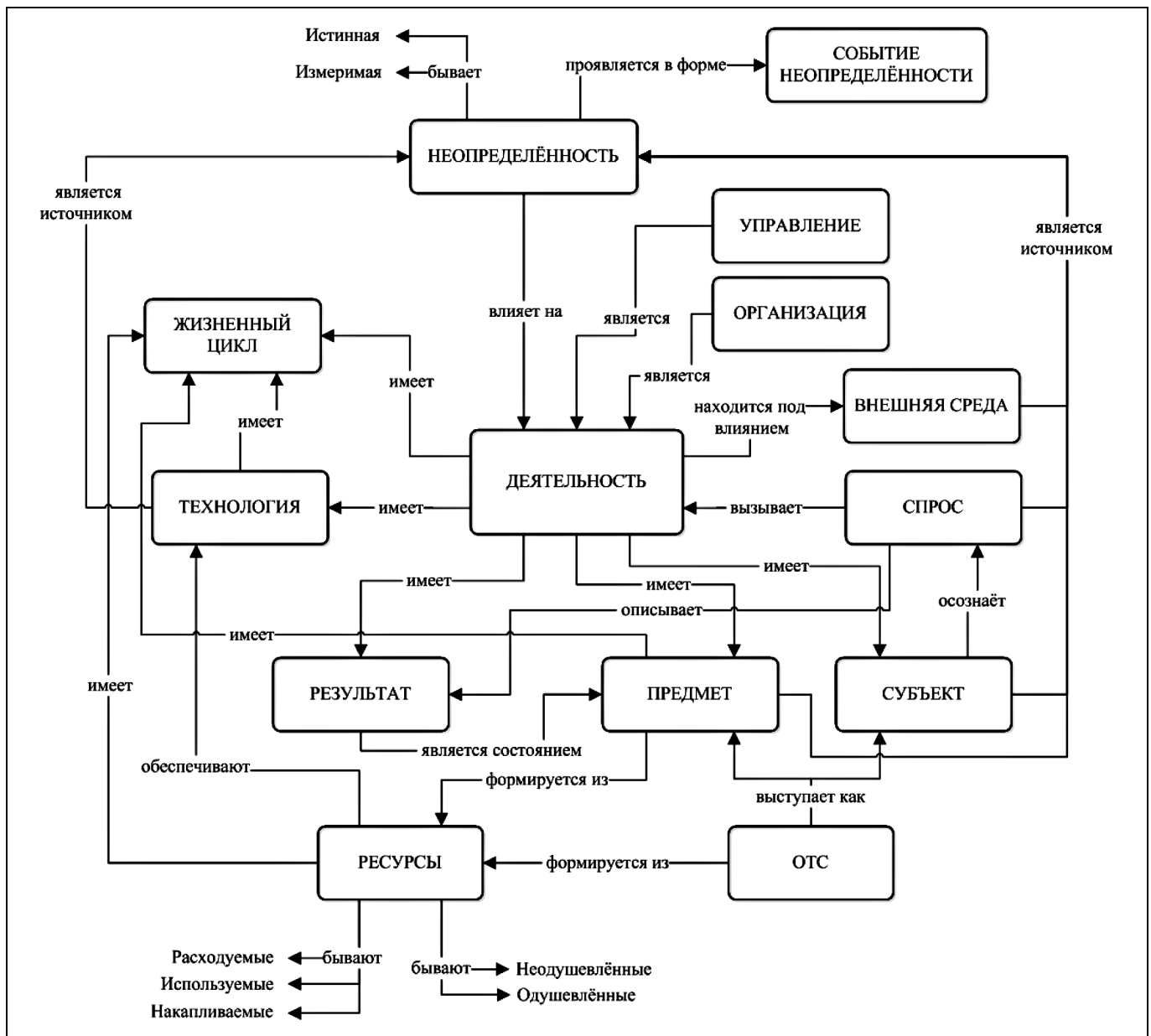


Рис. 1. Онтология теории комплексной деятельности

Классификация элементов КД по отношению к неопределенности и порождению новых элементов деятельности — регулярные, репликативные, креативные элементы КД — приведена в § 2.

Общие закономерности «жизненного цикла элемента КД» и его формальная модель — так называемая *процессная модель КД* — будут изложены во второй части настоящей работы. Там же будет представлен инструментарий ТКД, позволяющий описать и систематизировать процессы организации и управления КД и ОТС и сформулировать единый алгоритм организации и управления как комплексной деятельностью, так и любой осуществляющей ее системой, в состав которой входит человек.

В настоящий момент проблема организации и управления КД и ОТС в широком смысле в той или иной степени рассматривается различными областями знаний: теорией управления организационными системами, фирмами, организациями, проектами и т. д. [13, 14], теорией активных систем [15], теорией иерархических систем [16–18], менеджментом [19], теорией организации [20], методологией [21], кибернетикой [22], системотехникой (Systems Engineering) [7, 8, 23], теорией автоматического управления [24, 25], исследованием операций [26], и многими другими разделами современной фундаментальной и прикладной науки, каждый из которых насчитывает десятки и сотни разнообразных подходов и методов.

Однако существующие теории фокусируются в основном на элементах комплексной деятельности; например, получено большое количество результатов в теории организационных систем, разработано множество инструментов описания бизнес-процессов. Но при этом единая теория комплексной деятельности, рассматривающая деятельность как единую сложную систему, не создана.

Разработке единой ТКД в виде последовательности утверждений и совокупности общих и универсальных моделей и посвящена настоящая работа, которая продолжает методологическое направление, представленное трудами А.М. Новикова и Д.А. Новикова [10, 21] и их коллег. С другой стороны, комплексная деятельность рассматривается как сложная система, поэтому исследование опирается на и развивает подходы и методы теории систем (Systems Science) и ее прикладного расширения — системотехники (Systems Engineering) [7, 8, 11, 12, 23]. В качестве методов исследования (инструментов спецификации) применяются аппарат методологии, системотехники, системного анализа, а также средства и инструменты областей знаний, связанных с анализом и описанием бизнес-процессов (условно — «теорией бизнес-процессов») и управлением проектами, включающих

теоретические основы (см., например, обзор [27]), а также теорию взаимодействующих последовательных процессов Хоара [28, 29], исчисление взаимодействующих систем Милнера [30], алгебру взаимодействующих процессов [31], теорию сетей Петри, календарно-сетевое планирование и управление, дискретно-событийные системы, теорию частичных порядков, темпоральную логику, синхронно и асинхронно взаимодействующие автоматы, теорию процессов [32] и др., «процессный подход» [33] (соотношение ТКД и проектного подхода обсуждается в работе [34]), стандарты системы менеджмента качества (система стандартов ISO 9000:2000 [35]) и др.

1. СТРУКТУРЫ КОМПЛЕКСНОЙ ДЕЯТЕЛЬНОСТИ

Комплексная деятельность имеет «фрактальную» структуру: ее элементы декомпозируются на элементы, которые в свою очередь представляют собой КД, элементы КД образуют многоуровневые иерархии.

1.1. Структурный элемент деятельности

Рассмотрим описывающую элемент КД, модель на основе известной схемы [5, 21] процессуальных компонентов деятельности, отражающих «элементарный цикл деятельности», в котором в соответствии с *технологией* предпринимаются некоторые *действия*, приводящие к определенному *результату* (который в общем случае может и не совпадать с *целью*, как с желательным, предвосхищаемым образом результатом деятельности). Последовательное прохождение этого цикла — от потребности до результата — будем условно называть *реализацией деятельности* или реализацией элемента деятельности. Кроме указанных компонентов, модель элемента КД дополним *субъектом* и *предметом* КД. Для этого введем понятие *структурный элемент деятельности* (СЭД), который будем понимать как имеющий приведенную на рис. 2 структуру объект:

- созданный для достижения определенной цели/получения определенного результата (преобразования предмета деятельности);

- характеризующий деятельность (направленную на получение результата) в соответствии с определенной технологией над определенным предметом;

- субъектом которого является некоторая ОТС.

В терминах *системотехники* (см. рис. 2) деятельность 1 является целевой системой (System-of-Interest), предмет 3 — системой, выражающей операционное окружение (Systems Comprising Operational Environment), субъект 2 — системой, обеспе-

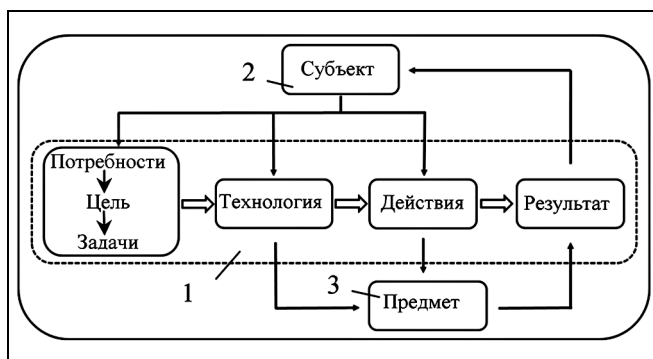


Рис. 2. Модель структурного элемента деятельности

чивающей целевую систему (Enabling System). Все элементы 1–3 образуют единую систему, ассоциированную с КД.

Каждый из элементов триады 1–3 является системой, поэтому триада 1–3 может рассматриваться и как комплексная система, и как *система систем* [36, 37], и как предприятие-система [37].

Поясним семантику стрелок на рис. 2. Стрелка от субъекта к целевому агрегату «потребности → цель → задачи» отражает тот факт, что именно субъект осуществляет целеполагание; стрелка от субъекта к технологии и действиям — что субъект реализует технологию совокупностью действий (действует соответственно технологии). Стрелка от результата к субъекту отражает оценку последним результата, а также саморегуляцию и рефлексию субъекта. Стрелки от технологии и от действий к предмету означают, что предмет изменяется в результате действий соответственно технологии, от предмета к результату — что результат является конечным состоянием предмета, его эволюции в процессе деятельности.

Далее будем пользоваться понятием СЭД в качестве унифицированного формализма, типовой модели элементов КД.

Для корректного построения моделей методически важен частный случай элемента КД, имеющий тривиальное структурное и процессное представления, для его обозначения употребляется термин *элементарная операция*. Он соответствует элементарной деятельности, имеющей единственную цель, не допускающей или не требующей дальнейшей детализации, и у которой все стадии жизненного цикла и/или все процессуальные компоненты, кроме действия, вырождены.

Таким образом, любой СЭД представляет собой композицию других элементов КД — СЭДов и/или элементарных операций, а любая элементарная операция — «вырожденную композицию, состоящую из самой себя».

1.2. Логическая структура комплексной деятельности

Прежде всего, необходимо выбрать основание для выделения структуры деятельности (*структура* системы — совокупность устойчивых связей между ее элементами, поэтому необходимо решить, какой вид связей будет рассматриваться в качестве «первичного»). Любая деятельность «начинается» с *потребности*, порождающей *цель*, и заканчивается *результатом* (как для элементарной операции, так и для КД любого уровня сложности). Потребность, цель и результат являются основными — «целевыми» — компонентами деятельности, поэтому структуру КД будем задавать именно их структурой, т. е. *структурой целей* (так как получаемый результат не всегда совпадает с целью, а потребности не всегда конкретны, и не они, а цель определяет задачи, технологию и действия). Технология вторична по отношению к задачам и целям, поэтому технологическая структура (будучи, тем не менее, более детальной) соответствует структуре целей. Задачи представляют собой декомпозицию целей, их детализацию в конкретных условиях деятельности. Действия также в своей структуре следуют структуре целей. Получаемый результат является следствием действий и технологии, поэтому и его структура отражает структуру целей.

Таким образом, структуры других компонентов КД «повторяют» структуру целей, поэтому будем пользоваться именно последней в качестве основания для логической структуры деятельности — состава СЭДов и «целевых» связей между ними, т. е. структуры СЭДов, которую и будем условно называть *логической структурой*.

Логическая структура КД определяется в виде конечного ациклического графа (рис. 3), отражающего тот факт, что каждый СЭД (и КД в целом как частный случай) декомпозирован на конечное число (в примере на рис. 3 — J) нижестоящих СЭДов и L элементарных операций Op_i .

При этом субъект СЭДа выполняет всех элементарные операции, на которые декомпозирован СЭД, будучи и их субъектом.

Образование иерархических структур СЭДов, декомпозиция одних СЭДов на совокупности нижестоящих по уровню других СЭДов и элементарных операций, отражает *общесистемные свойства* КД; *специфические особенности* сосредоточены в элементарных операциях — именно их содержание отличает одну конкретную КД от другой, имеющей одинаковую структуру.

Так как логическая структура отражает декомпозицию целей, ее граф имеет единственную корневую вершину. Это означает, что для достижения

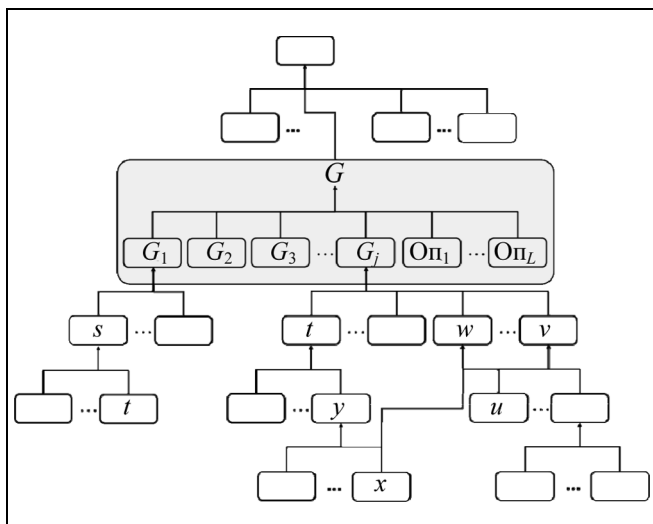


Рис. 3. Логическая структура СЭДа

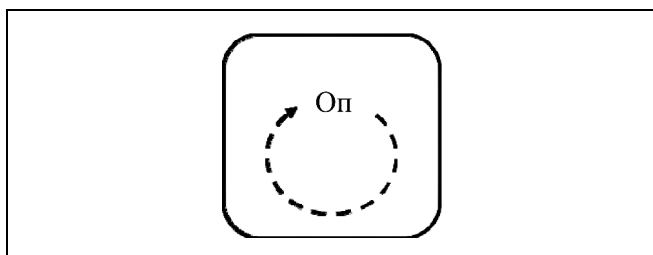


Рис. 4. Логическая структура элементарной операции

цели G необходимо обеспечить достижение подцелей $G_1, G_2, \dots, G_j$ (стрелки направлены от подцелей или операций к цели) для которых определены соответствующие СЭДы, а также некоторого числа подцелей $O_1, \dots, O_L$, достигаемых выполнением элементарных операций.

Не во всех случаях (не для всех СЭДов) достижение *всех* подцелей обязательны для достижения конечной цели G_N . Неопределенность и порождение новых элементов КД объясняют данную особенность, которая рассмотрена в § 2.

Для обеспечения полноты формализма вводится логическая структура элементарной операции, имеющая тривиальный вид (рис. 4). Дуга-петля носит условный характер, иллюстрируя возможность дальнейшей «декомпозиции» подцели O_L только лишь в саму себя.

Логическая структура отражает также отношения источника спроса (он же, как отмечалось выше, является потребителем результата деятельности) и субъекта деятельности: вышестоящий СЭД

и его субъект предъявляют спрос для нижестоящих СЭДов — задают требования к результатам нижестоящих — и являются потребителем их результатов. В ответ нижестоящие СЭДы, обладая определенными возможностями по удовлетворению спроса, накладывают ограничения на реализацию вышестоящих. Таким образом, связи логической структуры отражают также двухсторонние отношения «спрос — возможности (его удовлетворить)», или «*потребности — возможности*» в терминах работы [38].

Логическая структура КД, кроме структуры целей, отражает «управленческую» иерархию подчиненности СЭДов и ответственности субъектов за результаты — за достижение целей. Субъекты вышестоящих СЭДов вместе с ответственностью за достижение «своих» целей, также несут ответственность за достижение нижестоящих целей и, как следствие, за действия в рамках нижестоящих СЭДов. Если какая-либо цель является целью нижестоящего уровня по отношению только к одной цели, то за достижение цели соответствующего СЭДа отвечает соответствующий вышестоящий (непосредственно) СЭД.

Управленческая иерархия (отношение ответственности) применительно к комплексной деятельности и ее логической структуре означает необходимость для субъекта организовывать деятельность, контролировать ее реализацию, решать все возникающие при этом проблемы в зоне его ответственности. С другой стороны, подчиненные субъекты (субъекты нижестоящих СЭДов) обязаны следовать инструкциям вышестоящих и обеспечивать вышестоящих полной и своевременной информацией, в том числе, информировать о возникающих проблемах, которые требуют принятия решений на уровне вышестоящих субъектов, — *эскалировать проблемы* на вышестоящий уровень иерархии. Таким образом, управленческая иерархия задает отношения ответственности между субъектами СЭДов, отношения вложенности, принадлежности областей ответственности, областей эскалирования проблем.

1.3. Причинно-следственная структура комплексной деятельности

Элементы комплексной деятельности находят друг с другом в причинно-следственных отношениях, образующих временную структуру КД, определяемую технологией КД (с учетом неопределенности КД, рассматриваемой далее). Для отражения этих отношений введем причинно-следственную модель КД. На рис. 5 показан пример причинно-следственной модели СЭД, представленной в базовой нотации BPMN [39].

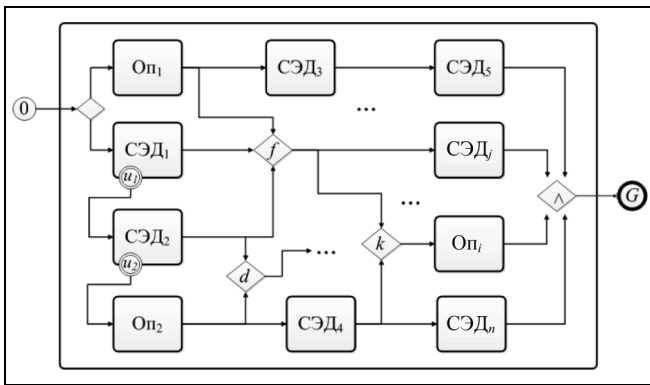


Рис. 5. Причинно-следственная модель комплексной деятельности

Прямоугольники с закругленными углами обозначают нижестоящие элементы деятельности — СЭДы ($СЭД_1, СЭД_2, \dots, СЭД_j$) и элементарные операции ($Оп_1, Оп_2, \dots, Оп_l$); круги — события неопределенности, происходящие во время выполнения элементов деятельности и влияющие на ход ее выполнения; ромбы (например, d, f и k) — контрольные точки (шлюзы), отражающие ветвления (распараллеливания) и слияния выполнения элементов деятельности. Элементы нотации связаны линиями со стрелками, которые отражают причинно-следственные связи и порядок выполнения элементов деятельности.

Так же, как и в нотации BPMN, принимаем концепцию *токенов* (token) [39], в соответствии с которой выполнение действий отражается возникновением и перемещением абстрактных объектов, токенов, от одного элемента причинно-следственной модели к другому, следуя стрелкам-связям.

Начало выполнения СЭДа в целом (см. рис. 5) моделируется появлением единственного токена в элементе модели, отражающем начальное событие (обозначено «0»), выполнению следующих за начальным событием элементов (в рассматриваемом примере это параллельное выполнение $Оп_1$ и $СЭД_1$) будет соответствовать исчезновению токена в «событии 0» и порождение токенов в элементах $Оп_1$ и $СЭД_1$. По завершению $Оп_1$ и $СЭД_1$ токены исчезают в них и порождаются в следующих за ними элементах, как бы двигаясь, следуя стрелкам, и т. д.

События u_1 и u_2 соответствуют реализации неопределенности в ходе выполнения элементов $СЭД_1$ и $СЭД_2$, порождению и параллельному выполнению элементов КД — появлению токенов в элементах u_1 и потом $СЭД_2$, u_2 и потом в $Оп_2$ со-

ответственно. Спецификация причинно-следственной модели включает в себя описание событий u_1 и u_2 .

Терминальное событие G (обозначено толстой окружностью) отражает достижение конечной цели моделируемого СЭДа и завершение его деятельности. Это событие наступит, когда все порожденные токены «переместятся» в элемент G .

Семантика причинно-следственной структуры предполагает, что необходимость последовательного достижения целей отражается их последовательным расположением — например, целей элементов $СЭД_3$ и $СЭД_5$. Условия достижения комбинаций нескольких параллельных целей задаются соответствующими правилами — конъюнктивно-дизъюнктивными формами. Например, для достижения цели f может потребоваться достижение пары целей $СЭД_1$ и $Оп_1$ или цели $СЭД_2$ ($(СЭД_1 \wedge Оп_1) \vee СЭД_2$). При этом последовательность достижения целей $СЭД_1$ и $Оп_1$ не регламентируется, так как в данном случае они представлены на диаграмме «параллельными» элементами. После того, как оказывается выполненной какая-либо из промежуточных целей, начинается выполнение целей — СЭДов и элементарных операций, следующих за ней (например, в соответствии с рис. 4, после выполнения цели f начнется выполнение цели $СЭД_j$).

Комплексность деятельности предполагает возможность параллельного выполнения множества ее элементов, поэтому в рамках СЭДа может параллельно выполняться несколько нижестоящих СЭДов и элементарных операций. Это отражается и наличием нескольких исходящих стрелок-связей из элементов, и возможностью наступления неопределенных событий (на рис. 4 — u_1 и u_2), которые могут быть инициированы в ходе выполнения СЭДа через механизм реакции на неопределенность (см. далее § 2).

Причинно-следственная структура отражает технологические связи комплексной деятельности, будучи фактически общесистемной технологией КД, в то время как элементарные операции (их содержание) и сами причинно-следственные связи представляют специфическую часть КД.

Все вершины логической структуры СЭДа, являющиеся целями или подцелями, должны быть отражены в причинно-следственной модели СЭДа для задания причинно-следственной логики их достижения. Отметим, что корневая вершина логической структуры совпадает с терминальным событием причинно-следственной модели. Дуги логической (отношения декомпозиции целей и подчиненности) и причинно-следственной (отно-

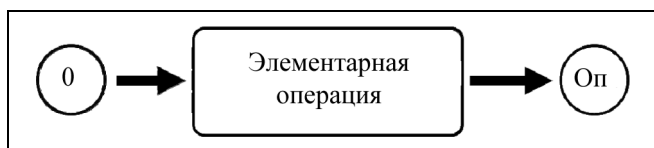


Рис. 6. Причинно-следственная модель элементарной операции

шения причинности и следования) структуры имеют в общем случае независимую друг от друга конфигурацию.

Причинно-следственная модель элементарной операции приведена на рис. 6.

Имеет смысл сравнить причинно-следственную модель КД с аналогичными моделями и инструментами из областей знаний, изучающих организационно-технические системы. К таким инструментам относятся многочисленные формальные модели *теории процессов*, модели производственных и бизнес-процессов и, как частный, но очень важный случай, *сетевые графики* (и их различные вариации).

Причинно-следственная модель КД аналогична упомянутым инструментам в том смысле, что, как и они:

- является структурной моделью и отражает причинно-следственные связи между элементами;
- допускает сложный характер самих элементов;
- для ее построения применяются модели теории графов.

Однако семантически она существенно отличается от сетевых графиков и инструментов теории процессов:

- формализм СЭДов более «богатый» (представляет цели, технологии, действия, субъект и предмет) по сравнению с элементами сравниваемых инструментов (которые описывают действия — работы, процессы);
- естественное свойство причинно-следственной модели заключается в ее изменчивости в ходе реализации КД, описываемой ею; фактически она «изменяет саму себя», что вызвано базовыми свойствами КД — неопределенностью и порождением элементов, их эволюцией во времени (см. § 2).

2. НЕОПРЕДЕЛЕННОСТЬ И ПОРОЖДЕНИЕ КОМПЛЕКСНОЙ ДЕЯТЕЛЬНОСТИ

Неопределенность имеет существенное значение для любой сложной системы, особенно для КД.

Неопределенностью КД будем называть возможность наступления в ходе КД таких событий, влияющих на реализацию КД и на ее результат, которые могут наступить, а могут и не наступить.

Следствием неопределенности КД является невозможность априори предсказать характеристики результата деятельности, момент его получения и усилия (ресурсы), которые будут для этого затрачены.

Следуя идеям F. Knight [40], будем, прежде всего, разделять измеримую и истинную неопределенности и далее рассматривать неопределенность как источник не только негативных проблем, но также и развития, в случае КД — как источник порождения спроса и, следовательно, элементов деятельности.

Измеримую неопределенность КД определим как возможность наступления описываемых некоторыми известными закономерностями событий. Для анализа таких событий могут быть применены количественные методы (например, вероятностные/статистические), основанные на предыдущих измерениях или фундаментальных законах (вместе с предположением о неизменности условий и закономерностей).

Истинной неопределенностью КД назовем возможность наступления уникальных (или редко повторяющихся) событий, которые не объясняются существующими фундаментальными законами и для которых нет априорных наблюдений.

Принципиальное отличие истинной неопределенности от измеримой заключается в том, что события первой из них возникают вследствие непознанных факторов (частый и важный, но частный случай — активный выбор индивида), в то время как события второй, хотя и непредсказуемы, описываются известными закономерностями.

Категория *неопределенности* (uncertainty) является предметом *теории систем*, наряду с категориями *сложности* (complexity) и *эмерджентности* (emergency), которые традиционно рассматриваются в качестве основных характерных особенностей сложных систем. Факторы сложности и эмерджентности свойственны также комплексной деятельности, а операционно они проявляются через неопределенность (в том числе свойств и поведения систем). Следовательно, категория неопределенности может рассматриваться достаточной для адекватного учета факторов сложности и эмерджентности, поэтому будем оперировать только категорией неопределенности.

Неопределенность КД может порождаться различными источниками, связанными со всеми процессуальными компонентами КД:

- неопределенность (вносимая элементами) внешней среды — неопределенность внешнего спроса и внешних условий, требований и норм;
- неопределенность (вносимая элементами) технологии и предмета — неопределенность средств, методов и факторов;

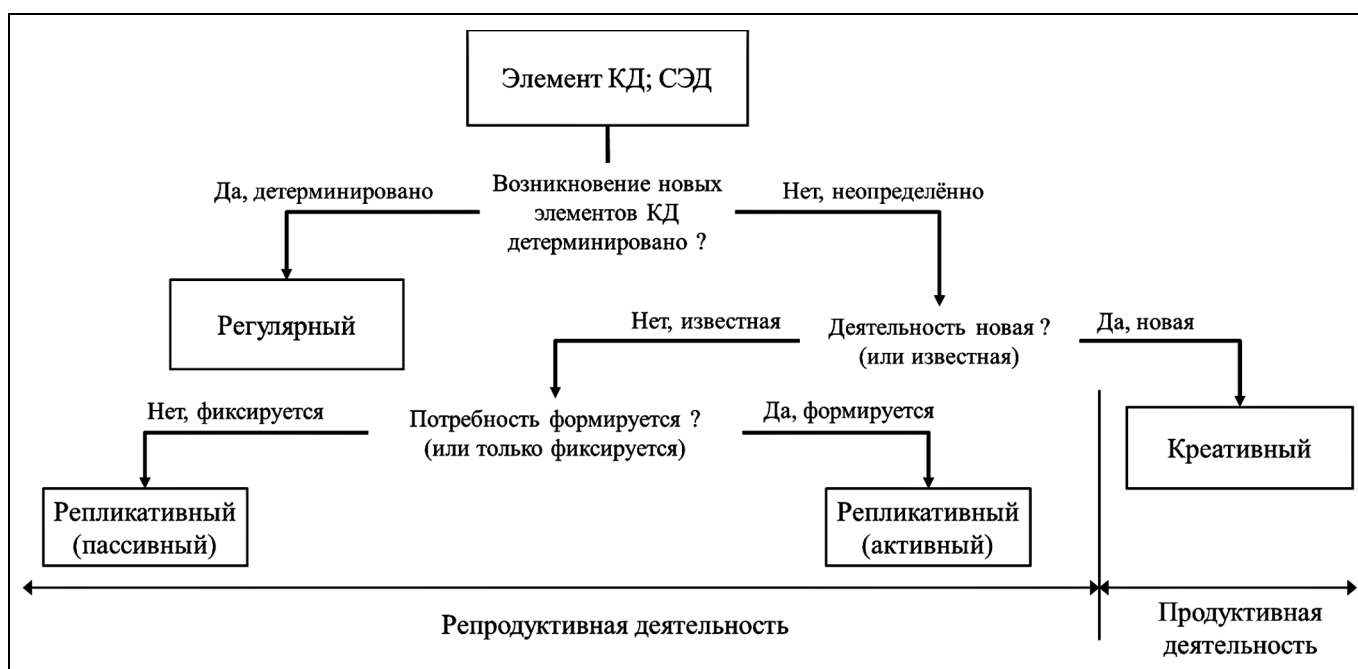


Рис. 7. Классификация КД и СЭДов по основанию порождения новых элементов КД

- неопределенность (активного поведения и рационального выбора, в том числе рефлексии) субъекта — неопределенность осознания внешней потребности, целеполагания, осуществления действий, оценивания результата, и наконец (или, прежде всего) принятия принципиального решения о том, выступать ли ему в роли субъекта той или иной конкретной КД.

В каждой из трех перечисленных групп могут встречаться случаи и измеримой, и истинной неопределенности (см. примеры в работе [5]).

При реализации комплексной деятельности одни элементы деятельности завершаются удовлетворением спроса/потребностей и не влекут каких-либо априори неизвестных последствий в смысле деятельности. В других случаях в ходе реализации одних СЭДов порождаются новые СЭДы и иерархии СЭДов под ними: в рамках выполнения некоторого СЭДа возникает спрос, далее в рамках другого СЭДа происходит актуализация новой потребности. Выявленная потребность влечет новое целеполагание — организацию новой деятельности — формирование нового СЭДа и иерархии нижестоящих в логической структуре СЭДов: «новая деятельность» может быть направлена или на удовлетворение потребности, или на выявление целесообразности (для субъекта деятельности) удовлетворения потребности (например, проверка прибыльности заказа). По мере завершения элементов деятельности, достижения результата, прекращается существование СЭДов.

Возникновение спроса может быть вызвано одним из двух факторов:

- осуществлением детерминированной декомпозиции уже существующего СЭДа соответственной известной технологии;
- наступлением неопределенных событий, вызывающих порождение элементов деятельности.

Приведем классификацию КД и СЭДов по общесистемному основанию: порождению новых элементов КД, новых СЭДов (рис. 7) [5].

1. «Регулярная» КД («регулярные» СЭДы) — новые элементы деятельности возникают только вследствие детерминированной декомпозиции вышестоящих СЭДов соответственно априори известной технологии (что составляет детерминированный спрос). Структура и технология регулярной КД детерминированные. Один из ключевых трендов текущего этапа развития человечества заключается в том, что любая деятельность целенаправленно подвергается максимальным, насколько это возможно, упрощению, стандартизации и регламентации. *Стандартизация и регламентация деятельности* — одно из основных средств повышения ее эффективности. Автоматизация и роботизация, в свою очередь, продолжают этот тренд, заменяя деятельность машинными операциями, вытесняя человека из области рутинных (регламентированных) операций и приводя к исчезновению многих профессий. В течение последних десятилетий таким образом исчезли профессии те-

лефонистки, машинистки, типографских рабочих, операторов на многих производствах.

Вероятно, всю регулярную деятельность можно рассматривать в качестве потенциально *полностью автоматизируемой*, при этом выбранное основание (детерминированная технология и порождение только априори известных элементов деятельности) определения регулярной деятельности может быть критерием разделения потенциально автоматизируемой и неавтоматизируемой деятельности.

Фактически, регуляризация КД служит «познанием КД» — фиксацией и систематизацией знаний о КД.

Технология регулярной деятельности стабильная, многократно проверенная и не требует изменений, поэтому технологическая неопределенность может быть вызвана только событиями неадекватности средств, методов, факторов задачам КД. На практике такие события сводятся к сбоям и отказам оборудования, получением некачественных партий сырья, материалов и комплектующих изделий и аналогичным случаям.

2. «Репликативная (пассивная)» КД и «репликативные (пассивные)» СЭДы. Деятельность, в результате которой порождается деятельность известного типа (известные потребности, цели и технологии), а нетривиальная составляющая деятельности заключается не в формировании, а в фиксации неопределенного спроса. Если спрос имеет характер массового, типового, повторяющегося явления, его не требуется создавать, а нужно только фиксировать, как это бывает для пассивной репликативной деятельности, то основная неопределенность деятельности измерима.

3. «Репликативная (активная)» КД и «репликативные (активные)» СЭДы. В ходе такой деятельности производится формирование неопределенного спроса, формируется новая потребность известного типа (возможно, у нового потребителя) и, как следствие, новая деятельность известного типа. В случае создания или формирования спроса (активная репликативная КД) он не может считаться типовым и повторяющимся, во всяком случае, на начальном этапе. Поэтому неопределенность в этом случае истинная, и основная неопределенность активной репликативной деятельности является истинной.

4. «Креативная» КД и «креативные» СЭДы — деятельность, в результате которой порождается неопределенный априори спрос на результаты неизвестной априори деятельности, технологию которой необходимо создать в ходе этой новой деятельности. Неизвестность технологии вызвана неопределенностью спроса и/или априорной неопределенностью спецификации результата де-

ятельности. Это — деятельность по получению результата, который не до конца специфицирован в начале деятельности. В этот класс попадает деятельность главных конструкторов и технологов, научных сотрудников, продюсеров фильмов и спектаклей, партнеров юридических фирм и др. Субъекты креативной КД самостоятельно определяют структуру и характеристики комплексного результата и, следовательно, структуру и технологию деятельности. Фактически они являются *архитекторами деятельности* (как системы) и архитекторами создаваемого результата (как системы). Принципиальное отличие креативной КД от репликативной и регулярной КД состоит в наличии в структуре первой, по крайней мере, одного фрагмента, предметом деятельности которого служит технология другого («нижестоящего») фрагмента КД, что является следствием необходимости создания новой технологии в рамках исполнения деятельности.

Ключевое свойство креативной КД заключается в том, что основная неопределенность является истинной и вызывается композицией неопределенностей внешней среды, технологии и субъекта. Априорная неизвестность технологии выражается в наличии в структуре креативного СЭДа специфических СЭДов, отражающих создание априори неизвестной технологии.

ЗАКЛЮЧЕНИЕ

Введено понятие комплексной деятельности, рассмотрены структуры комплексной деятельности, неопределенность и порождение комплексной деятельности.

Предложенные модели теории КД позволяют описать любую комплексную деятельность как организованную и объединенную общим целеполаганием, логической и причинно-следственной структурами совокупность элементов двух типов [5]:

- специфических элементарных операций (представляющих элементарную деятельность);

- единственного или нескольких управляющих элементов деятельности, которые реализуют (или постоянную, или многократную, или непрерывную) проверку наступления определенных условий и инициирование соответствующих специфических элементов деятельности, а также осуществляют установление связей между субъектом КД в целом, ресурсами и субъектами нижестоящих специфических элементов.

Во второй части статьи будут рассмотрены жизненные циклы комплексной деятельности, организация и управление как комплексная деятельность.



ЛИТЕРАТУРА

1. *Burkov V.*, et al. Mechanism Design and Management: Mathematical Methods for Smart Organizations / Ed. by prof. D. Novikov. — N.-Y.: Nova Scientific Publishing, 2013. — 204 p.
2. *Boardman J., Sauser B.* Systems Thinking: Coping with 21st Century Problems. — Boca Raton: CRC Press, 2008. — 240 p.
3. *Checkland P.* Systems Thinking, Systems Practice. — N.-Y.: John Wiley & Sons, 1999. — 424 p.
4. *Belov M.* General-System Modelling Framework of Complex Activity // 10th European Modelling Symposium on Mathematical Modelling and Computer Simulation. Ed. D. Al-Dabbas, V. Colla, M. Vannucci, A. Pantelous. IEEE Conference Record No. 40530. Pisa, 2016. — 241 p.
5. *Белов М.В., Новиков Д.А.* Методология комплексной деятельности. — М.: Ленанд, 2018. — 320 с.
6. *Белов М.В., Новиков Д.А.* Структура теории комплексной деятельности // Философия науки. — 2018 (в печати).
7. *ISO/IEC/IEEE 15288:2015* Systems and Software Engineering — System Life Cycle Processes.
8. *The Guide to the Systems Engineering Body of Knowledge (SEBoK)*, v. 1.2 / Ed. by A. Pyster, A. Olwell. Hoboken: The Trustees of the Stevens Institute of Technology, 2013. — URL: <http://www.sebokwiki.org> (дата обращения: 20.12.2016).
9. *White B.* Fostering Intra-Organizational Communication of Enterprise Systems Engineering Practices. MITRE Public Release Case No. 06-0351. National Defense Industrial Association, 9th Annual Systems Engineering Conference. San Diego, 2006. — 25 p.
10. *Новиков А.М., Новиков Д.А.* Методология научного исследования. — М.: Либроком, 2010. — 280 с.
11. *ISO/IEC/IEEE 4201:2011* Systems and Software Engineering — Architecture Description.
12. *INCOSSE Systems Engineering Handbook Version 3.2.2* — A Guide for Life Cycle Processes and Activities / Ed. by C. Haskins. — San Diego: INCOSSE, 2012. — 376 p.
13. *Бурков В.Н., Коргин Н.А., Новиков Д.А.* Введение в теорию управления организационными системами / под ред. чл.-корр. РАН Д.А. Новикова. — М.: Либроком, 2009. — 264 с.
14. *Новиков Д.А.* Теория управления организационными системами: 3-е изд. — М.: Физматлит, 2012. — 604 с.
15. *Бурков В.Н.* Основы математической теории активных систем. — М.: Наука, 1977. — 255 с.
16. *Гермейер Ю.Б.* Игры с противоположными интересами. — М.: Наука, 1976. — 327 с.
17. *Месарович М., Мако Д., Такахара И.* Теория иерархических многоуровневых систем. — М.: Мир, 1973. — 344 с.
18. *Моисеев Н.Н.* Математические задачи системного анализа. — М.: Наука, 1981. — 488 с.
19. *Гэлловэй Л.* Операционный менеджмент: Принципы и практика: пер. с англ. — СПб.: Питер, 2001. — 319 с.
20. *Богданов А.А.* Всеобщая организационная наука. — М.: Экономика, 1989. — Т. 1. — 304 с. — Т. 2. — 351 с.
21. *Новиков А.М., Новиков Д.А.* Методология. — М.: СИНТЕГ, 2007. — 668 с.
22. *Новиков Д.А.* Кибернетика: Навигатор. История кибернетики, современное состояние, перспективы развития. — М.: Ленанд, 2016. — 160 с.
23. *Rebovich G., White B.* Enterprise Systems Engineering: Advances in the Theory and Practice. — Boca Raton: CRC Press, 2011. — 459 p.
24. *Ким Д.П.* Теория автоматического управления: 2-е изд. — М.: Физматлит, 2007. — Т. 1: Линейные системы. — 310 с. — Т. 2: Многомерные, нелинейные, оптимальные и адаптивные системы. — 440 с.
25. *Bubnicki Z.* Modern Control Theory. — Berlin: Springer, 2005. — 423 p.
26. *Wagner H.* Principles of Operations Research: 2-nd ed. — NJ Upper Saddle River: Prentice Hall, 1975. — 1039 p.
27. *Калянов Г.Н.* Модели и методы теории бизнес-процессов (обзор) // Открытое образование. — 2015. — № 6 (113). — С. 4—9.
28. *Hoare C.* Communicating Sequential Processes. — N.-Y.: Prentice Hall, 1985. — 256 p.
29. *Hoare C.* Communicating Sequential Processes // Communications of the ACM. — 1978. — Vol. 21, iss. 8. — P. 666—677.
30. *Milner R.* A Calculus of Communicating Systems // Lecture Notes in Computer Science. — Heidelberg: Springer Verlag. — 1980. — Vol. 92. — 171 p.
31. *Bergstra J.* Process Algebra for Synchronous Communication // Information and Control. — 1984. — Vol. 60. — P. 109—137.
32. *Мионов А.М.* Теория процессов. — Переславль-Залесский: Университет г. Переславля, 2008. — 345 с.
33. *Репин В.В., Елиферов В.Г.* Процессный подход к управлению. Моделирование бизнес-процессов. — М.: Манн, Иванов и Фербер, 2013. — 544 с.
34. *Белов М.В.* Проблемы согласованного управления жизненными циклами организационно-технических систем // Управление большими системами. — 2018 (в печати).
35. *ISO 9000:2000* Системы менеджмента качества. Основные положения и словарь.
36. *Ackoff R.* Towards a Systems of Systems Concepts // Management Science. — 1971. — Vol. 17, N 11. — P. 661—671.
37. *Gorod A., Gandhi S., White B.*, et al. Modern History of System of Systems, Enterprises, and Complex Systems / Case Studies in System of Systems, Enterprise Systems, and Complex Systems Engineering. — Boca Raton: CRC Press, 2014. — P. 3—32.
38. *Rzevski G., Skobelev P.* Managing Complexity. — London: WIT Press, 2014. — 216 p.
39. *Business Process Model and Notation (BPMN)*, v. 2.0.2. — URL: <http://www.omg.org/spec/BPMN/2.0> (дата обращения: 27.04.2018).
40. *Knight F.* Risk, Uncertainty and Profit / Hart Schaffner and Marx Prize Essays. No. 31. — Boston — N.-Y.: Houghton Mifflin, 1921. — 381 p.

Статья представлена к публикации руководителем РРС А.А. Ворониным.

Белов Михаил Валентинович — канд. техн. наук, зам. ген. директора, компания ИБС, г. Москва, mbelov59@mail.ru,

Новиков Дмитрий Александрович — чл.-корр. РАН, директор, Институт проблем управления им. В.А. Трапезникова РАН, г. Москва, novikov@ipu.ru.

Не забудьте подписаться!

Подписку на журнал «Проблемы управления» можно оформить в любом почтовом отделении (подписной индекс 81708 в каталоге Роспечати или 38006 в объединенном каталоге «Пресса России»), а также через редакцию с любого месяца, при этом почтовые расходы редакция берет на себя. Отдельные номера редакция высылает по первому требованию.

ЗАДАЧА ОПТИМАЛЬНОГО РАЗМЕЩЕНИЯ ИНФОРМАЦИОННО-ТЕХНОЛОГИЧЕСКОГО РЕЗЕРВА В РАСПРЕДЕЛЕННЫХ СИСТЕМАХ ОБРАБОТКИ ДАННЫХ

С.К. Сомов

Отмечено, что информационно-технологический резерв — это новый вид информационного резервирования, применение которого в распределенных автоматизированных информационно-управляющих системах позволяет повысить эффективность функционирования таких систем при обработке типовых запросов пользователей. Сформулирована задача оптимального размещения идентичных копий информационно-технологического резерва по узлам распределенной системы в форме минисуммной задачи поиска p -медианы графа. Предложен алгоритм ее решения и приведен пример решения. Выполнен краткий анализ работы алгоритма решения.

Ключевые слова: распределенная система обработки данных, компьютерная сеть, информационно-технологический резерв.

ВВЕДЕНИЕ

Данная статья является продолжением предыдущей статьи [1]. В ней рассматривается проблема размещения копий информационно-технологического резерва (ИТР) по узлам компьютерной сети, на основе которой работает распределенная система обработки данных (РСОД). Информационно-технологический резерв представляет собой обобщение и дальнейшее развитие методов структурно-технологического резервирования, применение которого целесообразно в крупномасштабных автоматизированных информационно-управляющих системах с большим числом пользователей.

Характеристики и особенности обработки регулярных запросов пользователей систем данного класса позволяют разбить эти запросы на несколько различных групп, объединяющих запросы разного типа. Каждая такая группа запросов характеризуется некоторой инвариантной частью технологии обработки запросов, включенных в данную группу, и подмножеством данных, используемых при обработке этих запросов.

Инвариантную часть технологии можно оформить в виде подмножества процедур обработки данных, выполняемых в определенной последовательности. Это подмножество процедур, предназначенных для обработки регулярных запросов

определенного типа, можно выполнить заранее, до поступления в систему запросов. В результате будет сформирован промежуточный массив (массивы) данных, который может быть впоследствии использован при обработке поступивших запросов из группы соответствующего типа.

Подмножество процедур, а также сформированные ими промежуточные массивы данных составляют ИТР, используемый для ускорения обработки регулярных, типовых запросов пользователей. Копии ИТР, размещенные в узлах компьютерной сети, служат информационной основой для функционирования РСОД в будущем при обработке регулярных запросов пользователей. При этом обработка запросов будет вестись с использованием созданных заранее промежуточных массивов данных. Благодаря использованию данных промежуточных массивов ИТР обеспечивается повышение эффективности функционирования распределенной системы, так как значительно снижаются затраты времени на обработку регулярных запросов пользователей.

При регулярном создании полных копий (full backup) ИТР появляется возможность обработки средствами РСОД запросов пользователей к историческим данным. Кроме того, хранящиеся в узлах сети полные копии ИТР можно использовать также для восстановления данных и процедур ИТР в случае их частичного или полного разрушения.



Информационно-технологический резерв является составной частью системы обработки данных, поэтому проектирование и создание ИТР ведется в рамках процесса проектирования и создания всей РСОД.

Процесс проектирования и создания ИТР в рамках РСОД состоит из трех этапов [1].

На *первом этапе* выполняется анализ предметных областей пользователей системы, в ходе которого уточняются входные, промежуточные и итоговые данные, процедуры обработки этих данных и последовательность выполнения этих процедур при обработке типовых запросов пользователей.

На *втором этапе* на основе данных, полученных на первом этапе, осуществляется синтез информационного и программного обеспечения ИТР (массивы используемых входных, промежуточных и выходных данных и наборы процедур обработки данных).

На *третьем этапе* решается задача оптимального распределения ИТР (в виде определенного числа его копий) по узлам компьютерной сети, на базе которой работает РСОД. Копии ИТР размещаются по узлам компьютерной сети в целях повышения производительности, надежности и снижения стоимости функционирования системы (благодаря уменьшению затрат на обработку запросов).

Принципы и методы формирования структурно-технологического резерва подробно описаны в работе [1].

В настоящей статье описан третий этап процесса проектирования и создания ИТР для использования в распределенных системах. Дана формулировка и приведено решение задачи поиска оптимального размещения копий ИТР по узлам компьютерной сети. Оптимальное размещение копий ИТР позволяет предоставить пользователям системы кратчайший доступ к необходимой им информации и обеспечивает наименьшее время отклика системы на их запросы благодаря предварительно подготовленным промежуточным данным, хранящимся в ИТР.

1. ПОСТАНОВКА ЗАДАЧИ ОПТИМАЛЬНОГО РАЗМЕЩЕНИЯ ИНФОРМАЦИОННО-ТЕХНОЛОГИЧЕСКОГО РЕЗЕРВА В РАСПРЕДЕЛЕННОЙ СИСТЕМЕ ОБРАБОТКИ ДАННЫХ

Предположим, что РСОД работает на базе компьютерной сети, состоящей из J узлов (серверов), топология которой задана и которую можно представить в виде графа $G = (X, \Gamma)$. Каждый сервер x_j , $x_j \in X$, $j = \overline{1, J}$, $J = |X|$, сети может выступать в роли пользовательского сервера и обслуживать некоторое число пользователей системы (ввод пользователями запросов и выдача им результатов обработки этих запросов). Также сервер сети может вы-

полнять роль сервера ИТР, в котором размещена копия ИТР и который обрабатывает запросы пользователей, полученные от пользовательских серверов. При этом каждый запрос пользователя можно отнести к одному из N типов запросов. Допустим также, что в соответствии с некоторой процедурой определены маршруты, по которым передаются сообщения между узлами сети (запросы пользователей и ответы на эти запросы). Предполагается, что копии ИТР могут быть размещены в нескольких узлах сети. Необходимо определить такое подмножество $\tilde{X}_p$ из p узлов множества X всех узлов сети, размещение в которых p копий ИТР обеспечит минимум стоимости эксплуатации РСОД при обработке регулярных запросов пользователей.

Обозначим через $\Lambda = \|\lambda_j^n\|$ матрицу размерности $N \times J$, где λ_j^n — интенсивность запросов n -го типа, формируемых пользователями в j -м узле сети в единицу времени.

Каждый запрос пользователя посылается из узла i (пользовательский сервер) в узел j , $i, j = \overline{1, J}$, (сервер ИТР) при условии, что в j -м узле сети расположена копия ИТР. Результат обработки каждого запроса (ответ на запрос) посылается из узла j обратно в узел i . Предполагается, что РСОД работает в установившемся режиме и все сформированные пользователями запросы обрабатываются.

Введем обозначения:

$D = \{d_1, \dots, d_n, \dots, d_N\}$ — кортеж средних объемов данных, которыми обмениваются пользовательский сервер и сервер ИТР при обработке запроса n -го типа, $n = \overline{1, N}$;

$V = \{v_1, \dots, v_j, \dots, v_J\}$ — кортеж «весов» узлов компьютерной сети, в котором v_j — средний объем данных, который узел j передает и получает за единицу времени при обработке всех запросов, инициированных в этом узле:

$$v_j = \sum_{n=1}^N d_n \lambda_j^n;$$

$\text{dis}(x_i, x_j)$ — длина кратчайшего пути из узла x_i в узел x_j ;

s — стоимость передачи единичного объема данных по пути единичной длины;

$\text{cost}(x_j)$ — стоимость хранения копии ИТР, размещенной в узле x_j за единицу времени; она включает в себя затраты на использование технических средств для хранения информации и затраты на репликацию изменений в данных ИТР для поддержки в актуальном состоянии всех копий ИТР, размещенных в узлах РСОД (вопросы реализации

механизма репликации изменений в данных ИТР выходит за рамки настоящей статьи).

Для поиска оптимального размещения копий ИТР по узлам сети предположим, что граф $G = (X, \Gamma)$ неориентированный и сформулируем обобщенную минисуммную задачу о p -медиане, приняв в качестве критерия оптимизации минимум затрат на функционирование РСОД [2, 3].

Определим для узла x_i передаточное число

$$\sigma(x_i) = \sum_{x_j \in X} [v_j \text{dis}(x_i, x_j) s], \quad (1)$$

т. е. передаточное число представляет собой сумму стоимости обмена данными между узлом x_i со всеми другими узлами множества X при обработке запросов за единицу времени.

Вершина $\bar{x}$, для которой достигается минимум передаточного числа

$$\sigma(\bar{x}) = \min_{x_i \in X} \{\sigma(x_i)\} = \min_{x_i \in X} \left\{ \sum_{x_j \in X} [v_j \text{dis}(x_i, x_j) s] \right\},$$

называется медианой графа G .

Обозначим через X_p подмножество множества вершин X графа G , состоящее из p вершин, а через $d(X_p, x_j)$ — минимальное расстояние от вершины x_j множества X до одной из вершин множества X_p , т. е.

$$d(X_p, x_j) = \min_{x_i \in X_p} [\text{dis}(x_i, x_j)]. \quad (2)$$

Если обозначить через x_i^* вершину из множества X_p , на которой достигается минимум в выражении (2), то можно говорить о том, что точка x_j в выражении (2) «прикреплена» к точке x_i^* . Применительно к рассматриваемой задаче это означает, что запросы, возникающие в узле x_j , маршрутизируются для обработки в узел x_i^* с копией ИТР. Данный факт можно обозначить как $x_j \rightarrow x_i^*$.

По аналогии с формулой (1) определим передаточное число не для одной вершины, а для подмножества вершин X_p [2]:

$$\begin{aligned} \sigma(X_p) &= \sum_{x_j \in X} v_j s d(X_p, x_j) = \\ &= \sum_{x_j \in X} \left\{ v_j s \min_{x_i \in X_p} [\text{dis}(x_i, x_j)] \right\}. \end{aligned} \quad (3)$$

С учетом принятых обозначений задачу оптимального размещения ИТР по узлам РСОД можно сформулировать следующим образом: для неориентированного графа $G = (X, \Gamma)$ требуется найти такое подмножество $\tilde{X}_p$, состоящее из p вершин

этого графа, для которого достигается минимум функционала

$$F_p(\tilde{X}_p) = \sum_{x_i \in \tilde{X}_p} \text{cost}(x_i) + \sigma(\tilde{X}_p). \quad (4)$$

Первое слагаемое в выражении (4) определяет стоимость хранения копий ИТР, размещенных в p вершинах подмножества $\tilde{X}_p$ графа. Второе слагаемое представляет собой передаточное число подмножества $\tilde{X}_p$, которое определяется как

$$\sigma(\tilde{X}_p) = \min_{X_p \subseteq X} [\sigma(X_p)],$$

где, в свою очередь, передаточное число $\sigma(X_p)$ для подмножества X_p вычисляется по формуле (3).

В данной задаче могут быть ограничения. Например, может быть ограничение «сверху» на число p медиан, т. е. можно задать ограничение на максимальное число копий ИТР, размещенных в узлах компьютерной сети, на базе которой работает РСОД: $|\tilde{X}_p| \leq \bar{p}$.

Еще одним примером ограничения в данной задаче служит ограничение на средний объем информации, обрабатываемой сервером с копией ИТР за единицу времени, т. е. ограничение на объем данных в запросах всех N типов, полученных сервером за единицу времени из «прикрепленных» к нему узлов, и объем данных, возвращаемых сервером в виде ответов на эти запросы:

$$\sum_{x_j \in X} \sum_{x_i^* \rightarrow x_j} d_j \lambda_j^i \leq \bar{D}, \quad j = \overline{1, J}.$$

Для решения сформулированной задачи можно воспользоваться различными известными методами (например, методами целочисленного программирования, эвристическими методами и др.) [2–4].

Отметим, что поиск оптимального размещения копий ИТР по узлам компьютерной сети на основе сформулированной обобщенной минисуммной задачи о p -медиане не является единственным вариантом решения рассматриваемой проблемы. В литературе можно найти обзоры и описания других подходов к решению задачи оптимального размещения копий (реплик) данных в узлах компьютерных сетей, в таких областях, как, например, Content Delivery Network, Data Grid, Cloud Computing, распределенные базы данных [5–8].

2. АЛГОРИТМ РЕШЕНИЯ ЗАДАЧИ

Как уже отмечалось, для решения сформулированной минисуммной задачи о p -медиане мож-



но воспользоваться различными методами [9–11]. В качестве примера рассмотрим только один из возможных алгоритмов решения задачи оптимального размещения ИТР в РСОД. Он основан на эвристическом методе поиска p -медиан взвешенного графа, предложенном в работе [4]. Данный метод заключается в следующем.

Сначала произвольным образом из множества всех вершин графа X выбираются p вершин, которые образуют исходное множество X_p медианных вершин, которое с некоторой точностью аппроксимирует искомое p -медианное множество $\tilde{X}_p$. Затем делается проверка, имеет ли множество X такую вершину x_j , которая не входит в множество X_p , $x_j \in X \setminus X_p$, и которая может заменить некоторую медианную вершину $x_i \in X_p$. Для проверки возможности такой замены строится еще одно множество вершин $X'_p = (X_p \cup \{x_j\}) \setminus \{x_i\}$. Затем вычисляются и сравниваются между собой два передаточных числа для множества X'_p , в котором произошла замена вершины x_i на вершину x_j , и для исходного множества X_p : $\sigma(X'_p)$ и $\sigma(X_p)$.

Если при этом для передаточных чисел выполняется неравенство $\sigma(X'_p) < \sigma(X_p)$, то вершина x_i в множестве X_p заменяется на вершину x_j . Таким образом, множество X'_p становится новой версией исходного множества X_p , так как оно лучше аппроксимирует искомое множество вершин $\tilde{X}_p$. Затем аналогичным образом проверяется уже новая версия множества X_p .

Такие проверки и замены вершин в последовательно создаваемых аппроксимирующих множествах X_p проводятся до тех пор, пока не удастся найти ни одной вершины x_j , такой, чтобы замена на эту вершину любой из вершин текущего множества X_p приводила бы к улучшению (уменьшению) передаточного числа этого множества.

Сформулируем алгоритм эвристического поиска p -медиан графа.

Шаг 1. Случайным образом задается число p медиан графа. Затем случайным образом выбираются p вершин из множества X всех вершин графа. Выбранные вершины образуют первоначальную версию множества X_p . Все вершины множества X , не включенные во множество X_p , маркируются меткой «не протестирована».

Шаг 2. Случайным образом из множества $\{X \setminus X_p\}$ выбирается «не протестированная» вершина x_j . Если «не протестированных» вершин нет, то переход к шагу 6.

Шаг 3. В цикле для всех вершин x_i множества X_p вычисляется величина Δ_{ij} , на которую изменится значение функционала F_p (рассчитывается по формуле (4)) при замене вершины x_i в X_p на вершину x_j , выбранную на шаге 2):

$$\Delta_{ij} = F_p(X_p) - F_p(X_p^i), \text{ где } X_p^i = (X_p \cup \{x_j\}) \setminus \{x_i\},$$

т. е. изменение значения функционала

$$\Delta_{ij} = \left[\sum_{x_i \in X_p} \text{cost}(x_i) + \sigma(X_p) \right] - \left[\sum_{x_i \in X_p^i} \text{cost}(x_i) + \sigma(X_p^i) \right].$$

Передаточные числа $\sigma(X_p)$ и $\sigma(X_p^i)$ в этом выражении вычисляются по формуле (3).

Запоминается индекс i^* той вершины x_i множества X_p , для которой достигается максимальное изменение значения функционала при ее замене на вершину x_j , т. е.

$$\Delta_{i^*j} = \max_{x_i \in X_p} \Delta_{ij}.$$

Шаг 4. Если $\Delta_{i^*j} \leq 0$, т. е. замена какой-либо вершины множества X_p на «не протестированную» вершину x_j , выбранную на шаге 2, не улучшает значение функционала $F_p(X_p)$, то вершина x_j маркируется как «протестированная», и переходим на шаг 2.

Шаг 5. Если $\Delta_{i^*j} > 0$, то замена вершины x_{i^*} множества X_p на вершину x_j улучшает значение функционала задачи. Производится такая замена, и вершины x_j и x_{i^*} маркируются как «протестированные». В результате получаем новое множество $X_p = (X_p \cup \{x_j\}) \setminus \{x_{i^*}\}$. Переход к шагу 2.

Шаг 6. Конец работы алгоритма. Текущее множество X_p является аппроксимацией искомого p -медианного множества $\tilde{X}_p$ узлов РСОД для размещения копий ИТР, близкого к оптимальному.

Описанный алгоритм реализован в виде программного модуля (в среде MS Visual Studio на языке программирования C++), с помощью которого решалась задача оптимального размещения ИТР в РСОД — см. далее пример.

3. ПРИМЕР РЕШЕНИЯ ЗАДАЧИ

Рассмотрим работу описанного алгоритма при числе медиан $p = 3$ на примере РСОД, работающей на базе компьютерной сети из 12 узлов. Топология компьютерной сети с пронумерованными узлами и длинами дуг представлена в виде неориентированного графа на рис. 1.

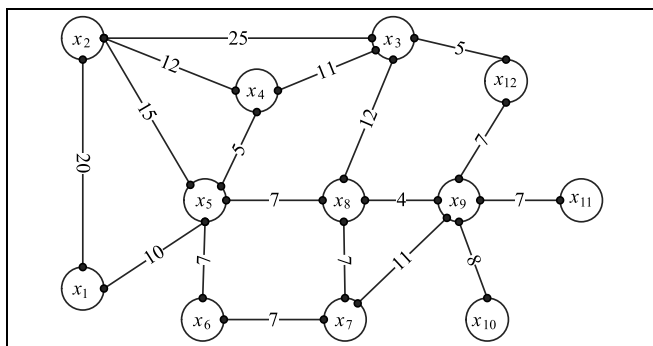


Рис. 1. Топология компьютерной сети

Характеристики РСОД и компьютерной сети, на базе которой она функционирует:

- РСОД обрабатывает запросы пользователей, относящиеся к одному из 10-ти типов запросов;
- стоимость s передачи единичного объема данных по пути единичной длины составляет 50 ед.;
- интенсивности λ_j^n запросов различного типа в узлах сети представлены в табл. 1;

— объемы данных d_n , передаваемых по сети при обработке запросов разного типа, приведены в табл. 2;

— «веса» v_j узлов сети (объем получаемой и передаваемой информации при обработке запросов) представлены в табл. 3, значения весов подсчитываются на основе данных в табл. 1 и 2;

— стоимость $\text{cost}(x_i)$ хранения ИТР в узлах сети представлена в табл. 4;

— матрица длин кратчайших путей показана в табл. 5.

В результате решения сформулированной задачи найдено близкое к оптимальному размещение $Xp1$ копий ИТР, данные которого представлены в табл. 6 и показано на графе (рис. 2), со значением функционала $F(Xp1)$ задачи, равного 4 750:

В таблице величины st_Xp , tr_Xp и st_Xp1 , tr_Xp1 это соответственно, стоимость хранения копий ИТР и передаточное число (см. формулу (4)) для исходного множества узлов Xp и найденного в результате решения задачи множества $Xp1$ (узлы x_3 , x_5 , x_9). Множество узлов Xp получено на шаге 1

Таблица 1

Матрица интенсивностей запросов различного типа в узлах сети

		Номер узла сети											
		1	2	3	4	5	6	7	8	9	10	11	12
Тип запроса	1	20	10	10	10	10	10	10	10	10	10	10	10
	2	10	5	5	30	5	5	20	5	20	5	5	5
	3	5	5	5	30	5	5	30	5	30	5	5	15
	4	10	15	10	20	15	10	20	15	20	15	10	15
	5	10	5	5	10	5	5	10	5	10	5	5	5
	6	5	10	5	10	10	10	10	10	10	10	10	10
	7	30	10	10	10	10	5	10	10	30	15	10	10
	8	20	20	10	15	20	10	15	20	15	20	10	20
	9	10	5	5	30	5	5	30	5	30	5	5	5
	10	10	5	10	20	5	10	20	5	20	5	10	5

Таблица 2

Объемы данных, передаваемых по сети при обработке запросов

Тип запроса	1	2	3	4	5	6	7	8	9	10
Объем данных	100	500	500	300	300	100	150	100	400	400

Таблица 3

Веса узлов сети

Номер узла сети	1	2	3	4	5	6	7	8	9	10	11	12
Вес узла	30 500	20 500	19 500	64 000	20 500	19 250	59 000	20 500	62 000	21 250	20 000	25 500

Таблица 4

Стоимость хранения копии ИТР в узлах сети

Номер узла сети	1	2	3	4	5	6	7	8	9	10	11	12
Стоимость	500	500	400	300	400	450	500	350	350	500	500	450

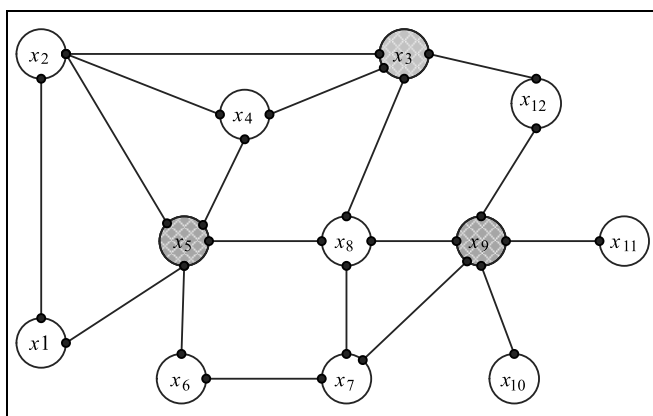


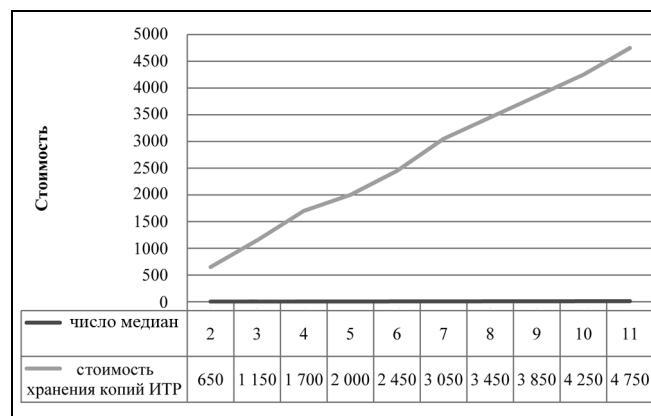
Рис. 2. Размещение копий ИТР в узлах компьютерной сети

описанного выше алгоритма. Для этого множества рассчитано значение функционала $F(Xp)$, равного сумме затрат st_Xp на хранение копий и передаточного числа tr_Xp , характеризующего затраты на обработку запросов.

В ходе экспериментов по выявлению зависимости значения функционала $F(Xp1)$ от числа медиан p (при неизменных характеристиках РСОД и компьютерной сети) были получены результаты, представленные далее в табл. 7 и графиках.

На рис. 3 и 4 показана близкая к линейной зависимость стоимости st_Xp1 хранения копий ИТР и затрат tr_Xp1 на обработку запросов пользователей от числа p .

В табл. 7 и на графике рис. 5 показана зависимость значений функционала задачи $F(Xp1)$ и его

Рис. 3. Зависимость стоимости st_Xp1 хранения копий ИТР от их числа p

компонент (стоимости st_Xp1 хранения копий ИТР и затрат tr_Xp1 на обработку запросов) от числа p копий ИТР, размещенных по узлам компьютерной сети.

Как видно из значений функционала задачи $F(Xp1)$, приведенных в табл. 7, и графика на рис. 5, функционал имеет минимальные значения при числе медиан графа от 4 до 6. При уменьшении числа медиан от 3 до 2 значение функционала растет. Аналогичная картина наблюдается при увеличении числа медиан от 7 до 11. Объяснение этому факту можно найти, посмотрев, как ведут себя графики затрат st_Xp1 на хранение копий ИТР и передаточного числа tr_Xp1 , характеризующего размер затрат на обработку запросов к копиям ИТР (см. рис. 3 и 4). При уменьшении числа копий

Матрица длин кратчайших путей

		Номер узла сети											
		1	2	3	4	5	6	7	8	9	10	11	12
Номер узла сети	1	0	20	26	16	10	17	24	17	21	29	31	28
	2	20	0	25	12	15	22	29	22	26	34	33	33
	3	26	25	0	11	16	23	19	12	12	20	19	5
	4	16	12	11	0	5	12	19	12	16	24	23	23
	5	10	15	16	5	0	7	14	7	11	19	18	18
	6	17	22	23	12	7	0	7	14	18	26	27	27
	7	24	29	19	19	14	7	0	7	11	19	18	18
	8	17	22	12	12	7	14	7	0	4	12	11	11
	9	21	26	12	16	11	18	11	4	0	8	7	7
	10	29	34	20	24	19	26	19	12	8	0	15	15
	11	31	33	19	23	18	27	18	11	7	15	0	14
	12	28	33	5	23	18	27	18	11	7	15	14	0

Таблица 6

Решение задачи размещения ИТР по узлам сети

Xp	$F(Xp)$	st_Xp	tr_Xp	$Xp1$	$F(Xp1)$	st_Xp1	tr_Xp1
x_1, x_3, x_5	6350	1300	5050	x_3, x_5, x_9	4750	1150	3600

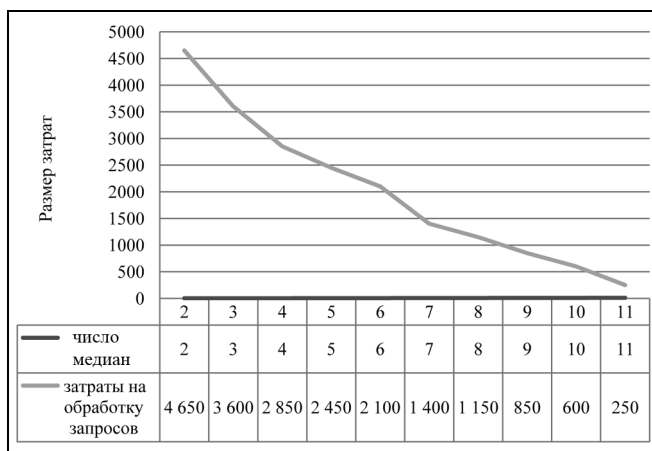


Рис. 4. Зависимость затрат tr_Xp1 на обработку запросов от числа копий ИТР

ИТР (числа медиан p) в сети происходит, с одной стороны, уменьшение затрат st_Xp1 на хранение ИТР, а с другой стороны, увеличиваются затраты tr_Xp1 на обработку запросов к копиям ИТР. Причем при заданных параметрах сети увеличение затрат на обработку запросов происходит быстрее, чем уменьшение стоимости хранения ИТР. При увеличении числа копий ИТР, начиная с некоторого их числа рост затрат на хранение превышает уменьшение затрат на обработку запросов.

Из приведенных результатов следует очевидный вывод о том, что для рассматриваемой задачи необходимо определить как оптимальное число копий ИТР, так и оптимальное размещение этих копий в узлах распределенной системы обработки данных.

Был также проведен эксперимент по проверке стабильности результатов, получаемых с помощью эвристического алгоритма поиска p -медиан графа.

При постоянном числе p медиан, равном 7, и неизменных характеристиках РСОД и компьютерной сети, алгоритм решения задачи запускался 12 раз. В результате получено множество значе-

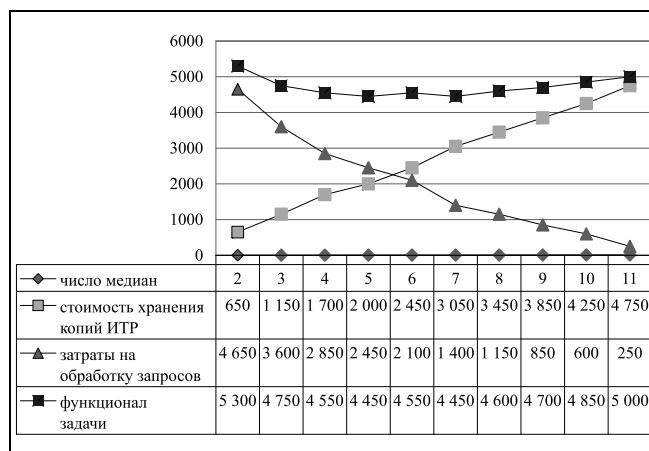


Рис. 5. Зависимость значения функционала и его компонент от числа медиан

ний функционала задачи $F(Xp1)$, представленное в табл. 8.

Среднее значение функционала задачи равно 4592. При этом отклонение значений функционала, полученных при 12-ти запусках алгоритма, от среднего значения не превышает 3,09 %, т. е. выбранный для решения задачи алгоритм дает достаточно стабильные результаты.

4. ОСНОВНЫЕ ВЫВОДЫ

Применение информационно-технологического резервирования в РСОД позволяет значительно увеличить эффективность функционирования систем благодаря существенному снижению времени обработки поступающих от пользователей запросов. Это обусловлено тем, что заранее сформированные промежуточные массивы данных ИТР и набор процедур обработки запросов размещаются в узлах компьютерной сети, расположенных как можно ближе к источникам запросов. Это обеспечивает возможности для кратчайшего доступа пользователей к необходимой им информации. Кроме того, процесс обработки запросов пользо-

Таблица 7

Зависимость значения функционала задачи от числа медиан графа

Число медиан	2	3	4	5	6	7	8	9	10	11
Стоимость хранения копий ИТР	650	1150	1700	2000	2450	3050	3450	3850	4250	4750
Затраты на обработку запросов	4650	3600	2850	2450	2100	1400	1150	850	600	250
Функционал задачи	5300	4750	4550	4450	4550	4450	4600	4700	4850	5000

Таблица 8

Значения функционала задачи

Номер расчета	1	2	3	4	5	6	7	8	9	10	11	12
Значение функционала	4550	4650	4500	4450	4600	4600	4650	4650	4700	4600	4650	4500



вателей ускоряется также благодаря предварительно подготовленным промежуточным массивам данных. Множество копий ИТР, размещенных в узлах сети, содержащих предварительно сформированные промежуточные массивы данных, служат информационной базой для функционирования РСОД в будущем при обработке запросов, поступивших от пользователей уже после создания ИТР.

Остановимся на некоторых важных вопросах, которым необходимо уделить особое внимание в процессе проектирования РСОД, использующих ИТР.

- Отметим, что при проектировании ИТР для крупномасштабных РСОД с огромным числом пользователей потребуется выполнить и большой объем работ по анализу предметных областей пользователей системы:

- территориальная распределенность пользователей;
- перечень запросов;
- их разбиение по различным типам;
- перечень запрашиваемой информации;
- частота возникновения запросов;
- выделение групп пользователей, запрашивающих аналогичные данные.

Возможно, этот процесс потребует выделения дополнительных ресурсов бизнес- и системных аналитиков.

- Очень важно решить сложную задачу обеспечения достоверности информации в массивах данных ИТР в случае изменения тех или иных исходных данных. Для ее решения можно применить различные алгоритмы, различающиеся по сложности и стоимости их реализации. Это в свою очередь обуславливает необходимость поиска оптимального варианта решения задачи. При этом, например, возможны следующие варианты:

- пересчет данных в одном узле и репликация изменений в данных в узлы с копиями ИТР;
- пересчет промежуточных данных в каждом из узлов с ИТР;
- рассылка сообщений-извещений о неактуальности данных в копиях и др.

Формальная постановка и поиск решения данной задачи составляет предмет дальнейших исследований.

- В случае большого числа распределенных источников исходных данных, используемых для формирования ИТР (например, таблиц территориально распределенной базы данных) системным архитекторам и аналитикам необходимо тщательно спроектировать эффективную технологию формирования ИТР. Так как процесс формирования ИТР на основе большого числа распределенных источников данных может потребовать слишком больших ресурсов компьютерной сети.

ЗАКЛЮЧЕНИЕ

Сформулированы задачи поиска оптимального размещения копий информационно-технологического резерва по узлам распределенной системы и представлен алгоритм их решения. Данный алгоритм реализован в виде программного модуля, который применялся для решения задачи оптимального размещения ИТР в РСОД.

Приведен пример решения задачи оптимизации по критерию минимума стоимости функционирования РСОД. На расчетных данных примера продемонстрирована зависимость стоимости эксплуатации системы, стоимости хранения ИТР и затрат на обработку запросов от числа копий ИТР, размещенных в узлах компьютерной сети.

Дано краткое описание направления дальнейших исследований по проблемам использования данного вида резерва в компьютерных сетях, неоднородных в смысле распределения частот возникновения в узлах сети запросов пользователей.

ЛИТЕРАТУРА

1. Сомов С.К. Создание информационно-технологического резерва в распределенных системах обработки данных // Проблемы управления. — 2018. — № 3. — С. 40–46.
2. Кристофидес Н. Теория графов. Алгоритмический подход. — М.: Мир, 1978. — 432 с.
3. Майника Э. Алгоритмы оптимизации на сетях и графах. — М.: Мир, 1981. — 322 с.
4. Teitz M.B., Bart P. Heuristic methods for estimating the generalized vertex median of a weighted graph // Operations Research. — 1968. — Vol. 16. — P. 955–961.
5. Sahoo J., Salahuddin M.A., Gliho R. A Survey on Replica Server Placement Algorithms for Content Delivery Networks // IEEE Communications Surveys & Tutorials. — 2016. — 30 p. — URL: <https://arxiv.org/ftp/arxiv/papers/1611/1611.01729.pdf> (дата обращения: 6.07.2017).
6. Souri A., Rahmani A. A Survey for Replica Placement Techniques in Data Grid Environment // Intern. Journal Modern Education and Computer Science. — 2014. — Vol. 5. — P. 46–51.
7. Mansouri N. Adaptive data replication strategy in cloud computing for performance improvement // Frontiers of Computer Science. — 2016. — Vol. 10, N 5. — P. 925–935.
8. Чернышев Г.А. Обзор подходов к организации физического уровня в СУБД // Тр. СПИИРАН. — 2013. — Вып. 1 (24). — С. 222–275.
9. Reese J. Methods for Solving the p -Median Problem: An Annotated Bibliography // Networks. — 2006. — Vol. 48, N 3. — P. 125–142.
10. Daskin M.S., Maass K.L. The p -Median Problem // Location Science. Springer. — 2015. — P. 21–45. — URL: <https://link.springer.com/book/10.1007/978-3-319-13111-5#toc> (дата обращения: 10.01.2018).
11. Mladenovic N., Brimberg J., Hansen P. The p -median problem: A survey of metaheuristic approaches // European Journal of Operational Research. — 2007. — Vol. 179, N 3. — P. 927–939.

Статья представлена к публикации членом редколлегии В.В. Кульбой.

Сомов Сергей Константинович — канд. техн. наук, вед. инженер, Институт проблем управления им. В.А. Трапезникова РАН, г. Москва, ✉ ssomov2016@ipu.ru.

БЕСКОНФЛИКТНАЯ САМОМАРШРУТИЗАЦИЯ ДЛЯ ТРЕХМЕРНОГО ПОЛНОГО МУЛЬТИКОЛЬЦА

В.С. Подлазов

Предложена системная сеть в виде неблокируемого трехмерного полного мультикольца. Разработана структура мультикольца и предложен алгоритм бесконфликтной прокладки прямых путей между ее узлами посредством динамической локальной самомаршрутизации в них.

Ключевые слова: системная сеть, обобщенный гиперкуб, полное мультикольцо, коммутационные свойства, неблокируемая сеть, прямые каналы, бесконфликтная самомаршрутизация, локальная динамическая самомаршрутизация.

ВВЕДЕНИЕ

Данная статья продолжает серию работ автора по созданию самомаршрутизируемых неблокируемых системных сетей для многопроцессорных вычислительных систем [1–3]. Эти сети обладают наибольшим быстродействием, так как передача данных в них ведется по прямым каналам без промежуточной буферизации.

Исходно эти сети имели топологию квазиполного графа или квазиполного орграфа [1, 2] и обладали квадратичной зависимостью числа узлов от их степени. Самомаршрутизация в этих сетях была статичной, при которой любой источник прокладывал весь бесконфликтный путь к приемнику. Это всегда можно было сделать для произвольной перестановки пакетов данных между узлами. Главный недостаток таких сетей — недостаточно большое число узлов в них.

Топологию квазиполного орграфа имели сети в виде двумерного обобщенного гиперкуба и двумерного полного мультикольца [2].

Обобщенный r -мерный p -ичный гиперкуб [4] со степенью узлов $m = rp$ и числом узлов $N = p^r$ используется как системная сеть в некоторых современных многопроцессорных вычислительных системах [5]. В работе [3] автором был построен неблокируемый трехмерный p -ичный гиперкуб, для которого был предложен алгоритм динамической локальной самомаршрутизации. Он позволяет параллельно прокладывать бесконфликтные пути при произвольной перестановке пакетов данных между узлами на основе только локальной инфор-

мации в промежуточных узлах без взаимодействия между ними.

В настоящей работе строится неблокируемое трехмерное p -ичное мультикольцо и предлагается алгоритм динамической локальной самомаршрутизации в нем для произвольной перестановки пакетов данных. Она ориентирована на создание сетей для однокристалльных процессоров-ускорителей с кольцевыми каналами, в которых необходимо иметь сотни и тысяч ядер.

1. НЕБЛОКИРУЕМОЕ ДВУМЕРНОЕ ПОЛНОЕ МУЛЬТИКОЛЬЦО

Примером одной из сетей, которая как неблокируемая, так и самомаршрутизируемая, служит двумерное p -ичное мультикольцо. Оно состоит из $2(p - 1)$ разных симплексных колец с шагами $\{1, 2, \dots, p - 1\}$, составляющих измерение X , и с шагами $\{p, 2p, \dots, p(p - 1)\}$, составляющих измерение Y , и имеет $N = p^2$ узлов. Каждое кольцо состоит из одинаковых дуг, длина которых задается разностью по $\text{mod } N$ номеров инцидентных узлов. На рис. 1 представлено двумерное троичное мультикольцо как симметричный орграф. В нем дуги с длиной 3 и 6 являются встречными дугами и обозначены пунктиром.

Однако двумерное мультикольцо можно представить и как двудольный орграф, если учесть, что каждый узел содержит абонента и коммутатор $p \times p$ и что абоненты связаны дугами только с коммутаторами своего или других узлов (рис. 2). Одну долю этого орграфа составляют коммутаторы, а другую — абоненты A_i (процессоры). Такой орграф мы называем квазиполным орграфом. Схема меж-

соединений в нем задается таблицей инцидентий (табл. 1), в которой в ячейках находятся номера инцидентных узлов.

Самомаршрутизация по таблице инцидентий осуществляется по номерам абонента-источника и абонента-приемника. Пусть для примера это будут узлы 2 и 8 соответственно, подчеркнутые в табл. 1. По правой части таблицы находятся номера коммутаторов, из которых есть дуги к приемнику. В примере это коммутаторы 2, 5 и 8, обозначенные двойным подчеркиванием. В левой части таблицы из этих коммутаторов находится тот коммутатор, к которому есть дуга от источника. В примере это коммутатор 2. По свойствам квазиполных орграфов [1, 2] такой коммутатор является единственным при любом p .

С другой стороны, любой путь длины L_2 в двумерном мультикольце представляет собой последовательность не более двух дуг, длины которых задаются как значения разрядов в p -ичной системе счисления $L_2 = i + jp$ ($0 \leq i, j \leq p - 1$). Здесь i задает длину дуги X , а jp — дуги Y .

2. ДВУМЕРНЫЕ МУЛЬТИКОЛЬЦА В ПОЛНОМ ТРЕХМЕРНОМ МУЛЬТИКОЛЬЦЕ

Рассмотрим мультикольцо, содержащие $N = p^3$ узлов, из каждого узла которого выходят три вида дуг с длинами $\{1, 2, \dots, p - 1\}$, $\{p, 2p, \dots, p(p - 1)\}$ и $\{p^2, 2p^2, \dots, p^2(p - 1)\}$ соответственно. Такое мультикольцо мы называем трехмерным полным p -ичным мультикольцом. Оно состоит из симплексных колец с разными шагами, каждое из которых состоит из дуг одной длины. Длины составляющих дуг задают длины шагов колец. Кольца с длинами шагов $\{1, 2, \dots, p - 1\}$ задают измерение X (мультикольцо X), кольца с длинами $\{p, 2p, \dots, p(p - 1)\}$ — измерение Y (мультикольцо Y) и кольца с длинами $\{p^2, 2p^2, \dots, p^2(p - 1)\}$ — измерение Z (мультикольцо Z).

На рис. 3 показан пример трехмерного троичного мультикольца. В нем кольца измерения X

Таблица 1

Инцидентии в квазиполном орграфе при $p = 3$

Коммутаторы	Дуги X от абонентов			Дуги Y к абонентам		
1	1	<u>2</u>	3	1	4	7
<u>2</u>	2	<u>3</u>	4	2	5	8
<u>3</u>	3	4	5	3	6	9
4	4	5	6	4	7	1
<u>5</u>	5	6	7	5	8	2
<u>6</u>	6	7	8	6	9	3
7	7	8	9	7	1	4
<u>8</u>	8	9	1	<u>8</u>	2	5
<u>9</u>	9	1	2	9	3	6

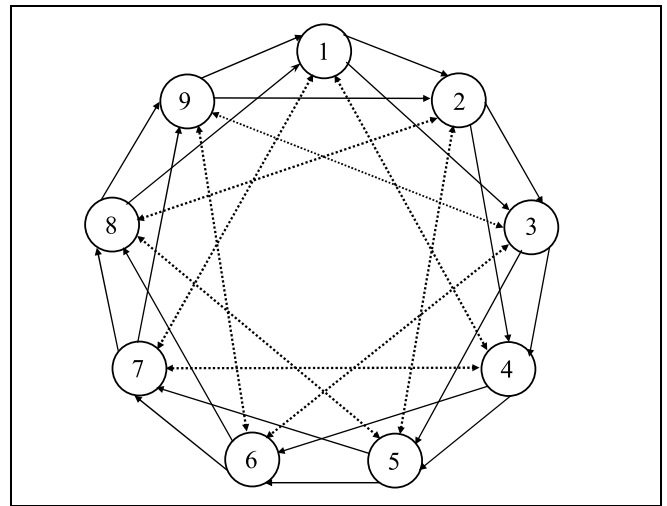


Рис. 1. Двумерное троичное мультикольцо с шагами $\{1, 2\}$ и $\{3, 6\}$

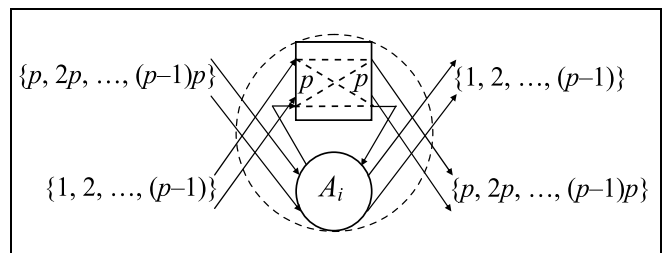


Рис. 2. Схема узла двумерного мультикольца

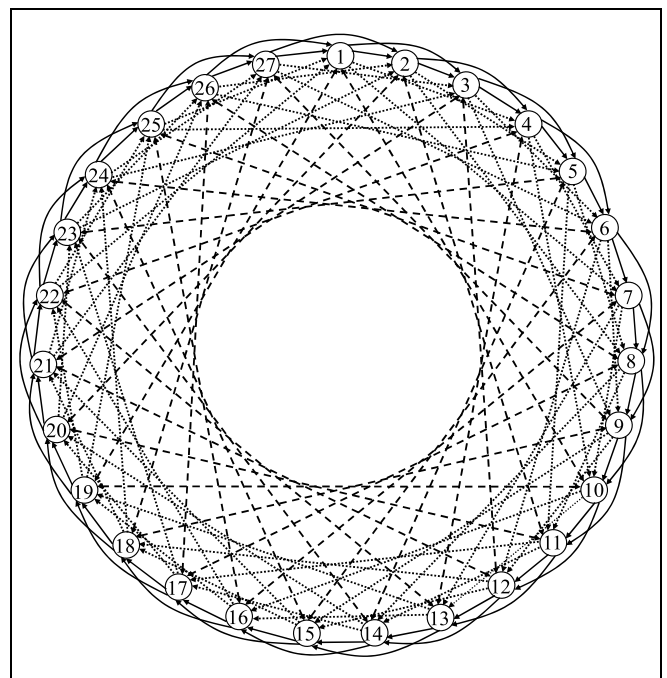


Рис. 3. Трехмерное троичное мультикольцо

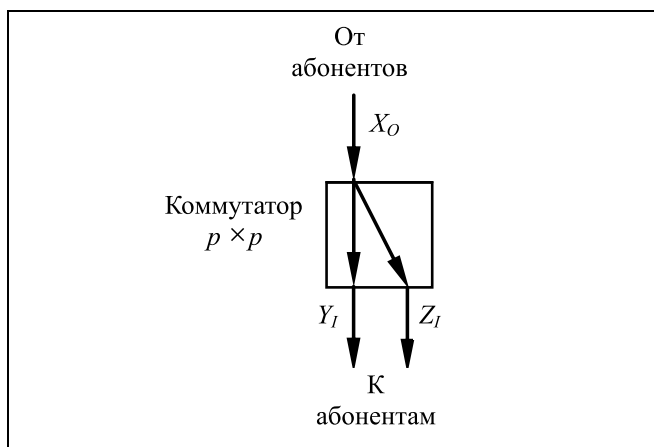


Рис. 4. Коммутатор для двухэтапных прямых путей

обозначены сплошными дугами, измерения Y — пунктирными дугами и измерения Z — штриховыми дугами.

Будем представлять любой кратчайший путь как последовательность дуг X , Y , Z , которые будем называть этапами пути. Примем, что любой узел имеет выходные дуги X_O от абонентов к коммутаторам и входные дуги Y_I и Z_I от коммутаторов к абонентам — $3p$ дуг. Из них $3(p-1)$ дуг являются межузловыми и три дуги — внутриузловыми.

Перечисленные дуги делают пары измерений X , Y и X , Z неблокируемыми двумерными мультикольцами или квазиполными орграфами XY и XZ соответственно. Это означает, что в них любой прямой путь, состоящий не более чем из двух дуг, прокладывается посредством самомаршрутизации только через один коммутатор, представленный на рис. 4. Он имеет схему $p \times 2p$ со сложностью $2p^2$ точек коммутации.

Схемы межсоединений в квазиполных орграфах XY и XZ задаются таблицами их инцидентий (табл. 2).

Рассмотренные выше дуги не позволяют рассматривать пару измерений Y , Z как неблокируемое двумерное мультикольцо, так как узлы не имеют дуг Y или дуг Z от абонентов к коммутаторам. Поэтому двухэтапные (двухдуговые) пути в измерениях Y и Z будем представлять как частный случай трехэтапных (трехдуговых) путей с локальными дугами на первом или на последнем этапах.

Для прокладки трехэтапных путей дополнительно потребуется иметь в каждом узле $p-1$ дуг мультикольца Y_M , проложенных только между коммутаторами узлов. Вообще кратчайший путь между любыми двумя абонентами (источником и приемником) можно представить как состоящий из дуги X_O от источника к коммутатору, дуги Y_M меж-

ду коммутаторами и дуги Z_I от коммутатора к приемнику. При этом двухэтапные и трехэтапные пути вообще не могут иметь конфликтов, так как дуги от источников и к приемникам при перестановке задаются единственным образом, а в середине трехэтапных путей используются дуги, которые отсутствуют в двухэтапных путях.

Рассмотрим кратчайший трехэтапный путь (рис. 5) от узла-источника s (горизонтальная штриховка) к узлу-приемнику d (квадратная штриховка) через промежуточные узлы t и i . Первый этап от источника в узле s в промежуточный узел t (левая косая штриховка) проходит по дуге X_O . Через узел t могут проходить пути от разных источников, которые на втором этапе используют разные дуги Y_M . Если каждая дуга проходит только один путь через узел i (правая косая штриховка), то он является бесконфликтным, и бесконфликтным же он останется на последнем этапе по разным дугам Z_I к приемникам d (клетчатая штриховка).

Ситуация, при которой на одну дугу Y_M претендуют несколько путей, считается конфликтом

Таблица 2

Инцидентии в квазиполных орграфах XY и XZ при $p = 3$

Коммутаторы	Дуги X от абонентов			Дуги Y к абонентам			Дуги Z к абонентам		
1	1	2	26	1	4	7	1	10	19
2	2	3	27	2	5	8	2	11	20
3	3	4	1	3	6	9	3	12	21
4	4	5	2	4	7	10	4	13	22
5	5	6	3	5	8	11	5	14	23
6	6	7	4	6	9	12	6	15	24
7	7	8	5	7	10	13	7	16	25
8	8	9	6	8	11	14	8	17	26
9	9	10	7	9	12	15	9	18	27
10	10	11	8	10	13	16	10	19	1
11	11	12	9	11	14	17	11	20	2
12	12	13	10	12	15	18	12	21	3
13	13	14	11	13	16	19	13	22	4
14	14	15	12	14	17	20	14	23	5
15	15	16	13	15	18	21	15	24	6
16	16	17	14	16	19	22	16	25	7
17	17	18	15	17	20	23	17	26	8
18	18	19	16	18	21	24	18	27	9
19	19	20	17	19	22	25	19	1	10
20	20	21	18	20	23	26	20	2	11
21	21	22	19	21	24	27	21	3	12
22	22	23	20	22	25	1	22	4	13
23	23	24	21	23	26	2	23	5	14
24	24	25	22	24	27	3	24	6	15
25	25	26	23	25	1	4	25	7	16
26	26	27	24	26	2	5	26	8	17
27	27	1	25	27	3	6	27	9	18

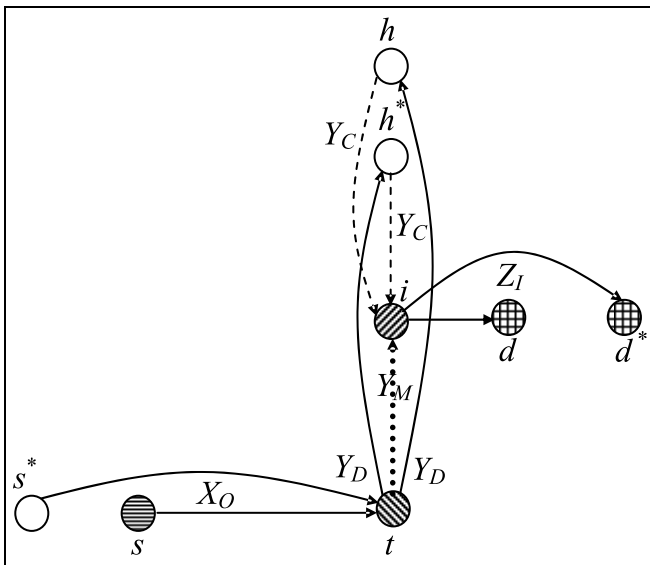


Рис. 5. Конфликт первого типа и его разрешение

первого типа на этой дуге. При этом один из этих путей имеет приоритет и проходит на выходную дугу Y_M коммутатора, а остальные остаются конфликтными в узле. Конфликт первого типа возникает, если на прохождение через узел t претендуют несколько путей (рис. 5) из разных узлов-источников s и s^* . Здесь и далее для краткости мы обозначаем только два таких узла, которых на самом деле может быть любое количество от двух до $p - 1$. Пусть эти пути через узел t претендуют на прохождение одной и той же дуги Y_M , ведущей в узел i . На рис. 5 конфликтная дуга обозначена жирным пунктиром, и она при разрешении конфликта не используется.

Для разрешения конфликта используется еще одно мультикольцо Y_D с дугами, проложенными только между коммутаторами узлов. Мультикольца Y_M и Y_D имеют одинаковые дуги, и последние обозначены на рис. 5 сплошными линиями. Использование дуг Y_D позволяет избежать повторных конфликтов на дугах измерения Y . Разрешение конфликта осуществляется следующим образом: из узла t прокладываются пути в разные вспомогательные узлы h и h^* (без штриховки на рис. 5) по заведомо разным дугам Y_D . Такими дугами Y_D являются аналоги разных дуг X_O в узел t из узлов s и s^* . Иначе говоря, если в узел t ведет дуга X_O с длиной k , то из узла t должна выходить дуга Y_D с длиной kp .

Затем осуществляется возврат в узел i по дугам еще одного мультиколыца Y_R с шагами $\{-p, -2p, \dots, -p(p-1)\}$. Дуги Y_R проложены только между ком-

мутаторами узлов. Если на дугах Y_R нет конфликтов, то проложенные пути бесконфликтные. Они останутся бесконфликтными и на дугах Z_I от коммутаторов узла i к приемникам в узлах d и d^* , так как они прокладываются по разным дугам.

При каких условиях возможен возврат в узел i (рис. 5 и 6)? Если дуга Y_M имеет длину jp , то возвратная дуга Y_R должна иметь длину $(k - j)p$. Такая дуга принадлежит к множеству дуг Y_R только при $k < j$. В противном случае приходится использовать дугу Y_R с длиной $(k - j)p - p^2$ (на рис. 6 она обозначена как $\underline{Y}_R$) и уменьшать на p^2 длину дуги Z_j , чтобы сохранить неизменной длину маршрута. Последняя дуга обозначается как $\underline{Z}_j$ (рис. 6).

Однако и на дугах Y_R могут возникать конфликты второго типа. Каждый из них происходит в том случае, если на проход через узел i на втором этапе претендует несколько путей из разных узлов t и t^* после первого этапа (рис. 7). Для его разрешения потребуется еще одно мультикольцо Z_M с дугами, проложенными только между коммутаторами узлов. Мультикольца Z_M и Z_I имеют одинаковые дуги.

Конфликт второго типа разрешается следующим образом. Из промежуточного узла h пути прокладывается по тем дугам Z_M , которые равны дугам Z_I . Все эти дуги Z_I различны, поэтому различны и соответствующие дуги Z_M . Поэтому пути бесконфликтно прокладываются в разные вспомогательные узлы e и e^* . Из них они завершаются бесконфликтно по дугам Y_I из узлов e и e^* .

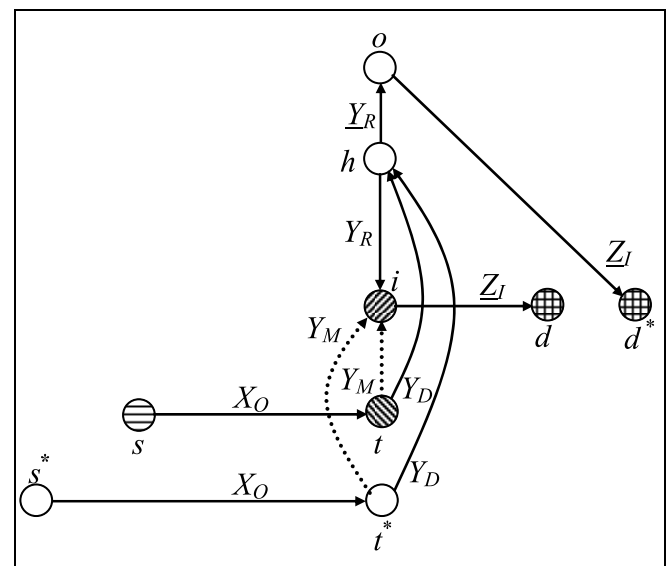


Рис. 6. Варианты разрешения конфликта первого типа

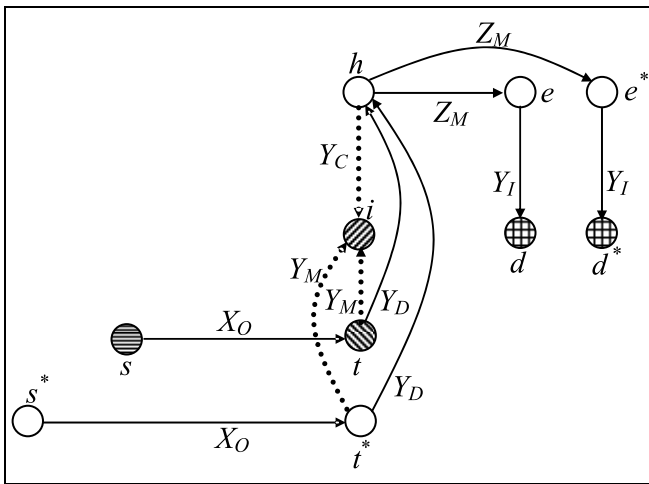


Рис. 7. Конфликт второго типа и его разрешение

Конфликты разных типов возникают при разных вариантах размещения узлов-источников в кольцах второго измерения — на одном или на разных кольцах. Других вариантов их размещения нет. Поэтому нет и других типов конфликтов.

Рассмотренная процедура задает общую схему прокладки бесконфликтных прямых путей посредством динамической самомаршрутизации, которая осуществляется только на основе локальной информации о конфликтах внутри каждого узла. Наличие такой процедуры позволяет сформулировать

Утверждение. Построенное трехмерное p -ичное мультикольцо с $M = 7N(p - 1)$ межузловыми дугами является неблокируемой сетью, для которой предложена процедура бесконфликтной локальной самомаршрутизации. Это мультикольцо в каждом узле имеет выходные дуги X_O от абонентов, входные дуги Y_R , Z_I к абонентам и дуги Y_M , Y_D , Y_C и Z_M между коммутаторами.

3. ПРОТОКОЛ САМОМАРШРУТИЗАЦИИ В ТРЕХМЕРНОМ МУЛЬТИКОЛЬЦЕ

Сначала представим алгоритм в терминах дуг измерений X , Y и Z , считая, что исходный маршрут имеет вид $X_O Y_M Z_I$, а длины составляющих дуг U обозначают как $|U|$. Обозначения узлов и дуг даются по рис. 5–7.

Алгоритм динамической локальной самомаршрутизации (общая схема)

1. Если X_O отсутствует ($|X_O| = 0$), то путь прокладывается по локальной дуге в узле s от источника к коммутатору. Переход к п. 3.

Если X_O присутствует ($|X_O| \neq 0$), то путь прокладывается от источника в узле s по дуге X_O к коммутатору узла t . Переход к п. 2.

2. Если Y_M отсутствует ($|Y_M| = 0$) и Z_I отсутствует ($|Z_I| = 0$), то путь прокладывается по локальной дуге в узле t от коммутатора к источнику. Конец алгоритма.

3. Если Y_M отсутствует ($|Y_M| = 0$), а Z_I присутствует ($|Z_I| \neq 0$), то путь прокладывается по дуге Z_I от коммутатора узла t к приемнику в узле d . Конец алгоритма.

Если Y_M присутствует ($|Y_M| \neq 0$), а Z_I отсутствует ($|Z_I| = 0$), то путь прокладывается по дуге Y_I от коммутатора узла t к приемнику в узле d . Конец алгоритма.

Если Y_M присутствует ($|Y_M| \neq 0$) и Z_I присутствует ($|Z_I| \neq 0$), то проверяется наличие конфликта на дуге Y_M .

Если конфликта нет, то путь прокладывается по дуге Y_M от коммутатора узла t к коммутатору узла i . Переход к п. 4.

Если на дуге Y_M конфликт имеет место, то путь прокладывается по уникальной для узла t дуге Y_D (разной для разных путей) от коммутатора узла t к коммутатору узла h . Переход к п. 5.

4. Путь прокладывается по дуге Z_I от коммутатора узла i к абоненту в узле-приемнике d . Конец алгоритма.

5. Проверяется наличие конфликта на дуге Y_R .

Если конфликта нет, то путь прокладывается по дуге Y_R от коммутатора узла h к коммутатору узла i или по дуге Y_R к промежуточному узлу o . Переход к п. 6.

Если же конфликт имеет место, то переход к п. 7.

6. Путь прокладывается по дуге Z_I или дуге Z_I от коммутатора узла i или узла o к абоненту в узле-приемнике d . Конец алгоритма.

7. Путь прокладывается по дуге Z_M от коммутатора узла h к коммутатору узла e . Переход к п. 8.

8. Путь прокладывается по дуге Y_I противоположной дуге Y_R от коммутатора узла e к приемнику в узле d . Конец алгоритма.

А теперь представим алгоритм в терминах путевой информации, которая содержится в пилотном пакете для реализации динамической локальной самомаршрутизации. Здесь также используются обозначения дуг и узлов на рис. 5–7.

Путевую информацию будем представлять номерами выходных дуг коммутаторов в каждом узле, которые должны составлять прямой путь. Дуги нумеруются для каждого измерения в порядке воз-



растания их длин числами от 0 до $p - 1$. Номер 0 задает либо внутриузловую дугу в узле-источнике и в узле-приемнике, либо отсутствие дуги между коммутаторами разных узлов. Таким образом, дуги с длиной m_1 измерения X задаются как числа m_1 ($0 \leq m_1 \leq p - 1$), дуги с длиной $m_2 p$ измерения Y задаются как числа m_2 ($0 \leq m_2 \leq p - 1$) и дуги с длиной $m_3 p^2$ измерения Z задаются как числа m_3 ($0 \leq m_3 \leq p - 1$).

Любой кратчайший путь задается тремя числами m_1, m_2, m_3 . Число m_1 задает дугу $X_O(m_1)$ от узла-источника s в промежуточный узел t . При $m_1 = 0$ узлы s и t совпадают. Число m_2 задает дугу $Y_M(m_2)$ от узла t к промежуточному узлу i . При $m_3 = 0$ узлы t и i совпадают. Число m_3 задает дугу $Z_I(m_3)$ от коммутатора узла i к узлу-приемнику d .

Для прокладки прямого пути пакет должен содержать пять чисел m_1, m_3, m_2 и m_2^*, m_3^* . Если $m_1 \leq m_2$, то $m_2^* = m_2 - m_1$ и $m_3^* = m_3$, в противном случае, $m_2^* = p - m_1 + m_2$ и $m_3^* = m_3 - 1$. Эти числа для каждого кратчайшего пути вычисляются каждым источником заранее. Числа m_2^* и m_3^* задают дуги $Y_R(m_2^*)$ или $\underline{Y}_R(m_2^*)$ и $Z_I(m_3^*)$ или $\underline{Z}_I(m_3^*)$.

Прокладка прямого пути осуществляется по следующему алгоритму.

Алгоритм динамической локальной самомаршрутизации (конкретная реализация)

1. Если $m_1 = 0$, то путь прокладывается по локальной дуге в узле s от источника к коммутатору. Переход к п. 3.

Если $m_1 > 0$, то путь прокладывается от источника в узле s по дуге $X_O(m_1)$ к коммутатору узла t . Переход к п. 2.

2. Если $m_3 = 0$ и $m_2 = 0$, то путь прокладывается по локальной дуге в узле t от коммутатора к источнику. Конец алгоритма.

3. Если $m_2 = 0$ и $m_3 \neq 0$, то путь прокладывается по дуге $Z_I(m_3)$ от коммутатора узла t к приемнику в узле d . Конец алгоритма.

Если $m_2 \neq 0$ и $m_3 = 0$, то путь прокладывается по дуге $Y_I(m_2)$ от коммутатора узла t к приемнику в узле d . Конец алгоритма.

Если $m_3 > 0$ и $m_2 > 0$, то проверяется наличие конфликта на дуге $Y_M(m_2)$.

Если конфликта нет или имеет место приоритет при конфликте, то путь прокладывается по дуге $Y_M(m_2)$ от коммутатора узла t к коммутатору узла i . Переход к п. 4.

Если на дуге Y_M конфликт имеет место, то путь прокладывается по уникальной для узла дуге $Y_D(m_1)$ от коммутатора узла t к коммутатору узла h . Переход к п. 5.

4. Путь прокладывается по дуге $Z_I(m_3)$ от коммутатора узла i к абоненту в узле-приемнике d . Конец алгоритма.

5. Проверяется конфликтность пути по дуге $Y_R(m_2^*)$.

Если конфликта нет или имеет место приоритет при конфликте, то путь прокладывается по дуге $Y_R(m_2^*)$ к коммутатору узла i или узла o . Переход к п. 6.

Если же конфликт на дуге Y_D имеет место, то переход к п. 7.

6. Путь прокладывается по $Z_I(m_3^*)$ от коммутатора узла i или узла o к абоненту в узле-приемнике d . Конец алгоритма.

7. Путь прокладывается по дуге $Z_M(m_3)$ от коммутатора узла t к коммутатору узла e . Переход к п. 8.

8. Путь прокладывается по дуге $Y_I(m_2)$ от коммутатора узла e к приемнику в узле d . Конец алгоритма.

4. ЭКСПЕРИМЕНТАЛЬНАЯ ПРОВЕРКА ПРОТОКОЛА БЕСКОНФЛИКТНОЙ ЛОКАЛЬНОЙ САМОМАРШРУТИЗАЦИИ

Была проведена экспериментальная проверка неблокируемости предложенного трехмерного p -ичного мультикольца. Была создана его имитационная модель, и в ней реализован алгоритм локальной самомаршрутизации. Была проведена частотная проверка конфликтности алгоритма на произвольных перестановках пакетов. Измерялись частоты f_1 и f_2 возникновения перестановок с хотя бы одним конфликтом первого типа и второго типа (после разрешения первого) соответственно. На рис. 8 приведены их значения для разных значений p .

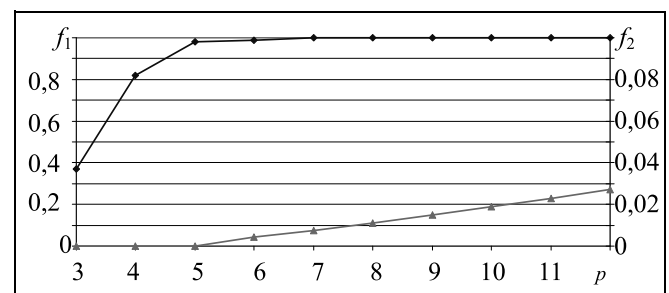


Рис. 8. Частоты конфликтных перестановок для алгоритма самомаршрутизации: —♦— f_1 ; —▲— f_2

Конечно, основной целью экспериментов было измерение частоты f_3 возникновения каких-либо конфликтов после разрешения конфликтов первого и второго типов. Эксперименты проводились при каждом p для нескольких миллионов (до десяти) случайных перестановок. Во всех случаях было зафиксировано значение $f_3 = 0$ (!). Последнее значение подтверждает с высокой степени достоверности незаблокируемость рассмотренного мультикольца. Естественно, что полную достоверность дал бы перебор всех перестановок. Однако он практически нереализуем при $p > 5$ и на суперкомпьютерах. Поэтому он и не проводился даже при $p < 5$ из-за неполноты достигаемого результата.

ЗАКЛЮЧЕНИЕ

Предложена новая структура трехмерного полного мультикольца на базе двумерных незаблокируемых мультиколец с минимальным числом дуг. Целью ее разработки было обеспечение возможности прокладки бесконфликтных путей посредством их локальной динамической самомаршрутизации в узлах мультикольца. Предложенная структура позволила исключить конфликты двухэтапных и трехэтапных путей и разработать алгоритм самомаршрутизации последних.

Предложенная структура и разработанные алгоритмы обеспечивают незаблокируемость мультикольца, т. е. возможность передачи пакетов данных по прямым путям без задержки на их буферизацию в промежуточных узлах. Построенное мультикольцо имеет семерной набор из $(p - 1)$ -й дуг, которые неравномерно распределены по его измерениям — одинарный набор в измерении X , четверной набор в измерении Y и двойной набор в измерении Z .

Рассмотренное в работе двумерное полное мультикольцо имеет топологию квазиполного орграфа. Однако подобные незаблокируемые сети могут иметь

и топологию квазиполного графа, которая позволяет разменивать число узлов на число разных прямых путей между узлами. Последнее свойство обеспечивает структурную отказоустойчивость сети.

Конечной целью автора было использование сетей с топологией квазиполного графа в качестве составных частей некоторой незаблокируемой трехмерной сети со свойством структурной отказоустойчивости. Однако графовая топология не имеет явного деления дуг и узлов по измерениям. Поэтому их надо ввести как-то «технически», и на этой основе построить указанную «трехмерную» сеть с кубической зависимостью числа узлов от их степени. Решение этой задачи и составит ближайшую цель автора.

ЛИТЕРАТУРА

1. *Каравай М.Ф., Подлазов В.С.* Метод инвариантного расширения системных сетей многопроцессорных вычислительных систем. Идеальная системная сеть // Автоматика и телемеханика. — 2010. — № 12. — С. 166—176.
2. *Каравай М.Ф., Подлазов В.С.* Распределенный полный коммутатор как «идеальная» системная сеть для многопроцессорных вычислительных систем // Управление большими системами. — 2011. — Вып. 34. — С. 92—116.
3. *Подлазов В.С.* Бесконфликтная самомаршрутизация для трехмерного обобщенного гиперкуба // Проблемы управления. — 2018. — № 3. — С. 26—32.
4. *Bhuyan L.N. and Agrawal D.P.* Generalized Hypercube and Hyperbus Structures for a Computer Network // IEEE Trans. on Computers. — 1984. — Vol. C-33. — N 4. — P. 323—333.
5. *Alverson R., Roweth D., Kaplan L., and Roweth D.* Cray[®] XCTM Series Network. — URL: <https://www.cray.com/sites/default/files/resources/CrayXCNetwork.pdf> (дата обращения 2.10.2017).

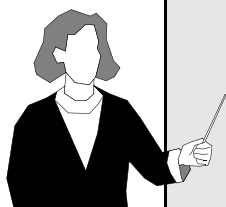
Статья представлена к публикации членом редсовета чл.-корр. РАН П.П. Пархоменко.

Подлазов Виктор Сергеевич — д-р техн. наук, гл. науч. сотрудник,

Институт проблем управления им. В.А. Трапезникова РАН, г. Москва, ✉ podlazov@ipu.ru.

Читайте в следующем номере

- ✓ **Бауэр В.П., Зацаринный А.А., Ильин Н.И.** и др. Прорывное ситуационное управление
- ✓ **Белов М.В., Новиков Д.А.** Основы теории комплексной деятельности. Ч. 2. Жизненные циклы комплексной деятельности. Организация и управление как комплексная деятельность
- ✓ **Васильев Г.А., Казаков В.Г., Тараканов А.Ф.** Теоретико-игровое моделирование противостояния сторон на основе рефлексивного управления
- ✓ **Вересников Г.С., Егоров Н.А., Кулида Е.Л., Лебедев В.Г.** Методы построения оптимальных очередей воздушных судов на посадку. Ч. 2. Методы приближенного решения
- ✓ **Иванов Н.Н.** Обобщенные стохастические сетевые графики с нестандартными дисциплинами свершения событий
- ✓ **Муранов А.А.** Анализ работы бортовых терминальных систем моноблочных жидкостных ракет-носителей при нештатном расходовании компонентов топлива
- ✓ **Постнов С.С.** Задачи оптимального управления для некоторых линейных систем дробного порядка, заданных уравнениями с производной Хильфера
- ✓ **Стецюра Г.Г.** Сетевая информационно-вычислительная поддержка взаимодействия подвижных роботов



РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ РЕЖИМОВ ПРИМЕНЕНИЯ ПРОТИВООПУХОЛЕВЫХ ВИРУСНЫХ ВАКЦИН НА ОСНОВЕ МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ

Н.А. Бабушкина, Е.А. Кузина, А.А. Лоос, Е.В. Беляева

Представлено математическое описание двухэтапной гибели опухолевых клеток в результате иммунного ответа организма на введение вирусной вакцины. Математическое описание в виде системы нелинейных дифференциальных уравнений реализовано как программный комплекс в системе MatLab-Simulink. В результате проведенного вычислительного эксперимента определены две стратегии эффективного применения противоопухолевой вирусной вакцины. Одна позволяет достичь полного уничтожения опухолевых клеток после однократного введения вакцины. Другая стратегия позволяет стабилизировать размер опухоли путем повторных введений вакцины. Предложенный подход к изучению эффективности метода вакцинотерапии может быть применен для различных видов экспериментальных опухолей и типов противоопухолевых вакцин.

Ключевые слова: математическая модель, опухолевые клетки, антитела, момент введения вакцины, эффективность вакцины, иммунная реакция, вирус, вакцинотерапия.

ВВЕДЕНИЕ

Применение для лечения онкологических заболеваний противоопухолевых вирусных вакцин изучается на протяжении нескольких десятилетий. Эффективность этого метода была доказана на экспериментальных моделях роста опухолей в работах [1–4]. Однако применение вакцинотерапии в клинической практике до настоящего времени не распространено и требует проведения дополнительных исследований.

Преимуществом метода вакцинотерапии по сравнению с лучевой и химиотерапией является отсутствие токсических эффектов, связанных с поражением нормальных здоровых тканей организма. Это позволяет расширять диапазон применяемых доз. Экспериментальный путь поиска наиболее эффективных стратегий введения вирусных вакцин требует больших затрат времени и средств на проведение исследований.

Затраты на экспериментальные исследования могут снижаться благодаря внедрению математи-

ческого моделирования на различных этапах проведения экспериментальных исследований, а также на этапе переноса полученных результатов в клинику.

Эффективность однократного введения вирусной вакцины оценивалась путем математического моделирования механизма гибели опухолевых клеток при вакцинотерапии, для чего применялись:

- математическая модель вакцинотерапии, описывающая механизм гибели опухолевых клеток в результате иммунного ответа на введение вируса [5–10];

- математическая модель Г.И. Марчука, описывающая динамику иммунного ответа организма на введение вируса [11, 12];

- математическая модель Н.Ф. Skipper, описывающая динамику снижения доли быстро делящихся опухолевых клеток по мере увеличения размеров опухоли [13];

- математическая модель противоопухолевой терапии с разрывными траекториями, оценивающая эффективность лечебных воздействий по тра-

екториям роста опухоли в эксперименте [10–14]; управление динамикой развития опухоли происходит дискретно в моменты введения вакцины извне.

Разработанная математическая модель вакцинотерапии служит адекватным и фундаментальным описанием механизма действия противоопухолевых вирусных вакцин, в этом заключается ее достоинство и отличие от других математических моделей, предложенных для исследования режимов применения вакцинотерапии [15–19].

Сложность динамических нелинейных математических моделей, описывающих реакции иммунной системы и механизм гибели опухолевых клеток при вакцинотерапии, не позволяет получить решение аналитическим путем без применения вычислительных программных комплексов. Разработанный программный модуль в среде MatLab-Simulink позволяет исследовать эффективность различных стратегий введения вирусных вакцин при широком диапазоне доз и моментов их введения. Это сокращает объем экспериментальных исследований и расходы на животных и препараты, что является экономически более эффективным.

Актуальность настоящего исследования заключается в том, что результаты моделирования позволяют обоснованно и целенаправленно дополнять результаты экспериментального и во многом эмпирического подхода к поиску эффективных стратегий лечения методом вакцинотерапии.

Цель данного исследования состоит в определении эффективности стратегии однократного введения вирусной вакцины, позволяющей максимизировать число погибших опухолевых клеток или достичь полного уничтожения опухолевых клеток.

Практическая значимость данной работы заключается в том, что математические модели, описывающие механизм гибели опухолевых клеток при введении вирусных вакцин были реализованы в виде программного комплекса для проведения компьютерных вычислительных экспериментов. Анализ полученных результатов позволяет получить зависимость эффективности иммунного ответа от дозы, описать динамику развития опухолевого процесса после введения вакцины. Полученные данные могут быть использованы для расчета оптимальных стратегий применения вирусных вакцин.

1. МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ДВУХЭТАПНОЙ ГИБЕЛИ ОПУХОЛЕВЫХ КЛЕТОК ПРИ ВАКЦИНОТЕРАПИИ

Процесс двухэтапной гибели опухолевых клеток в результате иммунного ответа организма на

введение вирусной вакцины представлен математической моделью противоопухолевой вакцинотерапии [5–10].

Процесс образования антител иммунной системой организма в ответ на вторжение чужеродных антигенов описан в математической модели инфекционного заболевания, которая разработана под руководством Г.И. Марчука [11, 12]. Данная модель служит базовой моделью, описывающей динамику развития иммунного ответа организма на введение любых вирусных вакцин. В работах [2–4, 20–25] было показано, что противоопухолевые вакцины не являются патогенными для организма. Следовательно, они не могут вызвать поражения здоровых органов и избирательно действуют только на инфицированные вирусом опухолевые клетки.

В связи с этим в модели Г.И. Марчука, вместо уравнения, описывающего динамику доли пораженного органа, была применена математическая модель вакцинотерапии, описывающая динамику двухэтапной гибели инфицированных вирусом опухолевых клеток. Параметры других трех уравнений модели инфекционного заболевания были адаптированы в соответствии с имеющимися траекториями роста экспериментальной опухоли после введения вирусной вакцины с вирусом ВЭЛ [2–10].

Таким образом, математическая модель инфекционного заболевания по теории Г.И. Марчука работает совместно с моделью вакцинотерапии.

1.1. Математическая модель вакцинотерапии

Математическая модель двухэтапной гибели опухолевых клеток после введения вирусной вакцины описывается дифференциальным уравнением вида [5–10]:

$$\begin{aligned} dN_V(t)/dt = & \lambda(t)N(t)\theta(t_V - t) + [\lambda(t - t_V) - \\ & - K_V I(\tau_1 - t) - K_{AV} A_V(t - t_V)]N_V(t - t_V)\theta(t_N - t) + \\ & + [\lambda(t - t_N) - K_{AN} A_N(t - t_N)]N_V(t - t_N), \end{aligned} \quad (1)$$

где $N(t)$ — численность популяции опухолевых клеток до введения вакцины, $\lambda(t)$ — темп размножения популяции опухолевых клеток в экспериментальной опухоли без введения вакцины, τ_1 — момент введения вакцины, Z_{CV} и Z_{CN} — время запаздывания иммунной реакции против вируса и против инфицированных опухолевых клеток, $A_V(t - t_V)$ — численность антител против вируса, $A_N(t - t_N)$ — численность антител против инфицированных опухолевых клеток, $t_V = \tau_1 + Z_{CV}$ — момент начала иммунной реакции против вирусов, $t_N = \tau_1 + Z_{CN}$ — момент начала иммунной реакции против опухолевых клеток, $\theta(t)$ — функция Хевизи



сайда, K_{AN} , K_{AV} и K_V — размерные коэффициенты [9, с. 50].

Первое слагаемое уравнения (1) описывает динамику роста опухолевых клеток до введения вирусной вакцины. Второе слагаемое описывает гибель инфицированных опухолевых клеток под действием вируса. Третье слагаемое уравнения описывает гибель опухолевых клеток под действием антител, которые образует иммунная система против инфицированных опухолевых клеток.

1.2. Математическая модель инфекционного заболевания по теории Г.И. Марчука

Для описания закономерностей развития реакции иммунной системы на появление чужеродных клеток была взята за основу математическая модель инфекционного заболевания [11–12]. Динамика численности вирусов согласно этой модели описывается уравнениями вида [9, с. 51]:

$$dV(t)/dt = \alpha_V V(t) - \beta_V A_V(t) V(t), \quad (2)$$

где $V_0 = V(\tau_1)$ — начальная доза вирусной вакцины, τ_1 — момент первого введения вирусной вакцины, α_V — темп размножения вирусов внутри клетки, β_V — темп гибели вирусов при их взаимодействии с антителами $A_V(t)$ [9, с. 51].

Начальное условие для решения уравнения (2) принято в модели вакцинотерапии в качестве управляющего параметра, характеризующего вводимую дозу вирусной вакцины $V_0 = V(\tau_1)$.

Первый этап иммунного ответа организма на введение вируса определяется численностью антител $A_V(t)$, которая вычисляется из уравнения:

$$dA_V(t)/dt = \alpha_A C_V(t - t_V) - \beta_{AV} A_V(t - t_V) V(t - t_V) - \beta_V A_V(t - \tau_1), \quad (3)$$

где α_A — темп образования антител из одной плазматической клетки, β_{AV} — темп убыли антител из-за взаимодействия с вирусами $A_V(t)$, β_V — темп уменьшения численности антител благодаря естественному разрушению.

Численность образования плазматических клеток $C_V(t)$ определяется из уравнения

$$\frac{dC_V(t)}{dt} = \alpha_C V(t - \tau_1) A_V(t - t_V) - \beta_{CV} [C_V(t - \tau_1) - C_{VN}], \text{ при } C_V(\tau_1) = C_{VN}, \quad (4)$$

где α_C — темп образования плазматических клеток, β_{CV} — размерный коэффициент, Z_{CV} — время запаздывания иммунной реакции на образование клона плазматических клеток. Второе слагаемое в

правой части этого уравнения отражает поддержание исходной численности плазматических клеток в норме C_{VN} [9, с. 51].

Второй этап иммунного ответа организма на образовавшиеся инфицированные клетки определяется численностью антител $A_N(t)$, которая вычисляется из уравнения:

$$\frac{dA_N(t)}{dt} = \alpha_{AN} C_N(t - t_N) - \beta_{AN} A_N(t - t_N) N_V(t - t_N) - \beta_{NN} A_N(t - t_N), \quad (5)$$

где α_{AN} — темп образования антител из одной плазматической клетки, β_{AN} — темп убыли антител $A_N(t)$ благодаря взаимодействию с инфицированными опухолевыми клетками $N_V(t)$, β_{NN} — темп уменьшения численности антител из-за естественного разрушения.

Численность образования плазматических клеток $C_N(t)$ определяется из уравнения

$$\frac{dC_N(t)}{dt} = \alpha_{CN} N_V(t) A_N(t - t_N) - \beta_{CN} [C_N(t - t_N) - C_{NN}], \quad (6)$$

при $C_N(\tau_1) = C_{NN}$, где α_{CN} — темп образования плазматических клеток, β_{CN} — размерный коэффициент, Z_{CN} — время запаздывания иммунной реакции для образования клона плазматических клеток против инфицированных опухолевых клеток [9, с. 51].

1.3. Кинетическая математическая модель роста опухоли по теории H.F. Skipper

Для адекватного описания эффективности действия вирусных вакцин необходимо учитывать, что один из факторов высокой избирательности действия вирусов по отношению к опухолевым клеткам состоит в высокой скорости их деления по сравнению с нормальными тканями организма. В измеряемом объеме опухоли присутствуют фракции быстро и медленно делящихся опухолевых клеток. По мере увеличения размера опухоли доля фракции быстро делящихся опухолевых клеток уменьшается, а доля фракции медленно делящихся клеток и временно не делящихся клеток напротив увеличивается, что описано в математической модели роста опухоли в работе H.F. Skipper [13].

Для описания снижения избирательности действия противоопухолевых вирусов в опухолях больших размеров в модель введена функция $P(t)$, которая описывает динамику снижения доли быстро

пролиферирующей фракции клеток по мере увеличения размеров опухоли [9, с. 51–52]:

$$P(t) = 1 - [1/K_p \pi^* \arctg(2\alpha_p \beta_p t / (1 - \beta_p^2 t^2))], \quad (7)$$

где α_p и K_p — постоянные параметры, t — текущее время роста популяции опухолевых клеток, $\beta_p = 1/t^*$, где t^* — момент времени, при котором численности фракций быстро и медленно пролиферирующих клеток равны.

Тогда число инфицированных клеток в измеряемом объеме опухоли вычисляется как $N_V(t_N) = N(t_N)P(t)$.

1.4. Математическая модель противоопухолевой терапии с разрывными траекториями для прогнозирования динамики роста опухоли

Математическая модель противоопухолевой терапии с разрывными траекториями [5–10] применяется для построения кинетических кривых роста опухоли после введения вирусной вакцины. Управление динамикой развития опухоли происходит дискретно в моменты введения вакцины в организм. Управляющим параметром служит доза вируса.

Дифференциальное уравнение, описывающее двухэтапную гибель опухолевых клеток, имеет вид:

$$\begin{aligned} dN(t)/dt = & \lambda(t)N(t)\theta(t_V - t) - S_V(V_0, \tau_1)N(t_V) \times \\ & \times \delta(t - t_V) + \lambda(t - t_V)N(t - \varepsilon_V(V_0))\theta(t_N - t) - \\ & - S_N(N_V, \tau_1)N(t_N)\delta(t - t_N) + \\ & + \lambda(t - t_N)N(t - \varepsilon_N(V_0)), \end{aligned} \quad (8)$$

где $N(t)$ — численность опухолевых клеток в измеряемом объеме опухоли в эксперименте в момент времени t , $\delta(t - t_V)$ и $\delta(t - t_N)$ — импульсная функция Дирака, описывающая мгновенную гибель опухолевых клеток на каждом из двух этапов стимуляции иммунной системы, $S_V(V_0, \tau_1)$ и $S_N(V_0, \tau_1)$ — относительное уменьшение размера опухоли в момент их гибели на каждом из двух этапов, $\varepsilon_V(V_0, t_V)$ и $\varepsilon_N(V_0, t_N)$ — задержка роста опухоли после гибели клеток на каждом из двух этапов иммунного ответа.

В модели вакцинотерапии (1)–(7) было принято допущение о том, что относительное уменьшение объема опухоли происходит только из-за гибели инфицированных вирусом опухолевых клеток $N_V(t)$. Число погибающих опухолевых клеток вычисляется как разность между максимальной и

минимальной численностью инфицированных клеток после каждого этапа гибели из уравнений:

$$\Delta N_V(t_V) = N_V(t_1^V) - N_V(t_2^V), \quad (9)$$

$$\Delta N_V(t_N) = N_V(t_1^N) - N_V(t_2^N), \quad (10)$$

где t_1^V , t_1^N и t_2^V , t_2^N — моменты начала и окончания иммунной реакции против вируса и против инфицированных опухолевых клеток соответственно.

Относительное уменьшение размера опухоли в моменты их гибели на каждом из двух этапов определяется из уравнений:

$$S_V(V_0, t_V) = \frac{\Delta N_V(t_V)}{N(t_V)}, \quad (11)$$

$$S_N(V_0, t_N) = \frac{\Delta N_V(t_N)}{N(t_N)}, \quad (12)$$

где $N(t_V)$ и $N(t_N)$ — численность опухолевых клеток в измеряемом объеме опухоли в эксперименте в моменты t_V и t_N .

Гибель опухолевых клеток вызывает задержку роста опухоли. Длительность задержки роста опухоли $\varepsilon_V(V_0, t_V)$ и $\varepsilon_N(V_0, t_N)$ определяется интервалом времени, в течение которого происходят гибель и последующее восстановление численности опухолевых клеток до исходных размеров в момент начала иммунного ответа. Значения длительности задержки роста опухоли $\varepsilon_V(V_0, t_V)$ и $\varepsilon_N(V_0, t_N)$ определялись из условия равенства размеров опухоли в момент начала иммунного ответа и через интервал времени, соответствующий задержке роста опухоли согласно уравнениям:

$$N(t_V) = N(t_V + \varepsilon_V(V_0, t_V)), \quad (13)$$

$$N(t_N) = N(t_N + \varepsilon_N(V_0, t_N)). \quad (14)$$

Согласно модели (8) кинетические кривые роста опухоли до и после каждого этапа гибели опухолевых клеток описываются уравнениями Гомпертца в виде (15) и (16), в которых учитывается сдвиг во времени на длительность задержки роста:

$$\begin{aligned} N(t - t_V) = \\ = RN_0^V \exp(\alpha_N(1 - \exp(-\beta_N(t_V - \varepsilon_V(V_0))))), \end{aligned} \quad (15)$$

$$\begin{aligned} N(t - t_N) = \\ = RN_0^N \exp(\alpha_N(1 - \exp(-\beta_N(t_N - \varepsilon_N(V_0))))), \end{aligned} \quad (16)$$

где $RN_0^V(V_0)$ и $RN_0^N(V_0)$ — численности выживших опухолевых клеток после первого и второго этапов



Значения параметров моделей

Уравнение	Параметр	Описание
(8)	$\alpha_N = 3,3613; \beta_N = 0,0332;$ $N_\infty = 23; N_0 = 0,76$	Параметры функции Гомперца, аппроксимирующей экспериментальные кривые роста популяции опухолевых клеток без введения вакцины (контроль)
(1)	$\tau_1 = 1$ $Z_{CV} = 4,5$ $Z_{CN} = 10,5$ $t_V = \tau_1 + Z_{CV}$ $t_N = \tau_1 + Z_{CN}$	Момент первого введения вирусной вакцины Время запаздывания иммунного ответа против вируса Время запаздывания иммунного ответа против инфицированных опухолевых клеток Момент начала иммунного ответа против вируса Момент начала иммунного ответа против инфицированных опухолевых клеток
(2)	$\alpha_V = 0,1$ $\beta_V = 15$ $V_0 = 0,015$	Темп размножения вирусов Темп гибели вирусов при их взаимодействии с антителами Доза вирусной вакцины
(3)	$\alpha_A = 100$ $\beta_{AV} = 70$ $\beta_A = 5$ $Av^{\max} = 1,05$	Темп образования антител из одной плазматической клетки Темп убыли антител из-за взаимодействия с вирусами Темп уменьшения численности антител из-за естественного разрушения Максимальное расчетное количество антител
(4)	$\alpha_C = 100$ $\beta_{CV} = 4,5$ $C_{VN} = 0,001$	Темп образования плазматических клеток Размерный коэффициент Исходная численность плазматических клеток
(5)	$\alpha_{AN} = 30$ $\beta_{AN} = 6,2$ $\beta_{NN} = 6,3$ $A_N^{\max} = 4,64$	Темп образования антител из одной плазматической клетки Темп убыли антител из-за взаимодействия с опухолевыми клетками Темп уменьшения количества антител за счет естественного разрушения Максимальное расчетное количество антител
(6)	$\alpha_{CN} = 76,677$ $\beta_{CN} = 38$ $C_{NN} = 0,0001$	Темп образования плазматических клеток Размерный коэффициент Исходная численность плазматических клеток
(1)	$K_V = 0,25;$ $K_{AV} = 0,8; K_{AN} = 0,8$	Постоянные коэффициенты уравнения динамики численности клеток после однократного введения вакцины
(7)	$\alpha_p = 0,3;$ $K_p = 0,95; \beta_p = 1/t^*$ $t^* = 35$ сут	Параметры функции $P(t)$, описывающей динамику снижения доли быстро пролиферирующих клеток по мере увеличения размера опухоли Момент времени, при котором численность фракций быстро и медленно пролиферирующих клеток равны
(8)	$\varepsilon_V = 4,3$ сут $\varepsilon_{CN} = 9,8$ сут $t_1^V = 5,82; t_1^N = 12,32$ $t_2^V = 7,13; t_2^N = 14,288$ $N_V^V(t_1^V) = 1,31;$ $N_V^N(t_1^N) = 1,15$	Задержка роста опухолевых клеток в результате их гибели под действием вируса Задержка роста опухолевых клеток в результате их гибели под действием антител против инфицированных опухолевых клеток Моменты достижения максимальной численности инфицированных опухолевых клеток к началу иммунной реакции Моменты достижения минимальной численности инфицированных опухолевых клеток к окончанию иммунной реакции Максимальные значения численности инфицированных опухолевых клеток перед началом иммунной реакции на каждом из двух этапов их гибели

гибели, способных возобновить рост опухоли. Их значения определяются из уравнений [10]:

$$RN_0^V = N(t_V) - \Delta N_V^V(V_0, t_V),$$

$$RN_0^N = N(t_N) - \Delta N_V^N(V_0, t_N).$$

При введении вирусной вакцины инфицируется не вся популяция опухолевых клеток, которая измеряется в эксперименте $N(t)$. Вирусы осаждаются преимущественно на фракцию быстро делящихся опухолевых клеток. Выжившие и не инфицированные опухолевые клетки RN_0^V и RN_0^N способны возобновить рост опухоли, который наблюдается в эксперименте.

Для оценки эффективности дозы вирусной вакцины, которая может уничтожить все клетки опухоли, в модель было введено пороговое значение минимальной численности опухолевых клеток N_0^{let} , которое способно дать начало процессу возобновления роста опухоли. Если траектория роста опухоли после введения вакцины опускается ниже порогового значения $RN_0^V < N_0^{let}$ и $RN_0^N < N_0^{let}$, это трактуется в модели как полное уничтожение опухолевых клеток.

Значения параметров модели вакцинотерапии и модели противоопухолевой терапии с разрывными траекториями приведены в таблице.

2. РЕЗУЛЬТАТЫ ВЫЧИСЛИТЕЛЬНОГО ЭКСПЕРИМЕНТА ДЛЯ ОЦЕНКИ ЭФФЕКТИВНОСТИ РАЗЛИЧНЫХ СТРАТЕГИЙ ВАКЦИНОТЕРАПИИ

Вычислительный эксперимент проводился в системе MatLab-Simulink, где был реализован комплекс программ, описывающих механизм гибели и процесс роста опухоли после введения вирусной вакцины. С помощью разработанного комплекса программ была оценена эффективность метода вакцинотерапии при введении вирусных вакцин в диапазоне доз от $V_0 = 0,01$ до $V_0 = 0,1$ и моментов их введения на интервале от 1 сут до $t^* = 35$ сут [9].

Эффективность метода оценивалась после каждого из двух этапов иммунного ответа по двум параметрам:

- численность образованных антител $A_V(V_0, \tau_1)$ и $A_N(V_0, \tau_1)$,
- численность погибающих опухолевых клеток $\Delta N_V(V_0, t_V)$ и $\Delta N_V(V_0, t_N)$.

В результате вычислительного эксперимента были получены расчетные графики изменения числа антител на первом и втором этапе иммунного ответа (рис. 1)) [9].

Эти графики отражают два принципиально разных типа зависимости численность образуемых антител от дозы вирусной вакцины. Динамика образования антител против вируса $A_V(t)$ имеет линейный характер, возрастает пропорционально введенной дозе вируса и не зависит от размера опухоли в момент введения вакцины (см. рис. 1, а)).

Динамика образования антител против инфицированных опухолевых клеток $A_N(t)$ имеет нелинейный характер и зависит от размера опухоли в момент введения вакцины (см. рис. 1, б)). Полученная нелинейная зависимость позволяет определить максимально эффективную дозу вируса

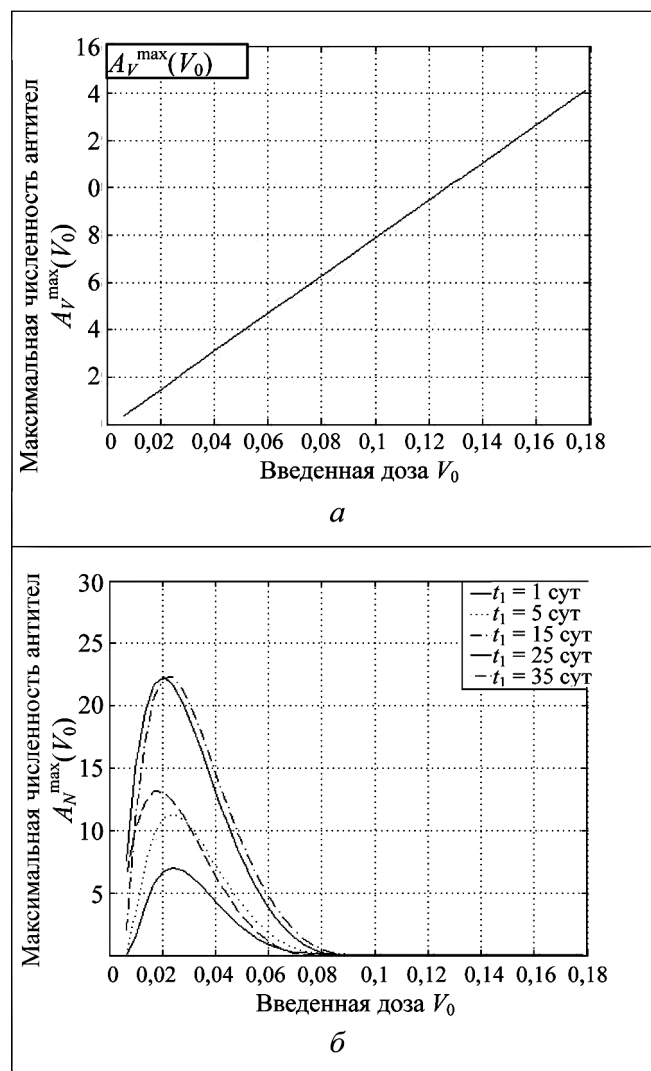


Рис. 1. Расчетные графики образования антител в зависимости от введенной дозы V_0 при разных моментах введения вакцины $\tau_1 = 1, 5, 15, 25, 35$ сут: а — $A_V(V_0)$ антитела против вируса (первый этап иммунного ответа); б — $A_N(V_0, \tau_1)$ антитела против инфицированных опухолевых клеток (второй этап иммунного ответа)

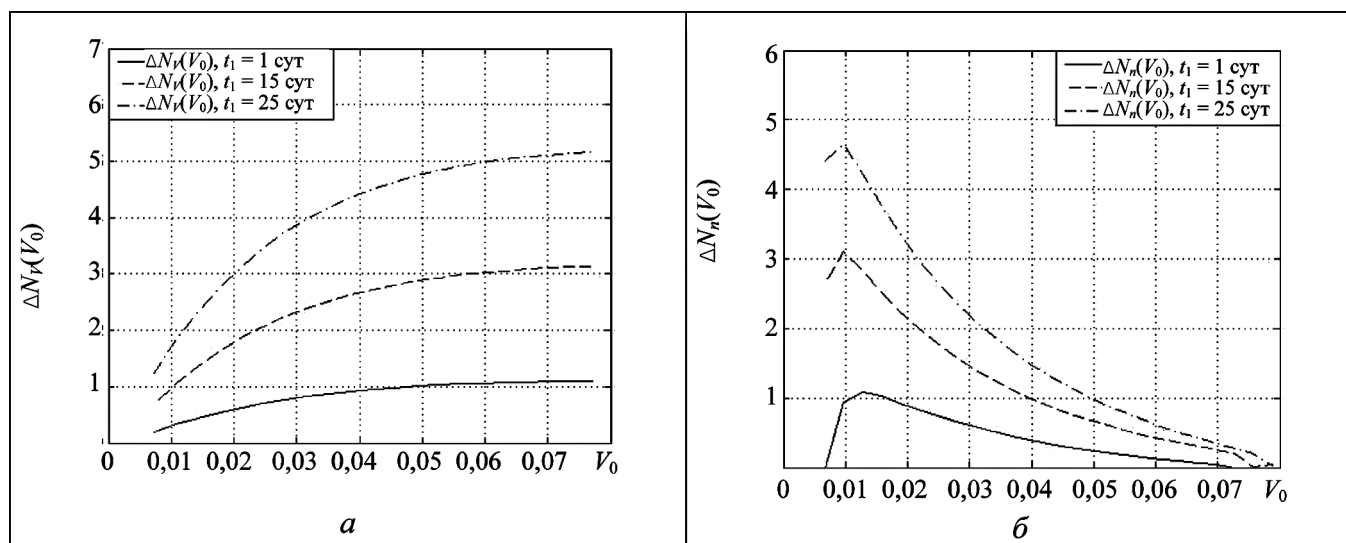


Рис. 2. Расчетный график числа погибающих инфицированных клеток в зависимости от дозы при введении вакцины на $t_1 = 1, 15, 25$ сут:

$a - \Delta N_V^V(V_0)$ (первый этап иммунного ответа); $b - \Delta N_V^N(V_0)$ (второй этап иммунного ответа)

$V_0^{\max} = 0,024$, при введении которой образуется максимальное число антител против инфицированных опухолевых клеток.

При введении незначительных доз меньше $V_0^{\max}$ после первого этапа иммунного ответа погибает незначительная часть опухолевых клеток, но образуется большое число инфицированных опухолевых клеток, которые способны вызывать иммунный ответ на втором этапе. При увеличении дозы свыше $V_0^{\max}$ растет число клеток, погибающих на первом этапе иммунного ответа. Это приводит к уменьшению образования инфицированных клеток. В результате снижается число образуемых антител $A_N(V_0, \tau_1)$ на втором этапе иммунного ответа.

При введении значительных доз вируса более $V_0 = 0,1$ гибнет практически вся популяция опухолевых клеток уже на первом этапе. Это приводит к быстрому и эффективному уничтожению опухоли, но при этом отсутствует второй этап иммунного ответа и не образуются антитела против инфицированных опухолевых клеток.

Как показывают результаты вычислительного эксперимента, образование антител против инфицированных опухолевых клеток $A_N(V_0, \tau_1)$ зависит не только от дозы, но и от момента введения τ_1 вирусной вакцины, а значит, от размера опухоли в момент введения. Динамика снижения доли быстро пролиферирующей фракции опухолевых клеток по мере увеличения размера опухоли описывается функцией $P(t)$ (7).

На основании зависимостей $A_V(t)$ и $A_N(V_0, \tau_1)$ определялась численность погибающих опухолевых клеток на каждом из двух этапов иммунного ответа (рис. 2).

Динамика гибели опухолевых клеток на каждом этапе принципиально различается. Число погибших клеток на первом этапе $\Delta N_V^V(V_0, t_V)$ при увеличении дозы растет, но не может превышать размера опухоли в момент введения вакцины. При увеличении размера опухоли в момент введения вакцины растет и число погибающих клеток (см. рис. 2, a). На втором этапе число погибающих клеток $\Delta N_V^N(V_0, t_N)$ при увеличении дозы вакцины снижается в результате уменьшения образования антител $A_N(V_0, \tau_1)$ (см. рис. 2, b).

3. КИНЕТИЧЕСКИЕ ТРАЕКТОРИИ РОСТА ОПУХОЛИ ПОСЛЕ ВВЕДЕНИЯ ВИРУСНОЙ ВАКЦИНЫ

Для прогнозирования динамики роста опухоли после введения различных доз вирусной вакцины в различные моменты введения авторы воспользовались математической моделью противоопухолевой терапии с разрывными траекториями согласно уравнениям (8)–(16) [5–10]. Объектом управления служит процесс изменения численности опухолевых клеток после введения вирусной вакцины, а управляющим параметром — доза вакцины при заданном размере опухоли в момент ее введения.

Оценка эффективности стратегии введения вирусных вакцин проводилась на основании дости-

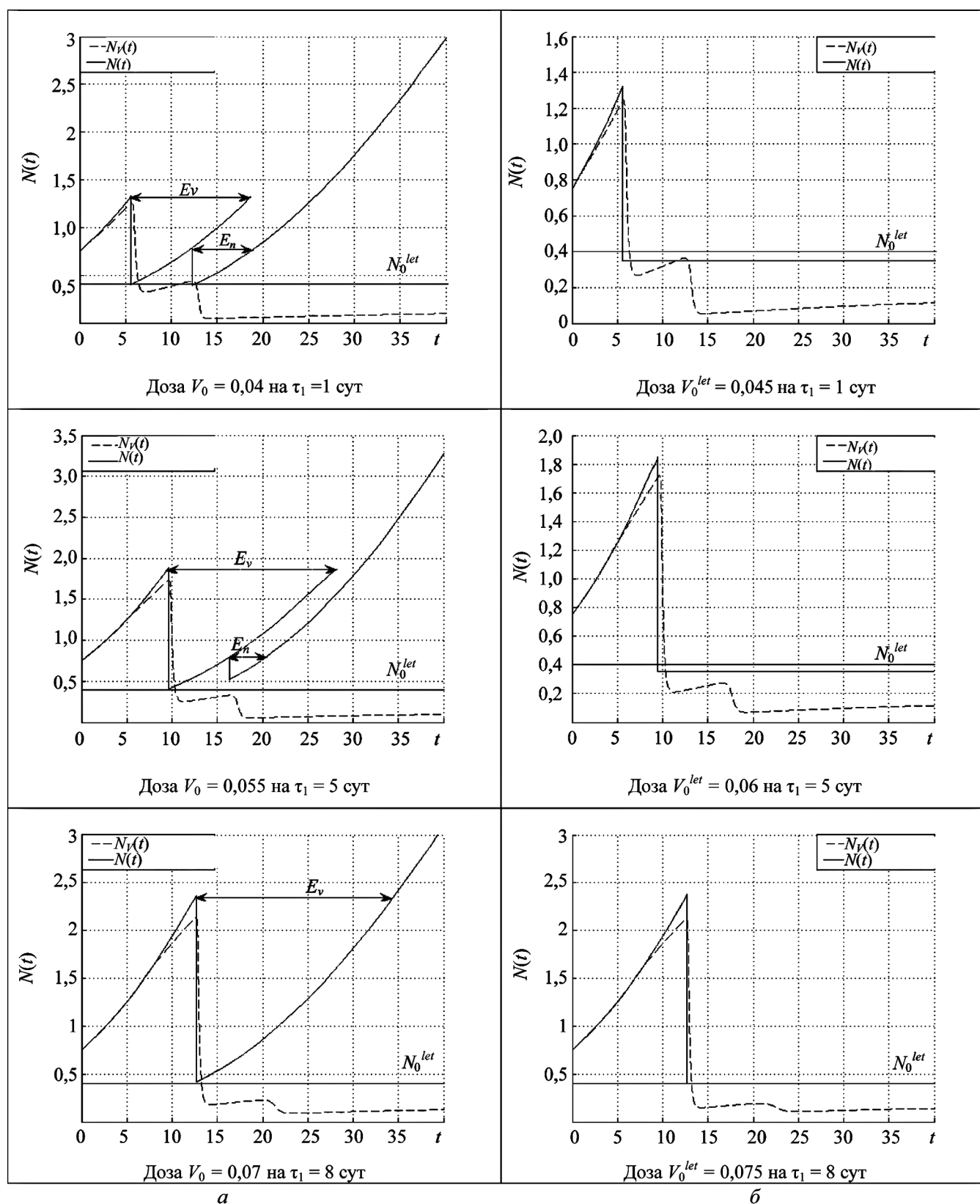


Рис. 3. Кинетические траектории роста опухоли при введении вирусной вакцины на $\tau_1 = 1, 5, 8$ сут: **а** — при дозах $RN_0^V \geq N_0^{let}$; **б** — при дозах $RN_0^V < N_0^{let}$

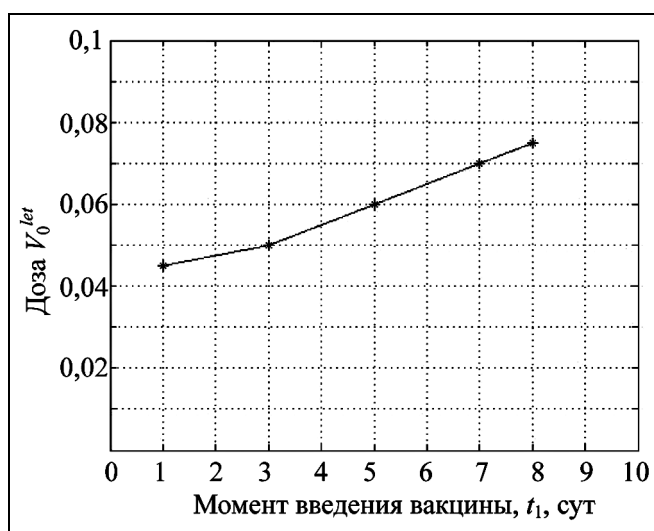


Рис. 4. Зависимость летальной дозы $V_0^{let}(\tau_1)$ от размера опухоли в момент введения вакцины

жения порогового значения N_0^{let} , которое определяет предел минимальной численности опухолевых клеток, способных дать начало процессу роста опухоли. Возобновление роста опухоли после каждого из двух этапов их гибели возможно при условии, что число выживших опухолевых клеток RN_0^V и RN_0^N выше порогового значения. Если число выживших опухолевых клеток ниже порогового значения, то возобновления роста опухоли не будет, что трактуется в модели как полное уничтожение опухолевых клеток. Длительность периода жизни условно принималась продолжительности жизни экспериментальных животных без опухоли.

По имеющимся экспериментальным данным [2–4] было принято, что значение $N_0^{let} = N_0/2 = 0,76/2 = 0,4$, где N_0 — начальная численность популяции опухолевых клеток.

Кинетические траектории роста опухоли, полученные в результате вычислительного эксперимента в системе MatLab-Simulink (рис. 3), позволили определить летальные дозы $V_0^{let}(\tau_1)$, при введении которых численность опухолевых клеток становится ниже порога, что трактуется в модели как полное уничтожение опухолевых клеток.

На основе этих траекторий была построена зависимость величины летальной дозы V_0^{let} от размера опухоли в момент введения вирусной вакцины (рис. 4).

Результаты исследования показали, что на данном виде опухоли полное уничтожение опухоле-

вых клеток при однократном введении вакцины возможно только для опухолей, продолжительность роста которых меньше 9 сут.

Для опухолей, продолжительность роста которых более 8 сут, достичь полного уничтожения опухолевых клеток при однократном введении вакцины в дозах исследуемого диапазона невозможно. Полученные результаты позволяют выделить две стратегии применения противоопухолевых вирусных вакцин.

Одна стратегия, при которой возможно снижение численности выживших клеток ниже порога N_0^{let} (см. рис. 3, б) на первом этапе иммунного ответа, может трактоваться как полное уничтожение опухолевых клеток. Однако эта стратегия возможна только для ограниченного размера опухоли (см. рис. 4).

Другая стратегия, при которой возможно стабилизировать размер опухоли, зафиксированный в момент начала лечения, заключается в повторных введениях вирусной вакцины. Стратегия стабилизации возможна для опухолей, продолжительность роста которых не превышает 25 сут [9].

Исследование стратегии стабилизации роста опухоли путем повторных введений вирусной вакцины является задачей дальнейших исследований.

ЗАКЛЮЧЕНИЕ

Представлено математическое описание сложных процессов двухэтапной гибели опухолевых клеток, которые происходят в результате иммунного ответа организма на введение противоопухолевой вирусной вакцины. Математическое описание представляет собой систему нелинейных дифференциальных уравнений, которые реализованы в виде программного комплекса в системе MatLab-Simulink. В результате вычислительного эксперимента получена оценка эффективности однократных введений вирусной вакцины в широком диапазоне доз при заданных размерах опухоли в момент ее введения.

Анализ результатов вычислительного эксперимента показал, что возможны две стратегии введения противоопухолевых вирусных вакцин.

Одна из них позволяет добиться полного уничтожения опухолевых клеток при однократном введении вакцины. Получен график зависимости дозы вирусной вакцины от размера опухоли в момент введения вакцины. Показано, что полное уничтожение опухолевых клеток возможно только для опухолей небольших размеров, продолжительность роста которых не превышает 8 сут. Полученный результат свидетельствует о том, что ранняя диагностика рака и выявление опухолей на ранней

стадии их роста позволяет добиться уничтожения опухолевых клеток при введении доз вакцины, соответствующих их размерам.

Другая стратегия позволяет стабилизировать размер опухоли путем повторных введений вакцины. Опухолевые клетки гибнут на каждом из двух этапов иммунного ответа, что приводит к образованию антител против собственных опухолевых клеток. Присутствие таких антител в организме служит регулятором процесса стабилизации роста опухоли. Стратегия стабилизации роста опухоли позволяет избежать хирургического вмешательства и применять вакцинотерапию в качестве регулярной прививки от собственных опухолевых клеток.

Разработка новых методов лечения онкологических заболеваний непосредственно связана со стоимостью проведения научных экспериментальных исследований. Снижение их стоимости может осуществляться путем активного внедрения методов математического моделирования и проведения компьютерных вычислительных экспериментов, которые позволят сократить расходы на экспериментальных животных и препараты.

ЛИТЕРАТУРА

1. Барышников А.Ю. Принципы и практика вакцинотерапии рака // Бюллетень СО РАМН / Москва. — 2004. — № 2 (112). — С. 59–63.
2. Громова А.Ю. Противоопухолевые свойства вакцинного штамма вируса венозного энцефаломиелита и его онколизата: дис... канд. мед. наук. — СПб., 1999. — 114 с.
3. Уразова Л.Н. Эффективность и механизмы противоопухолевого действия вирусных вакцин при экспериментальном онкогенезе: дис. ... д-ра биол. наук. — СПб., 2003. — 196 с.
4. Видяева И.Г. Вирусные вакцины и их онколизаты в терапии экспериментальных опухолей: дис. ... канд. мед. наук. — Томск, 2005. — 134 с.
5. Бабушкина Н.А. Оценка управляющих дозовых воздействий противоопухолевой вакцинотерапии с помощью математического моделирования // Проблемы управления. — 2013. — № 5. — С. 60–65.
6. Бабушкина Н.А., Глумов В.М. Математическое моделирование механизмов противоопухолевого действия вирусных вакцин / Тр. XI междунар. науч. конф. «Физика и радиоэлектроника в медицине и экологии» ФРЭМЭ 2014, Владимир — Суздаль. — Кн. 1. — С. 153–158.
7. Бабушкина Н.А., Кузина Е.А. Компьютерные технологии на основе математического моделирования в системной экспериментальной онкологии // Тр. восьмой междунар. конф. «Управление развитием крупномасштабных систем» (MLSD'2015), Москва. — С. 272–284.
8. Бабушкина Н.А., Глумов В.М., Кузина Е.А. Применение компьютерных технологий при экспериментальном изучении эффективности противоопухолевых вирусных вакцин // Тр. XII междунар. науч. конф. «Физика и радиоэлектроника в медицине и экологии» (ФРЭМЭ 2016), Владимир — Суздаль, 4–7 июля 2016 г. — Кн. 1. — С. 116–121.
9. Бабушкина Н.А., Глумов В.М., Кузина Е.А. Применение математического моделирования для оценки эффективности метода противоопухолевой терапии // Проблемы управления. — 2017. — № 3. — С. 49–56.
10. Babushkina N., Kuzina E. Analytical study of the antitumor viral vaccine introduction regimens based on mathematical modeling // Advances in Systems Science and Applications. — 2018. — Vol. 18, N 1. — P. 59–84.
11. Марчук Г.И. Математические модели в иммунологии. Вычислительные методы и эксперименты. — М.: Наука, 1991. — 304 с.
12. Романюха А.А. Математические модели в иммунологии и эпидемиологии инфекционных заболеваний. — М.: БИНОМ. Лаборатория знаний, 2011. — 293 с.
13. Skipper H.F. Kinetics of mammary tumor cell-growth and implications for therapy // Cancer. — 1971. — Vol. 28, N 6. — P. 1479–1499.
14. Бабушкина Н.А. Управление процессом химиотерапии с использованием ферромагнитных наночастиц // Проблемы управления. — 2011. — № 3. — С. 56–63.
15. Болодурина И.П., Луговская Ю.П. Оптимальное управление иммунологическими реакциями организма человека // Проблемы управления. — 2009. — № 5. — С. 44–52.
16. Русаков С.В., Чирков М.В. Математическая модель влияния иммунотерапии на динамику иммунного ответа // Проблемы управления. — 2012. — № 6. — С. 45–50.
17. Русаков С.В., Чирков М.В. Идентификация параметров и управление в математических моделях иммунного ответа // Российский журнал биомеханики. — 2014. — Т. 18, № 2. — С. 259–269.
18. Kogan Y., Halevi-Tobias K., Elishmeren M., et al. Reconsidering the Paradigm of Cancer Immunotherapy by Computationally Aided Real-time Personalization // Cancer Res. — 2012. — N 72. — P. 2218–2227.
19. Palladini, A., Nicoletti, G., Pappalardo, F., et al. In silico Modeling and In vivo Efficacy of Cancer-Preventive Vaccinations // Cancer Res. — 2010. — № 70. — P. 7755–7763.
20. Локтев В.Б., Иванкина Т.Ю., Немцов С.В., Чумаков П.М. Онколитические парвовирусы. Новые подходы к лечению раковых заболеваний // Вестник Российской академии медицинских наук. — 2012. — № 2. — С. 42–47.
21. Лежнин Ю.Н., Кравченко Ю.Е., Фролова Е.И. и др. Онкотоксические белки в противораковой терапии: механизмы действия // Молекулярная биология. — 2015. — Т. 49, № 2. — С. 264–278.
22. Hristov G., Krømer M., Li J., N., et al. Through its non structural protein NS1, parvovirus H-1 induces apoptosis via accumulation of reactive oxygen species // Journal of virology. — 2010. — Vol. 84, N 12. — P. 5909–5922.
23. Rommelaere J., Geletneky K., Angelova A.L., et al. Oncolytic parvoviruses as cancer therapeutics // Cytokine & growth factor reviews. — 2010. — Vol. 21. — N 2. — P. 185–195.
24. Grekova S.P., Aprahamian M., Daeffler L., et al. Interferon γ improves the vaccination potential of oncolytic parvovirus H-1PV for the treatment of peritoneal carcinomatosis in pancreatic cancer // Cancer biology & therapy. — 2011. — Vol. 12, N 10. — P. 888–895.
25. Angelova A.L., Aprahamian M., Balboni G., et al. Oncolytic rat parvovirus H-1PV, a candidate for the treatment of human lymphoma: In vitro and in vivo studies // Molecular Therapy. — 2009. — Vol. 17, N 7. — P. 1164–1172.

Статья представлена к публикации членом редколлегии А.И. Михальским.

Бабушкина Нина Александровна — канд. биол. наук, ст. науч. сотрудник, babushkina_na@mail.ru,

Кузина Екатерина Алексеевна — аспирант, инженер-программист, kate_k93@mail.ru,

Лоос Артем Александрович — инженер-программист, art23.a@yandex.ru,

Беляева Екатерина Владимировна — инженер-программист, katya_969696@mail.ru,

Институт проблем управления им. В.А. Трапезникова РАН, г. Москва.

КОМПЛЕКСИРОВАНИЕ РАСХОДОМЕРНОЙ И УРОВНЕМЕРНОЙ ИНФОРМАЦИИ В СИСТЕМАХ УПРАВЛЕНИЯ РАСХОДОВАНИЕМ ТОПЛИВА

В.К. Завадский, В.П. Иванов, Е.Б. Каблова, Л.Г. Кленовая

Предложены способ комплексирования уровнемерной и расходомерной информации и новые алгоритмы управления, позволяющие повысить точность работы системы управления расходом топлива в штатных условиях и при отказах бортовой аппаратуры.

Ключевые слова: комплексирование информации, фильтрация, точность управления, нештатная ситуация.

ВВЕДЕНИЕ

Тема статьи связана с бортовыми системами управления расходом топлива (СУРТ) жидкостных ракет-носителей (РН) и разгонных блоков (РБ), где в качестве датчиков в топливной системе одновременно применяются дискретные датчики уровня («измерительные точки») и датчики расхода компонентов топлива (типа «вертушка»). В качестве примеров можно привести топливные системы боковых и центрального блоков активно эксплуатируемой РН «Союз», применяемой, в том числе, и в пилотируемом варианте разрабатываемого в КБ «Салют» кислородно-водородном разгонном блоке тяжелого класса и др.

Задачи СУРТ состоят в обеспечении одновременного окончания рабочих запасов компонентов топлива — окислителя и горючего и своевременной выдачи сигнала об их окончании. Точность, с которой решаются указанные задачи, существенно влияет на потенциальную грузоподъемность РН (РБ) и безопасность их эксплуатации. Поэтому к точностным и надежностным характеристикам датчиков СУРТ предъявляются чрезвычайно высокие требования [1, 2].

Эффективное средство повышения качества работы бортовых систем заключается в применении фильтрации [3], основанной на осреднении избыточной измерительной информации. Источниками избыточности информации служат традиционно применяемое в ракетной технике троирование датчиков (структурная избыточность) и наличие функциональной связи между измерениями, поступающими в систему в различные моменты вре-

мени (функциональная избыточность, частным случаем которой является структурная избыточность). Предлагаемый способ повышения точности совместно применяемых датчиков также основан на возможности взаимного оценивания и осреднения погрешностей при комплексной обработке измерительной информации.

В настоящее время в существующих системах СУРТ такая возможность не используется или используется не полностью, что, по-видимому, связано с существенной разнородностью высокочастотных расходомерных и широкобазных измерений уровня, затрудняющей их комплексирование. Датчики расхода компонентов топлива обычно применяются в системе регулирования соотношения компонентов (РСК), поддерживающей требуемое значение коэффициента соотношения расходов компонентов и образующей внутренний контур СУРТ. Требуемое значение коэффициента соотношения формируется из условия обеспечения синхронизации моментов опорожнения баков во внешнем контуре управления по дискретной широкобазной уровнемерной информации об оставшихся объемах компонентов.

1. СПОСОБ КОМПЛЕКСИРОВАНИЯ УРОВНЕМЕРНОЙ И РАСХОДОМЕРНОЙ ИНФОРМАЦИИ

Поясним физическую сущность предлагаемого способа комплексирования (совместной обработки) уровнемерной и расходомерной информации на примере топливного бака, содержащего жидкий компонент топлива, подаваемый через питающий трубопровод в двигатель. Вдоль оси бака на задан-

ных уровнях установлены дискретные датчики — «измерительные точки», формирующие сигналы срабатывания в моменты прохождения зеркалом жидкости заданных уровней, соответствующих известным объемным запасам компонента. В питающей магистрали на входе в двигатель установлен датчик расхода компонента — «вертушка», формирующий высокочастотный импульсный сигнал с частотой, пропорциональной объемному расходу компонента.

В дальнейшем условно предполагается, что в полете вместо фиксации в моменты срабатывания «измерительных точек» бортового времени (как это делается на РН «Союз») или значений специальным образом формируемой оценки текущего запаса топлива (производимой на РН «Протон», «Зенит», «Ангара» и др.) в СУРТ фиксируется *суммарное число импульсов*, сформированных «вертушкой» от начала работы системы до момента срабатывания чувствительного элемента.

Суммарное число импульсов (интеграл от расхода) пропорционально объему жидкого компонента, израсходованного через двигатель. Поэтому номинальный начальный (на момент включения системы) объем рабочего запаса компонента и объемы под j -ми «измерительными точками» удобно выражать через расчетно-номинальное число импульсов N_j , сформированных «вертушкой» в интервале времени от момента начала работы системы ($j = 0$) или срабатывания j -й «измерительной точки», $j = 1, \dots, J$, до момента окончания рабочего запаса. Выраженные числом импульсов расходомера рабочие объемы под измерительными точками можно определять непосредственно при холодных стендовых испытаниях или во время специальным образом организованной тарировки баков с суммированием импульсов расходомера и фиксацией указанных сумм в моменты срабатывания «измерительных точек» и в момент окончания рабочего запаса компонента. Для исключения ряда методических погрешностей указанные мероприятия целесообразно проводить с помощью штатной бортовой измерительной аппаратуры (датчиков уровня и расходомеров). Поскольку условия стендовых испытаний отличаются от полетных условий, полученные при испытаниях значения N_j должны быть скорректированы с учетом различия типов (вязкости, плотности) и расходов компонентов, используемых при испытаниях и в полете, расчетных изменений объема из-за изменения среднебаковой температуры компонента в полете и его испарения, расчетной деформации баков под действием наддува и перегрузок и др.

Расчетно-номинальное число импульсов «вертушки» от момента начала работы системы до момента срабатывания j -й «измерительной точки», характеризующее израсходованный объем компо-

нента, представляется выражением $T_j = N_0 - N_j$, $j = 1, 2, \dots, J$.

Обозначим через n_j измеренное (фактическое) число импульсов, зафиксированное в полете от момента начала работы системы до момента срабатывания j -й «измерительной точки», а через $\Delta n_{измj}$ — рассогласование между измеренным и расчетно-номинальным числом импульсов, сформированных «вертушкой» в этом интервале времени: $\Delta n_{измj} = n_j - T_j$, $j = 1, 2, \dots, J$.

Основными причинами возникновения измеренных рассогласований в полете является ряд случайных факторов (возмущений), ухудшающих точность работы системы:

- отклонение от номинала начального значения объема компонента (связанное с ошибкой заправки компонента, изменением его температуры во время стоянки, разбросом расходов компонента до момента включения системы и др.);

- ошибки измерения объемов под «измерительными точками» (связанные с колебаниями уровня жидкого компонента в полете, ошибками измерения объемов N_j при тарировке, ошибками экстраполяции этих объемов на полетные условия и др.);

- отклонение значения коэффициента связи между измеряемым объемным расходом и частотой следования импульсов с датчика расхода (обусловленное различием характеристик экземпляров расходомеров, применяемых при тарировке и в полете, динамическими погрешностями измерения расхода компонента при его изменении и др.).

Связь между указанными факторами и измеренными рассогласованиями можно определить соотношением, образующим искомую функциональную зависимость между расходомерными и уровнемерными измерениями:

$$\Delta n_{измj} = \Delta n_0 + bT_j + \Delta n_{ошj}, \quad j = 1, 2, \dots, J, \quad (1)$$

где Δn_0 — выраженное числом импульсов бортового расходомера начальное отклонение объема компонента от расчетного значения, b — относительное отклонение от расчетного значения частоты следования импульсов бортового экземпляра датчика расхода при известном объемном расходе (ошибка настройки расходомера), $\Delta n_{ошj}$ — ошибка измерения израсходованного объема компонента, связанная с отклонением объема рабочего запаса компонента под j -й «измерительной точкой» в момент ее срабатывания (с ошибкой измерения уровня).

Ошибки измерения дискретных уровнемеров и ошибки расходомеров — главные факторы, определяющие точность работы СУРТ. Причем основная составляющая ошибки измерения на «измерительных точках», связанная с колебаниями уровня компонента в полете, *одинаково* (с точностью до



знака) действует на объем как *израсходованной*, так и *оставшейся* доли компонента.

Применяя к избыточной системе уравнений (1) (в предположении $\Delta n_{\text{ош } j} = 0$) метод наименьших квадратов (МНК), по совокупности измеренных рассогласований $\Delta n_{\text{изм } j}$ можно сформировать оценки возмущающих факторов $\Delta \hat{n}_{0j}$ и $\hat{b}_j$:

$$x_j = (A_j^T A_j)^{-1} A_j^T y_j,$$

где $x_j = [\Delta \hat{n}_{0j} \ \hat{b}_j]^T$ — вектор оценок возмущений размерности 2×1 , $y_j = [\Delta n_{\text{изм } 1} \ \Delta n_{\text{изм } 2} \ \dots \ \Delta n_{\text{изм } j}]^T$ —

вектор измерений размерности $j \times 1$, $A_j = \begin{bmatrix} 1 & T_1 \\ \dots & \dots \\ 1 & T_j \end{bmatrix}$ —

матрица размерности $j \times 2$.

Для уменьшения вычислительных затрат МНК-оценки возмущений $\Delta \hat{n}_{0j}$ и $\hat{b}_j$ удобно формировать рекуррентно [3] в форме фильтра Калмана [4], уточняя их по мере накопления измерительной информации о рассогласованиях $\Delta n_{\text{изм } j}$:

$$\begin{aligned} \Delta \hat{n}_{j/j-1} &= \Delta \hat{n}_{0j-1} + \hat{b}_{j-1} T_j, \\ \Delta \hat{n}_{0j} &= \Delta \hat{n}_{0j-1} + \alpha_j (\Delta n_{\text{изм } j} - \Delta \hat{n}_{j/j-1}), \end{aligned} \quad (2)$$

$$\hat{b}_j = \hat{b}_{j-1} + \beta_j (\Delta n_{\text{изм } j} - \Delta \hat{n}_{j/j-1}), \quad j = 1, 2, \dots, J,$$

где оптимальные по критерию наименьших квадратов значения коэффициентов фильтра Калмана α_j и β_j связаны с параметрами T_j соотношениями:

$$\begin{aligned} \beta_j &= W_j(2, j) = \\ &= \left(j T_j - \sum_{i=1}^j T_i \right) / \left(j \sum_{i=1}^j T_i^2 - \left(\sum_{i=1}^j T_i \right)^2 \right), \end{aligned} \quad (3)$$

$$\alpha_j = W_j(1, j) = 1/j - \beta_j \left(\sum_{i=1}^j T_i \right) / j, \quad (4)$$

а начальные условия — нулевые: $\Delta \hat{n}_{00} = 0$, $\hat{b}_0 = 0$.

Здесь $\Delta \hat{n}_{j/j-1}$ — прогноз j -го измерения по $j-1$ измерениям, $\Delta \hat{n}_{0j}$ и $\hat{b}_j$ — оценки возмущений Δn_0 и b по j измерениям, а $W_j = (A_j^T A_j)^{-1} A_j^T$.

Отметим, что МНК-оценки возмущений $\Delta \hat{n}_{0j}$, $\hat{b}_j$ сформированы без учета априорной информации о математических ожиданиях (как правило, равных нулю) и дисперсиях этих возмущений и могут быть уточнены с использованием этой информации: $\Delta \hat{n}_{0j} = C_j \Delta \hat{n}_{0j}$, $\hat{b}_j = B_j \hat{b}_j$.

Оптимальные значения коэффициентов C_j и B_j зависят от соотношения дисперсии ошибки априорной оценки D_0 и дисперсий D_j ошибок МНК-оценок возмущения и определяются отношением $D_0/(D_0 + D_j)$.

При условии некоррелированности ошибок измерения дискретных уровнемеров $\Delta n_{\text{ош } j}$ и постоянства их дисперсии $D_{\Delta n_{\text{ош } j}} = D_\varepsilon$ можно получить [4] явные выражения для ковариационной матрицы ошибок МНК-оценок вектора возмущений x : $K_j = (A_j^T A_j)^{-1} D_\varepsilon$ и, в частности,

$$\begin{aligned} D_{\hat{b}_j} &\rightarrow \infty \text{ для } j = 1; \\ D_{\hat{b}_j} &= j D_\varepsilon / \left(j \sum_{i=1}^j T_i^2 - \left(\sum_{i=1}^j T_i \right)^2 \right) \text{ для } j = 2, \dots, J. \end{aligned} \quad (5)$$

Как и МНК-оценки (2), оценки $\Delta \hat{n}_{0j}$ и $\hat{b}_j$, уточненные с учетом априорной информации, можно получать рекуррентно [4] по соотношениям, аналогичным соотношениям (2):

$$\begin{aligned} \Delta \hat{n}_{j/j-1} &= \Delta \hat{n}_{0j-1} + \hat{b}_{j-1} T_j, \\ \Delta \hat{n}_{0j} &= \Delta \hat{n}_{0j-1} + \tilde{\alpha}_j (\Delta n_{\text{изм } j} - \Delta \hat{n}_{j/j-1}), \end{aligned} \quad (6)$$

$$\hat{b}_j = \hat{b}_{j-1} + \tilde{\beta}_j (\Delta n_{\text{изм } j} - \Delta \hat{n}_{j/j-1}), \quad j = 1, 2, \dots, J,$$

где $\tilde{\alpha}_j = \tilde{W}_j(1, j)$, $\tilde{\beta}_j = \tilde{W}_j(2, j)$, $\tilde{W}_j = (P^{-1} D_\varepsilon + A_j^T A_j)^{-1} A_j^T$, а P — ковариационная матрица априорных оценок вектора $x_0 = [\Delta \hat{n}_{00} \ \hat{b}_0]^T$ возмущающих факторов: $P = M[x_0 x_0^T]$. Здесь и в дальнейшем $M[\cdot]$ — оператор математического ожидания.

С помощью улучшенных оценок возмущений $\Delta \hat{n}_{0j}$ и $\hat{b}_j$ и с учетом зависимости (1) можно сформировать оценки $\Delta \hat{n}_{\text{ош } j}$ фильтруемых (некоррелированных) составляющих ошибок измерения на «измерительных точках»:

$$\begin{aligned} \Delta \hat{n}_{\text{ош } j} &= \Delta n_{\text{изм } j} - \Delta \hat{n}_{0j} - \hat{b}_j T_j = \\ &= n_j - \Delta \hat{n}_{0j} - T_j(1 + \hat{b}_j). \end{aligned}$$

В итоге улучшенные оценки измеряемых координат объекта управления могут быть представлены соотношениями:

$$\begin{aligned} \hat{N}_j &= N_j(1 + \hat{b}_j) - \Delta \hat{n}_{\text{ош } j} = N_0(1 + \hat{b}_j) + \Delta \hat{n}_{0j} - n_j, \\ \hat{V}_j &= \dot{V}_{\text{изм } j} / (1 + \hat{b}_j). \end{aligned} \quad (7)$$

Здесь $\hat{N}_j$ — выраженная числом импульсов бортовой «вертушки» улучшенная оценка объема компонента под j -й «измерительной точкой» в момент ее срабатывания, $\dot{V}_{измj}$ — измеренное значение объемного расхода компонента топлива в интервале между $(j-1)$ -й и j -й измерительными точками, $\hat{V}_j$ — улучшенная оценка измерительного сигнала расходомера $\dot{V}_{измj}$.

В соотношении (7) учтено, что оценка ошибки уровнемера по расходомерной информации используется для оценки ошибки оставшегося запаса с противоположным знаком.

2. ВАРИАНТЫ АЛГОРИТМОВ УПРАВЛЕНИЯ

Сформированная по комбинированной информации улучшенная оценка (7) объема компонента под j -й «измерительной точкой» $\hat{N}_j$ может использоваться как для повышения точности определения момента окончания компонента, так и для улучшения синхронизации моментов опорожнения баков.

Действительно, признак $\Pi = 1$ окончания компонента естественно формировать в момент t выполнения условия: $\Pi = 1$, если $n(t) - n_j \geq \hat{N}_j$, где $n(t)$ — текущее значение суммы сформированных «вертушкой» импульсов, j — номер последней сработавшей «измерительной точки».

Аналогично, требуемое значение коэффициента $K_{РСКj}$ соотношения объемных расходов для системы РСК естественно формировать из условия равенства требуемого соотношения *объемных расходов* соотношению *объемных запасов* компонентов в момент их измерения, поскольку при выполнении такого условия обеспечивается одновременность окончания компонентов топлива

$$K_{РСКj} = (\hat{N}_j^0 - \Delta n_j^0) / (\hat{N}_j^r - \Delta n_j^r). \quad (8)$$

Здесь $\hat{N}_j^0$ и $\hat{N}_j^r$ — улучшенные с учетом априорной информации оценки (7) объемов окислителя и горючего под j -ми «измерительными точками» в момент их срабатывания; Δn_j^0 и Δn_j^r — поправки, учитывающие несинхронность срабатывания «измерительных точек» и равные числу импульсов, зафиксированных «вертушкой» в опережающем компоненте между моментом срабатывания j -й точки в этом компоненте и j -й точки в отстающем компоненте (для которого следует положить $\Delta n_j^{0(r)} = 0$).

Как следует из соотношений (7), в интервале времени между моментами срабатывания j -й и $(j+1)$ -й «измерительными точками», а, следовательно, и на всем участке работы СУРТ переменная величина

$$\begin{aligned} \hat{N}(t) &= \hat{N}_j - (n(t) - n_j) = \\ &= N_0(1 + \hat{b}_j) + \Delta \hat{n}_{0j} - n(t) \end{aligned} \quad (9)$$

является оценкой текущего объемного запаса компонента, выраженного числом импульсов бортовой «вертушки». В выражении (9) следует положить $\hat{b}_0 = \Delta \hat{n}_{00} = 0$. Соотношение (9) представляет собой *самонастраивающуюся* модель текущего объемного запаса компонента топлива с двумя переменными параметрами $\hat{b}_j$ и $\Delta \hat{n}_{0j}$, настраиваемыми по комбинированной информации (2), (6) в моменты срабатывания дискретных уровнемеров.

С учетом соотношения (9) алгоритм управления (8) может быть представлен в более общем виде:

$$K_{РСК(t)} = \hat{N}^0(t) / \hat{N}^r(t). \quad (10)$$

В отличие от алгоритма (2), (7) с явным оцениванием ошибки заправки $\Delta \hat{n}_{0j}$, рекуррентный алгоритм оценивания запаса (9) компонента может быть представлен в иной, эквивалентной форме с непосредственным формированием оценки $\hat{N}(t)$ текущего объемного запаса компонента топлива, выраженной числом импульсов расчетно-номинальной «вертушки» и настраиваемой по невязкам между фактическим и номинальным значениями этой оценки в моменты t_j срабатывания дискретных уровнемеров:

$$\begin{aligned} \hat{N}(t) &= \hat{N}_{j-1} - (n(t) - n_{j-1})\hat{c}_{j-1}, \\ \hat{N}_j &= \hat{N}(t_j) + \alpha_{mj}(\hat{N}(t_j) - N_j), \\ \hat{c}_j &= \hat{c}_{j-1} + \beta_{mj}(\hat{N}(t_j) - N_j), \end{aligned} \quad (11)$$

с начальными условиями $\hat{N}_0 = N_0$, $n_0 = 0$, $\hat{c}_0 = 1$. Здесь $c = (1 + b)^{-1} = f_{расч}/f$ — величина, обратная относительной частоте выходного сигнала бортовой «вертушки» и не равная единице из-за действия ошибки настройки b расходомера.

Можно показать, что оптимальные по модифицированному МНК-критерию (сформированному с учетом дополнительной информации о ковариационной матрице P априорных оценок возмущений) значения весовых коэффициентов α_{mj} и β_{mj} фильтра (11) могут быть вычислены с помощью



соотношений, аналогичных поясняющим соотношениям к формуле (6):

$$\alpha_{mj} = W_{mj}(1, j), \quad \beta_{mj} = W_{mj}(2, j),$$

$$W_{mj} = (P_j^{-1} D_\varepsilon + A_{mj}^T A_{mj})^{-1} A_{mj}^T,$$

$$A_{mj} = \begin{bmatrix} 1 & N_j - N_1 \\ & \dots \\ 1 & N_j - N_{j-1} \\ 1 & 0 \end{bmatrix}.$$

Здесь P_j — пересчитанная на момент j ковариационная матрица P вектора априорных начальных значений оцениваемых координат: $P_j = G_j P G_j^T$,

$$G_j = \begin{bmatrix} 1 & N_0 - N_j \\ 0 & 1 \end{bmatrix}.$$

При наличии более детальной достоверной статистической информации об ошибках измерения дискретных уровнемеров значения коэффициентов α_{mj} и β_{mj} фильтра (11) могут быть вычислены по приведенным выше соотношениям, но с использованием вместо W_{mj} более сложной матрицы [4], дополнительно включающей ковариационную матрицу R_j ошибок измерения: $R_j = M[(\Delta n_{\text{ош}1}, \Delta n_{\text{ош}2}, \dots, \Delta n_{\text{ош}j})^T (\Delta n_{\text{ош}1}, \Delta n_{\text{ош}2}, \dots, \Delta n_{\text{ош}j})]$.

Можно показать, что при условии диагональности матриц R_j рекуррентные оценки (11) оптимальны по критерию МНК с весами [4] и достаточно близки к оптимальным при наличии корреляции между ошибками измерения.

Как нетрудно заметить, матрицы $\tilde{W}_{mj}$ упрощаются до вида матриц W_{mj} в предположении некоррелированности и постоянства ошибок измерения дискретных уровнемеров

$$R_j = E_j D_\varepsilon, \quad (12)$$

где E_j — единичная матрица размера $j \times j$.

Отметим, что во многих практических случаях *целенаправленный* учет загроуляющего условия (12), соответствующий переходу от критерия МНК с весами [4] к упрощенному МНК-критерию, оказывается целесообразным, так как позволяет обеспечить *инвариантность* оптимальных оценок к, как правило, достоверно не известным корреляционным свойствам ошибок измерения ценой незначительного ухудшения точности оценок.

3. ОСОБЕННОСТИ ПРАКТИЧЕСКОЙ РЕАЛИЗАЦИИ

Из соображений уменьшения вычислительных погрешностей при бортовой программной реализации алгоритмов целесообразно вместо текущей

суммы импульсов $n(t)$, достигающей чрезмерно больших значений, воспользоваться ее относительной формой $n(t)/N_{\text{норм}}$, где $N_{\text{норм}} \geq N_0$ — нормирующее число импульсов. При этом формула (1) после почленного деления на $N_{\text{норм}}$ представляется в безразмерной относительной форме и, соответственно, изменятся значения параметров алгоритма T_j , N_j и β_j , а константы T_j и N_j и переменные $n(t)$ и $\hat{N}(t)$ могут интерпретироваться как относительные объемы компонента.

Предшествующие рассуждения для удобства проводились в неявном предположении о наличии в аппаратуре системы РСК счетчика импульсов $n(t)$, который в настоящее время отсутствует. В существующей системе [5] на вход БЦВМ передается сигнал f , пропорциональный текущему расходу. Поскольку число импульсов в момент срабатывания дискретного уровнемера потенциально может фиксироваться не с точностью до такта $\tau_{\text{ц}}$ работы БЦВМ ($\sim 0,03$ с), а существенно более высокой точностью, определяемой частотой сигнала на выходе расходомера (~ 800 Гц и более), то практическая реализация нового алгоритма могла бы сопровождаться доработкой аппаратной части системы РСК. Такая доработка потребовала бы применения специализированного высокочастотного процессора (в пределе — счетчика импульсов), осуществляющего интегрирование выходного сигнала расходомера (текущей частоты f) и фиксацию текущих значений интеграла в моменты срабатывания уровнемеров. Однако из соображений надежности на данном этапе можно ограничиться приближенным интегрированием с фиксацией интегрального сигнала расходомера в БЦВМ

$$n(t) \cong \sum_k f_k \tau_{\text{ц}} / N_{\text{норм}} = \sum_k R_v dV_{\text{норм}},$$

где f_k — выходной сигнал расходомера в k -м такте БЦВМ, $R_v = f_k / f_{\text{норм}}$ — текущий относительный объемный режим выработки компонента; $dV_{\text{норм}} = f_{\text{норм}} \tau_{\text{ц}} / N_{\text{норм}}$ — номинальный относительный объем компонента, вырабатываемый в элементарном цикле работы БЦВМ в номинальных условиях при $R_v = 1$. Откуда, с учетом дискретности работы БЦВМ, соотношение (9) представляется в виде

$$\hat{N}_k = \hat{N}_{k-1} - R_v dV_{\text{норм}},$$

а первое соотношение в (11) — в виде:

$$\hat{N}_k = \hat{N}_{k-1} - R_v dV_{\text{норм}} \hat{c}_{j-1}. \quad (13)$$

Предполагается, что указанные вычислительные операции реализуются в каждом такте, а в так-

тах срабатывания уровнемеров оценки $\hat{N}_k$ дополнительно корректируются в соответствии с алгоритмами (6) или (11).

Примечательная потенциальная особенность приведенных алгоритмов работы системы заключается в ослабленной, благодаря допущению (12), зависимости оптимальных значений параметров алгоритмов от априорной информации об ошибках дискретных уровнемеров, являющейся весьма неопределенной, особенно на начальных этапах создания системы (до проведения натурных испытаний). Действительно, значения коэффициентов α_j и β_j полностью определяются информацией о расстановке «измерительных точек» по объему бака (3), (4). Как показало моделирование, в связи с малостью дисперсии D_ϵ ошибок уровнемеров коэффициенты C_j , используемые для учета априорной информации по ошибке заправки, могут быть приняты равными единице без заметного ущерба для характеристик качества работы СУРТ. Это же относится к последним значениям (начиная с 5—7-го номера измерения) коэффициентов B_j . Начальные же значения этих коэффициентов, влияющие лишь на уровень первых управляющих команд для системы РСК, могут быть вычислены по формуле (5) или (6), исходя из соотношения между дисперсиями ошибок оценивания погрешности b настройки расходомера по уровнемерной информации и априорным значением дисперсии этого параметра.

По результатам статистического моделирования рациональное использование расходомерной информации во внешнем контуре системы позволяет заметно (на 1—2 % $K_{т\text{ ном}}$) уменьшить диапазон управления по коэффициенту соотношения расходов, потребный для решения задачи синхронизации моментов выработки компонентов топлива, и на 10—20 % уменьшить погрешность решения этой задачи и задачи определения момента окончания запасов топлива.

Однако преимущества нового алгоритма особенно заметно должны проявиться в нештатных ситуациях, связанных с отказами элементов в исполнительном и измерительном трактах системы и, как правило, сопровождающихся заметным изменением статистических характеристик возмущений и помех. Действительно, использование расходомерной информации во внешнем контуре позволяет решить задачу высокоточного определения момента окончания опережающего компонента и формирования команды на своевременное выключение двигателя в нештатных ситуациях, связанных с отказом исполнительного органа СУРТ (привода дросселя, изменяющего соотношение расходов компонентов), а также выходом

этого дросселя на упор вследствие утечек компонентов из-за повреждений в топливной системе и др.

Еще одно важное преимущество представленного алгоритма состоит в возможности «пересчета» измерительной информации о запасе компонента топлива в момент срабатывания в нем «измерительной точки» на произвольный момент времени t , соответствующий срабатыванию уровнемера в другом компоненте благодаря измерению и фиксации суммарного числа импульсов $n(t)$. Это позволяет формировать управляющие воздействия (10) и признаки окончания компонентов в случаях разнообразных сочетаний отказов дискретных уровнемеров и сохранить работоспособность системы в таких нештатных ситуациях с минимальными потерями точности. В частности, управляющие воздействия (10) могут формироваться при каждом срабатывании «измерительной точки» в любом компоненте, а не пары синхронно установленных точек (срабатывающих одновременно в номинальных условиях), как это делается обычно. Это же свойство позволяет снизить требования к точности геометрической расстановки «измерительных точек» по высоте баков и, в частности, делает возможным применение неодинакового числа дискретных уровнемеров в баках компонентов топлива.

Отказ одного из двух расходомеров также может быть скомпенсирован с минимальными потерями качества управления, благодаря реконфигурации алгоритма, построенной на основе учета поддержания постоянства (заданного циклограммного значения) тяги, и, следовательно, суммарного массового расхода топлива на ступени.

В заключение важно отметить что, как было указано выше, в алгоритмах СУРТ, реализованных на РН «Протон», «Зенит», «Ангара» и других, в моменты срабатывания j -х «измерительных точек» в баках окислителя и горючего фиксируются значения $G_j^{0(r)}$ формируемой в каждом элементарном цикле оценки текущего относительного запаса (выраженного в долях нормирующего запаса $G_{\text{норм}}$) суммарного топлива (окислитель плюс горючее) [2]:

$$G_k = G_{k-1} - R dG_{\text{норм}}(1 + \lambda_{j-1}), \quad (14)$$

где $R \cong \dot{m} / \dot{m}_{\text{ном}}$ — бортовая оценка текущего относительного режима расходования двигателем суммарного топлива, $dG_{\text{норм}}$ — номинальный относительный запас топлива, вырабатываемый в элементарном цикле работы БЦВМ, λ_j — переменный параметр модели (14) текущего запаса топлива, влияющий на темп списывания модельного запаса и настраиваемый по рассогласовани-



ям $\Delta G_j^{0(r)} = G_j^{0(r)} - G_{\text{ном}j}^{0(r)}$ между модельными $G_j^{0(r)}$ и номинальными $G_{\text{ном}j}^{0(r)}$ значениями относительного запаса в моменты срабатывания уровнемеров в соответствии с итеративным алгоритмом [1, 2]:

$$\begin{aligned}\lambda_j &= \lambda_{j-1} + \Delta\lambda_j, \\ \Delta\lambda_j &= K^0 \Delta\lambda_j^0 + K^r \Delta\lambda_j^r, \\ \Delta\lambda_j^{0(r)} &= A_j(B_j \Delta G_j^{0(r)} - \Delta G_{j-1}^{0(r)}).\end{aligned}\quad (15)$$

Поправки $\Delta\lambda_j^{0(r)}$ к оценкам относительных отклонений расходов компонентов дополнительно используются для вычисления управляющих воздействий синхронизации опорожнения баков на изменение относительного отклонения от номинала $\Delta\delta k_j$ коэффициента соотношения расходов компонентов топлива: $\Delta\delta k_j = K(\Delta\lambda_j^r - \Delta\lambda_j^0)$, где $K^{0(r)}$, K , A_j и B_j — константы алгоритма.

В отличие от соотношения (13) модель запаса (14) настраивается с помощью одной координаты λ_{j-1} из условия совпадения модельного и фактического запасов компонента в терминальный момент окончания процесса управления.

Из соображений сохранения преемственности в использовании отработанных и хорошо зарекомендовавших себя алгоритмических решений [1, 2] и с учетом очевидной связи между настраиваемыми моделями (13) и (14) можно предложить еще один перспективный вариант построения алгоритма работы СУРТ, где вместо модели (14) — две настраиваемых по расходомерной информации модели для каждого из компонентов топлива:

$$G_k^{0(r)} = G_{k-1}^{0(r)} - R_v^{0(r)} dV_{\text{ном}}(1 + \lambda_{j-1}^{0(r)}). \quad (16)$$

Для настройки этих моделей можно воспользоваться аналогичными алгоритму (15) итеративными соотношениями:

$$\begin{aligned}\lambda_j^{0(r)} &= \lambda_{j-1}^{0(r)} + \Delta\lambda_j^{0(r)}, \\ \Delta\lambda_j^{0(r)} &= A_j(B_j \Delta G_j^{0(r)} - \Delta G_{j-1}^{0(r)}), \\ \Delta\delta k_j &= K(\Delta\lambda_j^r - \Delta\lambda_j^0).\end{aligned}$$

Таким образом, в данном случае учет расходомерной информации в контуре синхронизации опорожнения баков свелся к формированию режима R_v расходования модельного запаса (16) каждого из компонентов по этой информации. Примечательно, что это полностью совпадает с известным вариантом [2] построения алгоритма работы СУРТ с дополнительным использованием инер-

циальной информации, в котором режим R в формуле (14) также формируется по инерциальной информации о наборе кажущейся скорости РН или РБ.

ЗАКЛЮЧЕНИЕ

Для систем управления расходом топлива с расходомерами предложены новый способ комплексования уровнемерной и расходомерной информации и новые алгоритмы управления, позволяющие заметно улучшить точностные и надежные показатели качества управления бортовых систем ракет-носителей и разгонных блоков.

Предложенные алгоритмы наиболее удобны для применения на разгонных блоках, для которых характерны применение расходомеров и наличие существенных ограничений на допустимый диапазон управления по коэффициенту соотношения расходов компонентов топлива.

ЛИТЕРАТУРА

1. Вакушин В.А., Жуков В.А., Завадский В.К. и др. Системы управления расходом топлива перспективных ракет-носителей (принципы построения, отказоустойчивые алгоритмы, технологии отработки бортового обеспечения) // Научно-технические разработки ОКБ-23 — КБ «Салют». — М.: Воздушный транспорт, 2006. — С. 579—592.
2. Завадский В.К., Иванов В.П., Каблова Е.Б., Кленовая Л.Г. Системы управления расходом топлива (назначение, принципы построения, алгоритмы управления) // Там же. — С. 243—253.
3. Жуков В.А., Завадский В.К., Иванов В.П. и др. Повышение точности и надежности работы систем управления расходом топлива на основе комплексования расходомерной и уровнемерной измерительной информации // Материалы 5-й всерос. науч.-техн. конф. с междунар. участием «Актуальные проблемы ракетно-космической техники» (V Козловские чтения, Самара 11—15 сент. 2017 г.). — Т. 1. — С. 262—269.
4. Брайсон А., Ю-Ши Хо. Прикладная теория оптимального управления. — М.: Мир, 1972.
5. Чадаев А.И., Тропова Е.И. Корректировка алгоритмов действия системы опорожнения баков боковых блоков ракеты-носителя «Союз-2» // Датчики и системы. — 2017. — № 2. — С. 25—28.

Статья представлена к публикации членом редколлегии Б.В. Павловым.

Завадский Владимир Константинович — канд. техн. наук, ст. науч. сотрудник, ✉ vladizav@yandex.ru,

Иванов Владимир Петрович — д-р техн. наук, зав. лабораторией, ✉ vladguc@ipu.ru,

Каблова Елена Борисовна — науч. сотрудник, ✉ vladguc@ipu.ru,

Кленовая Людмила Григорьевна — науч. сотрудник, ✉ vladguc@ipu.ru,

Институт проблем управления им. В.А. Трапезникова РАН, г. Москва.



Институт проблем управления им. В.А. Трапезникова РАН осуществляет подготовку (по очной и заочной форме) научных кадров в аспирантуре

Направление подготовки: 01.06.01 — «Математика и механика»

Специальность:

- 01.01.02 — Дифференциальные уравнения, динамические системы и оптимальное управление.

Направление подготовки: 09.06.01 — «Информатика и вычислительная техника»

Специальности:

- 05.13.01 — Системный анализ, управление и обработка информации (по отраслям) — по техническим наукам;
- 05.13.01 — Системный анализ, управление и обработка информации (в отраслях информатики, вычислительной техники и автоматизации) — по физико-математическим наукам;
- 05.13.05 — Элементы и устройства вычислительной техники и систем управления — по техническим наукам;
- 05.13.06 — Автоматизация и управление технологическими процессами и производствами (по отраслям) — по техническим наукам;
- 05.13.10 — Управление в социальных и экономических системах — по техническим наукам;
- 05.13.11 — Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей — по техническим наукам;
- 05.13.12 — Системы автоматизации проектирования (по отраслям) — по техническим наукам;
- 05.13.15 — Вычислительные машины, комплексы и компьютерные сети;
- 05.13.18 — Математическое моделирование, численные методы и комплексы программ — по техническим наукам.

Направление подготовки: 38.06.01 — «Экономика»

Специальности:

- 08.00.05 — Экономика и управление народным хозяйством (по отраслям и сферам деятельности, в том числе управление инновациями);
- 08.00.13 — Математические и инструментальные методы экономики.

Сроки обучения. В очной форме аспирантуры — 4 года (по направлению подготовки 38.06.01 — 3 года);
В заочной форме аспирантуры — 5 лет (по направлению подготовки 38.06.01 — 4 года).

Прием в аспирантуру. Все поступающие в аспирантуру проходят предварительное собеседование с предполагаемым научным руководителем. Результаты собеседования учитываются при решении вопроса о допуске поступающего к вступительным экзаменам.

Прием вступительных испытаний в аспирантуру проводится в конце августа — сентябре.

Прием заявлений и документов в аспирантуру — с 25 июня по 1 августа. Зачисление в аспирантуру до 1 октября. Документы и заявления подаются в Отдел докторантуры и аспирантуры ИПУ РАН. Вступительные экзамены по специальной дисциплине и иностранному языку проводятся до середины сентября. Вступительные экзамены могут быть проведены на русском и иностранных языках.

В Институте читаются лекции для аспирантов по следующим программам:

- «Эконометрический анализ экономического развития» — д-р экон. наук, профессор Нижегородцев Р.М.;
- «Теория автоматического управления» — д-р техн. наук, профессор С.А. Краснова;
- «Теория робастного управления» — д-р техн. наук, профессор А.П. Курдюков;
- «Теория многоагентных систем» — д-р техн. наук, профессор Р.П. Агаев;
- «Идентификация систем управления» — канд. техн. наук Макаров В.В.;
- «Методы дискретной оптимизации и управление проектами» — д-р физ.-мат. наук, профессор РАН М.В. Губко;
- «Управление и искусственный интеллект» — канд. техн. наук А.В. Макаренко;
- «Технологические основы управления производством продукции на промышленных предприятиях» — д-р техн. наук, профессор А.Л. Генкин;
- «Вычислительные машины, комплексы и компьютерные сети» — д-р техн. наук, профессор В.М. Вишневский;
- «Системы автоматизации проектирования» — д-р техн. наук, профессор А.В. Толоч.

МЕРОПРИЯТИЯ IFAC — Международной федерации по автоматическому управлению

Мероприятие	Дата	Место проведения
IFAC Conference on Technology, Culture and International Stability — 18th TECIS 2018	13—15 Sep 2018	Baku, AZERBAIJAN
INSTICC, IFAC et al. International Joint Conference on Computational Intelligence — 10th IJCCI 2018	18—20 Sep 2018	Seville, SPAIN
IFAC Conference on Engine and Powertrain Control, Simulation and Modeling — 5th E-COSM 2018	20—22 Sep 2018	Changchun, CHINA
IFAC Workshop on Control Applications of Optimization — 17th CAO 2018	15—19 Oct 2018	Yekaterinburg, RUSSIA
Conference on Latin American Conference on Automatic Control (in cooperation with IFAC) — XVIII CLCA 2018	24—26 Oct 2018	Quito, ECUADOR
OTM/IFAC/IFIP et al. International Workshop on Enterprise Integration, Interoperability and Networking — 13th EI2N 2018	24—25 Oct 2018	Valletta, MALTA
IFAC Conference on Analysis and Control of Chaotic Systems — 5th CHAOS 2018	30 Oct — 1 Nov 2018	Eindhoven, NETHERLANDS
IFAC Conference on Cyber-Physical and Human Systems — 2nd CPHS 2018	13—15 Dec 2018	Miami (FL), USA
IFAC Symposium on Dynamics and Control of Process Systems, including Biosystems — 12th DYCOPS 2019	23—26 Apr 2019	Florianopolis, BRAZIL
IFAC Workshop on Control of Systems Governed by Partial Differential Equations — 3rd CPDE 2019	22—24 May 2019	Oaxaca, MEXICO
SICE, ISCIE, JSME, IEEE, IFAC Conference on Asian Control Conference (in cooperation with IFAC) — 12th ASCC 2019	9—12 Jun 2019	Kitakyushu, Fukuoka, JAPAN
IFAC Workshop on Control of Smart Grid and Renewable Energy Systems — CSGRES 2019	10—12 Jun 2019	Jeju, REPUBLIC OF KOREA
IFAC Symposium on Advances in Automotive Control — 9th AAC 2019	25—28 Jun 2019	Orlans, FRANCE
EUCA, IFAC, IEEE CSS Conference on European Control Conference (in cooperation with IFAC) — ECC 2019	9—12 Jun 2019	Naples, ITALY
IFAC Symposium on Intelligent Autonomous Vehicles — 10th IAV 2019	3—5 Jul 2019	Gdansk, POLAND
IFAC Workshop on Thermodynamic Foundations for a Mathematical Systems Theory — 3rd TFMST 2019	8—10 Jul 2019	Louvain-la-Neuve, BELGIUM
Conference on American Control Conference (in cooperation with IFAC) — ACC 2019	10—12 Jul 2019	Philadelphia, Pennsylvania, USA
IFAC, IEEE, IFIP, IFORS et al. Conference on Manufacturing Modelling, Management and Control — 9th MIM 2019	28—13 Aug 2019	Berlin, GERMANY
IFAC Symposium on Advances in Automotive Control — 9th AAC 2019	25—28 Jun 2019	Orlans, FRANCE

Более полная информация доступна на сайте www.ifac-control.org

CONTENTS & ABSTRACTS

METHODS FOR SOLVING OF THE AIRCRAFT LANDING PROBLEM.

PART 1. EXACT SOLUTION METHODS. 2

Veresnikov G.S., Egorov N.A., Kulida E.L., Lebedev V.G.

It is noted that the aircraft landing problem is one of the main tasks in the planning and regulation of air traffic flows, which requires the use of a sufficiently serious mathematical apparatus. A brief overview of the main approaches to the formulation and solution of this problem is presented.

Keywords: air traffic, objective functions, the aircraft landing problem, linear programming, branch and bound method.

MINIMUM-TIME CONTROL PROBLEM FOR ELASTIC AND VISCO-ELASTIC INTERACTION BETWEEN BODY AND SURFACE 14

Galyaev A.A., Lysenko P.V.

Two minimum-time control problems are considered for the mechanical system consisting of material point and obstacle surface which interact with each other by the means of the elastic or visco-elastic spring. Interaction time interval is determined by physical conditions of contact. The dependences are studied of full interaction time and restitution coefficient on visco-elastic spring properties. The Python program is developed and the «material point — obstacle surface» system behavior modelling is conducted.

Keywords: mechanical system with impacts, Kelvin-Voigt media, minimum time control.

ROBUST STABILITY AND SELECTION OF REGULATORS SETTINGS FOR TECHNOLOGICAL OBJECTS WITH DELAY 21

Tsirlin A.M., Hovsepyan V.S.

The problem is considered of robust stability and choice of typical regulators parameters of single-loop linear systems with feedback for technological objects with delay. The set of possible parameters values of object transfer function is assumed to be closed and bounded, while the magnitude and the phase of amplitude-phase characteristic of an open system decrease monotonically with a frequency.

Keywords: systems with delay, robust stability, technological objects, robust tuning.

APPROXIMATIONS OF THE RENEWAL FUNCTION AND OPERATING COSTS MANAGEMENT STRATEGY . . 28

Rusev V.N., Skorikov A.V.

The approximation of the renewal function is considered for recurrent flows. The failure-free operation time is assumed to be described by the two parametrical Weibull-Gnedenko distribution. Analytical and numerical approaches (based on discretization methods) are proposed for the solution of Volterra integral equation. The relationships and algorithms obtained were verified by control cases using the Wolfram Mathematica mathematical software package. The results presented are applied to the «block replacement policy» to provide the reliable and effective functioning of technological objects.

Keywords: renewal function, the Weibull-Gnedenko distribution, the Volterra integral equation, moments generating function, discretization techniques of integral equations, reliability monitoring, block replacement policy, computer modelling in Wolfram Mathematica.

FOUNDATIONS OF THE COMPLEX ACTIVITY THEORY. PART 1. THE STRUCTURES OF COMPLEX ACTIVITY. UNCERTAINTY AND GENERATION OF COMPLEX ACTIVITY. 36

Belov M.V., Novikov D.A.

The foundations of the complex activity theory are described, extending the methodology to the case of any complex (having a non-trivial multi-level internal structure) human activity. The structural elements of complex activity are considered, its logical, causal and process structures are constructively described. It is shown that the system of co-

ordinated formal models ensures the practical application of the results both for practitioners and for scientists engaged in research of general principles of organization of activity (practical, scientific, etc.) and of management of organizational and technical systems. It is for the first time that the description formalism of the complex activity is proposed together with the subject — the organizational and technical system.

Keywords: complex activity, uncertainty, organization, control, management.

THE PROBLEM OF THE OPTIMAL PLACING OF THE INFORMATION-TECHNOLOGICAL BACKUP IN DISTRIBUTED DATA PROCESSING SYSTEMS 46

Somov S.K.

It is noted, that information-technological backup is a new type of information redundancy, the use of which in distributed automated informational control systems allows to increase the efficiency of their performance while processing the typical user requests. The task of optimal allocation of identical copies of information-technological backup throughout the nodes of the distributed system is formulated in the form of a minimum problem of finding the p -median of a graph. The algorithm is proposed for solving this problem and the example of the solution is given. The brief analysis of the solution algorithm results is performed.

Keywords: distributed data processing system, computer network, information-technological backup.

CONFLICT-FREE SELF-ROUTING FOR A THREE-DIMENSIONAL COMPLETE MULTIRING 54

Podlazov V.S.

The system-area network is proposed in the form of a non-blocking three-dimensional complete multiring. The structure of a multiring is developed and the algorithm is suggested of a conflict-free laying out direct paths between its nodes by means of dynamic local self-routing in them.

Keywords: system-area network, generalized hypercube, complete multiring, switching properties, non-blocking network, direct channels, conflict-free self-routing, local dynamic self-routing.

RESULTS OF THE ANTITUMOR VIRAL VACCINE INTRODUCTION REGIMENS STUDY BASED ON MATHEMATICAL MODELING 61

Babushkina N.A., Kuzina E.A., Loos A.A., Belyaeva E.V.

The mathematical description is presented of the two stages of tumor cells' death as a result of the body's immune response after antitumor viral vaccine introduction. The mathematical description in the form of a system of nonlinear differential equations is realized as a software set by means of the MatLab-Simulink system. As a result of the computing experiment, the two strategies of the effective application of the antitumor viral vaccine were determined. The first strategy leads to complete elimination of the tumor cells after a single-shot administration of the vaccine. The second strategy allows stabilizing tumor size through the recurrent introduction of the vaccine. The approach proposed to explore the effectiveness of vaccine therapy can be applied to different types of experimental tumors and antitumor vaccines.

Keywords: mathematical model, tumor cells, antibodies, moment of vaccine administration, vaccine effectiveness, immune response, virus, vaccine therapy.

INTEGRATION OF THE FLOW AND LEVEL MEASUREMENT INFORMATION FOR PROPELLANT CONSUMPTION CONTROL SYSTEMS 71

Zavatsky V.K., Ivanov V.P., Kablova E.B., Klenovaya L.G.

The method of integration the flow and level measurement information and new control algorithms allowing to increase the propellant consumption control system work accuracy in normal operating mode and in case of onboard equipment failure.

Keywords: the measuring information integration, filtration, control accuracy, the emergency situation.