

**Институт проблем управления
им. В.А. Трапезникова РАН**

УПРАВЛЕНИЕ БОЛЬШИМИ СИСТЕМАМИ

***Выпуск 58
Ноябрь 2015***

**СБОРНИК
ТРУДОВ**

ISSN 1819-2467

Регистрационный номер Эл. №ФС77-44158 от 09 марта 2011 г.

Москва – 2015

РОССИЙСКАЯ АКАДЕМИЯ НАУК
Институт проблем управления
им. В.А. Трапезникова

**УПРАВЛЕНИЕ
БОЛЬШИМИ
СИСТЕМАМИ**

СБОРНИК ТРУДОВ

Выпуск 58

Москва – 2015

КООРДИНАЦИОННЫЙ СОВЕТ

Академики РАН: Васильев С.Н., Емельянов С.В., Куржанский А.Б., Федосов Е.А., Черноусько Ф.Л.; члены-корреспонденты РАН: Желтов С.Ю., Каляев И.А., Пархоменко П.П., Попков Ю.С.; д-ра техн. наук: Дорофеев А.А., Кузнецов О.П., Кульба В.В., Лотоцкий В.А., Павлов Б.В., Поляк Б.Т., Рутковский В.Ю.

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Главный редактор: член-корр. РАН Новиков Д.А. **Зам. главного редактора:** д-р физ.-мат. наук Губко М.В.; **Отв. секретарь:** канд. техн. наук Базенков Н.И.; **Редактор:** канд. техн. наук Квинто Я.И. **Техн. редактор:** канд. техн. наук Куливец С.Г.

Д-ра техн. наук: проф. Алескеров Ф.Т. (ГУ ВШЭ), проф. Алчинов А.И. (ИПУ РАН), проф. Андриевский Б.Р. (ИПМ РАН), проф. Афанасьев В.Н. (МИЭМ), проф. Бахтадзе Н.Н. (ИПУ РАН), проф. Бурков В.Н. (ИПУ РАН), проф. Вишневский В.М. (ИПУ РАН), Галаев А.А. (ИПУ РАН), д-р физ.-мат. наук проф. Ерешко Ф.И. (ВЦ РАН), д-ра техн. наук Зоркальцев В.И. (ИСЭМ СО РАН), проф. Калашников А.О. (ИПУ РАН), проф. Калянов Г.Н. (ГУ ВШЭ), проф. Каравай М.Ф. (ИПУ РАН), д-р экон. наук, проф. Ключков В.В. (ИПУ РАН), д-ра техн. наук, Коргин Н.А. (ИПУ РАН), проф. Курдюков А.П. (ИПУ РАН), д-ра физ.-мат. наук, проф. Кушнер А.Г., проф. Лазарев А.А. (МФТИ), д-ра техн. наук: проф. Лебедев В.Г. (ИПУ РАН), проф. Мандель А.С. (ИПУ РАН), д-р биол. наук проф. Михальский А.И., д-р физ.-мат. наук, проф. Непейвода Н.Н. (ИПС РАН), д-р экон. наук, проф. Нижегородцев Р.М. (ИПУ РАН), д-ра техн. наук: проф. Орлов А.И. (МГТУ), д-ра физ.-мат. наук: проф. Рапопорт Л.Б. (ИПУ РАН), проф. Райгородский А.М. (МГУ), проф. Савватеев А.В. (РЭШ), д-ра техн. наук: проф. Самуйлов К.Е. (РУДН), проф. Сидельников Ю.В. (МАИ), Совлуков А.С. (ИПУ РАН) д-ра физ.-мат. наук: проф. Соловьев С.Ю. (МГУ), проф. Угольницкий Г.А. (ЮФУ), проф. Уткин В.А. (ИПУ РАН), проф. Хоботов Е.Н. (МГТУ), д-ра физ.-мат. наук: доцент Чеботарев П.Ю. (ИПУ РАН), проф. Чхартишвили А.Г. (ИПУ РАН), проф. Щербаков П.С. (ИПУ РАН).

РЕГИОНАЛЬНЫЕ РЕДАКЦИОННЫЕ СОВЕТЫ

Арзамас – д-р физ.-мат. наук проф. Пакшин П.В. **Волгоград** – д-ра физ.-мат. наук: проф. Воронин А.А., проф. Лосев А.Г. (ВолГУ); **Воронеж** – д-р техн. наук, проф. Баркалов С.А., д-р физ.-мат. наук, проф. Головинский П.А. (ВГАСУ), д-р техн. наук, проф. Подвальный С.Л. (ВГТУ); **Иркутск** – академик РАН Бычков И.В., д-р физ.-мат. наук, проф. Лакеев А.В. (ИДСТУ СО РАН); **Казань** – д-р физ.-мат. наук, проф. Маликов А.И., д-р техн. наук, проф. Сиразетдинов Р.Т. (КГТУ-КАИ); **Липецк** – д-ра техн. наук: проф. Погодаев А.К., Сараев П.В. (ЛГТУ); **Самара** – д-ра экон. наук: проф. Богатырев В.Д., проф. Гераськин М.И., д-р техн. наук, проф. Засканов В.Г. (СГАУ); **Петрозаводск** – д-р физ.-мат. наук, проф. Мазалов В.В., д-р техн. наук, доц. Печников А.А. (ИПМИ КарНЦ РАН); **Санкт-Петербург** – д-р физ.-мат. наук: проф. Петросян Л.А. (СПбГУ), д-р техн. наук проф. Фуртат И.Б. (ИПМ РАН); **Старый Оскол** – д-р техн. наук, проф. Еременко Ю.И. (СТИ).

Адрес редакции: 117997, г. Москва, ул. Профсоюзная, д. 65.

Адрес в Интернет: ubs.mtas.ru.

СОДЕРЖАНИЕ

Системный анализ

| | |
|---|----|
| Алгазин Г. И., Матюнин Е. В. <i>Об оптимальных стратегиях асимметрично информированных участников игровых взаимодействий...</i> | 6 |
| Выхованец В. С., Крыжановская А. В. <i>Совмещенные сети управления и данных</i> | 41 |
| Жилякова Л. Ю. <i>Ресурсная сеть с ограничением на ёмкость аттракторов.....</i> | 67 |

Информационные технологии в управлении

| | |
|--|-----|
| Ведешенков В. А., Курако Е. А., Лебедев В. Н. <i>О диагностируемости компонентов цифровых систем со структурой минимального квазиполного графа размера 7x7 с двумя путями.....</i> | 90 |
| Мелентьев В. А. <i>О топологической масштабируемости вычислительных систем</i> | 115 |

Управление в социально-экономических системах

| | |
|---|-----|
| Антоненко А. В., Лошкарев И. В., Панков В. С., Угольницкий Г. А. <i>Решение задачи стимулирования инноваций в электроэнергетике. Часть 1. Оптимизационные модели.</i> | 144 |
| Вожак А. В., Гитман М. Б., Столбов В. Ю. <i>Модели принятия коллективных решений в производственных системах.....</i> | 161 |
| Гераськин М. И., Егорова В. В. <i>Оптимальные механизмы планирования позаказного производства по финансовым и временным критериям</i> | 179 |
| Шумов В. В. <i>Модель безопасности государства</i> | 212 |

**Управление техническими системами и
технологическими процессами**

Лазарев А. А., Тарасов И. А.

Составление оптимального расписания движения поездов между двумя станциями, соединенными однопутной железной дорогой с разъездом

244

Управление подвижными объектами и навигация

Ватаманюк И. В., Панина Г. Ю., Ронжин А. Л.

Реконфигурация пространственного положения роя роботов

285

Яковлев К. С., Баскин Е. С., Андрейчук А. А.

Модели и алгоритмы минимизации ущерба от атмосферных выбросов промышленных предприятий

306

УДК 519.865.3
ББК 22.18

ОБ ОПТИМАЛЬНЫХ СТРАТЕГИЯХ АСИММЕТРИЧНО ИНФОРМИРОВАННЫХ УЧАСТНИКОВ ИГРОВЫХ ВЗАИМОДЕЙСТВИЙ

Алгазин Г. И.¹

(Алтайский государственный университет, Барнаул)

Матюнин Е. В.²

(ООО «МЕМ», Барнаул)

Рассматривается принятие оптимальных решений асимметрично информированными участниками игровых взаимодействий в условиях вероятностной неопределенности. Проводится исследование равновесия Байеса – Нэша в байесовых играх. Показывается, что в предложенных постановках байесовых игр нахождение равновесия Байеса – Нэша сводится к решению систем интегральных уравнений. Рассматриваются численные методы и предлагается алгоритм программной реализации решения данного типа задач.

Ключевые слова: асимметричная информированность, вероятностная неопределенность, байесовы игры, равновесие Байеса – Нэша, вариационное исчисление, численное решение.

1. Введение

Настоящая работа посвящена математическому обоснованию выбора решений при рассмотрении задач с вероятностной неопределенностью в том случае, когда не все

¹ Геннадий Иванович Алгазин, доктор физико-математических наук, профессор (algaz46@yandex.ru).

² Евгений Васильевич Матюнин, главный инженер (matyuninev@gmail.com).

взаимодействующие элементы располагают одинаковой информацией о существенных параметрах, влияющих на принятие решений. Такого рода задачи являются довольно распространенными в социальных и экономических системах. Причины, по которым важная для принятия решений информация, носящая случайный характер, известна одним участникам системы и недоступна другим, также весьма разнообразны.

В данной области исследования широкое применение получили теоретико-игровые подходы. В отечественной литературе эта область широко освещалась в работах, относящихся к теории активных систем [3–6, 13, 16, 18 и др.]. В зарубежной литературе задачи принятия решений в условиях вероятностной неопределенности исследовались в рамках теории контрактов, в качестве примера могут служить работы [23, 26, 29]. В обзорной работе [3] проводится сопоставление основных аспектов теории контрактов и теории активных систем. Авторы указывают на то, что данные области исследования развивались практически независимо, при этом были получены близкие результаты. Подробная классификация различных типов информированности элементов активных систем (в том числе асимметричной информированности) приводилась в [14].

Применительно к активным системам асимметрия информированности центра относительно типов агентов исследовалась, например, в работах [8, 12]. Рассматривалась внутренняя вероятностная неопределенность, где центру не известны типы агентов, но известно, что они описываются распределением Парето. Такого рода задачи возникают, например, при определении оптимального стимулирования или контроля персонала [11]. Также важной прикладной областью, привлекающей внимание исследователей, является разработка автоматизированных систем поддержки принятия решений. При проектировании алгоритмов работы в основу данных систем зачастую закладываются механизмы функционирования и принятия оптимальных решений в условиях асимметрии информированности [2]. В ряде работ уделяется внимание различным способам снижения асимметричности информации,

например, таким, как рыночные сигналы [31] или механизмы с сообщением информации [14].

Широкое применение в теоретико-игровом моделировании систем с вероятностной неопределенностью получили байесовы игры. В них структура информированности игроков задаётся с использованием дискретных либо непрерывных случайных параметров (определяющих типы игроков). Байесовы игры имеют приложение в таких областях, как построение оптимальных экономических механизмов, теория аукционов, теория организации промышленности [32]. Практическое применение теория байесовых игр с асимметрией информированности игроков нашла в описании логистических цепочек поставок [33], интернет-рекламе, проектировании беспроводных сетей и телекоммуникаций [24], инвестиционном менеджменте, где аналитику необходимо предоставить инвестору вид траекторий изменения доходности проекта от изменения некоторых значимых параметров системы [27], и многих других областях.

Исследование асимметрии информированности участников игровых взаимодействий в условиях вероятностной неопределенности относительно существенных параметров системы проводится также в рамках рефлексивных игр. В работе [20] рассматривается рефлексивная игра с асимметричным общим знанием участников о значении величины действий, выполнение объема которых влечет за собой выплату им вознаграждения. В статье [21] исследуется влияние взаимной информированности на выбор стратегий участниками одноходовых рефлексивных игр. Сопоставление подходов байесовых и рефлексивных игр рассматривалось в работе [15], где указывалось на то, что основная трудность исследования байесовых игр в общем случае связана с тем, что их структура имеет достаточно громоздкую конструкцию. Кроме того, были получены выводы о нецелесообразности использования бесконечной глубины структуры информированности для нахождения как информационного равновесия в рефлексивных играх, так и для нахождения равновесия Байеса – Нэша в байесовых играх.

Следует отметить, что универсального «аппарата» аналитического решения байесовых игр (в рамках равновесия

Байеса – Нэша) до сих пор не предложено. Также крайне мало работ, в которых применяются аналитические методы исследования, направленные на поиск общих закономерностей байесовых игр с асимметричной информацией. Поэтому в представленном исследовании авторы акцентируют внимание на математических аспектах принятия решений в условиях асимметрии информированности (относительно существенных вероятностных параметров системы), когда каждый участник взаимодействия вынужден определять стратегию поведения, предполагая, что оппонентам известна недоступная ему информация. При этом он может не представлять в точности, в какой мере его собственная информация известна другим.

Будем рассматривать решение непрерывных байесовых игр, используя в качестве математического аппарата исследования вариационное исчисление, так как стратегиями участников являются некоторые решающие правила (функции), определенные в пространстве допустимых решающих правил, доставляющие максимум целевым функционалам игроков и зависящие от входящих в рассматриваемую модель случайных параметров. Оптимальность выбора полученных стратегий обосновывается нахождением решения соответствующих задач вариационного исчисления с дополнительными условиями асимметрии информированности. Разработка методов решения байесовых игр с учетом этих условий игроками при обосновании оптимальности стратегий будет определять оригинальность проведенного авторами исследования.

Структура изложения материала настоящей статьи следующая: во втором разделе приводятся информационные формулировки байесовых игр. В третьем разделе предложено использование дополнительных условий асимметрии информированности в дифференциальной форме для байесовых игр. Четвертый раздел посвящен решению байесовой игры с полиномиальными целевыми функциями игроков. В пятом разделе рассматриваются численные методы нахождения равновесия Байеса – Нэша, сводящегося к решению систем интегральных уравнений Фредгольма. Шестой раздел посвящен рассмотрению примера конкурентного взаимодействия двух предприятий, принимающих решения по выбору оптимальных

планов производства в условиях асимметрии информированности. В заключении обсуждаются основные результаты и перспективы продолжения исследований. В приложении приводится доказательство теоремы о существовании равновесия Байеса – Нэша в рассматриваемых байесовых играх.

2. Общая теоретико-игровая модель с неполной информированностью участников

Приведем общую формулировку байесовой игры, следуя работе [25]. Байесова игра задается набором:

$$G = \{N, \Theta, X, F, P\},$$

где $N = \{1, \dots, n\}$ – множество игроков;

$$X = \prod_{i \in N} X_i \text{ – множество допустимых действий игроков}$$

(X_i – набор возможных действий i -го игрока);

$$\Theta = \prod_{i \in N} \Theta_i \text{ – множество всех типов игроков } (\Theta_i \text{ – набор}$$

возможных типов i -го игрока, типы задаются случайными параметрами);

$F: X \times \Theta \rightarrow R$ – множество всех функций выигрышей игроков;

$$P = \prod_{i \in N} P_i \text{ – множество представлений всех игроков о типах}$$

соперников (множество функций распределения типов игроков).

Стратегией i -го игрока в данном случае является отображение

$$x_i(\theta^i): \Theta_i \rightarrow X_i,$$

где θ^i – вектор случайных параметров, определяющих тип i -го игрока.

Принято выделять три информационные формулировки байесовых игр, зависящих от того, принимаются ли решения до наблюдения реализовавшихся значений типов игроков либо после [28]. При этом на момент принятия решения может иметь место как частичная, так и полная реализация значений всех случайных параметров, определяющих типы игроков.

Рассмотрим типы равновесий Байеса – Нэша, соответствующие различным информационным формулировкам байесовых игр (в настоящей работе будут исследоваться только ситуации равновесия в чистых стратегиях).

1. Сложившаяся игровая ситуация $x_1^*(\cdot), \dots, x_n^*(\cdot)$ называется ситуацией равновесия Байеса – Нэша в *ex ante* («предварительной») формулировке, если $\forall i \in N$ выполняется:

$$(1) \quad E_{[\theta^i]} f_i(x_i^*(\theta^i), x_{-i}^*(\theta^{-i}), \theta) \geq E_{[\theta^i]} f_i(x_i(\theta^i), x_{-i}^*(\theta^{-i}), \theta),$$

где $E_{[\cdot]}$ – оператор математического ожидания, f_i – функция выигрыша i -го игрока. В том случае, если типы игроков являются непрерывными случайными величинами, оптимальные стратегии игроков в рамках *ex ante* равновесия Байеса – Нэша (1) определяются в виде:

$$x_i^*(\theta^i) \in \text{Arg} \max_{x_i(\theta^i) \in \tilde{X}_i \ominus} \int f_i(x_i(\theta^i), x_{-i}^*(\theta^{-i}), \theta) dP(\theta),$$

где $x_{-i}^*(\theta^{-i})$ – оптимальные отклики всех игроков на действия i -го игрока; \tilde{X}_i – множество допустимых стратегий-функций i -го игрока; θ^{-i} – вектор случайных параметров, определяющих типы игроков отличных от i -го; θ – вектор всех типов игроков рассматриваемой системы.

С информационной точки зрения формулировка *ex ante* байесовой игры предполагает рассмотрение взаимодействий участников до получения игроками информации о реализовавшихся значениях случайных величин. Таким образом, участникам известны только функции распределения типов.

2. *Interim* («промежуточное») равновесие Байеса – Нэша – равновесие, при котором каждый игрок уже информирован о реализовавшемся значении своего типа, но не информирован о значениях типов других игроков. Состояние равновесия Байеса – Нэша для данной формулировки игры рассматривается в следующем виде:

$$E_{[\theta^{-i}]} f_i(x_i^*(\theta^i), x_{-i}^*(\theta^{-i}), \theta^i, \theta^{-i}) \geq E_{[\theta^{-i}]} f_i(x_i(\theta^i), x_{-i}^*(\theta^{-i}), \theta^i, \theta^{-i}).$$

Формализация данного вида равновесия Байеса – Нэша изначально закладывает асимметричную информированность участников игрового взаимодействия о типах друг друга.

3. *Ex post* («завершенная») формулировка байесовой игры рассматривается после получения значений исходов для каждого игрока. То есть после того как типы всех участников стали известны. *Ex post* равновесие Байеса – Нэша записывается в виде:

$$f_i(x_i^*(\theta^i), x_{-i}^*(\theta^{-i}), \theta^i, \theta^{-i}) \geq f_i(x_i(\theta^i), x_{-i}^*(\theta^{-i}), \theta^i, \theta^{-i}).$$

В работе не будет заостряться внимание на данном случае, так как такого рода игровая ситуация предполагает полную информированность участников обо всех параметрах системы.

3. Формализация условий асимметрии информированности в задачах принятия решений

В начале раздела рассмотрим содержательный пример информационного взаимодействия в условиях вероятностной неопределенности при асимметрии информированности. В данном примере также будет представлен переход от «предварительной» к «промежуточной» формулировке байесовой игры.

В качестве такого примера могут служить игровые взаимодействия в модели конкуренции двух фирм (например, модели конкуренции Курно, Чемберлина [30]), выходящих на рынок с однотипным товаром. При нахождении рациональных стратегий поведения может рассматриваться достаточно большое число параметров, которые влияют на ожидаемую прибыль каждой фирмы. Среди них могут быть параметры, о значениях которых участники взаимодействия информированы не полностью, например, один участник может быть не информирован о достаточно большем числе параметров, влияющих на определение себестоимости товара оппонентом. Такими параметрами могут быть тарифы электроэнергии, заработная плата персонала, эффективность используемого оборудования, различные накладные расходы (например, затраты на ведение рекламной компании), которые являются

существенными с точки зрения участников взаимодействия при выводе товара на рынок. При этом зачастую необходимо определить стоимость продукта и рынки сбыта до того момента, когда будут известны конкретные значения случайных параметров (на начальном этапе известны только вероятностные меры, полученные на основе анализа различных статистических данных), либо конкретные реализовавшиеся значения некоторых параметров так и не станут известны определенным игрокам. Такой информационный случай при рассмотрении задачи в виде байесовой игры соответствует как раз *ex ante* формулировке данного игрового взаимодействия. Переход от «предварительной» к «промежуточной» формулировке в модели конкуренции Чемберлина происходит в тот момент, когда каждая фирма получает информацию о том, конкретные реализовавшиеся значения каких случайных параметров она сможет наблюдать, а о значениях каких случайных величин так и не будет информирована. Принятие рациональных решений в такой информационной ситуации предполагает нахождение оптимальных стратегий участниками игрового взаимодействия в условиях асимметрии информированности.

Перейдем к формализации условий асимметрии информированности. Рассмотрим байесову игру двух лиц, введенную в разделе 2, где второй игрок не будет иметь информации о реализовавшихся значениях параметров $\theta^1 = (\theta_1, \dots, \theta_n)$, $\theta^1 \in \Theta^1$, а первый игрок не будет информирован о реализовавшихся значениях набора параметров $\theta^2 = (\theta_{n+1}, \dots, \theta_r)$, $\theta^2 \in \Theta^2$. Вектор информационных параметров $\theta \in \Theta$ имеет следующую структуру: $\theta = (\theta_1, \dots, \theta_n, \theta_{n+1}, \dots, \theta_r)$, $\Theta = \Theta^1 \cup \Theta^2$ – множество всех информационных параметров системы; множество всех индексов случайных параметров обозначим I ; множество индексов, отражающих номера компонент вектора параметров, неизвестных первому игроку, обозначим $I_2 \in I$, второму игроку – $I_1 \in I$. Решения участниками принимаются в предположении, что никакой дополнительной информации по реализации значений вектора параметров θ^1 для второго игрока, θ^2 для первого игрока, не ожидается [7]. Применяя принцип усреднения для устранения

неопределенности, формально считаем, что принимаемое оперирующей стороной решение зависит от того, какие значения примут неопределенные параметры, определяющие её тип, и от усредненных значений параметров оппонента. В работах [9, 10] асимметрию информированности игроков относительно существенных параметров рассматриваемой системы, применительно к задачам с вероятностной неопределенностью, предлагалось формализовывать введением дополнительных ограничений следующего вида:

$$(2) \quad \begin{aligned} \frac{\partial x_1(\cdot)}{\partial \theta_j} &= 0 \quad (j = h + 1, \dots, r); \\ \frac{\partial x_2(\cdot)}{\partial \theta_i} &= 0 \quad (i = 1, \dots, h). \end{aligned}$$

Качественно условия (2) означают, что стратегия конкретного игрока не зависит от параметров, точные значения которых так и не станут ему известны в ходе игрового взаимодействия, но могут быть известны оппоненту. Подобные условия (в виде не строгих неравенств) использовались в работах [1, 4], но они имели иной содержательный смысл.

Далее в работе будем рассматривать тип задач, удовлетворяющих следующим начальным предположениям:

1) на момент принятия решения о выборе стратегий игроки не знают реализовавшиеся значения своих типов (либо эти значения еще не реализовались);

2) каждый игрок сможет наблюдать значение своего типа, но не получит никакой дополнительной информации о реализовавшемся значении типа оппонента (условие асимметрии информированности (2));

3) для оценки своего выигрыша игроки на момент принятия решения рассматривают следующую функцию ожидаемой полезности:

$$u_i = \int_{\Theta} f_i(x_i(\theta^i), x_{-i}(\theta^{-i}), \theta^i, \theta^{-i}) dP(\theta) \rightarrow \max_{x_i(\theta^i)};$$

4) игрок i рассматривает функционирование своей системы в предположении, что он оперирует стратегией-функцией $x_i^*(\theta^i)$

и, соответственно, определяет ожидаемый выигрыш в предположении того, что придерживается данной стратегии.

Таким образом, будем рассматривать нахождение *ex ante* равновесия Байеса – Нэша с описанными выше условиями асимметрии информированности участников взаимодействия.

4. Равновесие Байеса – Нэша в играх двух лиц с асимметрией информированности участников

В качестве метода нахождения равновесия Байеса – Нэша рассмотрим сведение решения байесовой игры к задаче оптимального управления в случае несовпадающей информированности субъектов управления. Оператор управления состояниями субъекта функционирует в динамической случайной среде. Пусть θ^1, θ^2 – случайные векторы с функциями распределения $P_2(\theta^1), P_1(\theta^2)$. Множества индексов I_1, I_2 определяют структуру управляющих переменных $x_1(\cdot), x_2(\cdot)$.

Более подробно решение байесовой игры с асимметричной информацией рассмотрим для задачи с полиномиальными целевыми функциями, где $r = 1, h = 1$, т.е. $\theta = (\theta_1, \theta_2)$. Предположим, в целевую функцию первого игрока искомые стратегии $x_1(\cdot), x_2(\cdot)$ входят со степенями n, m , второго игрока – со степенями k, l , и целевые функции имеют вид:

$$f_1 = q_{11}(\tilde{x}_1 - a)^n + 2q_{12}(\tilde{x}_1 - a)^{\frac{n}{2}}(\tilde{x}_2 - b)^{\frac{m}{2}} + 2q_{13}(\tilde{x}_1 - a)^{\frac{n}{2}}\theta_1 + \dots + q_{44}\theta_2^2,$$

$$f_2 = s_{11}(\tilde{x}_1 - c)^k + 2s_{12}(\tilde{x}_1 - c)^{\frac{k}{2}}(\tilde{x}_2 - e)x_2^{\frac{l}{2}} + 2s_{13}(\tilde{x}_1 - c)^{\frac{k}{2}}\theta_1 + \dots + s_{44}\theta_2^2,$$

где $\tilde{x}_1 = x_1(\theta_1), \tilde{x}_2 = x_2(\theta_2), n, m, k, l > 1$ и $a, b, c, e \in R$.

Запись целевых функций в матричной форме выглядит следующим образом:

$$(3) \quad \begin{aligned} f_1(x_1(\theta_1), x_2(\theta_2), \theta_1, \theta_2) &= AQA^T, \\ f_2(x_1(\theta_1), x_2(\theta_2), \theta_1, \theta_2) &= BSB^T, \end{aligned}$$

где

$$Q = Q^T = (q_{ij}), \quad S = S^T = (s_{ij}), \quad s_{ij} \in R, \quad q_{ij} \in R$$

– некоторые коэффициенты ($i, j = 1, \dots, 4$),

$$A = \left((\tilde{x}_1 - a)^{\frac{n}{2}}, (\tilde{x}_2 - b)^{\frac{m}{2}}, \theta_1, \theta_2 \right), \quad B = \left((\tilde{x}_1 - c)^{\frac{k}{2}}, (\tilde{x}_2 - e)^{\frac{l}{2}}, \theta_1, \theta_2 \right).$$

Для устранения неопределенности будем использовать оператор математического ожидания.

Пусть выполнены следующие условия:

1) случайные величины θ_1, θ_2 , определяющие типы игроков, распределены на интервалах $[\underline{\theta}_1, \bar{\theta}_1]$, $[\underline{\theta}_2, \bar{\theta}_2]$ соответственно. Будем рассматривать случай, где

$$[\underline{\theta}_1, \bar{\theta}_1] = [\underline{\theta}_2, \bar{\theta}_2] = [\underline{\theta}, \bar{\theta}], \quad \underline{\theta} \in R, \quad \bar{\theta} \in R;$$

2) $x_1(\cdot), x_2(\cdot)$ – выпуклые и непрерывно дифференцируемые функции по θ_1, θ_2 на интервале $[\underline{\theta}, \bar{\theta}]$;

3) $P(\theta_1, \theta_2) = P_1(\theta_2) \cdot P_2(\theta_1)$ (т.е. θ_1, θ_2 – независимые случайные величины);

4) функции распределения $P_1(\theta_2), P_2(\theta_1)$ непрерывны по θ_1, θ_2 ;

$$5) \text{ выполняются условия: } \frac{\partial x_1(\cdot)}{\partial \theta_2} = 0, \quad \frac{\partial x_2(\cdot)}{\partial \theta_1} = 0;$$

$$6) \int_{\underline{\theta}}^{\bar{\theta}} p_1(\theta_2) d\theta_2 = 1, \quad \int_{\underline{\theta}}^{\bar{\theta}} p_2(\theta_1) d\theta_1 = 1, \text{ где } p_1(\theta_2), p_2(\theta_1) -$$

плотности распределения соответствующих случайных параметров.

Рассмотрим случай, где целевые функции игроков зависят и от θ_1 , и от θ_2 , при этом первый игрок находит стратегию, зависящую от параметра θ_1 (сможет наблюдать реализовавшееся значение случайной величины), а второй игрок находит стратегию, зависящую от параметра θ_2 . Условия вида (2) позволяют обосновать нахождение ситуации равновесия Байеса – Нэша для игры с асимметрией информированности участников. Ситуация равновесия Байеса – Нэша в данном случае записывается в следующей форме (не учитывается

момент реализации значений соответствующих случайных величин):

$$E_{[\theta_1]} f_1(x_1^*(\theta_1), x_2^*(\theta_2), \theta) \geq E_{[\theta_1]} f_1(x_1(\theta_1), x_2^*(\theta_2), \theta),$$

$$E_{[\theta_1]} f_2(x_1^*(\theta_1), x_2^*(\theta_2), \theta) \geq E_{[\theta_1]} f_2(x_1^*(\theta_1), x_2(\theta_2), \theta),$$

где $\theta = (\theta_1, \theta_2)$.

Так как стратегии игроков определяются в виде некоторых функций, зависящих от параметров θ_1, θ_2 , то для нахождения решения рассматриваемой игры необходимо найти решение следующей задачи вариационного исчисления:

$$u_1 = \int_{\Theta} f_1(\tilde{x}_1, \tilde{x}_2, \theta_1, \theta_2) dP(\theta) \rightarrow \max_{\tilde{x}_1},$$

$$u_2 = \int_{\Theta} f_2(\tilde{x}_1, \tilde{x}_2, \theta_1, \theta_2) dP(\theta) \rightarrow \max_{\tilde{x}_2},$$

(4) $\frac{\partial \tilde{x}_1}{\partial \theta_2} = 0,$

$$\frac{\partial \tilde{x}_2}{\partial \theta_1} = 0,$$

где u_1, u_2 – функции ожидаемой полезности игроков, определяющие ожидаемый выигрыш, в предположении того, что игроки будут оперировать стратегиями-функциями \tilde{x}_1, \tilde{x}_2 , усредняя параметры, значения которых соответствующим участникам известны не будут.

Полученные выражения рассматриваем как задачу оптимального управления, где \tilde{x}_1, \tilde{x}_2 – управление, а ограничения

$$\frac{\partial \tilde{x}_1}{\partial \theta_2} = 0, \quad \frac{\partial \tilde{x}_2}{\partial \theta_1} = 0$$

являются дифференциальной связью и выполняются во всех точках непрерывности управления на Θ . Таким образом, исследуемые задачи управления являются задачами со свободными концами и требуют нахождения стратегий, доставляющих экстремум функционалам среди всех

допустимых функций $\tilde{x}_1 \in C^1_{[\theta_1, \bar{\theta}_1]}$, $\tilde{x}_2 \in C^1_{[\theta_2, \bar{\theta}_2]}$, лежащих на заданных вертикалях.

Составим функции Лагранжа для задачи (4):

$$L_1 = \int_{\Theta} \lambda_1 f_1(\tilde{x}_1, \tilde{x}_2, \theta_1, \theta_2) p(\theta) + \mu_1(\theta_1) \left(\frac{\partial \tilde{x}_1}{\partial \theta_2} \right) d\theta,$$

$$L_2 = \int_{\Theta} \lambda_2 f_2(\tilde{x}_1, \tilde{x}_2, \theta_1, \theta_2) p(\theta) + \mu_2(\theta_2) \left(\frac{\partial \tilde{x}_2}{\partial \theta_1} \right) d\theta,$$

где $\lambda_1, \lambda_2, \mu_1(\cdot), \mu_2(\cdot)$ – множители Лагранжа, $p(\theta)$ – плотность распределения вектора θ .

Выпишем условия оптимального в слабом смысле стохастического процесса по \tilde{x}_1, \tilde{x}_2 :

$$\begin{aligned} \frac{\partial}{\partial \tilde{x}_1} \left(\lambda_1 f_1(\tilde{x}_1, \tilde{x}_2, \theta_1, \theta_2) p(\theta) + \mu_1(\theta_1) \left(\frac{\partial \tilde{x}_1}{\partial \theta_2} \right) \right) &= 0, \\ \frac{\partial}{\partial \tilde{x}_2} \left(\lambda_2 f_2(\tilde{x}_1, \tilde{x}_2, \theta_1, \theta_2) p(\theta) + \mu_2(\theta_2) \left(\frac{\partial \tilde{x}_2}{\partial \theta_1} \right) \right) &= 0; \end{aligned} \quad \Leftrightarrow$$

$$\lambda_1 \frac{\partial (f_1(\tilde{x}_1, \tilde{x}_2, \theta_1, \theta_2) p_1(\theta_2))}{\partial \tilde{x}_1} = \frac{\partial \mu_1(\theta_1)}{\partial \theta_2},$$

$$\lambda_2 \frac{\partial (f_2(\tilde{x}_1, \tilde{x}_2, \theta_1, \theta_2) p_2(\theta_1))}{\partial \tilde{x}_2} = \frac{\partial \mu_2(\theta_2)}{\partial \theta_1}.$$

Условия трансверсальности для свободных границ записываются в виде

$$\mu_1(\theta_1) \Big|_{\theta_1 \in [\theta_1, \bar{\theta}_1]} = 0,$$

$$\mu_2(\theta_2) \Big|_{\theta_2 \in [\theta_2, \bar{\theta}_2]} = 0.$$

Пусть $\lambda_1 = 1, \lambda_2 = 1$. Принимая во внимание, что

$$\mu_2(\theta_2)|_{\theta_2 \in [\underline{\theta}_2, \bar{\theta}_2]} = \int_{\underline{\theta}_2}^{\bar{\theta}_2} \frac{\partial(f_1(x_1(\theta_1), x_2(\theta_2), \theta_1, \theta_2)) p_1(\theta_2))}{\partial \tilde{x}_1} d\theta_2,$$

$$\mu_1(\theta_1)|_{\theta_1 \in [\underline{\theta}_1, \bar{\theta}_1]} = \int_{\underline{\theta}_1}^{\bar{\theta}_1} \frac{\partial(f_2(x_1(\theta_1), x_2(\theta_2), \theta_1, \theta_2)) p_2(\theta_1))}{\partial \tilde{x}_2} d\theta_1,$$

получаем необходимые условия экстремума следующего вида:

$$\int_{\underline{\theta}_2}^{\bar{\theta}_2} \frac{\partial(f_1(x_1(\theta_1), x_2(\theta_2), \theta_1, \theta_2)) p_1(\theta_2))}{\partial \tilde{x}_1} d\theta_2 = 0,$$

$$\int_{\underline{\theta}_1}^{\bar{\theta}_1} \frac{\partial(f_2(x_1(\theta_1), x_2(\theta_2), \theta_1, \theta_2)) p_2(\theta_1))}{\partial \tilde{x}_2} d\theta_1 = 0.$$

Следовательно, нахождение ситуации равновесия Байеса – Нэша игры (4) сводится к решению следующей системы интегральных уравнений:

$$\left\{ \int_{\underline{\theta}_2}^{\bar{\theta}_2} \frac{\left(\frac{(\tilde{x}_1 - a)^{n-1}}{q_{11}^{-1}} + (\tilde{x}_1 - a)^{\frac{n}{2}-1} \left(q_{12}(\tilde{x}_2 - b)^{\frac{m}{2}} + q_{13}\theta_1 + q_{14}\theta_2 \right) \right)}{(np_1(\theta_2))^{-1}} d\theta_2 = 0, \right.$$

$$\left. \int_{\underline{\theta}_1}^{\bar{\theta}_1} \frac{\left(\frac{(\tilde{x}_2 - e)^{l-1}}{s_{22}^{-1}} + (\tilde{x}_2 - e)^{\frac{l}{2}-1} \left(s_{12}(\tilde{x}_1 - c)^{\frac{k}{2}} + s_{23}\theta_1 + s_{24}\theta_2 \right) \right)}{(lp_2(\theta_1))^{-1}} d\theta_1 = 0. \right.$$

Таким образом, мы получили систему интегральных уравнений, решения которой являются искомыми стратегиями-функциями игроков в задаче (4).

Тип полученных интегральных уравнений зависит от степеней подынтегральных полиномов и значений коэффициентов. Например, при $n = m = k = l = 2$, полученное выражение является системой интегральных уравнений Фредгольма второго рода.

Приведем теорему о существовании равновесия Байеса – Нэша в задачах вида (4). Существование равновесия Байеса – Нэша в чистых стратегиях зависит от существования

неподвижной точки оператора в бесконечномерном пространстве.

Стратегия игрока в этом случае является измеримой функцией $x_i(\theta_i): \Theta_i \rightarrow X_i$, где $x_i(\theta_i) \in \tilde{X}_i$ – пространство измеримых функций, реализующих отображение из Θ_i в X_i с нормой следующего вида: $\|\tilde{x}_i\| = \sup_{\theta_i \in \Theta_i} |x_i(\theta_i)|$, $i = \{1, 2\}$.

Для рассмотрения вопросов существования равновесия Байеса – Нэша в игре вида (4) сформулируем следующую теорему о существовании неподвижной точки отображения метрического пространства в себя.

Теорема 1. Сложившаяся игровая ситуация $(x^*_1(\theta_1), x^*_2(\theta_2))$, где $x^*_1(\theta_1): \Theta_1 \rightarrow X_1$, $x^*_2(\theta_2): \Theta_2 \rightarrow X_2$, определяет равновесие Байеса – Нэша в байесовой игре (4), если выполнены следующие условия:

1) функции распределения $P(\theta_1 | \theta_2)$, $P(\theta_2 | \theta_1)$ непрерывны по θ_1 , θ_2 (т.е. представления игроков о типах друг друга непрерывны по соответствующим типам);

2) пространства типов Θ_1 , Θ_2 – не пустые, выпуклые и компактные подмножества евклидова пространства;

3) пространства действий X_1 , X_2 – не пустые, выпуклые и компактные подмножества евклидова пространства;

4) целевые функции игроков $f_1(x_1, x_2, \theta_1, \theta_2)$, $f_2(x_1, x_2, \theta_1, \theta_2)$ являются непрерывными по $x_1, x_2, \theta_1, \theta_2$;

5) $\forall \theta_i$ и измеримой функции $x_{-i}(\theta_{-i})$ функция ожидаемой полезности $E_{\theta_i}(f_i(x_i(\theta_i), x_{-i}(\theta_{-i}), \theta_i, \theta_{-i}))$ унимодальна по x_i ($i = \{1, 2\}$);

6) выполняется равномерная непрерывность семейства допустимых стратегий-функций $x_i(\theta_i) \in \tilde{X}_i$.

Доказательство теоремы приводится в Приложении и основывается на установлении существования неподвижной точки отображения функционального пространства в себя.

5. Численный алгоритм нахождения равновесия Байеса – Нэша в задачах с асимметричной информированностью участников

Приведем алгоритм нахождения ситуации равновесия Байеса – Нэша в байесовой игре с асимметричной информированностью участников. Рассмотрим подход к нахождению равновесия Байеса – Нэша, сводящийся к решению системы интегральных уравнений Фредгольма 2-го рода следующего вида:

$$(5) \quad \begin{cases} x_1(\theta_1) - \lambda_1 \int_{\Theta_2} x_2(\theta_2) K_1(\theta_1, \theta_2) d\theta_2 = y_1(\theta_1), \\ x_2(\theta_2) - \lambda_2 \int_{\Theta_1} x_1(\theta_1) K_2(\theta_1, \theta_2) d\theta_1 = y_2(\theta_2), \end{cases}$$

где $x_1(\theta_1)$, $x_2(\theta_2)$ – искомые стратегии-функции первого и второго игроков соответственно;

$K_1(\theta_1, \theta_2)$, $K_2(\theta_1, \theta_2)$ – ядра интегральных уравнений;

$y_1(\theta_1)$, $y_2(\theta_2)$ – правые части интегральных уравнений;

λ_1 , λ_2 – константы.

Запишем решение исходной задачи (4) в виде следующей системы интегральных уравнений:

$$\begin{cases} \left(\tilde{x}_1 - a \right)^{\frac{n}{2}} + \frac{q_{12}}{q_{11}} \int_{\bar{\theta}_2} (\tilde{x}_2 - b)^{\frac{m}{2}} p_1(\theta_2) d\theta_2 = -\frac{q_{13}}{q_{11}} \theta_1 - \frac{q_{14}}{q_{11}} \int_{\bar{\theta}_2} \theta_2 p_1(\theta_2) d\theta_2, \\ \left(\tilde{x}_2 - e \right)^{\frac{n}{2}} + \frac{s_{12}}{s_{22}} \int_{\bar{\theta}_1} (\tilde{x}_1 - c)^{\frac{k}{2}} p_2(\theta_1) d\theta_1 = -\frac{s_{24}}{s_{22}} \theta_2 - \frac{s_{23}}{s_{22}} \int_{\bar{\theta}_1} \theta_1 p_2(\theta_1) d\theta_1. \end{cases}$$

Данная система нелинейных интегральных уравнений сводится к системе интегральных уравнений Фредгольма 2-го рода вида (5) при выполнении следующих условий:

- 1) $n = m = l = k = 2$;
- 2) $\lambda_1 = -\frac{q_{12}}{q_{11}}$, $\lambda_2 = -\frac{s_{12}}{s_{22}}$;
- 3) $q_{11} \neq 0$, $s_{22} \neq 0$;

$$4) \quad y_1(\theta_1) = -\frac{q_{13}}{q_{11}} \theta_1 - \frac{q_{14}}{q_{11}} \int_{\underline{\theta}_2}^{\bar{\theta}_2} \theta_2 p_1(\theta_2) d\theta_2,$$

$$y_2(\theta_2) = -\frac{s_{24}}{s_{22}} \theta_2 - \frac{s_{23}}{s_{22}} \int_{\underline{\theta}_1}^{\bar{\theta}_1} \theta_1 p_2(\theta_1) d\theta_1;$$

$$5) \quad K_1(\cdot) = p_1(\theta_2), \quad K_2(\cdot) = p_2(\theta_1).$$

Численное решение представленной задачи будем рассматривать для общего случая, где ядра интегральных уравнений зависят от параметров θ_1 , θ_2 . Такая ситуация возможна, если случайные величины θ_1 , θ_2 имеют совместную плотность распределения. Предлагаемое решение также в полной мере применимо и для случая независимых плотностей распределения.

Проведем замену системы интегральных уравнений конечномерной системой алгебраических уравнений. Применяя для численного решения квадратурную формулу, представим систему интегральных уравнений (5) в следующем виде [17]:

$$(6) \quad \begin{cases} x_1(t_j) - \sum_{i=1}^n \frac{\bar{\theta}_2 - \theta_2}{n} x_2(s_i) K_1(t_j, s_i) = y_1(t_j) + v, j = 1, \dots, m, \\ x_2(t_j) - \sum_{i=1}^n \frac{\bar{\theta}_1 - \theta_1}{n} x_1(s_i) K_2(t_j, s_i) = y_2(t_j) + v, j = 1, \dots, m, \end{cases}$$

где v – ошибка, связанная с заменой интеграла конечной суммой, t_j и s_i – соответствующие значения разбиений случайных непрерывных параметров в узлах сетки (исследование задачи будем проводить в предположении, что $\underline{\theta}_1 = \underline{\theta}_2$, $\bar{\theta}_1 = \bar{\theta}_2$).

Для решения системы уравнений (6) составим матрицу коэффициентов её левой части:

$$(7) \quad A = \begin{pmatrix} E & A^2 \\ A^1 & E \end{pmatrix},$$

где E – единичная матрица соответствующей размерности,

$$A^1 = \begin{pmatrix} -K_1(t_{n/2+1}, s_1) & \cdots & -K_1(t_{n/2+1}, s_{n/2}) \\ \vdots & \ddots & \vdots \\ -K_1(t_n, s_1) & \cdots & -K_1(t_n, s_{n/2}) \end{pmatrix},$$

$$A^2 = \begin{pmatrix} -K_2(t_1, s_{n/2+1}) & \cdots & -K_2(t_1, s_n) \\ \vdots & \ddots & \vdots \\ -K_2(t_{n/2}, s_{n/2+1}) & \cdots & -K_2(t_{n/2}, s_n) \end{pmatrix}.$$

Таким образом, система (6) имеет матричную форму записи следующего вида:

$$(8) \quad \begin{pmatrix} E & A^2 \\ A^1 & E \end{pmatrix} \begin{pmatrix} x_2^1 \\ \vdots \\ x_2^n \\ x_1^1 \\ \vdots \\ x_1^n \end{pmatrix} = \begin{pmatrix} y_1(t_1) \\ \vdots \\ y_1(t_n) \\ y_2(t_1) \\ \vdots \\ y_2(t_n) \end{pmatrix}.$$

Решением рассматриваемого уравнения (8) является набор значений функций $x_1(t)$, $x_2(t)$, полученных в узлах сетки. Введем следующие обозначения: $\bar{x}_1 = (x_1^1, \dots, x_1^n)$, $\bar{x}_2 = (x_2^1, \dots, x_2^n)$. Интерполирование стратегий-функций $x_1(t)$ и $x_2(t)$ будем проводить нахождением полиномов некоторой степени с использованием многочленов Лагранжа:

$$(9) \quad x_1^a(t) = \sum_{i=0}^n x_1^i l_1^i(t),$$

$$x_2^a(t) = \sum_{i=0}^n x_2^i l_2^i(t).$$

В выражениях (9) $l_1^i(t)$, $l_2^i(t)$ – базисные полиномы, $x_1^a(t)$, $x_2^a(t)$ являются приближенными с некоторой точностью решениями системы (5).

Определим базисные полиномы следующим образом:

$$l_1^i(t) = \prod_{j=0, i \neq j}^n \frac{x_1 - x_1^j}{x_1^i - x_1^j} = \frac{x_1 - x_1^0}{x_1^i - x_1^0} \dots \frac{x_1 - x_1^{i-1}}{x_1^i - x_1^{i-1}} \frac{x_1 - x_1^{i+1}}{x_1^i - x_1^{i+1}} \dots \frac{x_1 - x_1^n}{x_1^i - x_1^n},$$

$$l_2^i(t) = \prod_{j=0, i \neq j}^n \frac{x_2 - x_2^j}{x_2^i - x_2^j} = \frac{x_2 - x_2^0}{x_2^i - x_2^0} \dots \frac{x_2 - x_2^{i-1}}{x_2^i - x_2^{i-1}} \frac{x_2 - x_2^{i+1}}{x_2^i - x_2^{i+1}} \dots \frac{x_2 - x_2^n}{x_2^i - x_2^n}.$$

Для определения погрешности численного решения уравнения (9) вычисляется функция невязки:

$$(10) R(t) = X^a(t) - \delta \int_{\ominus} K(t, s) X^a(s) ds - y(t),$$

где

$$X^a(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}, \quad X^a(s) = \begin{pmatrix} x_1(s) \\ x_2(s) \end{pmatrix},$$

$$y(t) = \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix}, \quad K(t, s) = \begin{pmatrix} K_1(t, s) & 0 \\ 0 & K_2(t, s) \end{pmatrix}.$$

Абсолютная величина невязки определяется следующим образом:

$$R^a = \max_{\ominus} |X^a(t) - \delta \int_{\ominus} K(t, s) X^a(s) ds - y(t)|.$$

Рассмотрим основные этапы программной реализации алгоритма решения задачи вида (5).

1. Инициализация входных параметров. В алгоритме задается число разбиений исходных интервалов, границы интегрирования, функции ядер, правая часть системы (5), интервалы для вывода графиков, параметры интерполяции.

2. Построение расчётной сетки.

3. Вычисление элементов матрицы коэффициентов (7). Заполнение матрицы коэффициентов системы уравнений (8) на основе выражений (9).

4. Решение уравнения (8).

5. Интерполяция решений. Для решений $\bar{x}_1(t)$ и $\bar{x}_2(t)$ уравнения (8) строится интерполяционный полином заданной степени.

6. Вычисление невязки. По формуле (10) вычисляется функция невязки для полученных полиномов, определяющая порядок точности проводимых вычислений.

7. Вывод результата.

Уточним, что представленный алгоритм допустим для систем интегральных уравнений с несингулярными ядрами.

Для демонстрации применения рассмотренного численного метода к нахождению ситуации равновесия Байеса – Нэша в предложенных задачах приведем расчетный пример описанного алгоритма с несингулярными ядрами интегральных уравнений, зависящими и от параметра θ_1 , и от параметра θ_2 :

$$K_1(\theta_1, \theta_2) = \frac{(1 + \theta_1)}{\left(1 - \frac{e}{2}(3 - e^2)\right) e^{-(\theta_1 + \theta_2 + \theta_1 \cdot \theta_2)}};$$

$$K_2(\theta_1, \theta_2) = \frac{(1 + \theta_2)}{\left(1 - \frac{e}{2}(3 - e^2)\right) e^{-(\theta_1 + \theta_2 + \theta_1 \cdot \theta_2)}}.$$

Рассмотрение данного расчетного примера возможно, если случайные величины θ_1 , θ_2 имеют экспоненциальные на интервале плотности распределения. Предположим, что множество непрерывных случайных параметров определено на интервале $[0, 1]$. Пусть правая часть первого уравнения системы (5) является функцией, зависящей от параметра θ_1 , и имеет вид: $2 - \theta_1/3$. Правая часть второго уравнения системы (5) является функцией, зависящей от параметра θ_2 , и имеет вид: $12/5 - \theta_2$.

Интерполирование проведем с использованием многочленов Лагранжа 4-й степени. Результат, полученный в ходе реализации представленного алгоритма в программной среде математического моделирования *Maple*, проиллюстрируем с помощью графиков (см. рис. 1 и 2).

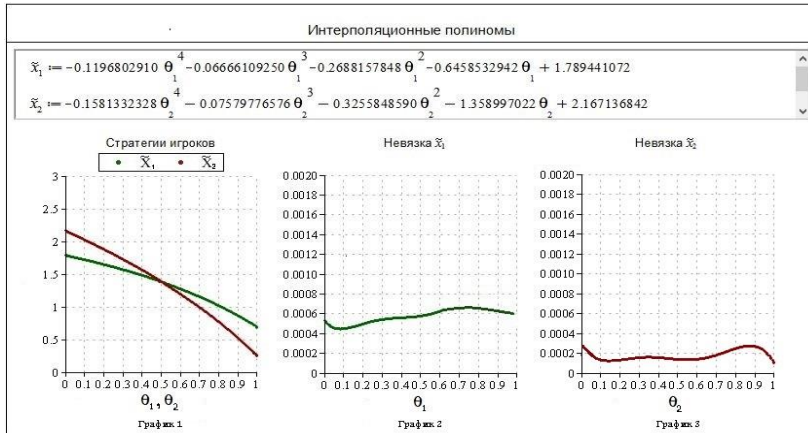


Рис. 1. Реализация алгоритма нахождения равновесия Байеса – Нэша в игре с асимметричной информацией. Окно вывода результата (число разбиений исходного интервала $n = 15$)

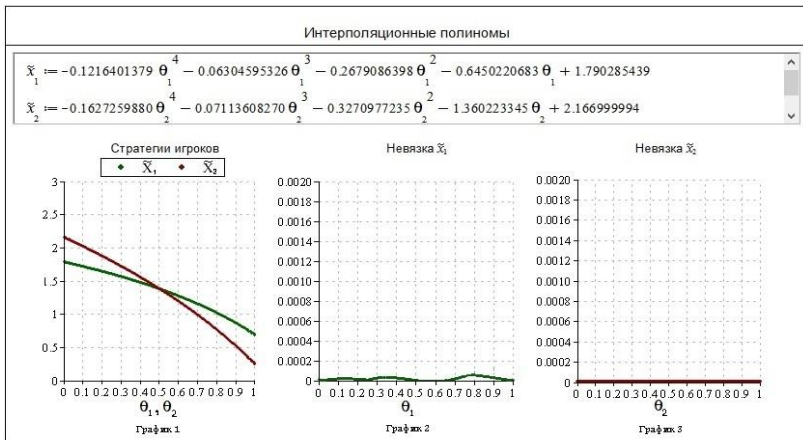


Рис. 2. Реализация алгоритма нахождения равновесия Байеса – Нэша в игре с асимметричной информацией. Окно вывода результата (число разбиений исходного интервала $n = 40$)

В верхней части рис. 1 и 2 представлены полученные интерполяционные полиномы, определяющие численное решение системы интегральных уравнений (5), их вид также

представлен на графике 1. В рассматриваемой теоретико-игровой интерпретации данные полиномы определяют стратегии игроков в виде функциональных зависимостей. На графиках 2 и 3 представлена величина погрешности применяемого численного метода в виде функций невязок, определяющих точность проводимых вычислений для различного исходного числа узлов разностной сетки. Анализ приведенных зависимостей показывает, что при изменении числа разбиений исходного интервала с 15 до 40, абсолютная величина невязки уменьшается для $x_1^a(t)$ с 0,00064 до 0,000081, для $x_2^a(t)$ с 0,00025 до 0,000042. Таким образом, рассмотренный пример также иллюстрирует, что предложенный численный метод применим для нахождения равновесных ситуаций в исследуемых задачах.

6. Пример конкурентного взаимодействия двух предприятий в условиях асимметрии информированности

Рассмотрим в качестве примера модель дуополии Курно [30] в виде байесовой игры (с конкретными целевыми функциями, стратегиями, типами), для которой сначала выпишем интегральные уравнения, затем получим их численное решение и проинтерпретируем его в терминах исходной задачи.

В исследуемой модели каждый из двух конкурирующих друг с другом производителей, производящих однотипный товар, пытается максимизировать свою прибыль, которая зависит от его собственных действий, от действий конкурента и от некоторых параметров рассматриваемой системы. При этом каждый производитель покрывает своей продукцией некоторую часть рыночного спроса при единой цене. Оптимальное решение будет определяться выбором плана выпуска продукции каждым предприятием в зависимости от действий конкурента и от ожидаемых значений случайных параметров, определяющих себестоимость производимой продукции, при асимметричной информированности предприятий о точных значениях этих параметров.

Введем описание того, при каких условиях взаимодействия предприятий возникают подобные задачи с асимметрией информированности участников. Для построения торговых стратегий и планирования продаж дилерам необходимо знать оценочную стоимость либо количество производимой продукции задолго до выпуска нового вида товара. Поэтому у предприятий возникает необходимость сообщать дилерам, реализующим производимый товар, стоимость продукции либо объем производимой продукции до момента начала выпуска несмотря на то, что в системе может присутствовать множество неопределенных (случайных) параметров, оказывающих влияние на себестоимость продукции. Производитель в некоторых случаях может так и не узнать точные значения всех параметров, влияющих на определение конечной стоимости продукции, но эти значения могут быть известны конкуренту. В рассматриваемой модели каждый производитель на момент начала выпуска будет информирован о себестоимости своей продукции, но не будет информирован о себестоимости однотипной продукции конкурента. Асимметрично информированные производители в ходе определения своих оптимальных стратегий могут проводить усреднение по всем случайным параметрам, но также могут воспользоваться доступной им информацией, выбирая стратегии в виде функциональных зависимостей от возможных значений случайных параметров. В ряде случаев это приносит существенное увеличение прибыли каждого участника, а также увеличение общей отраслевой прибыли.

На приведенном примере рассмотрим также эффективность применения стратегий-функций (от параметров, значения которых станут известны производителям на момент реализации решения) по сравнению со стратегиями, полученными в виде числовых значений при усреднении ожидаемых полезностей производителей по случайным параметрам.

Модель Курно представим в виде байесовой игры (в соответствии с изложенной в разделе 2 формулировкой) со следующими функциями ожидаемой полезности:

$$u_1 = \frac{a - x_1 - x_2}{b} x_1 - \theta_1 x_1,$$

$$u_2 = \frac{a - x_1 - x_2}{b} x_2 - \theta_2 x_2,$$

где a, b – положительные константы;

x_1, x_2 – стратегии игроков (количество производимых единиц продукции);

θ_1, θ_2 – типы игроков (стоимость производства единицы продукции).

Участники взаимодействия стремятся максимизировать свои функции ожидаемой полезности:

$$u_1(x_1, x_2, \theta_1) \rightarrow \max_{x_1},$$

$$u_2(x_1, x_2, \theta_2) \rightarrow \max_{x_2}.$$

В модели стоимость производства единицы продукции является не постоянной величиной, а может изменяться в зависимости от множества факторов, таких как технологии производства, уровень заработной платы персонала, стоимость сырья, комплектующих и т.д.

Найдем стратегии, полученные в виде числовых значений при усреднении функций ожидаемой полезности по непрерывным случайным параметрам (случай I). Трактовать данную ситуацию взаимодействия будем таким образом, что на момент начала производства предприятие либо не будет иметь точной информации о себестоимости продукции, либо использование стратегий-функций для определения плана производства по каким-либо причинам недопустимо. Функции ожидаемой полезности в данном случае имеют вид:

$$\int_{\theta_1}^{\bar{\theta}_1} \frac{a - x_1 - x_2}{b} x_1 - \theta_1 x_1 dP_1(\theta_1) \rightarrow \max_{x_1},$$

$$\int_{\theta_2}^{\bar{\theta}_2} \frac{a - x_1 - x_2}{b} x_2 - \theta_2 x_2 dP_2(\theta_2) \rightarrow \max_{x_2}.$$

Предположим, что случайные величины θ_1 , θ_2 распределены равномерно на интервалах $[\underline{\theta}_1, \bar{\theta}_1]$, $[\underline{\theta}_2, \bar{\theta}_2]$ соответственно, т.е. $p_1(\theta_1) = \frac{1}{\bar{\theta}_1 - \underline{\theta}_1}$, $p_2(\theta_2) = \frac{1}{\bar{\theta}_2 - \underline{\theta}_2}$. Выполнение условий оптимальности первого порядка приводит к необходимости решения системы уравнений:

$$(11) \quad \begin{cases} \frac{a - 2x_1 - x_2}{b} - \int_{\underline{\theta}_1}^{\bar{\theta}_1} \theta_1 p_1(\theta_1) d\theta_1 = 0, \\ \frac{a - x_1 - 2x_2}{b} - \int_{\underline{\theta}_2}^{\bar{\theta}_2} \theta_2 p_2(\theta_2) d\theta_2 = 0. \end{cases}$$

Пусть случайные параметры имеют следующие множества допустимых значений: $\theta_1, \theta_2 \in [1,28; 3,01]$, константы $a = 9500$, $b = 350$. Определяя решение системы уравнений (11), получаем оптимальные стратегии производителей: $x_1 = x_2 = 2916,42$. Ожидаемые прибыли равны: $u_1 = u_2 = 24301,39$. Суммарная общая прибыль следующая: $U = u_1 + u_2 = 48602,35$.

Рассмотрим случай, где предприятиям на момент начала выпуска может стать доступна некоторая дополнительная информация о существенных параметрах системы (случай II). В частности, на момент реализации решения производителям будет известна себестоимость производства собственной продукции. При этом первый производитель не информирован и не будет информирован о точном значении параметра θ_2 (себестоимость единицы продукции конкурента), а второй производитель не информирован и не будет информирован о точном значении параметра θ_1 . Функции ожидаемых полезностей производителей зависят как от выбора собственной стратегии, так и от выбора стратегии конкурентом. Рассматривая описанный случай взаимодействия, запишем функции ожидаемой полезности следующим образом:

$$\int_{\underline{\theta}_1}^{\bar{\theta}_1} \int_{\underline{\theta}_2}^{\bar{\theta}_2} \frac{a - x_1(\theta_1) - x_2(\theta_2)}{b} x_1(\theta_1) - \theta_1 x_1(\theta_1) dP_1(\theta_1) dP_2(\theta_2) \rightarrow \max_{x_1(\theta_1)},$$

$$\int_{\underline{\theta}_1}^{\bar{\theta}_1} \int_{\underline{\theta}_2}^{\bar{\theta}_2} \frac{a - x_1(\theta_1) - x_2(\theta_2)}{b} x_2(\theta_2) - \theta_2 x_2(\theta_2) dP_1(\theta_1) dP_2(\theta_2) \rightarrow \max_{x_2(\theta_2)}.$$

Введем условия асимметрии информированности производителей:

$$\begin{cases} \frac{\partial x_1(\theta_1)}{\partial \theta_2} = 0, \\ \frac{\partial x_2(\theta_2)}{\partial \theta_1} = 0. \end{cases}$$

В соответствии с изложенным в разделе 4 методом решения, нахождение ситуации равновесия Байеса – Нэша сводится к решению системы интегральных уравнений:

$$\begin{cases} \int_{\underline{\theta}_2}^{\bar{\theta}_2} \frac{p_2(\theta_2)}{a^{-1}b} d\theta_2 - \left(\frac{2x_1(\theta_1)}{b} + \theta_1 \right) \int_{\underline{\theta}_2}^{\bar{\theta}_2} p_2(\theta_2) d\theta_2 - \int_{\underline{\theta}_2}^{\bar{\theta}_2} \frac{x_2(\theta_2)}{b} p_2(\theta_2) d\theta_2 = 0, \\ \int_{\underline{\theta}_1}^{\bar{\theta}_1} \frac{p_1(\theta_1)}{a^{-1}b} d\theta_1 - \left(\frac{2x_2(\theta_2)}{b} + \theta_2 \right) \int_{\underline{\theta}_1}^{\bar{\theta}_1} p_1(\theta_1) d\theta_1 - \int_{\underline{\theta}_1}^{\bar{\theta}_1} \frac{x_1(\theta_1)}{b} p_1(\theta_1) d\theta_1 = 0. \end{cases}$$

Принимая во внимание свойства плотности распределения:

$$\int_{\underline{\theta}_1}^{\bar{\theta}_1} p_1(\theta_1) d\theta_1 = 1, \quad \int_{\underline{\theta}_2}^{\bar{\theta}_2} p_2(\theta_2) d\theta_2 = 1,$$

получаем следующую систему интегральных уравнений:

$$\begin{cases} x_1(\theta_1) + \frac{1}{2} \int_{\underline{\theta}_2}^{\bar{\theta}_2} x_2(\theta_2) p_2(\theta_2) d\theta_2 = \frac{a - b\theta_1}{2}, \\ x_2(\theta_2) + \frac{1}{2} \int_{\underline{\theta}_1}^{\bar{\theta}_1} x_1(\theta_1) p_1(\theta_1) d\theta_1 = \frac{a - b\theta_2}{2}. \end{cases}$$

Применяя численный метод решения, рассмотренный в разделе 5, получаем решение системы интегральных уравнений, определяющее оптимальные стратегии-функции предприятий:

$$x_1(\theta_1) = 3291,79 - 175,00\theta_1, \quad x_2(\theta_2) = 3291,79 - 175,00\theta_2.$$

Ожидаемые прибыли в данном случае равны: $u_1 = u_2 = 24323,21$,
 $U = u_1 + u_2 = 48646,42$.

Для наглядного сравнения полученных результатов приведем таблицу 1.

Таблица 1. Значения ожидаемых полезностей предприятий

| Принцип нахождения стратегий | $u_1(\cdot), u_2(\cdot)$ | $U(\cdot)$ |
|------------------------------|--------------------------|------------|
| I | 24301,39 | 48602,78 |
| II | 24323,21 | 48646,42 |

Таким образом, если при взаимодействии предприятия с заказчиком допустимо представление плана производства в виде стратегий-функций (используя информированность о значениях себестоимости продукции), то производитель имеет возможность получать дополнительную прибыль. Реализация данной ситуации также позволяет увеличить возможную общую отраслевую прибыль по сравнению с реализацией решения, получаемого в виде числовых значений при усреднении функций ожидаемой полезности по случайным параметрам системы.

7. Заключение

В работе рассмотрен подход к формализации асимметрии информированности участников социальных и экономических процессов и его применение в байесовых играх для нахождения ситуаций равновесия. Исследованы случаи сведения задач принятия решений с вероятностной неопределенностью к решению систем интегральных уравнений. Рассмотрены численные методы и программная реализация решения байесовой игры в условиях асимметрии информированности игроков. На основе изложенных методов проведено исследование взаимодействия асимметрично информированных игроков на примере конкуренции двух предприятия, производящих однотипный товар.

Представленный подход к асимметрии информированности участников игрового взаимодействия и рассмотренные методы решения игр с неполной информацией, на наш взгляд, расширяют прикладную область байесовых игр, а также являются важным инструментом для обоснования выбора правил рационального взаимодействия участников для достаточно широкого класса конфликтных ситуаций. Предложенные методы применимы при решении прикладных задач проектирования оптимальных экономических механизмов, задач теории управления, теории аукционов и других смежных дисциплин.

Перспективным для исследования задач принятия решений с асимметрией информированности представляется подробный анализ и выработка методов нахождения Парето-оптимальных стратегий игроков, исследование методов нахождения совершенного равновесия Байеса – Нэша для сигнальных игр с континуальными пространствами типов, разработка программного инструментария поддержки принятия решений в условиях асимметрии информированности игроков, а также исследование зависимости решения байесовых игр от конечных иерархий представлений игроков.

8. Приложение

Доказательство теоремы 1. Определим соответствующие вспомогательные отображения. Лучший отклик игрока i на сложившуюся информационную обстановку запишем в виде: $z(\tilde{x}_{-i}): \tilde{X}_{-i} \rightarrow \tilde{X}_i$.

$z_i(\theta_i, \tilde{x}_{-i})$ – отображение, определяющее возможные оптимальные действия \tilde{x}_i , зависящие от θ_i ($z_i(\theta_i, \tilde{x}_{-i}) \in \tilde{X}_i$).

$C = C_1 \times C_2$ – обозначим топологическое произведение пространств непрерывно дифференцируемых функций, где $\tilde{X}_1 \in C_1$, $\tilde{X}_2 \in C_2$, $\tilde{X}_1 \times \tilde{X}_2 \in \tilde{X} \in C$.

По условию постановки игры (4) типы игроков распределены на пространстве $\Theta \in R^2$, конкретные значения

возможных действий игроков являются элементами пространств $X_1 \in R, X_2 \in R$. Рассматривается несингулярный случай, поэтому функции ожидаемой полезности вида (3) являются непрерывными на всей области определения. Таким образом, пункты 2–4 рассматриваемой теоремы для задачи (4) выполняются. Пункт 1 теоремы выполняется, так как представления i -го игрока о типах других игроков непрерывны по конкретному частному информационному параметру, известному i -му игроку. Информационные параметры, определяющие типы игроков, являются независимыми случайными величинами. Предполагается, что представления игроков о типах друг друга непрерывны по соответствующим типам.

Условия 2, 4 теоремы означают, что $\forall \theta_i \in \Theta_i$ и $\forall \tilde{x}_{-i} \in \tilde{X}_{-i}$ нахождение решения выражения (4) включает оптимизацию непрерывной функции по не пустому, выпуклому, компактному пространству, а условия 1 и 3 предполагают, что целевые функции непрерывны по параметрам θ_i ($i = \{1, 2\}$).

Для выполнения условия 5 требуется, чтобы целевая функция была унимодальна по всем переменным. Так как \tilde{X}_i выпуклое, то оптимальное действие единственно $\forall \tilde{x}_i \in \tilde{X}_i$ и $\forall \tilde{x}_{-i} \in \tilde{X}_{-i}$. Таким образом, $\forall \tilde{x}_{-i} \in \tilde{X}_{-i}$ лучший отклик игрока i – непрерывная функция, зависящая от θ_i , и решение выражения (4) – непрерывная функция $z_i(\theta_i, \tilde{x}_{-i}): \Theta_i \rightarrow X_i$.

Представим $z_i(\tilde{x}_{-i}): \tilde{X}_{-i} \rightarrow \tilde{X}_i$ в виде $z_i(\tilde{x}_{-i}): \tilde{X}_{-i} \rightarrow C_i$. Следовательно, неподвижная точка оператора $z(\tilde{x}): C \rightarrow C$ в (4) определяет ситуацию равновесия Байеса – Нэша.

Теорема Шаудера о неподвижной точке [19] гласит: оператор $z(\tilde{x}): T \rightarrow T$ имеет неподвижную точку, если:

- 1) T – не пустое, замкнутое, ограниченное, выпуклое подмножество банахова пространства;
- 2) $z(\tilde{x}): T \rightarrow T$ – компактный оператор.

Выполнение условия 1 теоремы Шаудера вытекает из свойств пространства C .

Рассмотрим подробнее пункт 2 теоремы Шаудера. По определению оператор компактен, если он непрерывен и отображает ограниченные множества в относительно компактные множества. По теореме Арцела – Асколи множество $\tilde{X} \subset C$ относительно компактно тогда и только тогда, когда для семейства функций $x(\theta_i) \in \tilde{X}$ выполняется равномерная ограниченность и равностепенная непрерывность.

Так как рассматривается оператор, действующий из одного банахова пространства в другое, то достаточно доказать, что $z(\tilde{x})$ вполне непрерывный оператор. Применение условий асимметрии информированности (2) сводит задачу (4) к рассмотрению оператора вида:

$$X(\theta_i) = \int_{\Theta_i} K(\theta_i, \theta_j) X(\theta_j) d\theta_j,$$

где

$$X(\theta_i) = \begin{pmatrix} x_1(\theta_i) \\ x_2(\theta_i) \end{pmatrix}, K(\theta_i, \theta_j) = \begin{pmatrix} K_1(\theta_i, \theta_j) & 0 \\ 0 & K_2(\theta_i, \theta_j) \end{pmatrix}, f = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}.$$

В силу условий 1, 4 теоремы $K(\theta_1, \theta_2)$ – непрерывное на квадрате $\{\underline{\theta} \leq \theta_1, \theta_2 \leq \bar{\theta}\}$ ядро. $x(\theta_i)$ – ограниченное множество функций из $C_{[\underline{\theta}, \bar{\theta}]}$, $\|x_i\| = \sup_{\theta_i \in \Theta_i} |x(\theta_i)| \leq k$.

Покажем, что данный оператор вполне непрерывен.

Докажем равностепенную непрерывность семейства функций $x(\theta_i)$. В силу равномерной непрерывности ядра $K(\theta_1, \theta_2) \forall \varepsilon > 0$ найдется такое $\delta = \delta(\varepsilon) > 0$, что

$$|K(\theta_1^1, \theta_2) - K(\theta_1^0, \theta_2)| < \frac{\varepsilon}{k(\bar{\theta} - \underline{\theta})}.$$

При $|\theta_1^1 - \theta_1^0| < \delta$ и любых $\theta_2 \in [\underline{\theta}, \bar{\theta}]$, где θ_1^1, θ_1^0 – фиксированные значения параметра θ_1 . Поэтому как только $|\theta_1^1 - \theta_1^0| < \delta$, так сразу для всех функций $x(\theta_i)$

$$|X(\theta_i^1) - X(\theta_i^0)| \leq \int_{\Theta_i} |K(\theta_i^1, \theta_j^1) - K(\theta_i^0, \theta_j^0)| X(\theta_i) d\theta_i < \varepsilon.$$

Это и означает равномерную непрерывность семейства функций $x(\theta_i)$.

Теперь покажем, что функции

$$X(\theta_i) = \int_{\Theta_i} K(\theta_i, \theta_j) X(\theta_i) d\theta_j$$

равномерно ограничены.

Так как по условию постановки *теоремы* X_i и Θ_i компактны, то по теореме Вейерштрасса любая непрерывная вещественная функция на компактном пространстве ограничена и достигает своих наибольших и наименьших значений. Действительно, для любых $\theta_1, \theta_2 \in [\underline{\theta}, \bar{\theta}]$

$$|x(\theta_i)| \leq Mk(\bar{\theta} - \underline{\theta}), \text{ где}$$

$$M = \max_{a \leq \theta_1, \theta_2 \leq b} |K(\theta_1, \theta_2)|.$$

По определению оператор является вполне непрерывным, если он отображает всякое ограниченное множество банахова пространства в относительно компактное множество банахова пространства. Следовательно, по теореме Арцела-Асколи, множество $\tilde{X} \subset C$ относительно компактно. То есть оператор $z(\tilde{x})$ – компактный оператор.

Так как мы показали, что $z(\tilde{x}): C \rightarrow C$ – компактный и непрерывный оператор и $\forall i \in N$ C_i является замкнутым, ограниченным и выпуклым пространством из произведения пространств C , то по теореме Шаудера, неподвижная точка в $z(\tilde{x}): C \rightarrow C$ существует. Что и требовалось доказать.

Литература

1. АЛГАЗИН Г.И., АЛГАЗИНА Д.Г. *Моделирование сетевого взаимодействия на конкурентных рынках* // Управление большими системами. – 2013. – №43. – С. 172–216.
2. БЕЛЯЕВА М.А., БУРЕШ О.В., ШАТАЛОВА Т.Н. *Разработка интегрированной системы поддержки принятия решений по управлению проектами в условиях неопределенности* // Вестник Оренбургского государственного университета. – 2011. – №13(132). – С. 43–48.
3. БУРКОВ В.Н., ЕНАЛЕЕВ А.К., НОВИКОВ Д.А. *Механизмы стимулирования в вероятностных моделях социально-экономических систем* // Автоматика и телемеханика. – 1993. – №11. – С. 3–30.
4. БУРКОВ В.Н., КОРГИН Н.А., НОВИКОВ Д.А. *Введение в теорию управления организационными системами*. – М.: Либроком, 2009. – 264 с.
5. ГОРЕЛИК В.А., ГОРЕЛОВ М.А., КОНОНЕНКО А.Ф. *Анализ конфликтных ситуаций в системах управления*. – М.: Радио и связь, 1991. – 288 с.
6. ГУБКО М.В., НОВИКОВ Д.А. *Теория игр в управлении организационными системами*. – М.: Синтег, 2002. – 148 с.
7. ЖАРИКОВ А.В. *Равновесие Нэша в игре двух лиц для вариантов информированности игроков* // Известия Алтайского государственного университета. – 2007. – №1/2(73). – С. 55–59.
8. КОРГИН Н.А., НОВИКОВ Д.А. *Задача стимулирования в условиях внутренней неопределенности о типах агентов, описываемых распределением Парето* // Системы управления и информационные технологии. – 2006. – №4(26). – С. 66–69.
9. МАКСИМОВ А.В. *Расчет двухканальной системы в условиях несовпадающей информированности* // Класс управления, использующий принцип противоречия, его проявления в живых системах и возможности применения в технике: Тезисы докладов научно-практической конференции. Иркутск: СЭИ СО РАН, 1984. – С. 75–77.
10. МАКСИМОВ А.В., ОСКОРБИН Н.М. *Многопользовательские информационные системы: основы*

теории и методы исследования: монография. – Барнаул: Изд-во Алт. ун-та, 2005. – 250 с.

11. МАМЧЕНКО О.П., ОСКОРБИН Н.М. *Моделирование иерархических систем: учебник для вузов.* – Барнаул: Изд-во Алт. ун-та, 2007. – 317 с.

12. НОВИКОВ Д.А. *Задача стимулирования Парето-агента // Автоматика и телемеханика.* – 2007. – №1. – С. 137–146.

13. НОВИКОВ Д.А. *Теория управления организационными системами. 3-е изд.* – М.: Издательство физико-математической литературы, 2012. – 604 с.

14. НОВИКОВ Д.А., ПЕТРАКОВ С.Н. *Курс теории активных систем.* – М.: Синтег, 1999. – 108 с.

15. НОВИКОВ Д.А., ЧХАРТИШВИЛИ А.Г. *Рефлексивные игры.* – М.: Синтег, 2003. – 149 с.

16. НОВИКОВ Д.А., ЧХАРТИШВИЛИ А.Г. *Рефлексия и управление: математические модели.* – М.: Издательство физико-математической литературы, 2013. – 412 с.

17. ПОЛЯНИН А.Д., МАНЖИРОВ А.В. *Справочник по интегральным уравнениям.* – М.: ФИЗМАТЛИТ, 2003. – 384 с.

18. ФЕДЯНИН Д.Н., ЧХАРТИШВИЛИ А.Г. *Модель информационного управления в активных сетевых структурах при неполной информированности центра // Проблемы управления.* – 2012. – №6. – С. 13–18.

19. ХАТСОН В., ПИМ ДЖ. *Приложения функционального анализа и теории операторов.* – М.: Мир, 1983. – 432 с.

20. ЧХАРТИШВИЛИ А.Г. *Рефлексивная игра «аккордная оплата труда» // Управление большими системами.* – 2004. – №6. – С. 143–149.

21. ЧХАРТИШВИЛИ А.Г. *Теоретико-игровое моделирование информационного управления // Управление большими системами.* – 2006. – №12-13. – С. 161–171.

22. BERGE S. *Topological spaces.* – New York: Macmillan Co., 1963. – 91 p.

23. BOLTON P., FAURE-GRIMAUD A. *Satisficing contracts // Review of Economic Studies.* – 2010. – Vol. 77(3). – P. 937–971.

24. HAN Z., NIYATO D., SAAD W. et al. *Game theory in wireless and communication Networks: Theory, models, and application.* – Cambridge: University Press, 2012. – 554 p.

25. HARSANYI J.C. *Games with incomplete information played by "Bayesian" players* // Management Science. – Part I: 1967. – Vol. 14, No. 3. – P. 159–182; Part II: 1968. – Vol. 14, No. 5. – P. 320–334; Part III: 1968. – Vol. 14, No. 7. – P. 486–502.
26. HOLMSTROM B., MILGROM P. *Multi task principal–agent analyses: incentive contracts, asset ownership, and job design* // Journal of Law, Economics, and Organization. – 1991. – №7. – P. 24–52.
27. LELAND H.E., PYLE D.H. *Informational asymmetries, financial structure, and financial intermediation* // The Journal of Finance. – 1977. – Vol. 32, №2. – P. 371–387.
28. MYERSON R.B. *Probability Models for Economic Decisions*. – CA.: Duxbury Press, 2005. – 354 p.
29. NEEMAN Z., PAVLOV G. *The value of information in a principal-agent model with moral hazard: The ex post contracting case* // Games and Economic Behavior. – 2012. – Vol. 74, Iss. 1. – P. 352–365.
30. RASMUSEN E. *Games and information. An introduction to game theory. Fourth edition*. – Oxford: Blackwell Publishers, 2006. – 484 p.
31. SPENCE M. *Signaling in retrospect and the informational structure of markets* // The American Economic Review. – 2002. – Vol. 92, №3. – P. 434–459.
32. TIROLE J. *The theory of industrial organization*. – MA.: MIT Press, 1988. – 479 p.
33. WU H., PARLAR M. *Games with incomplete information: A simplified exposition with inventory management applications* // International Journal of Production Economics. – 2011. – Vol. 133, Iss. 2. – P. 562–577.

OPTIMAL STRATEGIES OF PLAYERS UNDER ASYMMETRIC INFORMATION

Gennady Algazin, Altai State University, Barnaul, Doctor of Science, professor (algaz46@yandex.ru).

Eugeny Matyunin, MEM LLC, Barnaul, chief engineer (matyuninev@gmail.com).

Abstract: Optimal decisions of players under asymmetric information and probabilistic uncertainty are studied using the concept of the Bayesian Nash equilibrium. Calculation of the Bayesian Nash equilibrium for the class of games considered is reduced to solving a system of integral equations. We suggest numerical methods and provide the algorithm to find the solution in such games.

Keywords: asymmetric information, probabilistic uncertainty, Bayesian games, Bayesian Nash equilibrium, calculus of variations, numerical calculation.

*Статья представлена к публикации
членом редакционной коллегии А.Г. Чхартишвили*

*Поступила в редакцию 10.06.2015.
Опубликована 30.11.2015.*

УДК 517.977
ББК 32.96

СОВМЕЩЕННЫЕ СЕТИ УПРАВЛЕНИЯ И ДАННЫХ

Выхованец В. С.¹

(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, Москва)

Крыжановская А. В.²

(ОАО «Яндекс», Москва)

Рассмотрено динамическое управление процессами на основе использования совмещённых сетей управления и данных. Благодаря заданию на диаграмме процесса потока данных и потока управления появляется возможность автоматической генерации управляющей программы процесса. Для спецификации типов данных в совмещенной сети управления и данных разработана теория типов, позволяющая представлять данные в виде вложенных друг в друга именованных списков. Показано, что совмещенная сеть управления и данных является универсальной алгоритмической моделью со строгой типизацией данных. Доказано, что синтез такой сети по описанию выполняемого ею преобразования типов данных является алгоритмически неразрешимой проблемой. Использование совмещенных сетей управления и данных позволяет повысить качество управления технологическими, производственными и организационными процессами на предприятиях.

Ключевые слова: моделирование процессов, динамическое управление процессами, совмещенные сети управления и данных, алгоритмические модели, теория типов, неразрешим-

¹ Валерий Святославович Выхованец, доктор технических наук, доцент (valery@vykhovanets.ru).

² Александра Валерьевна Крыжановская, руководитель проектов (aleksundra@gmail.com).

мые проблемы, универсальная алгоритмическая модель со строгой типизацией данных.

1. Введение

Процессно-ориентированный подход, заключающийся в представлении объекта управления как многоуровневого набора взаимосвязанных процессов, широко используется в современных системах технологического, производственного и организационного управления.

Входами процесса обычно являются выходы других процессов. Процесс описывается диаграммой, представляющей собой сеть, состоящую из процессов (узлов) и связей управления (дуг), задающих частичный порядок выполнения подчинённых процессов [11]. Данные процессов, как правило, не отражаются на диаграммах и связываются с процессами на этапе реализации с помощью атрибутов процесса, через специальные объекты внутри процесса, путем адресной передачи и приема данных через общую для всех процессов шину данных. Такая организация работы с данными исторически наследуется от самых первых методологий анализа и обусловлена тем, что выделение потока данных на этапе моделирования сложно реализовать и методически, и технически [5]. Отсутствие явно задаваемого потока данных приводит к невозможности автоматического создания программ управления процессами, а также затрудняют актуализацию уже запущенных программ.

Целью настоящей статьи является улучшение качества управления процессами путём использования такой их модели, которая основана на явной трассировке данных и управления между процессами. Это позволяет решить проблемы автоматической генерации управляющей программы по её описанию в виде иерархически организованных диаграмм процессов, а также улучшает выразительные возможности модели.

2. Проблемы моделирования процессов

Современное моделирование процессов возникло на основе методологии структурного анализа и проектирования.

2.1. СТРУКТУРНОЕ МОДЕЛИРОВАНИЕ

Методологии структурного анализа и проектирования SADT (Structured Analysis and Design Technique) [5] и стандарт IDEF0 (Integrated Computer-Aided Manufacturing DEfinition) [13] предназначены для описания процессов в функциональном аспекте: они состоят из диаграмм, текстов и глоссария, имеющих ссылки друг на друга.

Диаграммы – главные компоненты модели, все функции и интерфейсы в них представлены как узлы (функциональные блоки) и дуги. Каждый узел может быть декомпозирован на другой диаграмме. Следует отметить, что нотации в методологии SADT и IDEF0 предназначены для описания только состава процессов, а не их последовательности.

2.2. ПОТОКИ ДАННЫХ

Для частичного устранения недостатков IDEF0 используется описание потоков данных в нотации DFD (Data Flow Diagram) [15]. Цель такого представления – продемонстрировать, как каждый процесс преобразует свои входные данные в выходные. Основными компонентами диаграмм потоков данных являются внешние и внутренние сущности, процессы, накопители данных и потоки данных. Каждый процесс может быть детализирован при помощи другой диаграммы или спецификации. Языки спецификации, как правило, не стандартизованы и могут варьироваться от естественного языка до визуальных языков моделирования.

Дальнейшее развитие методологии IDEF, а именно стандартов IDEF1 (методология моделирования информационных потоков внутри системы) и IDEF1X (методология построения реляционных структур данных), преследовало цель устранить недостатки, присущие моделям потока данных.

2.3. ДИНАМИЧЕСКИЕ МОДЕЛИ

Модели в ранее рассмотренных нотациях являются статическими по определению и не позволяют отразить функционирование моделируемых объектов во времени. Проблему создания динамических диаграмм пытались решить с помощью IDEF2 (Simulation Model Design) – методологии динамического

моделирования развития. Однако ввиду сложности использования этой методологии разработка соответствующего стандарта быстро прекратилась.

Следующим шагом в моделировании процессов стала разработка и использование стандарта IDEF3 [12], предназначенного для описания потоков работ. Модели IDEF3 применяли для детализации функциональных блоков IDEF0, не имеющих диаграмм декомпозиции. Выразительные средства стандарта IDEF3 оказались близки алгоритмическим блок-схемам.

2.4. ПОТОКИ РАБОТ

Необходимость не только описать процессы, но и полностью или частично их автоматизировать привела к появлению концепции потока работ Workflow [14], согласно которой данные и задания передаются от одного исполнителя другому для выполнения определенных действий в соответствии со сводом процедурных правил.

Данные в Workflow не перемещаются вместе с управлением, а содержатся в глобально доступных переменных и локальных переменных блоков. Последнее вызывает значительные трудности контроля и синхронизации данных.

2.5. ИНТЕГРИРОВАННЫЕ МЕТОДОЛОГИИ

Неоднозначная семантика и низкая выразительность диаграмм в стандартах IDEF и DFD привела к объединению известных методов моделирования в форме создания интегрированных методологий, одной из которых является ARIS (Architecture of Integrated Information Systems) [11]. Модель процесса в ARIS является в некотором смысле расширением IDEF3 и представляет собой поток последовательно выполняемых работ, расположенных в порядке их выполнения.

Ввиду многообразия выразительных средств в ARIS введен механизм методологических фильтров, позволяющих в рамках одного процесса применять только определенный набор объектов и связей. Для разработки таких фильтров требуются значительное количество времени и разработчик высокой квалификации. В конечном итоге методология ARIS дает возможность описать процессы, провести анализ полученных моделей

и определить требования к управляющей информационной системе. Однако эта методология не обеспечивает полного, корректного и однозначного описания моделируемых процессов.

2.6. ДИНАМИЧЕСКОЕ УПРАВЛЕНИЕ

Негибкость создаваемых моделей, их неспособность обеспечить оперативное реагирование на постоянные изменения в среде являются основными недостатками методологий SADT, IDEF, DFD, ARIS и др., которые стимулировали разработку концепции следующего поколения – BPM (Business Process Management) [17].

В методологии BPM на смену радикальному реинжинирингу процессов приходит динамическое управление процессами, заключающееся в возможности корректировки уже автоматизированных процессов в ответ на изменения в структуре и организации предприятия. В этом случае инструменты моделирования позволяют создавать и внедрять новые процессы «на лету», а само моделирование больше похоже на моделирование в Workflow.

Для разработки исполняемых моделей в BPM используют нотацию BPMN (Business Process Model and Notation) и соответствующие языки моделирования: язык моделирования процессов BPML (Business Process Modeling Language), язык исполнения процессов BPEL (Business Process Execution Language) и язык определения процессов XPDЛ (XML Process Definition Language). Однако построение моделей непосредственно на этих языках неудобно для разработчиков по причине их синтаксической и семантической сложности. По этой причине наибольшую сложность для разработчиков программных средств в методологии BPM является конвертация моделей процессов в их исполняемые формы.

До настоящего времени применение BPEL в основном заключается в автоматизированном создании шаблона для «ручной доводки» модели до исполняемого вида. В частных случаях, когда доступны исполняемые формы всех используемых процессов, а также согласованы и специфицированы механизмы передачи между ними данных, возможно автоматическое созда-

ние исполняемой модели составного процесса. Однако для массового распространения методологии BPM пока не хватает технологических возможностей, позволяющих создавать и автоматически исполнять модели автоматизируемых процессов.

2.7. ВЫВОДЫ

Таким образом, блок-схема – распространенный вид графических моделей, описывающих алгоритмы или процессы, в которых отдельные шаги или подпроцессы изображаются в виде блоков, соединенных между собой линиями, указывающими последовательность их выполнения. Однако попытка реализации процессов с помощью блок-схем не увенчалась успехом, так как известные их нотации не обладают возможностями универсальных алгоритмических моделей.

По этой причине основной проблемой современных средств моделирования процессов является наличие труднопреодолимого семантического разрыва между графической (аналитической) и исполняемой формами процесса. Семантический разрыв проявляется в том, что понятия, объекты и структуры данных, которыми оперирует аналитик, не совпадают, а порой даже не согласуются с понятиями, объектами и структурами данных, которые должен использовать разработчик исполняемой модели процессов. Для преодоления указанного семантического разрыва требуется обеспечить безусловное автоматическое преобразование исходной графической модели процесса в его исполняемую форму.

3. Совмещенная сеть управления и данных

Появление проблем, свойственных современным средствам автоматизированного управления технологическими, производственными и организационными процессами, следует связать с недостаточностью данных, отражаемых на графических моделях процессов, их нечеткой пространственно-временной структурированностью и несогласованностью типов.

Для преодоления этих проблем предлагается использовать графические модели процессов, содержащие не только связи процессов по управлению, но и источники данных и пути их

передачи. Предлагается объединить выразительные возможности двух моделей – моделей для описания потоков работ и моделей для описания потоков данных [8, 19].

3.1. ГРАФИЧЕСКАЯ НОТАЦИЯ

Графическая нотация, основанная на представлении процесса в виде совмещенной сети управления и данных, позволяет явно указать на одной диаграмме как потоки управления, так и потоки данных, связанные с моделируемым процессом (рис. 1).

Через каждый узел сети (показан в виде прямоугольника) проходят поток управления и поток данных. Поток управления, входящий в узел, называется активатором (показан в виде ромба). Получение активации означает начало работы узла. Выход потока управления из узла называется событием (показан в виде квадрата). Событие, созданное узлом, может активировать один или несколько других узлов.

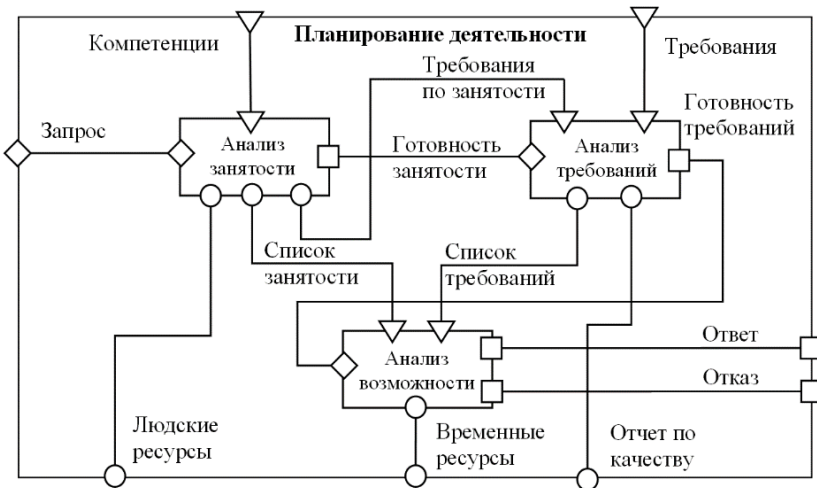


Рис. 1. Совмещенная сеть управления и данных

Потоки данных, входящие в узел (показаны в виде треугольников), передают узлу структурированные входные данные. После того как узел закончил работу, в исходящие потоки данных (показаны в виде окружностей) поступают структуриро-

ванные выходные данные, которые могут передаваться следующему узлу в качестве входных.

При срабатывании одного или нескольких активаторов узел активируется, считываются входные данные и выполняются действия, привязанные к этому узлу. Затем формируются одно или несколько событий и выходные данные.

Часть узлов, которые не имеют диаграммы, называются базовыми. Составной узел отличается от базового узла тем, что не является последним в иерархии, т.е. декомпозируется в совмещенную сеть и состоит из других узлов, каждый из которых, в свою очередь, может быть либо базовым, либо составным.

3.2. ИСПОЛНЕНИЕ СОВМЕЩЕННЫХ СЕТЕЙ

К каждому базовому узлу привязывается вычислительная семантика в виде последовательности действий (операций) на языке некоторого внешнего интерпретатора. В качестве интерпретаторов могут выступать компиляторы или интерпретаторы языков программирования, различные виртуальные машины, вызовы внешних программ с передачей им входных и получением от них выходных данных, какие-либо исполнительные устройства.

В свою очередь, составной узел строится из других узлов, а его вычислительная семантика однозначно определяется внутренними узлами и соединяющими их дугами.

Базовые узлы выполняются непосредственно тем интерпретатором, который определён во время их конструирования, а их вычислительная семантика задаётся явно. Исполнение составного узла – это исполнение всех его внутренних узлов.

При активации одного или нескольких активаторов составного узла в соответствии с его внутренними связями активизируются внутренние узлы. После окончания функционирования внутренних узлов их события по внутренним связям активизируют другие внутренние узлы. Функционирование составного узла завершается, когда все внутренние узлы перестают активироваться и заканчивают свое функционирование.

Аналогично активации происходит передача данных между внутренними узлами и формирование выходных данных составного узла.

3.3. СИНХРОНИЗАЦИЯ И РАСПАРАЛЛЕЛИВАНИЕ

Нотации многих известных методологий моделирования процессов включают в себя специальные элементы для явной (статической) организации их параллельного выполнения. Для этого, например, используются такие элементы как «логическое И», многоэкземплярные циклы, шлюз «ИЛИ» с неэксклюзивным условием и т.п. [11].

Моделирование процессов совмещёнными сетями управления и данных позволяет реализовать динамическое (неявное) распараллеливание, заключающееся в том, что в каждый конкретный момент времени и в зависимости от предыстории развития моделируемого процесса принимается решение о том, какие внутренние процессы подлежат параллельному исполнению, а какие нет.

Может оказаться так, что при одном и том же описании процесса в один момент времени параллельному исполнению подлежит одно подмножество внутренних процессов, а в другой момент времени – другое.

Таким образом, совместная и отдельная трассировка данных и управления между процессами предоставляет гибкий механизм синхронизации моделируемых процессов, позволяет реализовать естественное их распараллеливание и эффективное выполнение.

4. Типизация данных

В известных подходах к моделированию процессов структуры и типы данных возникают только на заключительном этапе реализации модели и входят в зону ответственности разработчика исполняемой программы.

Использование явной трассировки данных на диаграммах процессов требует применения достаточно развитого аппарата структуризации, типизации, объединения, сравнения, разделения и контроля данных на всех этапах работы с моделью процесса. По этой причине рассмотрим формальную теорию типов, необходимую для описания потоков данных в совмещённых сетях.

4.1. ТЕОРИЯ ТИПОВ

Зададим алфавит теории типов T , состоящий из конечного множества знаков простых типов $\Omega = \{\tau_1, \tau_2, \dots, \tau_m\}$ и фигурных скобок. Определим перечислимое множество формул теории следующим образом:

- знак простого типа данных $\tau \in \Omega$ является типом;
- если φ – тип, то и $\{\varphi\}$ является типом;
- если φ и ψ – типы, то и $\varphi\psi$ – тип.

Выводом в теории T является порождение нового типа из имеющихся типов-посылок с помощью правил вывода вида

- (1) $\varphi \Rightarrow \{\varphi\}$;
- (2) $\varphi, \psi \Rightarrow \varphi\psi$,

где \Rightarrow – знак вывода (подстановки).

Правило (1) задает специализацию типа, а правило (2) – агрегацию типов. При специализации из исходного типа создается новый тип, рассматриваемый как выделенный. При агрегации из двух типов создается новый тип, включающий в себя агрегируемые типы как составные части – подтипы.

Пример 1. Пусть задано множество простых типов $\Omega = \{b\}$, где b – логический тип, принимающий значения на множестве из двух элементов $\{0, 1\}$. Тогда с помощью правил вывода могут быть получены следующие типы:

$$c = \underbrace{bb \dots b}_{16} = b^{16} \text{ – тип символа с разрядностью 16;}$$

$$i = b^{32} \text{ – целочисленный тип с разрядностью 32;}$$

$f = b^{53}\{b^{11}\}$ – тип чисел с плавающей запятой, где 53 разряда отводится под мантиссу, а 11 разрядов – под порядок числа;

$$s = c^{256} \text{ – строковый тип с длиной строк 256 знаков;}$$

$$p = b^{1024} \text{ – тип битовых последовательностей данных с}$$

длиной 1024 разряда. ♦

Как видно из примера 1, для порождения множества типов достаточно одного простого типа.

Аксиоматика теории типов состоит из следующих аксиом:

- (3) $\exists \tau (\tau = \tau)$;

- (4) $\forall \varphi \forall \psi \exists \chi (\chi = \varphi \psi)$;
 (5) $\forall \varphi \forall \psi (\psi \varphi = \varphi \psi)$;
 (6) $\forall \varphi \forall \psi (\varphi \psi > \varphi \wedge \varphi < \varphi \psi)$;
 (7) $\forall \varphi \exists \chi (\chi = \{\varphi\} \wedge \neg(\chi = \varphi) \wedge \neg(\chi > \varphi) \wedge \neg(\chi < \varphi))$,

где для выражения аксиом использовано исчисление предикатов с кванторами существования \exists и всеобщности \forall , логической связкой «И» \wedge , логической связкой «НЕ» \neg , а также двуместными предикатами сравнения типов: равно $=$, больше $>$, меньше $<$. Круглые скобки использованы для выделения высказываний, на которые действуют стоящие перед ними кванторы.

Аксиома существования (3) утверждает, что существует хотя бы один тип, который равен самому себе. В свою очередь аксиома агрегации (4) утверждает существование типов-агрегатов и порождает натуральные числа вида τ , $\tau\tau$, $\tau\tau\tau$ и так далее. Аксиома равенства (5) разделяет типы на классы эквивалентности, а аксиома порядка (6) задает некоторый порядок типов. И, наконец, аксиома специализации (7) позволяет выделить (специализировать) любой тип таким образом, что он становится не равным себе и не сравнимым с собой. По своей сути аксиома специализации утверждает существование структурированных типов.

4.2. ОСНОВНЫЕ ТЕОРЕМЫ

Определение 1. Два типа φ и ψ называются несравнимыми, если и только если неверно, что они равны, неверно, что φ больше ψ , и неверно, что φ меньше ψ :

$$\varphi \# \psi \leftrightarrow \neg(\varphi = \psi) \wedge \neg(\varphi > \psi) \wedge \neg(\varphi < \psi),$$

где $\#$ – двуместный предикат несравнимости типов; \leftrightarrow – логическая связка двухстороннего следования.

Теорема 1. Существуют несравнимые типы, т.е.

$$\exists \varphi \exists \psi (\neg(\varphi > \psi) \wedge \neg(\varphi < \psi) \wedge \neg(\varphi = \psi)).$$

Доказательство. Пусть $\varphi = \tau$. Существование типа φ следует из аксиомы (3). Тогда из аксиомы (7) следует существование типа $\psi = \{\tau\}$ такого, что $\neg(\varphi = \psi)$, $\neg(\varphi > \psi)$ и $\neg(\varphi < \psi)$. ♦

Теорема 2. Любая перестановка подтипов φ_i ($i = 1, \dots, N$) типа $\varphi = \varphi_1 \varphi_2 \dots \varphi_N$ является тождественным преобразованием

ем этого типа, т.е. для любой последовательности неповторяющихся индексов i, j, \dots, k таких, что $1 \leq i, j, \dots, k \leq N$ и $i \neq j \neq \dots \neq k$ выполнимо $\varphi_1 \varphi_2 \dots \varphi_N = \varphi_i \varphi_j \dots \varphi_k$.

Доказательство. Пусть $\varphi = \varphi_1 \dots \psi \dots \varphi_N$, где $\psi = \varphi_i \varphi_{t+1}$, $1 \leq t < N$. Из аксиомы (5) находим, что $\psi = \varphi_{t+1} \varphi_t$. Следовательно, перестановка двух соседних подтипов является тождественным преобразованием типа. Пусть теперь $i \neq j \neq \dots \neq k$ и $1 \leq i, j, \dots, k \leq N$. Тогда (i, j, \dots, k) есть перестановка чисел от 1 до N . Известно, что любая перестановка может быть получена путем перестановки соседних элементов. Отсюда получаем утверждение теоремы. ♦

Теорема 3. Два типа φ и ψ равны тогда и только тогда, когда существует такая перестановка подтипов типа φ , которая преобразует его в тип ψ .

Доказательство. В соответствии с аксиомой (4) представим типы φ и ψ как агрегации одинакового числа подтипов: $\varphi = \varphi_1 \varphi_2 \dots \varphi_N$ и $\psi = \psi_1 \psi_2 \dots \psi_N$. Это можно сделать всегда и, возможно, многими способами.

Пусть типы φ и ψ равны. Тогда из теоремы 2 следует, что существует агрегация подтипов этих типов и перестановка (i, j, \dots, k) подтипов типа φ такие, что $\psi_1 = \varphi_i$, $\psi_2 = \varphi_j$, ..., $\psi_N = \varphi_k$. Доказано прямое утверждение теоремы.

Верно и обратное утверждение. Пусть существует перестановка (i, j, \dots, k) подтипов типа φ такая, что $\psi_1 = \varphi_i$, $\psi_2 = \varphi_j$, ..., $\psi_N = \varphi_k$. Тогда в соответствии с теоремой 2 типы φ и ψ равны. Что требовалось доказать. ♦

Следует обратить внимание на рекурсивный характер определения равенства типов. Если некоторый подтип φ_i является специализацией других подтипов, то соответствующий ему тип ψ_j также должен быть специализацией, а сравнение таких подтипов должно осуществляться отдельно и независимо.

Пример 2. Пусть заданы простые типы $\Omega = \{b, c, i, f, s, p\}$. Два типа $\varphi = f^2 s \{cbi\} p$ и $\psi = p f s \{bic\} f$ равны, так как подтип $\{bic\}$ преобразуется в подтип $\{cbi\}$ перестановкой $(2, 3, 1)$, а тип ψ преобразуется в тип φ перестановкой $(2, 5, 3, 4, 1)$. ♦

Таким образом, все типы теории типов T разбивается на классы эквивалентности. В один класс эквивалентности попа-

дают типы, заданные с точностью до перестановки агрегированных в них подтипов.

Теорема 4. Тип φ меньше типа ψ тогда и только тогда, когда существует такой тип ω , что $\psi = \varphi\omega$ или $\psi = \omega\varphi$.

Доказательство. Пусть $\varphi < \psi$. Тогда из аксиомы (6) следует существование некоторого типа ω такого, что $\psi = \varphi\omega$, а из аксиомы (5) выводится $\psi = \omega\varphi$. Пусть теперь $\psi = \varphi\omega$ или $\psi = \omega\varphi$. Из аксиом (6) и (5) следует $\varphi < \varphi\omega$ или $\varphi < \omega\varphi$. Откуда получаем $\varphi < \psi$. ♦

Пример 3. Пусть задано множество простых типов из примера 2. Тогда тип $\varphi = fs\{cbi\}p$ меньше типа $\psi = pfs\{bic\}f$, так как из тождественных им типов $fsp\{cbi\}$ и $fsp\{cbi\}f$ находим, что $\psi = \varphi f$. ♦

В итоге имеем, что типы частично упорядочены и не существует наименьшего и наибольшего типа.¹ На типах может быть задана иерархия (рис. . 2), аналогичная иерархии типов в логических теориях [9] и диаграммам наследования классов в объектно-ориентированном программировании [1].

Значением типа является значение переменной, имеющей этот тип. Для получения значения типа введем операцию разыменования, которую будем обозначать квадратными скобками.

¹ Не существование минимального типа следует из теоремы 1. Это позволяет не ограничиваться одним простым типом при построении прикладной теории типов. В свою очередь не существование наибольшего типа выводится из аксиомы агрегации и аксиомы порядка методом от противного. Это позволяет рассматривать теорию типов как конечную (финитную, конструктивную) теорию и, тем самым, избежать проблем, связанных с бесконечными построениями.

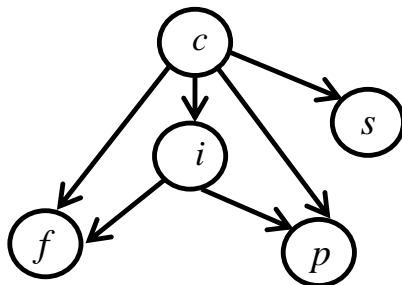


Рис. 2. Иерархия типов из примера 1

Пример 4. Пусть задан тип $c^3s\{bi\}$ над множеством простых типов из примера 2. Тогда значением $[c^3s\{bi\}]$ этого типа будет структура, состоящая из массива трех переменных с типом c , переменной типа s и структуры bi , состоящей, в свою очередь, из переменной типа b и переменной типа i .

Аналогично можно разыменовывать не весь тип, а некоторый его подтип, например, $c^3[s]\{bi\}$ – выделение из переменной типа $c^3s\{bi\}$ значения агрегированной в нее переменной типа s . ♦

4.3. РАЗРЕШИМОСТЬ ТЕОРИИ ТИПОВ

При моделировании процессов совмещенными сетями управления и данных возникает две задачи, связанные с типами: на этапе моделирования процесса – это проверка допустимости трассировки данных от одного узла сети к другому, а на этапе выполнения процесса – это проверка достаточности поступивших данных. Так как значение большего типа может быть преобразовано в значение меньшего типа, то обе эти задачи сводятся к простому сравнению типов.

Теорема 5. Сравнение типов разрешимо, т.е. для любых двух типов φ и ψ конструктивными средствами может быть установлена истинность одного и только одного из следующих высказываний: $\varphi \# \psi$, $\varphi = \psi$, $\varphi > \psi$, $\varphi < \psi$.

Доказательство. Будем рассматривать произвольный тип как конечную строку в алфавите, состоящем из знаков простых типов и фигурных скобок. Распознавание строки,

выражающей произвольный тип, как и любого скобочного выражения, выполняется магазинным автоматом [7, с. 251]. Следовательно, распознавание типов, выводимых в теории типов, разрешимо.

Пусть $\varphi = \varphi_1\varphi_2 \dots \varphi_N$ и $\psi = \psi_1\psi_2 \dots \psi_M$, где $N \leq M$, а выделенные подтипы – это простые или специализированные подтипы, отсортированные в лексикографическом порядке. Последнее возможно ввиду того, что лексикографическая сортировка строки типа является его тождественным преобразованием (теорема 3) и задача сортировки строки разрешима, например, методом «пузырьковой» сортировки [3, с. 144], для которого может быть построена соответствующая машина Тьюринга [7, с. 328].

Тогда каждый подтип φ_i ($i = 1, \dots, N$) будем проверять на равенство соответствующему ему подтипу ψ_i . Если подтип ψ_i не равен φ_i , то $\varphi \# \psi$. Если ψ_i равен φ_i , то переходим к следующему подтипу φ_{i+1} . Если все подтипы φ_i просмотрены, а в типе ψ еще остались непроверенные подтипы ψ_j ($j > N$), то $\varphi < \psi$. В противном случае $\varphi = \psi$.

Существование и единственность результата сравнения типов непосредственно следует из описания процедуры сравнения. Очевидно, для этой процедуры может быть построена машина Тьюринга. Следовательно, сравнение типов разрешимо. ♦

Таким образом, декомпозицию типов на подтипы и поиск перестановок подтипов, переводящих один тип в другой (условие теоремы 3), можно осуществить путем сортировки простых и специализированных подтипов в некотором лексикографическом порядке с последующим их поэлементным сравнением.

Пример 5. После сортировки подтипов типов φ и ψ из примера 2 в лексикографическом порядке, задаваемом, например, порядком перечисления простых типов во множестве Ω , имеем $\varphi = f^2ps\{bci\}$ и $\psi = f^2ps\{bci\}$. Откуда находим, что типы φ и ψ равны. ♦

Для представления именованных списков вида (Имя, Значение 1, Значение 2, ...), где любым из значений может быть другой именованный список, во множество простых типов Ω включается специальный тип α . Тогда любой именованный тип

φ задается как $\{\alpha^x \varphi\}$, где $x > 1$ – индекс имени типа, а именованный тип – как $\{\alpha \varphi\}$.

5. Теория совмещенных сетей

Для определения выразительных возможностей совмещенных сетей управления и данных при моделировании процессов рассмотрим их формальную теорию, наследуемую от ранее описанной теории типов.

5.1. ФОРМАЛИЗМ СОВМЕЩЕННЫХ СЕТЕЙ

Определение 2. Совмещенной сетью управления и данных S называется упорядоченное множество, состоящее из четырех подмножеств N , E , D и T :

$$(8) \quad S = (N, E, D, T),$$

где $N = \{n_i \mid i = 0, \dots, p\}$ – множество внутренних узлов сети с числом элементов p , куда включен внешний интерфейсный узел n_0 ; $E = \{(n_i, n_j, t_k) \mid n_i, n_j \in N; t_k \in T\}$ – множество дуг управления; $D = \{(n_i, n_j, t_k) \mid n_i, n_j \in N; t_k \in T\}$ – множество дуг данных; T – множество типов данных; дуги заданы упорядоченными множествами из трех элементов $(n_i, n_j, t_k) \in N \times N \times T$, где $n_i \in N$ – начальный узел дуги, $n_j \in N$ – ее конечный узел, $t_k \in T$ – тип дуги, \times – операция декартова произведения множеств.

Интерфейсный узел n_0 имеет специальное назначение (рис. 3). Дуги управления (начинаются ромбом), идущие из него, являются активаторами сети, а входящие в него – событиями. В свою очередь дуги данных (начинаются кругом), связанные с этим узлом, задают входные и выходные данные. При включении совмещенной сети в другую сеть в качестве узла она замещается своим интерфейсным узлом.

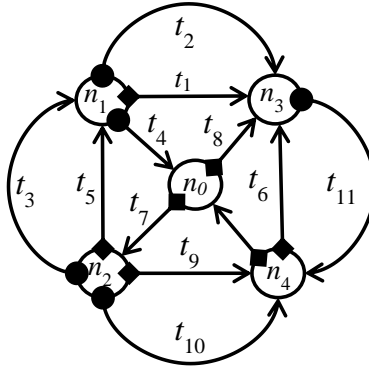


Рис. 3. Пример совмещенной сети управления и данных

Пусть задан узел n некоторой совмещенной сети, имеющий r активаторов a_1, a_2, \dots, a_r и s входных данных i_1, i_2, \dots, i_s , а также генерирующий u событий e_1, e_2, \dots, e_u и v выходных данных o_1, o_2, \dots, o_v . Тогда этот узел на внешнем уровне может быть описан как функция

$$(9) \quad n(\{\{\alpha a_1 a_2 \dots a_r\} \{i_1 i_2 \dots i_s\}\}) \rightarrow \{\{\alpha e_1 e_2 \dots e_u\} \{o_1 o_2 \dots o_v\}\}$$

с аргументами типа $\{\{\alpha a_1 a_2 \dots a_r\} \{i_1 i_2 \dots i_s\}\}$, которая возвращает результат типа $\{\{\alpha e_1 e_2 \dots e_u\} \{o_1 o_2 \dots o_v\}\}$ при каждом своем запуске (вычислении). Специальный простой тип α здесь использован для того, чтобы отличить аргументы управления от аргументов данных.

Определение 3. Выражение (9) называется интерфейсным описанием сети (узла сети).

Следует обратить внимание на то, что обычно на практике дуги управления имеют логический тип. Однако в соответствии с (8) и (9) нет никаких препятствий рассматривать эти дуги как пути передачи структурированных управляющих данных.

5.2. НЕРАЗРЕШИМЫЕ ПРОБЛЕМЫ

Несмотря на простоту формализма и доказанных ранее теорем, в рассматриваемой теории совмещенных сетей управления и данных имеются и неразрешимые проблемы.

Теорема 6. Задача построения совмещенных сетей по их интерфейсному описанию является алгоритмически неразрешимой.

Доказательство. Пусть интерфейсное описание совмещенной сети задано формулой (9). Задача построения сети по ее интерфейсному описанию в качестве элементарной операции будет содержать трассировку типизированных данных между узлами строящейся сети. Отсюда, не нарушая общности, сведем задачу построения искомой сети к построению некоторого типа выходных данных

$$\Lambda = \{e_1 e_2 \dots e_u\} \{o_1 o_2 \dots o_v\}$$

из подтипов, принадлежащих типу некоторых входных данных Δ :

$$\Delta = \{a_1 a_2 \dots a_r\} \{i_1 i_2 \dots i_s\}.$$

Представим типы Λ и Δ , их подтипы, а также подтипы специализированных подтипов и т.д., как списки строк конечной длины Y и X соответственно, отсортированные в лексикографическом порядке. Например, для типа

$$\{ba\{b^2 a\{a^2 b\}\}\}\{a^2\}b\{ba\}$$

строится следующий список подтипов длины 6:

$$(ab, ab^2, a^2b, b, a^2, ab).$$

Для выравнивания длины списков Y и X (возможно, и суммарной длины их строк), в каждый из списков добавим некоторое конечное число строк произвольных типов. При дополнении списков будем учитывать возможные преобразования входных типов данных (использовать подстроки исходных типов данных), а также типы и подтипы данных внутренней памяти конструируемой сети, если таковая имеется. Пусть длина этих списков станет равна L .

Тогда задача построения искомой сети эквивалентна комбинаторной проблеме Поста: поиску такой последовательности неповторяющихся индексов i, j, \dots, k ($1 \leq i, j, \dots, k \leq L$), для которых совпадут строки $x_1 x_2 \dots x_L$ и $y_i y_j \dots y_k$, где x_1, x_2, \dots, x_L и $y_i y_j \dots y_k$ — строки списков X' и Y' соответственно. Известно, что комбина-

торная проблема Поста неразрешима [16].¹ Следовательно, не существует алгоритма, который по заданному интерфейсному описанию узла сможет согласовать аргументы внутренних и внешних узлов сети и, в конечном итоге, построить сеть, реализующую это описание. Что требовалось доказать. ♦

На содержательном уровне теорема б утверждает, что в формализме совмещенных сетей управления и данных не существует алгоритма синтеза программ, т.е. не существует универсального алгоритма, который по описанию преобразования данных на языке теории типов строит совмещенную сеть, реализующую это преобразование.

Также следует заметить, что построить искомую сеть не помогает и дополнение множеств исходных и результирующих типов любым конечным множеством производных типов.

Может показаться, что теорема б является следствием низких выразительных качеств совмещенных сетей управления и данных. Однако это не так.

5.3. УНИВЕРСАЛЬНАЯ АЛГОРИТМИЧЕСКАЯ МОДЕЛЬ

Известно несколько алгоритмических моделей, устанавливающих эквивалентность между неформальным и формальным определением алгоритма. Основным результатом общей теории алгоритмов [2] – это эквивалентность универсальных алгоритмических моделей. Так, формализм частично-рекурсивных функций [4] эквивалентен другим алгоритмическим моделям: машине Тьюринга [18], нормальным алгоритмам Маркова [6], лямбда-исчислению Черча [10].

Теорема 7. Всякая частично-рекурсивная функция вычислима совмещенной сетью управления и данных.

¹ Здесь использовано следствие из комбинаторной проблемы Поста, согласно которому из неразрешимости задачи построения индексов $1 \leq i, j, \dots, k \leq L$ для обеспечения равенства строк $x_i x_j \dots x_k$ и $y_i y_j \dots y_k$ для любых списков X и Y достаточно большой длины следует и неразрешимость этой задачи для строк $x_1 x_2 \dots x_L$ и $y_1 y_2 \dots y_k$, так как списки X и Y предполагаются произвольными.

Доказательство. Для доказательства теоремы необходимо показать, что константа 0, функция следования $f(x) = x + 1$ и функции тождества

$$I_m^n(x_1, x_2, \dots, x_n) = x_m \quad (m = 1, \dots, n),$$

а также оператор суперпозиции

$$S_m^n(h, g_1, g_2, \dots, g_m) = h(g_1, g_2, \dots, g_m),$$

оператор примитивной рекурсии:

$$\begin{cases} R_n(x_1, x_2, \dots, x_n, 0) = g(x_1, x_2, \dots, x_n), \\ R_n(x_1, x_2, \dots, x_n, y+1) = h(x_1, x_2, \dots, x_n, y, R_n(x_1, x_2, \dots, x_n, y)); \end{cases}$$

и неограниченный оператор минимизации

$$\mu_y(P(x_1, x_2, \dots, x_n, y)) = \min y \mid P(x_1, x_2, \dots, x_n, y)$$

реализуются совмещенными сетями управления и данных, где y, x_1, x_2, \dots, x_n – переменные; h, g_1, g_2, \dots, g_m – функции; P – предикат (функция, возвращающая логическое значение).

Реализация перечисленных выше функции и операторов совмещенными сетями приведена на рис. 4, 5, 6 и 7, где все дуги помечены типом натуральных чисел произвольной сколь угодно большой разрядности и использованы вспомогательные базовые блоки:

– мультиплексирования данных M :

$$M(a_1, a_2, x_1, x_2) = \begin{cases} x_1, & \text{если } a_1, \\ x_2, & \text{если } a_2; \end{cases}$$

– коммутации управления K ($x \neq 0$):

$$z = \begin{cases} 0, & \text{если } x = 0, \\ 1, & \text{если } x \neq 0; \end{cases} \quad \begin{cases} e_1, & \text{если } x = 0, \\ e_2, & \text{если } x \neq 0; \end{cases}$$

и один составной блок C – счетчик со сбросом:

$$y = \begin{cases} 0, & \text{если } a_1, \\ y + 1, & \text{если } a_2; \end{cases}$$

определенные на рис. 8. Из приведенных рисунков непосредственно следует утверждение теоремы. ♦

Теорема 7 утверждает, что совмещенные сети управления и данных имеют выразительные возможности универсальных алгоритмических моделей, однако отличаются от последних

наличием строгой типизации данных и полностью графической формой описания.

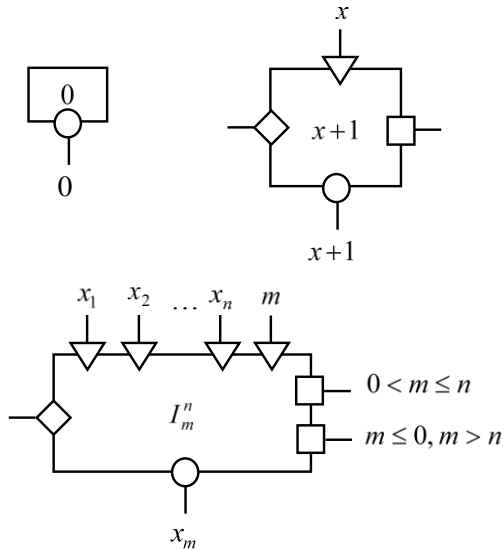


Рис. 4. Элементарные функции

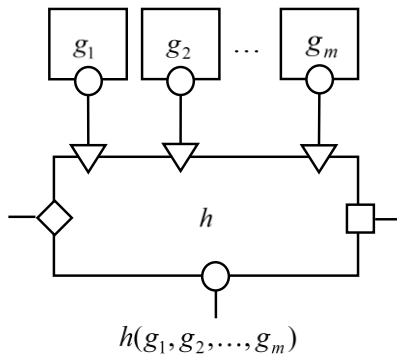


Рис. 5. Суперпозиция функций

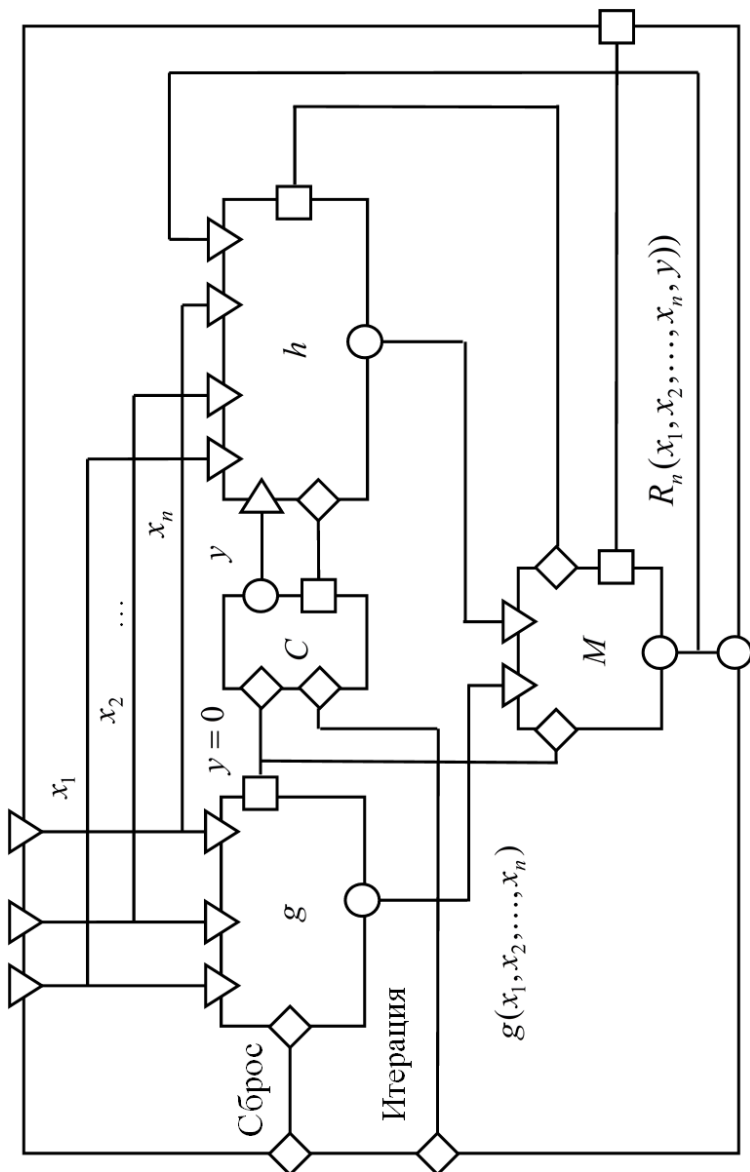


Рис. 6. Прimitивная рекурсия

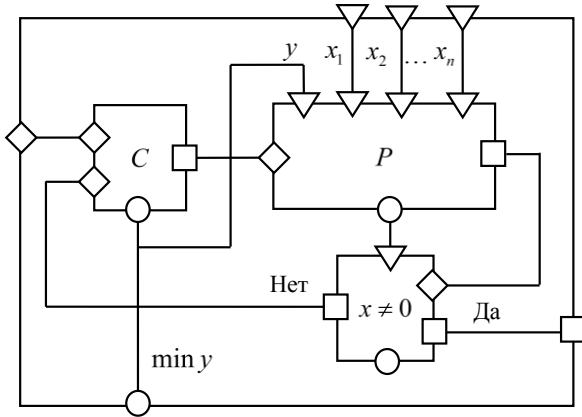


Рис. 7. Оператор минимизации

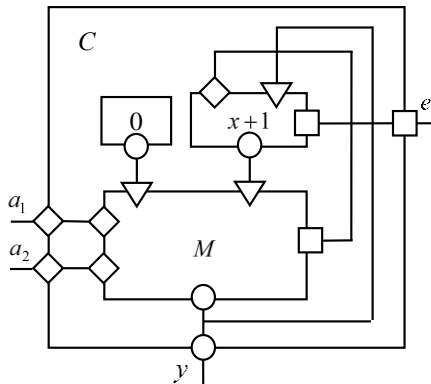
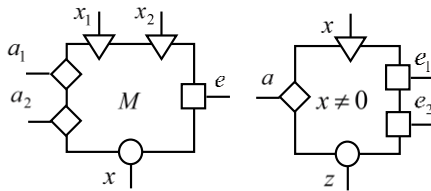


Рис. 8. Вспомогательные операторы: мультимплексор M , коммутатор K ($x \neq 0$), счетчик C

6. Заключение

Применение совмещенных сетей управления и данных позволяет преодолеть накопившиеся проблемы в области динамического управления процессами. В частности, на основе графической модели процесса в нотации совмещенных сетей управления и данных появляется возможность автоматически создавать управляющую программу процесса, а использование строгой типизации данных позволяет выполнить эффективную верификацию создаваемых моделей.

Совместная трассировка данных и управления между узлами сети предоставляет гибкий механизм синхронизации моделируемых процессов, обеспечивая естественное их распараллеливание в среде исполнения. Реализуемая при этом параллельность близка к естественному параллелизму моделируемых процессов.

С теоретической точки зрения, совмещенные сети управления и данных являются универсальной алгоритмической моделью со строгой типизацией данных. Для сравнения типов данных найдена эффективная процедура, сводимая к операциям над строками конечной длины. Для этой алгоритмической модели также доказано, что синтез программ является одной из неразрешимых проблем общей теории алгоритмов.

Как и у других алгоритмических моделей, неразрешимыми проблемами для совмещенных сетей управления и данных является распознавание любых нетривиальных свойств вычислимых функций и, в частности, распознавание по описанию процесса в виде совмещенной сети остановится ли этот процесс при заданных входных данных или не остановится (проблема остановки).

Литература

1. БУЧ Г., РАМБО Д., ДЖЕКОБСОН А. *Язык UML: Руководство пользователя*. – М., СПб.: ДМК «Пресс Питер», 2004. – 432 с.
2. КЛИНИ С.К. *Введение в метаматерику*: Пер. с англ. – М., 1957. – 526 с.
3. ЛЕВИТИН А.В. *Алгоритмы: введение в разработку и анализ*. – М.: Вильямс, 2006. – 576 с.

4. МАЛЬЦЕВ А.И. *Алгоритмы и рекурсивные функции*. – М.: Наука, 1986. – 366 с.
5. МАРКА Д., МАКГОУЭН К. *Методология структурного анализа и проектирования SADT*. – М.: Метатехнология, 1993. – 247 с.
6. МАРКОВ А.А. *Теория алгоритмов*. – М.: Изд-во АН СССР, 1954. – 376 с.
7. ХОПКРОФТ Д., МОТВАНИ Р., УЛЬМАН ДЖ. *Введение в теорию автоматов, языков и вычислений*: Пер. с англ. – М.: Издательский дом Вильямс, 2002. – 528 с.
8. ЯЦУТКО А. В., ВЫХОВАНЕЦ В.С. *Динамическое управление бизнес-процессами на основе совмещенных сетей управления и данных* // Инженерный журнал: наука и инновации. – 2013. – №2(14). – [Электронный ресурс] – URL: <http://engjournal.ru/catalog/mathmodel/social/530.html>. (дата обращения 20.11.2015).
9. CHURCH A. *A formulation of the simple theory of types* // The Journal of Symbolic Logic. – 1940. – Vol. 5, No.2. – P. 56–68.
10. CHURCH A. *A set of postulates for the foundation of logic* // Annals of Mathematics. – 1932. – Series 2, No. 33. – P. 346–366.
11. DAVIS R. *Business Process Modeling with ARIS: A Practical Guide*. – New York: Springer-Verlag, 2005. – 531 p.
12. *Information integration for concurrent engineering (IICE)* / R.J. Mayer, C.P. Menzel, M.K. Painter, et al. // IDEF3 Process Description Capture Method Report. – Ohio: Air Force Material Command, 1995. – 235 p.
13. *Integration definition for function modeling (IDEF0)*. – Washington: National Institute of Standards and Technology, 1993. – 116 p.
14. FISCHER L. *Workflow handbook 2002*. – Association with the Workflow Management Coalition (WfMC), 2002. – 428 p.
15. GANE C., SARSON T. *Structured systems analysis: Tools and techniques*. – New York: Prentice-Hall, 1979. – 452 p.
16. POST E.L. *A variant of a recursively unsolvable problem* // Bulletin of American Mathematical Society. – 1946. – Vol. 52, No 4. – P. 264–268.
17. SMITH H., FINGAR P. *Business process management (BPM): The Third Wave*. – Meghan-Kiffer Press, 2002. – 312 p.

18. TURING A.M. *Computability and λ -Definability* // Journal of Symbolic Logic. – 1937. – Vol. 2, No. 4. – P. 153–163.
19. VYKHOVANETS V., YATSUTKO A. *Dynamic business process management based on the combined control and data networks* // Preprints of the 2013 IFAC Conference on Manufacturing Modelling, Management, and Control. Saint Petersburg, Russia, June 19–21, 2013. – P. 672–677.

COMBINED NETWORKS OF CONTROL AND DATA

Valeriy Vykhovanets, Institute of Control Sciences of RAS, Moscow, Dr. Sc., associative professor (valery@vykhovanets.ru).

Aleksandra Kryzhanovskaya, Yandex Company, Moscow, project manager (aleksundra@gmail.com).

Abstract: An approach is considered to dynamic process management based on the use of combined networks of control and data. Drawing a data flow and a control flow in a joint process diagram allows automatically generating the executable code for process control. A theory of types is developed to care for data types. The data are stored in the form of nested named lists. A combined network of control and data is shown to be a universal algorithmic model with strong data typing. Synthesis of a network for the given transformation of data types is proved to be an algorithmically unsolvable problem. The use of combined networks of control and data can improve the quality of control in technological, manufacturing and organizational processes.

Keywords: process modeling, dynamic control of processes, combined networks of control and data, algorithmic models, theory of types, unresolved problems, universal algorithmic model of strong typing data.

Статья представлена к публикации членом редакционной коллегии М.Ф. Караваем

Поступила в редакцию 14.02.2015.

Опубликована 30.11.2015.

УДК 519.177

ББК 22.18

РЕСУРСНАЯ СЕТЬ С ОГРАНИЧЕНИЯМИ НА ЕМКОСТЬ АТТРАКТОРОВ¹

Жилякова Л. Ю.²

(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, Москва)

Рассматриваются регулярные несимметричные ресурсные сети, в которых вершины-аттракторы, аккумулирующие ресурс сверх порогового значения, имеют ограничения на емкость. В этом случае излишки ресурса перераспределяются в некоторое множество вершин, названных вторичными аттракторами. Изучается функционирование сети с ограничениями на емкость аттракторов при разных значениях суммарного ресурса и определяются пороговые значения, при которых сеть изменяет поведение.

Ключевые слова: ресурсная сеть, графовая динамическая пороговая модель, аттракторы, ограничения на емкость.

1. Введение

Ресурсная сеть, впервые предложенная в [8], – графовая динамическая потоковая модель, в которой в дискретном времени происходит перераспределение ресурса между вершинами. Сеть представлена ориентированным взвешенным графом; веса ребер обозначают их пропускные способности.

Вершины распределяют ресурс в исходящие ребра по двум разным правилам с пороговым переключением. Переключение

¹ Работа выполнена при частичной финансовой поддержке РФФИ (проекты № 14-01-00422а, 15-07-02488а).

² Людмила Юрьевна Жилякова, доктор физико-математических наук (zhilyakova.ludmila@gmail.com)

происходит в зависимости от количества ресурса, содержащегося в вершине. Если это количество превосходит сумму весов всех исходящих из нее ребер, в каждую смежную вершину передается ресурс, равный пропускной способности соответствующего ребра. В противном случае вершина отдаст весь свой ресурс, разделив его пропорционально пропускным способностям. Модель параллельна: на каждом такте дискретного времени t вершины, имеющие ресурс, отдают его одновременно.

Для такой модели получены результаты, позволяющие для каждой топологии сети и пропускных способностей ее ребер, для любого суммарного ресурса и его начального распределения по вершинам определить предельное состояние и предельный поток. Определена формула, которая позволяет для произвольной сети вычислить агрегированный показатель – пороговое значение ресурса T [6]. При суммарном ресурсе в сети, не превосходящем T , ее функционирование, начиная с некоторого конечного такта, описывается однородной цепью Маркова [4]. Более интересные результаты получены для ресурсов выше порогового значения. В этом случае процесс перераспределения ресурса описывается неоднородной цепью Маркова и в сетях с разными топологиями возникают различные эффекты.

При *малых* ресурсах (меньших порогового значения T) функционирование сети полностью определяется стохастической матрицей, полученной из матрицы пропускных способностей нормированием строк [4]. Предельное состояние для регулярных сетей (которые исследуются в настоящей работе) всегда существует и единственно; вектор предельного состояния пропорционален вектору предельных вероятностей однородной цепи Маркова. На сети с малым ресурсом переносятся все результаты, полученные для случайных блужданий и процессов рассеяния на графах [3, 7, 9, 14, 17]. Модели, основанные на однородных цепях Маркова, используются во многих предметных областях, в частности, для решения задач о нахождении консенсуса в многоагентных системах и социальных сетях [1, 2], при исследовании распространения эпидемий, в моделировании регуляторных сетей экспрессии генов и многих других.

Разнообразные прикладные приложения, основанные на модели рассеяния на графах, описаны в книге [14].

При *больших* ресурсах ресурсная сеть наследует часть свойств целочисленных пороговых моделей, таких как, в частности, игра выстреливания фишек, в которой вершины графа обмениваются фишками с выполнением закона сохранения. Вершина может выстрелить, если ее суммарное количество фишек больше количества выходных ребер. Тогда она посылает по одной фишке вдоль каждого ребра. Классическая игра представляет собой последовательные выстреливания вершин [12, 13]; очередность вершин выбирается случайным образом. Были исследованы и параллельные игры выстреливаний [18]. Эта математическая модель нашла применение при описании процессов, получивших название «самоорганизованная критичность», таких, в частности, как сход лавин [10, 11, 15, 16].

Ресурсная сеть при больших ресурсах, объединяя в себе свойства моделей рассеяния и пороговых моделей, демонстрирует проявление новых свойств. К ним относится способность вершин отдавать ресурс независимо от его количества (таким образом, обеспечивается устойчивость и сходимости, которых нет в целочисленных моделях); способность некоторых вершин накапливать излишки ресурса (чего нельзя достичь в однородных марковских процессах), при этом все остальные вершины имеют ресурс, пропорциональный компонентам вектора предельных вероятностей соответствующей однородной цепи Маркова [4–6].

2. Основные определения, обозначения и предварительные результаты

В настоящем разделе будут конспективно перечислены основные определения и предварительные результаты, полученные в [8, 4–6] и др.

Ресурсная сеть представляет собой ориентированный граф $G = (V, E)$, $|V| = n$. Ребра e_{ij} имеют неотрицательные веса r_{ij} , которые обозначают их пропускные способности; $R = (r_{ij})_{n \times n}$ – матрица пропускных способностей.

Вершины сети содержат ресурс, который перераспределяется в дискретном времени: $q_i(t) \geq 0$ – количество ресурса в вершине v_i в момент времени t . Состояние сети в момент t описывается вектором $Q(t) = (q_1(t), \dots, q_n(t))$. Предельное состояние сети будем обозначать через Q^* .

Обозначим через r_i^{in} и r_i^{out} сумму всех пропускных способностей входящих и исходящих ребер соответственно. Будем называть эти величины входной и выходной пропускными способностями вершины v_i .

Правила функционирования сети. В момент t вершина v_i по ребру e_{im} отдает в смежную ей вершину v_m :

r_{im} единиц ресурса, если $q_i(t) > r_i^{out}$ (правило 1);

$\frac{r_{im}}{r_i^{out}} q_i(t)$ единиц ресурса, если $q_i(t) \leq r_i^{out}$ (правило 2).

Суммарный ресурс сети обозначается через W . При функционировании выполняется закон сохранения: $W = \text{const}$.

Множество вершин, для которых $q_i(t) \leq r_i^{out}$, называется зоной $Z(t)$. Вершины из $Z(t)$ функционируют по правилу 2. $Z^+(t)$ – множество вершин, ресурс которых больше их выходной пропускной способности, они функционируют по правилу 1.

T – пороговое значение ресурса, такое что при $W \leq T$ все вершины, начиная с некоторого t' , переходят в зону $Z(t)$; при $W > T$ зона $Z^+(t)$ не пуста и не изменяется, начиная с некоторого t'' . Для несимметричных сетей существование и единственность T доказаны в *теореме 3* [4], в дальнейших работах было показано, что T существует и единственно в любой эргодической сети (сети без стоковых вершин).

Для предельного состояния Q^* обозначим эти зоны через Z^* и Z^{+*} .

Доказано, что при $W \leq T$ сеть любой топологии эквивалентна однородной цепи Маркова, заданной стохастической матрицей R' , полученной из матрицы пропускных способностей R нормированием по строкам (этот результат для регулярных сетей был получен в [4] и впоследствии перенесен на остальные классы сетей). Процессы перераспределения ресурса в регуляр-

ных сетях при малом ресурсе всегда сходятся. Предельное состояние существует и единственно (не зависит от начального распределения ресурса по вершинам).

Вектор предельного состояния сети при $W = 1$ совпадает с вектором предельных вероятностей соответствующей цепи Маркова. Обозначим этот вектор через Q^{1*} .

При $W > T$ в некоторых вершинах начинают накапливаться излишки ресурса.

Введем обозначение:

$$(1) \quad \Delta r_i = r_i^{in} - r_i^{out}.$$

По знаку Δr_i вершины делятся на три класса:

- 1) *вершины-приемники*, для которых $\Delta r_i > 0$;
- 2) *вершины-источники*, для которых $\Delta r_i < 0$;
- 3) *нейтральные вершины*, для которых $\Delta r_i = 0$.

Вершины, способные при $W > T$ оказаться в зоне Z^{+*} , называются *аттракторами*. В роли аттракторов могут выступать некоторые приемники и некоторые нейтральные вершины (критерий аттрактивности будет приведен ниже после введения всех требуемых обозначений), причем первые способны притягивать ресурс и поэтому названы *активными аттракторами*, вторые могут лишь сохранить ресурс, если он был достаточно большим в начальном состоянии. Такие аттракторы называются *пассивными*.

Поток в ресурсной сети

Ресурс, выходящий из вершины v_i по ребру e_{ij} в момент t , приходит в вершину v_j в момент $t + 1$; между моментами t и $t + 1$ он находится в ребре e_{ij} . Этот ресурс назовем потоком $f_{ij}(t)$. Общий поток сети описывается матрицей $F(t) = (f_{ij}(t))_{n \times n}$.

Величиной потока будем называть сумму

$$f_{sum} = \sum_{i=1}^n \sum_{j=1}^n f_{ij}(t).$$

Введем следующие обозначения.

$$\sum_{j=1}^n f_{ij}(t) = f_i^{out}(t); \quad \sum_{i=1}^n f_{ij}(t) = f_j^{in}(t+1).$$

Будем полагать, что $f_j^{in}(0) = 0$.

$F^{out}(t) = (f_1^{out}(t), \dots, f_n^{out}(t))^T$ – вектор исходящего потока;

$F^{in}(t) = (f_1^{in}(t), \dots, f_n^{in}(t))$ – вектор входящего потока;

F^* – матрица предельного потока;

$f_{sum}^* = \sum_{i=1}^n \sum_{j=1}^n f_{ij}^*$ – суммарный предельный поток; F^{in*} и F^{out*} –

векторы двух предельных потоков.

Входное и выходное соседство.

Назовем *входными соседями* вершины v_i множество $V_i^{in} = \{v_k | r_{ki} > 0\}$; *выходными соседями* вершины v_i будем называть множество $V_i^{out} = \{v_k | r_{ik} > 0\}$.

3. Описание сети с ограничениями на емкость аттракторов

3.1. МОТИВАЦИЯ ИССЛЕДОВАНИЯ

В работах [5, 6] исследовано поведение регулярных несимметричных сетей с ресурсом, превышающим пороговое значение T . Напомним, что *регулярной* называется сеть, представленная сильно связным графом, у которого длины всех циклов взаимно просты, а *несимметричной* – сеть, в которой присутствует хотя бы пара вершин, для которых Δr_i (формула (1)) отлична от нуля. То есть в несимметричной сети необходимо существует хотя бы один источник и хотя бы один приемник. При этом хотя бы один приемник будет в такой сети активным аттрактором.

Несимметричные сети с несколькими активными аттракторами – как правило, искусственные конструкции. В них для нескольких вершин должен выполняться критерий аттрактивности ([6]):

$$j = \arg \min_{i \in \{1, \dots, n\}} \frac{r_i^{out}}{q_i^{1*}},$$

где q_i^{1*} – элементы вектора предельных вероятностей однородной цепи Маркова Q^{1*} .

В противном случае, даже если в сети много приемников, активным аттрактором из них может быть только один.

При любом начальном распределении суммарного ресурса $W > T$ в регулярной несимметричной сети с одним аттрактором будет единственное предельное состояние, в котором аттрактор забирает все излишки, сколь большими они бы ни были. (Если в сети два и более аттрактора, излишки ресурса распределятся между ними.) Распределение ресурса между аттракторами зависит от начального состояния. В остальных вершинах предельные значения ресурса постоянны при сколь угодно большом количестве суммарного ресурса и его начальном распределении по вершинам – не-аттрактивные вершины в предельном состоянии всегда содержат ресурс, строго меньший своей выходной пропускной способности, даже если это вершины-приемники [5].

Новая модель, представленная в настоящей статье, предполагает, что аттракторы должны «делиться» излишками ресурса с другими вершинами сети. Это достигается тем, что в аттракторах вводится *ограничение на емкость* – максимальное количество ресурса, которое может получить аттрактор v_i сверх величины r_i^{out} (величине r_i^{out} равна соответствующая компонента вектора предельного состояния при $W = T$).

Если ресурс в сети достаточно велик, то после насыщения аттракторов он начнет распределяться в другие вершины. Новые вершины, принимающие излишки ресурса, назовем *вторичными аттракторами*. Ясно, что процесс ограничения можно продолжать и дальше, вводя в сети аттракторы все более высоких рангов. «Изначальные» аттракторы (аттракторы сети без ограничений) будем называть первичными. Поскольку процесс ограничения первичных, вторичных и т.д. аттракторов одинаков, рассмотрим лишь переход от первичных аттракторов к вторичным.

3.2. ПОСТАНОВКА ЗАДАЧИ

Как уже было отмечено, вектор предельного состояния при $W = 1$ является вектором предельных вероятностей цепи Маркова, заданной графом сети. Для него введено обозначение Q^{1*} .

Вектор предельного состояния при $W = T$ обозначим через \tilde{Q} .

Введем такую нумерацию вершин сети, при которой первичные аттракторы имеют номера от 1 до l . В [5] было доказано, что $\tilde{Q} = (r_1^{out}, \dots, r_l^{out}, \tilde{q}_{l+1}, \dots, \tilde{q}_n)$, причем, $\tilde{q}_j < r_j^{out}$, $j = l + 1, \dots, n$.

Из единственности предельного состояния следует, что вектор \tilde{Q} пропорционален вектору Q^{1*} с коэффициентом T : $\tilde{Q} = Q^{1*} T$. При этом компоненты с номерами от $l + 1$ до n не изменяются при любом $W > T$, и все излишки распределяются только среди аттракторов (Следствие 5 из теоремы 1 [5]). Вектор предельного состояния для $W > T$ имеет вид

$$Q^* = (r_1^{out} + \Delta q_1^*, \dots, r_l^{out} + \Delta q_l^*, \tilde{q}_{l+1}, \dots, \tilde{q}_n), \text{ где } \sum_{i=1}^l \Delta q_i^* = W - T.$$

Предположим теперь, что для всех аттракторов введено ограничение на емкость p , такое что каждый аттрактор может получить лишь ресурс, равный $r_i^{out} + p$ и дальнейшее накопление в этой вершине становится невозможным.

Ясно, что при этом появятся новые вершины, накапливающие излишки.

Перечислим основные задачи исследования получившейся модели:

1. Исследование динамики сети и переходных процессов при разном суммарном ресурсе.

2. Нахождение пороговых значений количества суммарного ресурса, при которых происходят качественные изменения функционирования сети – насыщение первичных аттракторов и начало накопления во вторичных. Описание функционирования в каждом из интервалов между двумя соседними пороговыми значениями.

3. Исследование устойчивости: существование предельных потоков и предельных состояний в каждом из интервалов.

4. Нахождение критерия «вторичной аттрактивности».

5. Нахождение векторов предельных состояний и потоков.

В данной работе мы остановимся на первых двух пунктах (и частично третьем), но сначала опишем алгоритм ограничения емкости аттракторов и обоснуем его использование.

4. Алгоритм функционирования сети

Ограничение на емкость аттракторов формулируется просто и достаточно естественно: если представить вершины в виде некоторых резервуаров ресурса, то у аттракторов на высоте p от значения r_i^{out} ставится заслонка и нового ресурса вершина принять не может.

Сделаем одно важное замечание относительно начальных состояний. Если в начальном состоянии в первичном аттракторе содержится ресурс, превосходящий его ограничение, то будем полагать, что такой аттрактор ограничить нельзя. А поскольку заранее неизвестно, какие вершины в сети являются первичными аттракторами, будем рассматривать такие начальные состояния $Q(0)$, в которых ресурс выше значений $r_i^{out} + p$ может содержаться только в вершинах-источниках или в тех нейтральных вершинах, которые гарантированно не будут первичными аттракторами.

4.1. ФУНКЦИОНИРОВАНИЕ СЕТИ ПОСЛЕ НАСЫЩЕНИЯ ПЕРВИЧНЫХ АТТРАКТОРОВ

Представим фазу функционирования сети с ограничениями, на которой происходит накопление ресурса в аттракторах. На этой фазе входной поток в аттракторы превосходит выходной. При наличии ограничений в вершинах-аттракторах на каждом такте должно проверяться условие, не превышены ли их емкости. Если емкость вершины превышена, она должна принять не весь ресурс, а лишь его часть. При этом нужно решить, что делать с оставшимся ресурсом.

При моделировании физического объекта логично было бы оставить излишки ресурса в ребрах. Такое решение хорошо еще и тем, что на каждом последующем такте входящий в первичный аттрактор поток будет строго равен выходящему из нее потоку, потому что остатки ресурса во входящих ребрах пропорционально уменьшили бы их пропускную способность, частично «закупорив» их собой. Тем самым в аттракторе будет поддерживаться стабильный ресурс $r_i^{out} + p$. Однако в модели выполняется закон сохранения. На каждом такте времени весь ресурс должен быть в вершинах, и этот выход не является приемлемым. Поэтому перед каждым тактом вершины анализируют

свой потенциальный входящий поток. Если в некоторой вершине на некотором ненулевом такте t входной поток избыточен, т.е. при принятии всего потока выполнится неравенство $q_i(t) > r_i^{out} + p$, она оповещает свое входное соседство, и все вершины $v_k \in V_i^{in}$ по соответствующим ребрам отдают не весь планируемый поток, а пропорционально убавляют его на некоторую часть с коэффициентом $c_i(t) \in (0, 1)$. Суммарный входной поток уменьшается с этим же коэффициентом:

$$f_i^{in-new}(t) = f_i^{in}(t) - c_i(t) \cdot f_i^{in}(t) = (1 - c_i(t)) \cdot f_i^{in}(t).$$

При этом, поскольку вершина функционирует по правилу 1, на каждом такте она отдает ресурс, равный своей выходной пропускной способности r_i^{out} . Чтобы удержать в ней стационарный ресурс, равный $r_i^{out} + p$, дивергенция потока (разность между входным и выходным потоком) в этой вершине должна быть равна 0, и, следовательно, входной поток тоже должен быть равен r_i^{out} .

Имеем:

$$(1 - c_i(t)) \cdot f_i^{in}(t) = r_i^{out};$$

откуда коэффициент пропорциональности

$$(2) \quad c_i(t) = 1 - \frac{r_i^{out}}{f_i^{in}(t)} = \frac{f_i^{in}(t) - r_i^{out}}{f_i^{in}(t)} = \frac{\Delta q_i(t)}{f_i^{in}(t)},$$

где $\Delta q_i(t)$ – потенциальный избыток ресурса в аттракторе v_i .

Коэффициент $c_i(t)$ зависит только от входного потока в вершину-аттрактор v_i . Неотданный ресурс остается в этих вершинах на такте t .

В следующем разделе формула для коэффициента $c_i(t)$ претерпит некоторые изменения.

4.2. МАРКИРОВКА ПЕРВИЧНЫХ АТТРАКТОРОВ

Для того чтобы функционирование сети с ограничениями на аттракторы было корректным, первичные аттракторы необходимо пометить до начала процесса перераспределения ресурса. Это условие, на первый взгляд, выглядит искусственным и усложняет модель. Однако на практике оказывается, что в сети, которая о себе ничего не знает (до начала ее работы неизвестно,

какие вершины являются первичными аттракторами), не всегда возможно корректно поставить ограничения на емкость.

Рассмотрим две иллюстрации этого утверждения.

1. Представим себе, что в сети имеется более одного первичного аттрактора, и рассмотрим некоторую пару из множества первичных аттракторов. Пусть количество ресурса настолько велико, что начиная с некоторого такта t оба они набрали бы максимально допустимое количество ресурса, и при этом в момент t в некоторых источниках еще содержался бы избыток ресурса, который они отдают в аттракторы. Поскольку (по определению первичных аттракторов) при таких процессах их входной поток превосходит выходной, на каждом такте в обоих аттракторах появляются потенциальные излишки, которые должны остаться в тех вершинах, откуда они пришли. Представим теперь, что эти аттракторы связаны двусторонней парой ребер. Тогда первый оставляет свои излишки у второго, а второй – у первого. При этом ни первый, ни второй не могут их оставить у себя, потому что имеют максимально возможное количество ресурса $r_i^{out} + p$. Эти излишки должны быть перераспределены между другими вершинами. Но другие вершины уже один раз получили излишки назад и пересчитали свой ресурс. Процесс пересчета ресурса у входящих соседей придется повторить. Избежать такого итеративного распределения излишков на каждом такте можно лишь пометив аттракторы. Тогда будет достаточно не возвращать ресурс из первичных аттракторов в смежные с ними первичные аттракторы, а распределять излишек между не-аттрактивными вершинами из их входного соседства. Формула (2) для расчета коэффициента пропорциональности $c_i(t)$ в этом случае изменится:

$$c_i(t) = \frac{\Delta q_i(t)}{\sum_{k=l+1}^n f_{ki}(t)},$$

если помеченные первичные аттракторы имеют номера от 1 до l .

Весь потенциальный избыток ресурса в аттракторе будет пропорционально делиться лишь между не-аттрактивными вершинами.

2. Если аттракторы в сети не помечены, они определяются во время функционирования следующим образом. Предполага-

ется, что если вершина в начальном состоянии имела малый ресурс, а за время функционирования ей удалось набрать ресурс выше предельной емкости, то эта вершина – активный аттрактор – по-другому определить аттрактор, не зная вектора предельного состояния, невозможно [6]. В большинстве случаев это предположение верно. Однако в некоторых случаях динамика сети может быть такой, что определенные вершины могут сначала набирать, а затем терять ресурс (т.е. на протяжении значительного времени демонстрировать поведение аттракторов, не являясь таковыми). Рассмотрим пример, иллюстрирующий такую динамику.

Пример 1. Пусть сеть с пятью вершинами задана матрицей пропускных способностей:

$$R = \begin{pmatrix} 1 & 3 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 100 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

Это полный граф, два ребра которых имеют пропускные способности, отличные от остальных.

Напомним, что в матрице R i -я строка соответствует пропускным способностям ребер, выходящих из вершины v_i , а i -й столбец – пропускным способностям ребер, входящих в вершину v_i .

В этой сети два приемника – вершины v_1 и v_2 , один источник – v_3 и две нейтральных вершины v_4 и v_5 . Аттрактор этой сети – вершина v_2 , несмотря на то, что у вершины v_1 входная пропускная способность значительно выше (v_2 «паразитирует» на v_1).

Рассмотрим функционирование этой сети *без ограничения на аттракторы*. Пусть вектор начального состояния имеет вид $Q(0) = (0, 0, 0, 0, 100)$, т.е. ресурс 100 находится в нейтральной вершине.

Вектор предельного состояния $Q^* = (6,266; 84,104; 3,210; 3,210; 3,210)$.

Из рис. 1 видно, что прежде чем уйти в зону $Z(t)$, вершина v_1 более ста сорока тактов держится в зоне $Z^+(t)$ вместе с вершиной v_2 , при этом больше семидесяти тактов количество ресурса в ней возрастает. До некоторого такта $t' \approx 75$ обе эти вершины неотличимы по поведению и годятся на роль аттракторов. Если накладывать ограничения на емкость, не зная заранее, сколько в сети аттракторов и каковы они, то ограничения придется наложить на обе вершины, хотя вершина v_1 является вторичным аттрактором этой сети, и именно она должна получить все излишки ресурса.

Единственным первичным аттрактором является вершина v_2 – в конечном счете, несмотря на немонотонную динамику сети, она притянула к себе все излишки, и именно ее одну нужно ограничивать. Если поставить на нее ограничение $p = 10$, получим динамику, представленную на рис. 2.

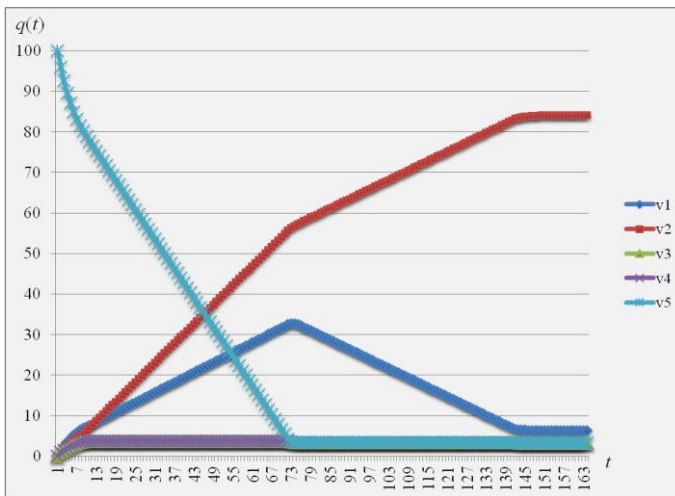


Рис. 1. Функционирование сети без ограничений на аттракторы при большом ресурсе

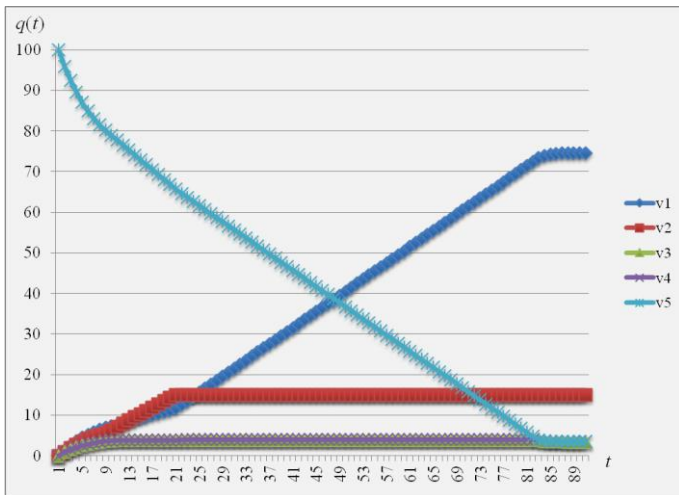


Рис. 2. Функционирование сети с ограничением на емкость аттрактора $p = 10$ (вершина v_2 имеет ресурс, равный 15; излишки ресурса – в вершине v_1)

Предельное состояние при функционировании сети с ограничением будет $Q^* = (74,557; 15; 3,438; 3,503; 3,503)$.

Однако если предварительно не иметь информации о том, какая из вершин v_1 и v_2 является первичным аттрактором, нет никакой возможности узнать, какую из них нужно ограничивать, когда ресурс в них достигнет величины, равной $r_i^{out} + p$, – для вершины v_1 эта величина равна 17, для вершины v_2 , соответственно, 15.

Единственный выход, позволяющий избежать некорректного задания ограничений, как и в предыдущем случае, состоит в том, чтобы, используя критерий аттрактивности, определять первичные аттракторы сети и ставить на них ограничения до начала функционирования.

Замечание. Как видно из сравнения двух векторов предельного состояния, в сети с ограничениями изменились все компоненты вектора Q^* . Вектор предельного состояния в сети с $W = 100$ без ограничений: $Q^* = (6,266; 84,104; 3,210; 3,210; 3,210)$, с ограничениями: $Q^* = (74,557; 15; 3,438; 3,503; 3,503)$. Ресурсы в не-аттрактивных вершинах увеличились. Объяснение

причин увеличения компонент вектора предельного состояния и, как следствие, суммарного потока в сети с ограничениями будет дано в следующем разделе.

5. Исследование потока при ограничениях на аттракторы

В сети без ограничений на аттракторы при $W > T$ для вычисления компонент вектора предельного потока количество избыточного ресурса (сверх T) не имеет значения. При $W = T + \Delta W$ для любого сколь угодно большого значения ΔW величина суммарного предельного потока в регулярной сети неизменна и равна T (теорема 1 [5]).

Пусть по-прежнему нумерация вершин сети такова, что первичные аттракторы имеют номера от 1 до l . В [5] доказано, что при $W = T$ предельное состояние описывается вектором

$$(3) \quad \tilde{Q} = (r_1^{out}, \dots, r_l^{out}, \tilde{q}_{l+1}, \dots, \tilde{q}_n),$$

независимо от начального состояния. Этим же вектором описываются входной и выходной предельные потоки при любом $W = T + \Delta W$, $\Delta W \geq 0$:

$$F^{in*} = F^{out*T} = (r_1^{out}, \dots, r_l^{out}, \tilde{q}_{l+1}, \dots, \tilde{q}_n); f_{sum}^* = T.$$

Для доказательства утверждений о величине потока в сетях с ограничениями сформулируем лемму о свойстве активных аттракторов сети без ограничений. Напомним, что активным аттрактором называется вершина-приемник, способная притягивать излишки ресурса при $W > T$.

Лемма 1. В несимметричной регулярной сети до каждого активного аттрактора существует путь хотя бы из одной не-аттрактивной вершины.

Доказательство. Рассмотрим произвольный активный аттрактор v_i . Поскольку v_i – приемник, для него выполняется неравенство $r_i^{in} > r_i^{out}$. Если его входное соседство V_i^{in} содержит не-аттрактивную вершину, лемма доказана. Если V_i^{in} не содержит не-аттрактивных вершин, то все входные соседи v_i сами являются аттракторами (активными или пассивными), т.е. это приемники или нейтральные вершины. А это означает, что для

вершин, принадлежащих V_i^{in} , должны существовать свои входные соседи, которые будут компенсировать ресурс, отдаваемый ими в активный аттрактор v_i , и служить для них источниками. Таким образом, существует по крайней мере один источник, из которого имеется путь в v_i . \square

Сформулируем утверждения об оценке величины суммарного потока в сетях с ограничением при разных количествах суммарного ресурса.

Утверждение 1. Пусть в регулярной ресурсной сети с ограничением на аттракторы $r_i^{out} + p$, $i = 1, \dots, l$, суммарный ресурс равен $W = T + \Delta W$, $\Delta W > 0$, и в начальном состоянии ни один аттрактор не имеет ресурс, больший своего ограничения. Тогда:

- 1) если $\Delta W \leq lp$, предельный поток существует и $f_{sum}^* = T$;
- 2) если $\Delta W > lp$, выполняется неравенство:

$$(4) \quad f_{sum}(t) \Big|_{t \rightarrow \infty} > T.$$

Доказательство. 1) Пусть $\Delta W \leq lp$. По определению функционирования сети с ограничениями, ограничение на аттрактор включается лишь тогда, когда его ресурс достиг величины $r_i^{out} + p$. До этого момента вершина-аттрактор функционирует так же, как если бы ограничения не было.

1.1) Если аттрактор в сети единственен, то условие $\Delta W \leq lp$ редуцируется до $\Delta W \leq p$, предельное состояние описывается вектором $Q^* = (r_1^{out} + \Delta W, \tilde{q}_2, \dots, \tilde{q}_n)$; предельный поток существует и $f_{sum}^* = T$.

1.2) Рассмотрим сеть с двумя аттракторами v_1 и v_2 . Докажем, что при $t \rightarrow \infty$ каждый аттрактор будет функционировать без ограничений. Если ни в какой момент ресурс в v_1 и v_2 не достигает ограничения, утверждение доказано. Предельное состояние описывается вектором

$$Q^* = (r_1^{out} + \Delta W_1, r_2^{out} + \Delta W_2, \tilde{q}_3, \dots, \tilde{q}_n),$$

$$\Delta W_1 \leq p, \Delta W_2 \leq p, \Delta W_1 + \Delta W_2 = \Delta W;$$

предельный поток существует и $f_{sum}^* = T$.

Пусть в некоторый момент времени t' ресурс в одном из аттракторов (не ограничивая общности, положим, что это v_1)

достиг ограничения $r_1^{out} + p$. Тогда этот аттрактор возвращает излишки в соседние не-аттрактивные вершины. Но поскольку сеть регулярна, из этих не-аттрактивных вершин существуют пути в вершину v_2 , которая по определению первичного аттрактора притягивает ресурс. При этом вершина v_2 может, как максимум, набрать ресурс, равный $r_2^{out} + p$, потому что общий излишек ресурса в сети не превосходит величины $2p$. Если $\Delta W = 2p$, то вектор предельного состояния будет иметь вид $Q^* = (r_1^{out} + p, r_2^{out} + p, \tilde{q}_3, \dots, \tilde{q}_n)$, и аттракторы достигнут верхней границы своей емкости, но не будут принимать, и, следовательно, отдавать излишки.

1.3) Для l аттракторов рассуждения аналогичны. Ресурс $\Delta W \leq lp$ может какое-то время перераспределяться между l аттракторами. Но поскольку хотя бы один аттрактор в сети будет не насыщенным, он будет притягивать к себе ресурс, пока в сети не установится сбалансированный поток $f_{sum}^* = T$.

2. Докажем что при $\Delta W > lp$ выполняется неравенство (4).

Процессы стабилизации потоков в регулярной несимметричной сети без ограничений при $W > T$ описаны в теореме 1 [5]. Весь ресурс ΔW в сети без ограничений должен распределиться между вершинами-аттракторами. Суммарный поток при этом равен T . При $\Delta W > lp$ в сети без ограничений гарантированно существует хотя бы один аттрактор v_i , в котором ресурс будет больше ограничения на емкость $r_i^{out} + p$ (принцип Дирихле).

Тогда при наличии ограничения в этом аттракторе появится избыток входного ресурса, который он будет возвращать своим входным соседям. По лемме 1 в регулярной несимметричной сети в этот аттрактор существует путь от не-аттрактивной вершины. Если эта вершина находится во входном соседстве аттрактора, это означает, что к ней будет возвращаться часть ресурса из вершины v_i . А поскольку она с какого-то момента будет функционировать по правилу 2 и отдавать весь свой ресурс, то выходящий поток из нее будет больше, чем в сети без ограничений за счет возврата от аттрактора, и, соответственно, возрастет суммарный поток в сети. Если во входном соседстве v_i не существует не-аттрактивных вершин, то ресурс будет возвращаться к не-аттрактивной вершине вдоль пути, описанного в лемме 1 (в обратном направлении). И в этой вершине за счет

этого возврата поток будет больше, чем в сети без ограничений. Таким образом, при $\Delta W > lp$ суммарный поток в сети строго больше значения T при $t \rightarrow \infty$. \square

Легко доказать, что верно обратное утверждение.

Следствие. Пусть в регулярной ресурсной сети с ограничением на аттракторы $r_i^{out} + p$, $i = 1, \dots, l$, суммарный ресурс равен $W = T + \Delta W$, $\Delta W > 0$, и в начальном состоянии ни один аттрактор не имеет ресурс, больший своего ограничения. Тогда:

- 1) если $f_{sum}^* = T$, то $\Delta W \leq lp$;
- 2) если $f_{sum}(t)|_{t \rightarrow \infty} > T$, то $\Delta W > lp$.

Доказательство следует из доказательства утверждения 1.

Замечание. Величина p берется одной и той же для всех аттракторов, потому что в сети ни количество аттракторов, ни сами аттракторы не известны заранее. Они помечаются перед началом функционирования, но уже после распределения ресурса по вершинам. Поэтому разные ограничения можно задавать лишь случайным образом, однако это обобщение не усложнит модель.

Покажем, что в сети с ограничениями существует еще одна интегральная характеристика – второе пороговое значение ресурса T^{II} .

Утверждение 2. В регулярной ресурсной сети с ограничением на аттракторы $r_i^{out} + p$, $i = 1, \dots, l$, существует дополнительное пороговое значение (помимо T) T^{II} такое, что в зависимости от количества ресурса в сети существует четыре режима функционирования с тремя пороговыми переключениями, задающимися константами T , lp и T^{II} :

1) при $W \leq T$ с некоторого конечного момента t вся сеть функционирует по правилу 2 и представляет собой однородную регулярную цепь Маркова;

2) при $T < W \leq T + lp$ суммарный поток в сети не изменяется с ростом W , предельный поток существует, $F^{in*} = F^{out*\Gamma} = \tilde{Q}$, $f_{sum}^* = T$;

3) при $T + lp < W \leq T^{\text{II}} + lp$ суммарный поток возрастает с ростом W , суммарный предельный поток существует и $f_{sum}^* = W - lp$;

4) при $W > T^{\text{II}} + lp$ суммарный поток не изменяется с ростом W , суммарный предельный поток существует и $f_{sum}^* = T^{\text{II}}$.

Доказательство.

1) Очевидно, при малом ресурсе функционирование сети с ограничениями эквивалентно функционированию сети без ограничений, для которых пункт 1) выполняется.

2) Доказательство этого пункта следует из утверждения 1.

3) Рассмотрим поток в сети при $W > T + lp$ как функцию от W : $f_{sum} = f_{sum}(W, t)$. При $W = T + lp$ во всех аттракторах наступает насыщение. Предельный поток существует. Будем постепенно увеличивать W . При появлении избытка сверх $T + lp$ хотя бы в одном из аттракторов «срабатывает» ограничение. Тогда он принимает из вершин-источников не весь ресурс, а его часть; ресурс в источниках, смежных с этим аттрактором (или связанных с ним путем (лемма 1)), а, следовательно, и суммарный поток в сети начинает возрастать. Поскольку сеть регулярна, ограничения срабатывают во всех аттракторах, сколь бы ни было мало значение $W - (T + lp)$. Так как все вершины, кроме аттрактивных, функционируют по правилу 2, они отдают весь имеющийся у них ресурс на каждом такте t , и он становится потоком. Таким образом, суммарный поток в сети выше, чем при $T < W \leq T + lp$, и он растет с ростом величины W .

Ресурс в аттракторах стабилизируется на значениях $r_i^{out} + p$. Все остальные вершины функционируют по правилу 2, следовательно, весь их ресурс является потоком. Тогда независимо от существования предельного потока в сети, суммарный предельный поток существует и $f_{sum}^*(W) = W - lp$. Это весь ресурс вершин, за исключением накопленных излишков в первичных аттракторах. При росте W это равенство выполняется до тех пор, пока в предельном состоянии ресурс некоторой вершины, не бывшей аттрактором, не достигнет пороговой величины r_k^{out} . Тогда эта вершина при дальнейшем увеличении W будет накапливать ресурс (таких вершин может быть несколько). Эти вершины являются вторичными аттракторами. Значение W , при

котором для некоторых не-аттрактивных вершин выполнится $q_k^* = r_k^{out}$, является пороговым. Обозначим его через T^{II} .

4) при $W > T^{\text{II}}$ вторичные аттракторы функционируют по правилу 1 и притягивают к себе все излишки ресурса. Суммарный предельный поток при этом: $f_{sum}^* = T^{\text{II}}$. □

Замечание 1. В сетях без ограничений при любом, сколь угодно большом ресурсе, суммарный предельный поток постоянен и равен пороговому значению T . Именно при этом значении аттрактор получает в предельном состоянии ресурс, равный своей выходной пропускной способности. Аналогично, в сети с ограничениями на аттракторы таким пороговым значением является величина T^{II} . При $W > T^{\text{II}}$ весь ресурс сверх T^{II} поглотится вторичными аттракторами.

Замечание 2. Существование T^{II} доказано в утверждении 2. Вопрос о его единственности пока остается открытым, так же как и вопрос о существовании предельных потоков и состояний сети при $W > T + lp$.

Замечание 3. При $p = 0$ режим 2) отсутствует, и при возрастании суммарного ресурса вплоть до значения T^{II} суммарный предельный поток все время возрастает без промежуточного плато. При $p > 0$ суммарный предельный поток стабилизируется при возрастании W до насыщения емкостей аттракторов.

6. Заключение

В работе предложена модификация графовой динамической пороговой модели «ресурсная сеть». При большом суммарном ресурсе в исходной модели существовали вершины, накапливающие в себе основную часть ресурса при его перераспределении. Как правило, в произвольной сети такая вершина единственна. Для того чтобы изменить это положение вещей и распределить излишки более равномерно, в модели вводятся ограничения на емкости вершин-аттракторов. Тогда при их насыщении излишки ресурса начинают переходить в другое множество вершин. Эти вершины названы вторичными аттракторами.

Настоящая работа посвящена описанию и всестороннему анализу предложенной модели, а также исследованию ее устойчивости. В сети с ограничениями на емкости аттракторов для любого суммарного ресурса суммарный предельный поток определяется однозначно. Доказано существование второго порогового значения суммарного ресурса T^{II} , при переходе через которое сеть изменяет поведение. Показано, что в сети с ограничениями для значения суммарного ресурса существует четыре промежутка, на которых сеть демонстрирует различное поведение: $(0, T]$, $(T, T + lp]$, $(T + lp, T^{\text{II}} + lp]$, $(T^{\text{II}} + lp, \infty)$.

В ходе дальнейших исследований планируется доказать, что в сети существуют предельные состояния и потоки (не только суммарный поток стационарен); доказать единственность T^{II} и найти формулу для его нахождения, найти критерий вторичной аттрактивности, найти векторы предельных потоков и состояний.

Автор выражает глубокую благодарность д.ф.-м.н. Р.П. Агаеву за конструктивные замечания и полезные обсуждения в ходе подготовки статьи.

Литература

1. АГАЕВ Р.П., ЧЕБОТАРЕВ П.Ю. *Метод проекции в задаче о консенсусе и регуляризованный предел степеней стохастической матрицы* // Автоматика и телемеханика. – 2011. – №12. – С. 38–59.
2. АГАЕВ Р.П., ЧЕБОТАРЕВ П.Ю. *Сходимость и устойчивость в задачах согласования характеристик (обзор базовых результатов)* // Управление большими системами. – 2010. – №30.1 – С. 470–505.
3. ГАНТМАХЕР Ф.Р. *Теория матриц*. – М.: Физматлит. 2004. – 560 с.
4. ЖИЛЯКОВА Л.Ю. *Несимметричные ресурсные сети. I. Процессы стабилизации при малых ресурсах* // Автоматика и телемеханика. – 2011. – №4. – С. 133–143.

5. ЖИЛЯКОВА Л.Ю. *Несимметричные ресурсные сети. II. Потоки при больших ресурсах и их стабилизация* // Автоматика и телемеханика. – 2012. – №6. – С. 103–118.
6. ЖИЛЯКОВА Л.Ю. *Несимметричные ресурсные сети. III. Исследование предельных состояний* // Автоматика и телемеханика. – 2012. – №7. – С. 67–77.
7. КЕМЕНИ ДЖ., СНЕЛЛ ДЖ. *Конечные цепи Маркова*. – М.: Наука, 1970. – 272 с.
8. КУЗНЕЦОВ О.П. *Однородные ресурсные сети. I. Полные графы* // Автоматика и телемеханика. – 2009. – №11. – С. 136–147.
9. РОБЕРТС Ф.С. *Дискретные математические модели с приложениями к социальным, биологическим и экологическим задачам*. – М. Наука, 1986. – 496 с.
10. ВАК P., TANG C., WIESENFELD K. *Self-organized criticality* // Phys. Rev. A. – 1988. – Vol. 38, No. 1. – P. 364–374.
11. ВАК P., CHEN K. *Self-organized criticality* // Scientif. Am. – 1991. – No. 264. – P. 46–53.
12. BJÖRNER A., LOVASZ L., SHOR P. *Chip-firing games on graphs* // Europ. J. Comb. – 1991. – No. 12. – P. 283–291.
13. BJÖRNER A., LOVASZ L. *Chip-firing game on directed graphs* // J. Algebraic Combinatorics. – 1992. – No. 1. – P. 305–328.
14. BLANCHARD PH., VOLCHENKOV D. *Random Walks and Diffusions on Graphs and Databases: An Introduction* (Springer Series in Synergetics). – Springer-Verlag, Berlin–Heidelberg. 2011. – 262 p.
15. DHAR D. *The abelian sandpile and related models* // Physica A: Statist.l Mechan. Appl. – 1 February, 1999. – Vol. 263, No. 1–4. – P. 4–25.
16. DHAR D., SADHU T., CHANDRA S. *Pattern formation in growing sandpiles* // Eur. Phys. Lett. – 2009. – Vol. 85, No. 4. – 48002. 2009. arXiv:0808.1732 [cond-mat.stat-mech]
17. LOVASZ L., WINKLER P. *Mixing of Random Walks and Other Diffusions on a Graph* // Surveys in Combinatorics / Ed. P. Rowlinson. London Math. Soc. Lecture Notes Series 218, Cambridge Univ. Press, 1995. – P. 119–154.

18. PRISNER E. *Parallel Chip Firing on Digraphs // Complex Systems.* – 1994. – No. 8. – P. 367–383.

RESOURCE NETWORK WITH LIMITED CAPACITY OF ATTRACTOR VERTICES

Ludmila Zhilyakova, Institute of Control Sciences of RAS, (Moscow, Profsoyuznaya st., 65), Doctor of Science, Leader Researcher, zhilyakova.ludmila@gmail.com.

Abstract: An asymmetric regular resource network with limited capacity of attractor vertices is considered. Provided attractors' capacity is unlimited they accumulate all the resource of a network, except some small volumes left in the rest vertices. On the other hand, attractors of limited capacity are saturated, and then another set of vertices starts attracting the rest resource. We refer to this phenomenon as to secondary attractivity, and such vertices, respectively, are called secondary attractors. We study the flows in a network with different total amount resource and search the threshold values which lead to a new behavior of a network.

Keywords: resource network, graph dynamic threshold model, attractors, capacity limitation.

Статья представлена к публикации членом редакционной коллегии П.Ю. Чеботарёвым

*Поступила в редакцию 20.08.2015.
Опубликована 30.11.2015.*

УДК 004.942 + 004.272.43
ББК 32.973.202

**О ДИАГНОСТИРУЕМОСТИ КОМПОНЕНТОВ
ЦИФРОВЫХ СИСТЕМ СО СТРУКТУРОЙ
МИНИМАЛЬНОГО КВАЗИПОЛНОГО ГРАФА
РАЗМЕРА 7×7 С ДВУМЯ ПУТЯМИ
МЕЖДУ ДВУМЯ АБОНЕНТАМИ**

Ведешенков В. А.¹, Курако Е. А.², Лебедев В. Н.³
(ФГБУН Институт проблем управления РАН, Москва)

Представлен подход к оцениванию величин диагностируемости компонентов (абонентов, коммутаторов и линий связи) цифровых систем (ЦС) с новой структурой – минимального квазиполного графа размера 7×7 с двумя путями между двумя абонентами. Оценки величин диагностируемости компонентов анализируемых ЦС определены на основе анализа результатов диагностирования различных отказовых ситуаций в системе из 7 абонентов и 7 коммутаторов. Показано, что использованные методы и средства диагностирования гарантируют правильное диагностирование не больше чем двух неисправных абонентов и двух неисправных коммутаторов, или одного неисправного абонента и трех неисправных коммутаторов, или двух неисправных линий связи.

Ключевые слова: цифровая система, минимальный квазиполный граф размера 7×7 , абонент, коммутатор, линия связи, диагностируемость.

¹ Виктор Алексеевич Ведешенков, доктор технических наук, старший научный сотрудник (vva@ipi.ru, Москва, ул. Профсоюзная, д.65, тел. (495) 334-75-90).

² Евгений Александрович Курако, научный сотрудник (kea@ipi.ru, Москва, ул. Профсоюзная, д.65, тел. (495) 334-75-90).

³ Виталий Николаевич Лебедев, кандидат технических наук, доцент (lebvn@ipi.ru, Москва, ул. Профсоюзная, д.65, тел. (495) 334-92-81).

1. Введение.

Минимальный квазиполный граф образуется на основе однородного двудольного графа, одну долю которого составляют коммутаторы $m \times m$, а другую – m -портовые абоненты. В одной доле имеется N коммутаторов, а в другой – N абонентов. Для минимального квазиполного графа параметры N и m связаны соотношением $N = m(m - 1) / \sigma + 1$ и не могут быть взяты произвольно. Значение m выбирается минимальным, при котором любые два узла в одной доле связаны σ путями длины два через разные узлы в другой доле. Каждый такой путь проходит через один коммутатор, и разные пути проходят через разные коммутаторы [5, 6].

В числе возможных областей применения графов с подобной новой структурой называют отказоустойчивые многомашинные вычислительные системы (МВС) реального времени, где, например, подмножество вершин одной доли представляет совокупность процессорных элементов или вычислительных машин, а подмножество вершин другой доли – коммутаторы [8].

Пример такого графа приведен на рис. 1 для $N = 7$, $m = 4$, $\sigma = 2$.

Цифровыми системами (ЦС) называются модели многомашинных или многопроцессорных вычислительных систем, отражающие необходимые диагностические свойства и параметры анализируемых МВС.

Целью работы является получение оценок величин диагностируемости компонентов (абонентов, коммутаторов, линий связи) по результатам однократного диагностирования (без ремонта) ЦС со структурой минимального квазиполного графа размера 7×7 с 2 путями между 2 абонентами.

Максимальное число одновременно неисправных и произвольно расположенных компонентов, однозначно диагностируемых по результатам однократного тестирования системы (без ремонта), называется величиной t диагностируемости системы [9]. Знание оценок величин диагностируемости компонентов анализируемой системы позволяет оценить достоверность ре-

зультатов, полученных системой диагностирования, а также задавать разумные требования к точности диагностирования на этапе проектирования системы.

Аналитические оценки величин диагностируемости ЦС, построенных с использованием трех типов компонентов, неизвестны. С целью единообразия в приводимых ниже соотношениях и таблице 3 абоненты названы компонентами 1-го типа, коммутаторы – 2-го типа, линии связи – 3-го типа.

Представляется целесообразным для подобных ЦС ввести следующие оценки величин диагностируемости:

t_{120} – оценка величин диагностируемости смеси компонентов 1-го и 2-го типов,

t_{003} – оценка величины диагностируемости компонентов 3-го типа.

Для получения искомым оценок выполним в рассматриваемой ЦС серию процедур диагностирования отказовых ситуаций, включающих неисправные компоненты различных типов и кратности. Сравнение полученного диагноза с составом исходной отказовой ситуации позволит заключить: диагноз является правильным (при совпадении диагноза с составом отказовой ситуации) или нет: состояние некоторых компонентов не идентифицировано. По результатам такого сравнения для выполненной серии процедур диагностирования определим оценки величин диагностируемости – количество компонентов различных типов, для которых независимо от их расположения данная система диагностирования дает гарантированно правильный диагноз.

2. Постановка задачи

Исходная МВС со структурой минимального квазиполного графа $S_{N,m,\sigma}$ представлена диагностическим графом, у которого N вершин одной доли представляют абонентов (процессоры), N вершин другой доли – коммутаторы, причем каждый абонент связан с каждым другим абонентом двумя линиями связи, проходящими через 2 разных коммутатора.

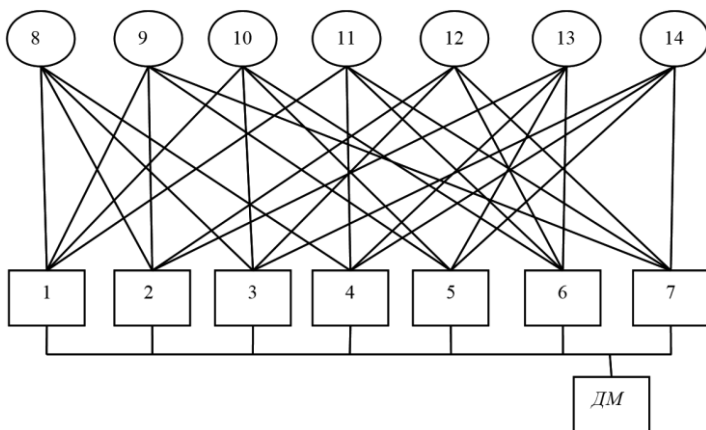


Рис. 1. Схема цифровой системы со структурой минимального квазиполного графа $S_{7,4,2}$ размера 7×7 с двумя путями между двумя абонентами: 1–7 – абоненты; 8–14 – коммутаторы, ДМ – диагностический монитор

Будем считать, что допускаются устойчивые отказы ограниченного числа компонентов: абонентов, коммутаторов и линий связи, причем неисправности компонентов таковы, что прекращают работу тех компонентов, в которых они возникли, и не влияют на работоспособность смежных компонентов.

В диагностировании ЦС участвует исправный диагностический монитор (ДМ), который находится в модуле, внешнем по отношению к диагностируемой ЦС.

Требуется определить оценки величин диагностируемости компонентов (абонентов, коммутаторов, линий связи) по результатам однократного диагностирования (без ремонта) ЦС со структурой минимального квазиполного графа размера 7×7 с 2 путями между 2 абонентами.

3. Подход к диагностированию компонентов ЦС со структурой минимального квазиполного графа размера 7×7 с 2 путями между 2 абонентами

Для проверки работоспособности компонента с предполагаемым характером неисправности (работа – отказ) со стороны проверяющего модуля достаточно послать к нему запрос, на который проверяемый компонент должен дать ответ в течение заданного интервала времени Δt . Отсутствие ответа в течение интервала Δt является признаком проявления (обнаружения) неисправности компонента.

Использование более сложных моделей неисправности компонентов потребует соответствующего усложнения проверок, необходимых для обнаружения таких неисправностей. Но и в этом случае результатом работы системы диагностирования будет диагноз: номера исправных компонентов и компонентов, подозреваемых в неисправности.

Процесс диагностирования компонентов ЦС включает такие процедуры [1, 3]:

- исполнение исправным модулем функций проверяющего модуля;
- нахождение пути от одного исправного модуля к другому исправному для передачи ему функций проверяющего модуля;
- раскодирование результатов выполненных проверок.

Для реализации этих процедур в ЦС со структурой минимального квазиполного графа размера 7×7 с 2 путями между 2 абонентами поступим следующим образом.

Проверяющим модулем m_j назначим абонент A_j ($j = 1, \dots, 7$), а в проверяемую подсистему $U_{i,k}$ включим следующие компоненты: коммутатор p_i , абонента A_k , линию связи l_{i-k} между ними и линию связи l_{j-i} , соединяющую A_j с p_i , причем $k \neq j$. Например, для системы $S_{7,4,2}$ на рис. 1 для реализации одной из проверок абонент 5 будет исполнять функции проверяющего модуля m_5 , а в проверяемую подсистему будут входить коммутатор 9, абонент 2 и линии связи l_{5-9} и l_{9-2} .

Таким образом, индекс (i, k) подсистемы $U_{i,k}$ состоит из номера i ($i = 8, \dots, 14$) проверяемого коммутатора и номера k

($k = 1, \dots, 7$) проверяемого абонента, а индекс линии связи l_{i-k} (например, линии связи l_{9-2}) состоит из номера i коммутатора и номера k абонента, которые она соединяет (в примере – коммутатор 9 и абонент 2). Аналогично формируется индекс линии связи l_{j-i} .

Тогда отдельная проверка (тест $p_{j,i,k}$) будет выполняться следующим образом. Получив управление в процессе диагностирования, проверяющий абонент A_j передает через коммутатор p_i и линии связи l_{j-i} и l_{i-k} проверяемому абоненту A_k запрос о его состоянии. Исправный абонент A_k передает ответ через коммутатор p_i и линии связи l_{i-k} и l_{j-i} проверяющему абоненту A_j , который формирует оценку $r_{i,k} := 0$; если в течение интервала Δt ответ не придет, то $r_{i,k} := 1$.

На основании полученных оценок абонент A_j выбирает исправную подсистему U_{i^*,k^*} , абоненту A_{k^*} из состава которой передает управление и необходимую информацию для продолжения процесса диагностирования. Здесь следует подчеркнуть, что передача управления исправному абоненту A_{k^*} (из состава проверенной подсистемы U_{i^*,k^*}) приводит к тому, что следующим проверяющим абонентом не будет назначен абонент, попавший в число подозреваемых в отказе (получивший оценку 1). Вследствие этого в число тестов, на которых диагностируется конкретная отказовая ситуация, не могут попасть тесты, проверяющий абонент которых попал в число подозреваемых в отказе.

В таблице 1 показана информация о составе проверяемых подсистем $U_{i,k}$ и номерах проверяющих их абонентов A_j . В таблице 1 для абонента A_j ($j = 5, 2, \dots, 7$) записаны номера компонентов и линий связи, входящих в состав подсистем U_{i_1,k_1} , U_{i_2,k_2} , проверяемых абонентом A_j при исполнении проверок строки T_{j1} , T_{j2} , T_{j3} . (Проверки, относящиеся к одному проверяющему абоненту, в таблице 1 разбиты на три части из-за невозможности размещения в одной строке 6 подсистем, проверяемых одним абонентом.) Отметим, что в клетках строк T_{51}, \dots, T_{73} таблицы 1 показаны номера индексов проверяющего абонента и проверяемых подсистем без символов абонента A , проверяемой подсистемы U и линии связи l .

Таблица 1. Таблица состава проверяемых подсистем

| Строки T_{j1} , T_{j2} , T_{j3} | Абонент A_j | Проверяемые подсистемы $U_{i,k}$, оценки результатов их проверок $r_{i,k}$ | | | |
|--|------------------|--|--------------|--------------------------|--------------|
| | | $U_{i1, k1}$ | $r_{i1, k1}$ | $U_{i2, k2}$ | $r_{i2, k2}$ |
| T_{51} | 5 | (5-9), 9, (9-2), 2 | | (5-10), 10, (10-6), 6 | |
| T_{52} | 5 | (5-14), 14, (14-3), 3 | | (5-9), 9, (9-7), 7 | |
| T_{53} | 5 | (5-13), 13, (13-4), 4 | | (5-10), 10, (10-1), 1 | |
| T_{21} | 2 | (2-8), 8, (8-1), 1 | | (2-13), 13, (13-5), 5 | |
| T_{22} | 2 | (2-9), 9, (9-1), 1 | | (2-12), 12, (12-3), 3 | |
| T_{23} | 2 | (2-13), 13, (13-6), 6 | | (2-8), 8, (8-4), 4 | |
| T_{11} | 1 | (1-10), 10, (10-3), 3 | | (1-9), 9, (9-7), 7 | |
| T_{12} | 1 | (1-8), 8, (8-4), 4 | | (1-9), 9, (9-5), 5 | |
| T_{13} | 1 | (1-10), 10, (10-6), 6 | | (1-11), 11, (11-7), 7 | |
| T_{31} | 3 | (3-12), 12, (12-6), 6 | | (3-14), 14, (14-4), 4 | |
| T_{32} | 3 | (3-10), 10, (10-5), 5 | | (3-8), 8, (8-1), 1 | |
| T_{33} | 3 | (3-14), 14, (14-7), 7 | | (3-8), 8, (8-2), 2 | |
| T_{61} | 6 | (6-13), 13, (13-4), 4 | | (6-11), 11, (11-1), 1 | |
| T_{62} | 6 | (6-12), 12, (12-7), 7 | | (6-12), 12, (12-2), 2 | |
| T_{63} | 6 | (6-10), 10, (10-3), 3 | | (6-13), 13, (13-5), 5 | |

| Строки $T_{j1},$ $T_{j2},$ T_{j3} | Абонент A_j | Проверяемые подсистемы $U_{i,k}$, оценки результатов их проверок $r_{i,k}$ | | | |
|--|------------------|--|--------------|--------------------------|--------------|
| | | $U_{i1, k1}$ | $r_{i1, k1}$ | $U_{i2, k2}$ | $r_{i2, k2}$ |
| T_{41} | 4 | (4-14), 14, (14-7), 7 | | (4-8), 8, (8-3), 3 | |
| T_{42} | 4 | (4-13), 13, (13-2), 2 | | (4-11), 11, (11-6), 6 | |
| T_{43} | 4 | (4-11), 11, (11-1), 1 | | (4-14), 14, (14-5), 5 | |
| T_{71} | 7 | (7-14), 14, (14-5), 5 | | (7-12), 12, (12-2), 2 | |
| T_{72} | 7 | (7-11), 11, (11-6), 6 | | (7-11), 11, (11-4), 4 | |
| T_{73} | 7 | (7-9), 9, (9-2), 2 | | (7-12), 12, (12-3), 3 | |

Отметим также, что в рассматриваемой ЦС линии связи являются двунаправленными. В таблице 1 показаны индексы линий связи того направления, которое нужно для реализации данной проверки. Для обеспечения полноты проверки линий связи таблицу 1 нужно было бы расширить, добавив проверки с инверсным направлением линий связи, это удвоило бы размерность таблицы 1 и трудоемкость последующих операций. Поскольку одной из целей данной работы является получение оценок величин диагностируемости компонентов рассматриваемой ЦС, то для определения порядка этих оценок можно ограничиться анализом таблицы меньшей размерности (с общим номером для двунаправленных линий связи).

Как видно из таблицы 1, каждый абонент A_j проверяет других абонентов 6 раз в составе различных подсистем, что в 6 раз больше, чем нужно для контроля работоспособности. Но приведенные 42 проверки необходимы для исключения исправного коммутатора из числа подозреваемых в отказе при неисправности двух абонентов, которые он соединяет.

В целом диагностирование состояния компонентов ЦС со структурой минимального квазиполного графа размера 7×7 с 2 путями между 2 абонентами организовано следующим образом.

По сигналу об ошибке, полученному от средств функционального контроля, диагностический монитор (ДМ) выдает запрос о состоянии абоненту A_{j1} , выбранному заранее (например, абоненту 5 – первому абоненту в таблице 1). Если в пределах интервала Δt ДМ не получает ответа на свой запрос, он отправляет запрос следующему абоненту A_{j2} (например, абоненту 2 в таблице 1).

После получения ответа в пределах интервала Δt ДМ выдает этому же абоненту A_{j1} (A_{j2}) команду на начало диагностических процедур.

По команде ДМ о начале диагностирования начальный абонент A_{j1} (в примере – абонент 5) выполнит следующие действия (для упрощения в последующих двух абзацах вместо символьных индексов используются цифровые индексы из таблицы 1):

– образует промежуточную таблицу PT_5 , переписет в нее из таблицы 1 строки T_{51} , T_{52} , T_{53} , найдет в них индексы (9, 2), (10, 6), (14, 3), (9, 7), (13, 4), (10, 1) проверяемых подсистем $U_{9,2}$, $U_{10,6}$, $U_{14,3}$, $U_{9,7}$, $U_{13,4}$, $U_{10,1}$, передаст абоненту 2 (6, 3, 7, 4, 1) запрос о состоянии соответствующей подсистемы, ответ на который абонент 2 (6, 3, 7, 4, 1) должен вернуть абоненту 5.

– В соответствии с полученными и не полученными ответами абонент 5 формирует оценки проверенных подсистем, записывает их в соответствующие клетки строк T_{51} , T_{52} , T_{53} , выбирает среди них исправную подсистему U_{i^*, k^*} , (с оценкой 0), абоненту A_{k^*} которой передает строки T_{51} , T_{52} , T_{53} с оценками и управление для продолжения процесса диагностирования.

Абонент A_{k^*} (и каждый последующий исправный абонент из проверенных подсистем) выполнит действия, аналогичные действиям абонента 5, и передаст таблицу PT_5 с накопленными оценками и управление продолжением процесса диагностирования следующему исправному абоненту.

Наконец, абонент $A_{k^{**}}$, выполнив функции проверяющего и не обнаружив в строках таблицы PT_5 ни одного исправного абонента, не исполнявшего функций проверяющего, закончит процесс диагностирования и перейдет к раскодированию полученных оценок проверок. Результаты раскодирования абонент $A_{k^{**}}$ через ДМ передает обслуживающему персоналу для проведения восстановительных работ (замены исправными компонентами, подозреваемых в неисправности).

Цель раскодирования – преобразовать полученные оценки проверок в списки исправных компонентов и компонентов, подозреваемых в неисправности.

Для раскодирования результатов проверок и формирования диагноза воспользуемся алгебро-логическим методом, основы которого опубликованы в [2, 4, 7].

А. Обозначим переменной \bar{m}_n исправное состояние компонента с номером n , а переменной m_n неисправное состояние компонента с номером n , т.е. для этих переменных справедливы логические соотношения:

- компонент n исправен – $\bar{m}_n = 1, m_n = 0$;
- компонент n неисправен – $\bar{m}_n = 0, m_n = 1$.

Функцию технического состояния компонентов с номерами n_1, n_2, \dots, n_k :

– при нулевом результате их проверки тестом r_j (обозначим оценку \bar{r}_j) запишем так:

$$\bar{F}_j = \bar{m}_{n_1} \wedge \bar{m}_{n_2} \dots \wedge \bar{m}_{n_k} \text{ (все компоненты исправны);}$$

– при единичном результате их проверки тестом r_j (обозначим оценку r_j) запишем так:

$F_j = m_{n_1} \wedge m_{n_2} \wedge \dots \wedge m_{n_k}$ (хотя бы один из компонентов неисправен).

Б. Предлагаемый метод раскодирования результатов тестирования ЦС на множестве проверок p_1, p_2, \dots, p_k состоит из следующих этапов.

1. Для каждого результата \bar{r}_j (r_j) записать функцию \bar{F}_j (F_j), в которую подставить номера компонентов, взятые из состава проверки p_j .

2. Построить функцию $\bar{\Phi}$ как конъюнкцию всех функций \bar{F}_j .

3. Построить функцию Φ как конъюнкцию всех функций F_j .

4. Вместо нахождения конъюнкции $\Phi \wedge \bar{\Phi}$ упростить отдельные дизъюнктивные члены функции Φ , вычеркнув те переменные, для которых в функции $\bar{\Phi}$ есть одноименные переменные с отрицанием (с чертой сверху).

5. Произвести конъюнктивное перемножение оставшихся членов функций F_j , выполнить поглощение членов большей длины членами меньшей длины. При этом будут полезны известные формулы:

$$\bar{m}_n \wedge m_n = 0, \bar{m}_n \vee m_n = 1, \bar{m}_n \cdot (\bar{m}_n \vee \bar{m}_p) = \bar{m}_n$$

$$\bar{m}_n \wedge m_n = 0, \bar{m}_n \vee m_n = 1, \bar{m}_n \wedge (\bar{m}_n \vee \bar{m}_p) = \bar{m}_n.$$

6. Выделить из полученного логического выражения члены минимальной длины. Они определяют диагноз – подмножество компонентов, подозреваемых в неисправности по результатам выполненных проверок.

4. Пример

Предположим, что из-за большого перерыва между циклами контроля в рассматриваемой ЦС накопились отказы абонентов 3, 6, коммутатора 8 и двух линий связи (9-2) и (5-9).

После подтверждения исправности абонента 5 по сигналу ДМ о начале диагностирования начальный абонент 5 выполнит следующие действия:

– образует промежуточную таблицу PT_5 , переписет в нее из таблицы 1 строки T_{51}, T_{52}, T_{53} , найдет в них индексы (9-2), (10-6), (14, 3), (9, 7), (13, 4), (10, 1) проверяемых подсистем $U_{9,2}, U_{10,6}, U_{14,3}, U_{9,7}, U_{13,4}, U_{10,1}$, передаст абоненту 2 (6, 3, 7, 4, 1) запрос о состоянии своей подсистемы, ответ на который абонент 2 (6, 3, 7, 4, 1) должен вернуть абоненту 5.

В соответствии с полученными ответами абонент 5 формирует оценки $r_{9,2} = 1, r_{10,6} = 1, r_{14,3} = 1, r_{9,7} = 1, r_{13,4} = 0, r_{10,1} = 0$, записывает их в соответствующие клетки строк T_{51}, T_{52}, T_{53} (для

сокращения числа таблиц будем записывать результаты проверок, находящиеся в разных модулях, в общую таблицу 2), анализирует сформированные оценки: так как среди этих оценок есть нулевые, то передает абоненту 4 подсистемы $U_{13,4}$ таблицу PT_5 (со строками T_{51} , T_{52} , T_{53}) и управление для продолжения процесса диагностирования. (Следует отметить, что в строке T_{53} и в последующих полужирным шрифтом в результатах проверки выделен 0 того исправного абонента, которому передается управление продолжением процесса диагностирования.)

Абонент 4 выполнит следующие действия:

– в полученную таблицу PT_5 переписет из таблицы 1 строки T_{41} , T_{42} , T_{43} , найдет в ней индексы (13, 2), (11, 6), (11, 1), (14, 5), (14, 5), (12, 2) проверяемых подсистем, передаст в эти подсистемы запрос об их состоянии, ответ на который абонент 2 (6, 1, 5, 5, 2) должен вернуть абоненту 4;

– абонент 4 сформирует соответствующие оценки (показаны в строках T_{41} , T_{42} , T_{43}) и абоненту с нулевой оценкой (абоненту 2) передаст таблицу PT_5 и управление процессом диагностирования.

Получив управление, абонент 2 выполнит действия, аналогичные действиям абонентов 5 и 4. Так как только для подсистемы (13, 5) сформирована нулевая оценка и абонент 5 уже выполнял функции проверяющего модуля, то абонент 2 должен найти среди ранее проверенных другого абонента для выполнения этих функций. Таким модулем является абонент 1 (отмечен $\underline{0}$ в строке T_{42} таблицы 2). Поэтому абонент 2 через коммутатор 13 передаст таблицу PT_5 и управление процессом диагностирования абоненту 4, который через коммутатор 11 передаст таблицу PT_5 и управление процессом диагностирования абоненту 1.

Получив управление, абонент 1, а затем абонент 7 выполняют действия, аналогичные действиям абонентов 5 и 4, сформированные оценки запишут в строки T_{11} , T_{12} , T_{13} и T_{71} , T_{72} , T_{73} таблицы 2 соответственно.

После выполнения своих проверок абонент 7 по записям в таблице 2 не обнаружит исправных абонентов, не выполнявших функций проверяющих, закончит процесс диагностирования и

перейдет к раскодированию полученных оценок проверок. Результаты раскодирования абонент 7 через ДМ передаст обслуживающему персоналу для проведения восстановительных работ (замены исправными компонентов, подозреваемых в неисправности).

Таблица 2. Информация к примеру

| Строки $T_{j1},$ $T_{j2},$ T_{j3} | Проверяющий абонент A_j | Проверяемые подсистемы $U_{i,k}$, оценки результатов их проверок $r_{i,k}$ | | | |
|--|------------------------------|--|----------------|-----------------------------|----------------|
| | | $U_{i1,k1}$ | $r_{i1,k1}$ | $U_{i2,k2}$ | $r_{i2,k2}$ |
| <i>Таблица PT_5</i> | | | | | |
| T_{51} | 5 | (5-9), 9, (9-2), 2 | 1 ₁ | (5-10), 10, (10-6), 6 | 1 ₂ |
| T_{52} | 5 | (5-14), 14, (14-3), 3 | 1 ₃ | (5-9), 9, (9-7), 7 | 1 ₄ |
| T_{53} | 5 | (5-13), 13, (13-4), 4 | 0 ₁ | (5-10), 10, (10-1), 1 | 0 ₂ |
| T_{41} | 4 | (4-13), 13, (13-2), 2 | 0 ₃ | (4-11), 11, (11-6), 6 | 1 ₅ |
| T_{42} | 4 | (4-11), 11, (11-1), 1 | 0 ₄ | (4-14), 14, (14-5), 5 | 0 ₅ |
| T_{43} | 4 | (7-14), 14, (14-5), 5 | 0 ₆ | (7-12), 12, (12-2), 2 | 0 ₇ |
| T_{21} | 2 | (2-8), 8, (8-1), 1 | 1 ₆ | (2-13), 13, (13-5), 5 | 0 ₈ |
| T_{22} | 2 | (2-9), 9, (9-1), 1 | 1 ₇ | (2-12), 12, (12-3), 3 | 1 ₈ |

| Строки $T_{j1},$ $T_{j2},$ T_{j3} | Проверяющий абонент A_j | Проверяемые подсистемы $U_{i,k}$, оценки результатов их проверок $r_{i,k}$ | | | |
|--|------------------------------|--|-----------------|-----------------------------|-----------------|
| | | $U_{i1,k1}$ | $r_{i1,k1}$ | $U_{i2,k2}$ | $r_{i2,k2}$ |
| <i>Таблица PT_5</i> | | | | | |
| T_{23} | 2 | (2-13), 13, (13-6), 6 | 1 ₉ | (2-8), 8, (8-4), 4 | 1 ₁₀ |
| T_{11} | 1 | (1-10), 10, (10-3), 3 | 1 ₁₁ | (1-9), 9, (9-7), 7 | 0 ₉ |
| T_{12} | 1 | (1-8), 8, (8-4), 4 | 1 ₁₂ | (1-9), 9, (9-5), 5 | 1 ₁₃ |
| T_{13} | 1 | (1-10), 10, (10-6), 6 | 1 ₁₄ | (1-11), 11, (11-7), 7 | 0 ₁₀ |
| T_{71} | 7 | (7-14), 14, (14-5), 5 | 0 ₁₁ | (7-12), 12, (12-2), 2 | 0 ₁₂ |
| T_{72} | 7 | (7-11), 11, (11-6), 6 | 1 ₁₅ | (7-11), 11, (11-4), 4 | 0 ₁₃ |
| T_{73} | 7 | (7-9), 9, (9-2), 2 | 1 ₁₆ | (7-12), 12, (12-3), 3 | 1 ₁₇ |

Применим алгебро-логический метод для раскодирования результатов проверок, показанных в таблице PT_5 (таблица 2). (С целью сокращения объема текста некоторые из производимых операций будем опускать.)

В таблице 2 пятнадцать строк, содержащих 30 проверок: результаты 13 из них получили оценку 0 (проверенные подсистемы исправны), а 17 проверок получили оценку 1 (некоторые из компонентов этих подсистем неисправны). Для того чтобы в последующем упростить ссылки на эти проверки и их оценки, в клетках с оценками таблицы 2 в качестве нижнего индекса

указан порядковый номер оценки, причем нулевые и единичные оценки подсчитаны отдельно.

Для некоторых из этих проверок приведем функции \bar{F}_j и F_j , где в качестве индекса j указан порядковый номер оценки:

$$\bar{F}_1 = \overline{l_{5-13}} \cdot \overline{l_{13-4}} \cdot \overline{c_{13}} \cdot \overline{a_4}; \quad \bar{F}_2 = \overline{l_{5-10}} \cdot \overline{l_{10-1}} \cdot \overline{c_{10}} \cdot \overline{a_1};$$

$$\bar{F}_3 = \overline{l_{4-13}} \cdot \overline{l_{13-2}} \cdot \overline{c_{13}} \cdot \overline{a_2}; \quad \bar{F}_4 = \overline{l_{4-11}} \cdot \overline{l_{11-1}} \cdot \overline{c_{11}} \cdot \overline{a_1};$$

$$\bar{F}_{12} = \overline{l_{7-12}} \cdot \overline{l_{12-2}} \cdot \overline{c_{12}} \cdot \overline{a_2}; \quad \bar{F}_{13} = \overline{l_{7-11}} \cdot \overline{l_{11-4}} \cdot \overline{c_{11}} \cdot \overline{a_4};$$

$$F_1 = l_{5-9} \vee l_{9-2} \vee c_9 \vee a_2; \quad F_2 = l_{5-10} \vee l_{10-6} \vee c_{10} \vee a_6;$$

$$F_3 = l_{5-14} \vee l_{14-3} \vee c_{14} \vee a_3; \quad F_4 = l_{5-9} \vee l_{9-7} \vee c_9 \vee a_7;$$

$$F_{16} = l_{7-9} \vee l_{9-2} \vee c_9 \vee a_2; \quad F_{17} = l_{7-12} \vee l_{12-3} \vee c_{12} \vee a_3.$$

Для построения функции $\bar{\Phi}$ нужно образовать конъюнкцию 13 функций \bar{F}_j . Ниже показана часть полученной функции $\bar{\Phi}$, содержащая переменные абонентов, коммутаторов и тех линий связи, которые оказались полезными для последующего упрощения функции Φ :

$$\bar{\Phi} = \overline{a_1} \cdot \overline{a_2} \cdot \overline{a_4} \cdot \overline{a_5} \cdot \overline{a_7} \cdot \overline{a_9} \cdot \overline{a_{10}} \cdot \overline{a_{11}} \cdot \overline{a_{12}} \cdot \overline{a_{13}} \cdot \overline{a_{14}} \cdot$$

$$(1) \overline{l_{5-10}} \cdot \overline{l_{4-5}} \cdot \overline{l_{11-7}} \cdot \overline{l_{9-7}} \cdot \overline{l_{1-9}} \cdot \overline{l_{2-13}} \cdot \overline{l_{4-11}} \cdot \overline{l_{10-1}} \cdot \overline{l_{7-12}} \cdot \overline{l_{2-12}}$$

(отсутствуют a_3, a_6, a_8).

После построения функции Φ как конъюнкции 17 функций F_j (часть которых показана выше), проведено упрощение полученного выражения: вычеркнуты те переменные абонентов, коммутаторов и линий связи, для которых в функции $\bar{\Phi}$ есть одноименные переменные с отрицанием, проведено поглощение членов большей длины членами меньшей длины. Ниже показана часть преобразованной функции Φ , содержащая несколько членов наименьшей и близкой к ней длины.

$$\Phi = l_{9-2} \cdot l_{5-9} \cdot a_3 \cdot a_6 \cdot c_8 \vee$$

$$(2) \vee l_{9-2} \cdot l_{5-9} \cdot a_3 \cdot a_6 \cdot (l_{2-8} \cdot l_{1-8} \vee l_{2-8} \cdot l_{4-8} \vee l_{1-8} \cdot l_{4-8})$$

$$\vee l_{9-2} \cdot l_{5-9} \cdot a_3 \cdot c_8 \cdot l_{10-6} \cdot l_{11-6} \cdot l_{13-6} \vee \\ \vee l_{9-2} \cdot l_{5-9} \cdot a_6 \cdot c_8 \cdot l_{10-3} \cdot l_{12-3} \cdot l_{14-3} .$$

Поскольку в функции Φ только один член минимальной длины (показан первым), то возможно, что полученный диагноз является правильным: подозреваются в неисправности абоненты 3 и 6, коммутатор 8, линии связи (9–2) и (5–9) (в данном случае он совпадает с составом отказовой ситуации, рассмотренной в начале примера). Вместе с тем этот вариант диагноза не удовлетворяет показанным ниже оценкам величин диагностируемости из соотношений (3), т.е. достоверность полученного диагноза не гарантируется.

У обслуживающего персонала есть два варианта действий:

– поверить полученному диагнозу: заменить исправными все компоненты, подозреваемые в неисправности, повторить диагностику для подтверждения работоспособности восстановленной системы;

– для повышения достоверности диагноза заменить исправными часть компонентов, подозреваемых в неисправности, и повторить диагностирование частично восстановленной системы. В качестве заменяемых целесообразно брать компоненты, входящие во все или большинство вариантов. В данном примере целесообразно заменить линии связи l_{9-2} и l_{5-9} (входят во все варианты).

5. Подход к оцениванию величин диагностируемости

Известно следующее

Определение 1. Система S из n модулей является однократно диагностируемой относительно t отказов, если все отказавшие модули можно идентифицировать по результатам раскодирования результатов диагностирования при условии, что число отказавших модулей не превышает t [9].

В дальнейшем будет полезно определение 2, противоположное определению 1.

Определение 2. Система S из n модулей не является однократно диагностируемой относительно t отказов, если диагноз

хотя бы одной отказовой ситуации мощностью t является неполным, т.е. состояние некоторых компонентов не идентифицировано, или неоднозначным.

Представленные ниже оценки величин диагностируемости компонентов (t_{100} , t_{020} , t_{003} , t_{120}) ЦС со структурой минимального квазиполного графа размера 7×7 с 2 путями между абонентами получены на основе анализа серии результатов диагностирования типичных отказовых ситуаций. Для диагностирования использован подход к диагностированию компонентов рассматриваемой ЦС, приведенный в разделе 3.

В таблице 3 показаны результаты диагностирования отказовых ситуаций различной кратности рассматриваемой системы. В столбцах q_1 , q_2 , q_3 таблицы 3 приведены числа неисправных компонентов 1-го, 2-го и 3-го типов в отказовой ситуации, номер которой показан в левом столбце; в столбце, озаглавленном «Неисправные компоненты», приведены номера отказавших компонентов; в столбце «Диагноз» – номера компонентов, подозреваемых в отказе при дешифрации результатов диагностирования этой отказовой ситуации, и комментарии. В клетках на пересечении столбца «Диагноз» и строк 20, 23 (с неоднозначным диагнозом) знак \vee использован для разделения возможных результатов диагноза данной отказовой ситуации.

1) В строках 8–24 таблицы 3 показаны примеры некоторых отказовых ситуаций различной мощности t , для которых получен неполный или неоднозначный диагноз.

Диагноз отказовых ситуаций строк 8–11, 13 отнесен к неполному, так как в соответствующих отказовых ситуациях остались не идентифицированными технические состояния компонентов 13, 13, 10, 10, 4 и 8, что отмечено знаком тильда (\sim) над этими номерами.

Поясним результаты диагноза для ситуации 8. Структура рассматриваемой ЦС такова, что для проверки любого коммутатора нужно два исправных абонента из четырех, присоединенных к данному коммутатору. К коммутатору 13 присоединены абоненты 2, 4, 5, 6, при неисправности трех абонентов (в отказовой ситуации 8 – это 2, 4, 6) для коммутатора 13 нельзя найти ни одной проверки с двумя исправными абонентами из этих

четырёх. При диагностировании ситуации 8 коммутатор 13 тестируется в составе подсистемы [(5–13), 13, (13–4), 4] (строка T_{53} в таблице 1) вместе с неисправным абонентом 4; по результатам тестирования эта подсистема получает оценку 1. При последующем раскодировании оценок тестирования коммутатор 13 исключается из числа подозреваемых в неисправности, так как входящий в эту проверяемую подсистему абонент 4 получает также единичные оценки в составе подсистем, где коммутаторы исправны.

Поэтому коммутатор 13 остается не идентифицированным, что отмечено тильдой в клетке «Диагноз» для ситуации 8. В строке 9 приведены результаты диагностирования для ситуации, отличающейся от ситуации 8 добавлением неисправного коммутатора 13. Одинаковый диагноз для отказовых ситуаций 8 и 9 подтверждает факт влияния трех неисправных абонентов на отсутствие идентификации состояния соответствующего коммутатора (коммутатора 13 в примерах ситуаций 8 и 9).

Аналогичный диагноз (не идентифицирован коммутатор 10) показан для отказовых ситуаций 10 и 11, включающих 3 неисправных абонента: 1, 3, 6.

Таблица 3. Таблица результатов диагностирования

| NN п/п | q_1 | q_2 | q_3 | Неисправные компоненты | Диагноз |
|-----------|-------|-------|-------|---------------------------|--------------------------|
| 1 | 1 | 1 | 0 | 7, 14 | 7, 14 |
| 2 | 2 | 0 | 0 | 4, 6 | 4, 6 |
| 3 | 2 | 1 | 0 | 1, 6, 11 | 1, 6, 11 |
| 4 | 0 | 2 | 0 | 8, 13 | 8, 13 |
| 5 | 0 | 3 | 0 | 8, 12, 14 | 8, 12, 14 |
| 6 | 1 | 2 | 0 | 4, 8, 13 | 4, 8, 13 |
| 7 | 2 | 2 | 0 | 4, 6, 10, 12 | 4, 6, 10, 12 |
| 8 | 3 | 0 | 0 | 2, 4, 6 | 2, 4, 6 13 – неполный |
| 9 | 3 | 1 | 0 | 2, 4, 6, 13 | 2, 4, 6 13 – неполный |

| NN п/п | q_1 | q_2 | q_3 | Неисправные компоненты | Диагноз |
|-----------|-------|-------|-------|---------------------------|--|
| 10 | 3 | 1 | 0 | 1, 3, 6, 11 | 1, 3, 6, 11 10 – неполный |
| 11 | 3 | 2 | 0 | 1, 3, 6, 11, 14 | 1, 3, 6, 11, 14 10, – неполный |
| 12 | 1 | 3 | 0 | 2, 10, 13, 14 | 2, 10, 13, 14 |
| 13 | 2 | 3 | 0 | 1, 2, 11, 13, 14 | 1, 2, 11, 13, 14 4, $\bar{8}$ – неполный |
| 14 | 2 | 3 | 0 | 1, 2, 10, 12, 14 | 1, 2, 10, 12, 14 |
| 15 | 0 | 4 | 0 | 8, 11, 13, 14 | 8 • 11 • 13 • 14 4 – неполный |
| 16 | 0 | 1 | 1 | 9, (9–1) | 9 9–1 – неполный |
| 17 | 0 | 1 | 1 | 9, (14–4) | 9 • (14–4) |
| 18 | 1 | 1 | 1 | 1, 9, (8–1) | 1 • 9 8–1 – неполный |
| 19 | 1 | 1 | 1 | 1, 9, (14–4) | 1 • 9 • (14–4) |
| 20 | 2 | 1 | 2 | 1, 7, 9, (8–1), (14–4) | 1 • 7 • 9 • (14–4) ✓ 1 • 7 • (9–2) • (14–4) ✓ 1 • 7 • (9–5) • (14–4) неоднозначный, 8–1 – неполный |
| 21 | 2 | 1 | 2 | 3, 6, 8, (9–2), (9–5) | 3 • 6 • 8 • (9–2) • (9–5) |
| 22 | 0 | 0 | 2 | (9–1), (9–2) | (9–1) • (9–2) |
| 23 | 0 | 0 | 3 | (9–1), (9–2), (9–7) | 9 ✓ (9–1) • (9–2) • (9–7) ✓ (9–1) • (9–2) • (9–5) ✓ (9–1) • (9–5) • (9–7) ✓ (9–2) • (9–5) • (9–7) неоднозначный |
| 24 | 0 | 0 | 3 | (2–8), (2–9), (2–13) | (2–8) • (2–9) • (2–13) |

Подобное положение имеет место, если в отказовую ситуацию входит 4 неисправных коммутатора: из-за особенностей структуры эта четверка блокирует пути к одному из абонентов и исключает возможность его тестирования. Пример – отказовая ситуация 15, в которой коммутаторы 8, 11, 13, 14 блокируют возможность идентификации абонента 4.

Различие диагноза для ситуаций 13 и 14 подтверждает то, что диагноз зависит не только от количества неисправных компонентов, но также от номеров этих компонентов (их взаимного расположения).

2) В процессе проведенных сеансов диагностирования замечено, что для исключения исправного коммутатора из числа подозреваемых при неисправности двух абонентов, которые он соединяет, нужна конкретная проверка, включающая данный коммутатор и 2 исправных абонента из тех четырех, которые присоединены к данному коммутатору. Из четырех абонентов, присоединенных к одному коммутатору, можно образовать 6 различных пар абонентов, каждая из которых требует проверки, отличной от других.

Таким образом, диагностический тест, различающий каждый из 7 исправных коммутаторов от двух неисправных абонентов, которые он соединяет, для данной системы должен включать $P = 7 \times 6 = 42$ различные проверки. Возвращаясь к параметрам структуры, значение P можно записать так:

$$P = N \cdot m(m - 1) / 2 = 7 \cdot 4 \cdot 3 / 2 = 42.$$

Эти 42 различные проверки участвуют при тестировании проверяемых подсистем, показанных в таблице 1.

3) В строках 16–24 приведены отказовые ситуации, в состав которых входит различное количество неисправных линий связи (компонентов 3-го типа).

Хорошо известен факт: последовательно включенные неисправные модуль и линию связи нельзя разделить по результатам одной проверки. В таких ситуациях предполагается, что более вероятна неисправность модуля, в диагноз вставляется номер модуля, который «маскирует» прилегающую неисправную линию связи. С учетом этой гипотезы получен диагноз для

ситуаций 16, 18, 20. В ситуациях 17, 19, 21 неисправные линии связи не прилегают ни к одному из неисправных модулей: результат диагностирования правильный (и полный).

Для уточнения диагноза в подобных ситуациях нужны дополнительные проверки, в которых участвует только один из компонентов проверяемой ситуации.

Если отказовая ситуация включает только неисправные линии связи, то при двух неисправных линиях (строка 22) диагноз правильный (и полный), а при трех неисправных линиях диагноз может быть как правильным (строка 24), так и неоднозначным (строка 23) – в зависимости от расположения этих компонентов. В ситуации 23 три неисправные линии примыкают к общему коммутатору 9. Диагноз этой ситуации с учетом приведенной ранее гипотезы (о приоритете отказа модуля по сравнению с отказом линии связи) оказывается неоднозначным. Диагноз содержит один член длины 1 (коммутатор 9) и 4 члена длины 3: различные сочетания трех из четырех линий связи, примыкающих к коммутатору 9, включая исходную отказовую ситуацию.

Как отмечалось, чтобы различить исправный коммутатор от неисправной линии связи, нужна проверка, включающая данный коммутатор и две исправные линии связи, подходящие к нему. В рассматриваемой ситуации к коммутатору 9 прилегают три неисправные линии связи, а одной оставшейся исправной линии связи недостаточно для организации искомой проверки.

В отличие от проверок коммутатора для проверки абонента с получением правильного диагноза достаточно одной исправной линии, что имеет место в ситуации 24.

4) На основе анализа данных таблицы 3 можно заключить, что достоверный диагноз отказовых ситуаций в ЦС со структурой минимального квазиполного графа размера 7×7 на выбранной системе тестов гарантирован:

– для ситуаций, содержащих неисправные компоненты 1-го и 2-го типов;

– для ситуаций, содержащих только неисправные компоненты 3-го типа;

причем и в том, и другом случаях количество неисправных компонентов должно удовлетворять следующим соотношениям:

$$\begin{aligned} t_{120}^1 &= (q_1 \leq t_1 = 2; q_2 \leq t_2 = 2; q_3 = 0); \\ (3) \quad t_{120}^2 &= (q_1 \leq t_1 = 1; q_2 \leq t_2 = 3; q_3 = 0); \\ t_{003} &= (q_1 = q_2 = 0; q_3 \leq t_3 = 2). \end{aligned}$$

(Отказовые ситуации, подтверждающие эти цифры, показаны в строках 7, 12, 22 таблицы 3.)

Сравнение количеств компонентов первого (q_1), второго (q_2) и третьего (q_3) типов в смеси компонентов, подозреваемых в отказе по результатам диагностирования, с оценками из (3) позволяет оценить достоверность полученных результатов:

– если количества компонентов, подозреваемых в отказе, не превышают соответствующих оценок, то полученный диагноз является правильным;

– если же хотя бы для одного из компонентов количество подозреваемых в отказе превышает соответствующую оценку, то возможно, что диагноз неправилен, нужны дополнительные проверки для его уточнения.

Очевидные и достаточно эффективные действия – замена исправными части компонентов, подозреваемых в неисправности в полученном диагнозе. К вопросу о том, в каком количестве и какие именно компоненты нужно заменить в первую очередь, возможны различные подходы. Первое, что приходит в голову, – заменить компоненты, входящие в состав члена наименьшей длины в полученных вариантах диагноза. Но варианты диагноза отказовой ситуации 23 опровергают правильность этого соображения, если использовать только его.

Диагноз для ситуации 23 содержит один член длины 1 (коммутатор 9) и 4 члена длины 3: различные сочетания трех из четырех линий связи, примыкающих к коммутатору 9, включая исходную отказовую ситуацию). Понятно, что в этом случае замена на исправный коммутатора 9 и повторный запуск системы диагностирования приведет к тем же вариантам диагноза.

Зная, что подобный диагноз является характерным для отказа трех линий связи, примыкающих к одному (общему) коммутатору, напрашивается другой вариант: заменить на исправные 2 линии связи, например, (9-5) и (9-7). Последующий запуск системы диагностирования даст единственный вариант диагноза: отказали линии связи (9-1) и (9-2). Этот диагноз является правильным и

достоверным, что подтверждают записи для отказовой ситуации 22 в таблице 3 и соотношение $q_3 \leq t_{003} = 2$ из (3).

6. Заключение.

Представлен подход к оцениванию величин диагностируемости компонентов (абонентов, коммутаторов и линий связи) цифровых систем (ЦС) с новой структурой – минимального квазиполного графа размера 7×7 с 2 путями между 2 абонентами. Оценки величин диагностируемости компонентов анализируемых ЦС установлены на основе анализа серии результатов диагностирования отказовых ситуаций различной кратности в системе из 7 абонентов и 7 коммутаторов.

Как показал анализ, оценки величин диагностируемости компонентов анализируемых ЦС зависят от ряда факторов:

- структурных особенностей анализируемой ЦС, в частности, от параметра m – числа портов абонентов;
- полноты диагностического теста относительно числа и видов обнаруживаемых неисправных компонентов;
- выбранного метода диагностирования;
- выбранного метода раскодирования результатов выполненных проверок.

Установлено, что выбранные (и показанные в тексте) таблицы проверок, методы диагностирования и раскодирования обеспечивают в рассматриваемой ЦС достоверное правильное однократное (без ремонта) диагностирование не больше:

- 2-х неисправных абонентов и 2-х неисправных коммутаторов, или
- 1-го неисправного абонента и 3-х неисправных коммутаторов, или
- 2-х неисправных линий связи.

Приведенные величины оценок определены для «вредных» (как правило, смежных по взаимному расположению) неисправностей различных компонентов. Это не означает, что не могут быть получены правильные диагнозы при более разбросанном размещении неисправных компонентов по структуре ЦС.

Приведен пример диагностирования кратного отказа двух абонентов, коммутатора и двух линий связи.

Литература

1. ВЕДЕШЕНКОВ В.А., КУРАКО Е.А., ЛЕБЕДЕВ В.Н. *О диагностировании цифровых систем со структурой минимального квазиполного графа размера 7×7* // Проблемы управления. – 2014. – №6. – С. 68–76.
2. ВЕДЕШЕНКОВ В.А. *Алгебро-логический метод дешифрации результатов тестирования цифровых систем* // Сборник научных трудов «Международной научно-практической конференции «Актуальные вопросы образования и науки». Тамбов: ООО «Консалтинговая компания Юком», 2014, Часть 2. – С. 33–36.
3. ВЕДЕШЕНКОВ В.А. *Организация диагностирования цифровых систем со структурой симметричного двудольного графа* // Проблемы управления. – 2009. – №6. – С. 59–67.
4. ВЕДЕШЕНКОВ В.А., НЕСТЕРОВ А.М. *О двух методах дешифрации результатов диагностирования цифровых систем* // Электронное моделирование. – 1981. – Т. 3, №2. – С. 53–58.
5. КАРАВАЙ М.Ф., ПОДЛАЗОВ В.С. *Распределенный полный коммутатор как «идеальная» системная сеть для многопроцессорных вычислительных систем* // Управление большими системами. – 2011. – №34. – С. 92–116.
6. КАРАВАЙ М.Ф., ПАРХОМЕНКО П.П., ПОДЛАЗОВ В.С. *Комбинаторные методы построения двудольных однородных минимальных квазиполных графов (симметричных блок-схем)* // Автоматика и телемеханика. – 2009. – №2. – С. 153–170.
7. ЛОБАНОВ А.В. *Обнаружение и идентификация неисправностей в распределенных управляющих вычислительных системах с программно-управляемой сбое- и отказоустойчивостью* // Автоматика и телемеханика. – 1998. – №1. – С. 155–164.

8. ALVERSON R., ROWETH D., KAPLAN L. *Cray XC® Series Network* [Электронный ресурс] – URL: <http://www.cray.com/Assets/PDF/products/xc/CrayXC30Networking.pdf>. (дата обращения 18.11.2015)
9. PREPARATA F.P., METZE G., CHIEN R.T. *On the connection assignment problem of diagnosable systems* // IEEE Trans. Electr. Comput. – 1967. –Vol. 16, №6. – P. 848–854.

ON DIAGNOSABILITY OF COMPONENTS IN DIGITAL SYSTEMS WITH STRUCTURE OF MINIMAL QUASICOMPLETE 7×7 GRAPH WITH 2 PATHS CONNECTING 2 TERMINALS

Victor Vedeshenkov, Institute of Control Sciences of RAS, Moscow, Doctor of Science, senior researcher associate (Moscow, Profsoyuznaya st., 65, vva@ipu.ru).

Eugene Kurako, Institute of Control Sciences of RAS, Moscow, researcher associate (Moscow, Profsoyuznaya st., 65, kea@ipu.ru).

Vitaly Lebedev, Institute of Control Sciences of RAS, Moscow, Cand.Sc., associate professor (Moscow, Profsoyuznaya st., 65, lebvini@ipu.ru).

Abstract: An approach is suggested to estimate diagnosability of components (a terminal, a router, a communication line) of digital systems (DS) with the novel structure of minimal quasicomplete graph for 7 terminals and 7 routers with 2 paths between 2 terminals. The analysis of typical failures in the system allows proposing estimated component diagnosability metrics. It is shown that diagnosing techniques and methods used guarantee the correct diagnosis for no more than: two faulty terminals and two faulty routers, or one faulty terminal and three faulty routers, or two broken communication lines.

Keywords: digital systems, minimal quasicomplete graph by dimension 7×7 , abonent, commutator, communication line, diagnosability.

Статья представлена к публикации членом редакционной коллегии М.Ф. Караваем

Поступила в редакцию 15.04.2015.

Опубликовано 30.11.2015.

УДК 021.8 + 025.1
ББК 78.34

О ТОПОЛОГИЧЕСКОЙ МАСШТАБИРУЕМОСТИ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ¹

Мелентьев В. А.²

(ФГБУН Институт физики полупроводников
им. А.В. Ржанова СО РАН, Новосибирск)

Большая часть опубликованных в научной литературе результатов, посвященных исследованию масштабирования параллельных задач и систем, локализованы под конкретные их реализации. Эксклюзивность, обусловленная различиями в классах решаемых задач, в техническом, технологическом и топологическом воплощении систем, во-первых, не позволяет напрямую использовать эти результаты в анализе и синтезе иных параллельных систем и задач, и во-вторых, не дает цельной картины взаимной обусловленности заданных параметров системы с востребованными показателями ее функционирования. На основе предложенной в настоящей работе модели предпринята попытка восполнить имеющийся в этом отношении пробел, и прежде всего, оценить влияние топологии на масштабируемость параллельных систем и решаемых на них задач.

Ключевые слова: топологии, сетевые технологии систем, масштабируемость параллельных вычислительных систем и задач.

1. Введение

Понятие масштабируемости может быть отнесено к любой системе как свойство, характеризующее зависимость критиче-

¹ Работа выполнена при поддержке Российского фонда фундаментальных исследований, проект №14-07-00169.

² Виктор Александрович Мелентьев, кандидат технических наук (melva@isp.nsc.ru)

ски важных (существенных) показателей качества функционирования от числа составляющих ее конструктивно и/или функционально самостоятельных элементов. При этом потребности потенциальных пользователей системы определяют множество оцениваемых показателей, а составные элементы, способы их соединения и взаимодействия – вариативное множество «измерений», в которых следует оптимизировать эти показатели. Как правило, одним из основных показателей качества вычислительной системы (ВС) считается эффективность использования всей совокупности ее процессоров. Однако такая эффективность представляет интерес прежде всего для собственников системы. Соответствие системы потребностям конкретных пользователей оценивают эффективностью использования не всех, а только тех процессоров ВС, которые задействованы в образуемых для реализации соответствующих приложений подсистемах.

Достижение наивысших значений эффективности использования компонентов системы далеко не всегда является целью оптимизации ее архитектуры или прикладных алгоритмов. Например, общесистемным задачам непременно сопутствует требование оперативности их решения, а актуализация приложений является причиной создания специализированных систем; в коммерческих системах реального времени определяющим фактором также часто является оперативность пользовательских приложений, даже в ущерб эффективности использования ресурсов. Однако в этих случаях тоже можно говорить об эффективности, но уже не в части использования ресурсов, а в отношении повышения актуальности получаемых результатов.

Целью данной работы является формализация обусловленной топологией вычислительной системы зависимости ее потенциальных в отношении параллелизма возможностей при горизонтальном масштабировании¹. Естественно, что исследование такой топологической обусловленности предполагает абстрагирование от ограничений, связанных с присутствием в

¹ Под горизонтальным масштабированием ВС понимается наращивание ее возможностей увеличением числа вычислительных модулей.

параллельном алгоритме скалярных (нераспараллеливаемых) фрагментов, учитываемых классическим вариантом закона Амдала, названного в [5] «бичом параллельных вычислений». При исключении из рассмотрения фактора «скалярности» основным обстоятельством, определяющим реальную производительность вычислительной системы, становится ее зависимость от межпроцессных обменов, связанных как с топологией системной сети, так и с ее технологическими характеристиками – темпом выдачи сообщений, задержкой, пропускной способностью, и эта зависимость тем существеннее, чем больше размер системы [1].

В работе предложена модель параллельных вычислений, на базе которой проведено размежевание топологических и технологических факторов по их влиянию на реальную производительность системы и на предельный (при директивных критериях эффективности) порядок подсистем; даны формальные определения функций топологической масштабируемости параллельных задач и системы, основанные на использовании показателей плотности графа ВС с достижимостью, лимитированной используемой в системе сетевой технологией, параметрами задачи и заданными критериями эффективности их решения.

2. Описание модели масштабируемой ВС

Понятно, что оценка влияния сетевой топологии на потенциальную масштабируемость задач в масштабируемой ВС будет правомерной, если количественный критерий оценки такого влияния не меняет относительную упорядоченность сопоставляемых топологий при допускаемых предназначением ВС изменениях класса решаемых задач и обрабатываемых данных. В принципе, сетевая топология ВС может быть спроектирована под определенную задачу в смысле биективности ее информационно-логической структуры физической структуре системы – это позволяет добиться достаточно высоких показателей качества таких систем, но только при решении ими соответствующих специализации задач. Отметим также, что специализация систем и эффективность их использования в задачах, выходя-

щих за рамки специализации, антагонистичны по своей сути. Поэтому случай узкой специализации в данной работе мы не затрагиваем и далее будем вести речь о системах универсального назначения, пригодных для решения достаточно широкого класса задач и увеличивающих, таким образом, круг потенциальных потребителей и коммерческую эффективность их использования.

Известная в научной литературе концепция неограниченно-параллелизма, с вербальным описанием которой можно ознакомиться, например, в [3] и [4], вводит ряд допущений в свойства параллельной системы в целом, не разделяя при этом свойства распараллеливаемого приложения и свойства реализующей его вычислительной системы. Для выявления влияния на параллелизм общесистемных архитектурных особенностей ВС (в нашем случае – топологии) мы далее абстрагируемся от приложений, считая их неограниченно распараллеливаемыми при том, что все ограничения параллелизма обусловлены только архитектурой вычислительной системы.

Из обобщенного формального описания такой постановки выделим часть, отнесенную только к задаче. Здесь W и w – измеряемые временем объемы вычислений при решении произвольной задачи на одном и на p процессорах вычислительной системы. Соответствующие числу задействованных процессоров p объемы подлежащих обмену данных обозначим через Q и q , измеряя их при этом информационными единицами (байтами):

1. Задача допускает разбиение на произвольное число p информационно-связанных параллельных ветвей, $1 \leq p \leq \infty$. Информационный граф W_p распараллеленной на p ветвей задачи может быть нерегулярным, но обязательно связан.

2. Масштабирование данных в задаче с коэффициентом m увеличивает объем вычислений W и объем Q подлежащих обмену данных в m раз.

3. Общий объем вычислений W и объем Q подлежащих обмену данных при разбиении задачи на p параллельных ветвей не зависят от числа процессоров p и распределяются по ним равномерно: $w = W/p$ и $q = Q/p$.

4. Параллельный алгоритм не содержит скалярных фрагментов – это свойство вытекает из сформулированных выше п.п. 1 и 3.

Понятно, что если бы в системе могла быть реализована полносвязная топология с реактивностью сетевой технологии¹, сравнимой с быстродействием процессора, то это соответствовало бы концепции неограниченного параллелизма системы в целом, и проблемы наращивания быстродействия ВС не возникало бы. Однако, во-первых, полная связность процессоров при неограниченном наращивании числа процессоров реализована быть не может как из экономических соображений, так и из-за отсутствия технической возможности неограниченного наращивания числа портов в процессорах. Во-вторых, из чисто физических ограничений (скорости света и несравнимости расстояний внутри процессора и между многими процессорами системы) даже предельное быстродействие NT не будет адекватным быстродействию процессора и его оперативной памяти, – к примеру, возможности высокоскоростной передачи данных по медным проводникам ограничены 40 Гбит/с, что соответствует единицам сантиметров на плате и примерно одному метру при передаче по кабелю [11]. Поэтому сдерживающим фактором линейного наращивания вычислительной мощности является наличие задержек в реализации межпроцессорных обменов. В связи с этим данный выше перечень свойств задачи с неограниченным параллелизмом дополним свойствами реализующей ее вычислительной системы, влияющими на величину таких задержек:

5. Все процессоры системы идентичны, их общее число n достаточно для реализации на них p параллельных ветвей, и первоначальное предварительное распределение входных данных по задействованным в параллельном приложении процессорам не требуется.

6. Топология ВС является регулярной² и неполносвязной.

¹ Далее в тексте статьи словосочетание «сетевая технология» заменено аббревиатурой NT (*Network Technology*).

² Степени всех вершин графа ВС одинаковы.

7. Общие объемы W и Q не зависят от топологии сети связи и от используемой NT , и ограничения на минимальные объемы w и q отсутствуют.

8. Вычислительные и коммуникационные элементы ВС допускают совмещенную во времени работу.

9. Временные затраты на обмены пропорциональны расстояниям между информационно-связанными задачей вершинами графа ВС.

10. Совокупность используемых в вычислительной системе топологии и NT гарантирует отсутствие сетевых коллизий и связанных с ними задержек.

Перечисленные здесь свойства неограниченно распараллеливаемой задачи (п.п. 1-4) и свойства (п.п. 5-10) предназначенной для решения таких задач вычислительной системы в совокупности определяют используемую далее модель параллельных вычислений.

3. Технологический и топологический аспекты масштабируемости параллельных вычислений

Предпоследнее из свойств используемой нами модели параллельных вычислений является тривиальным в практике построения и использования коммуникационных сетей: как правило, время передачи сообщений между наиболее удаленными элементами оценивают диаметром соответствующих графов [10], характеризующим коммуникационную задержку в худшем случае [9]. Учитывая это, представим функцию $T_{ND}(p)$ сетевых задержек (ND – Network Delay) произведением определяемой топологией функции расстояния $L(p)$ между наиболее удаленными в графе ВС и задействованными в решении задачи процессорами на определяемую NT и зависящую¹ от $q = Q/p$ функцию задержки $t_{NT}(q)$: $T_{ND}(p) = L(p) \times t_{NT}(q)$. Отметим при этом, что неперенным свойством функции $t_{NT}(q)$ является ее обратная от p и прямая от Q зависимость: $p_1 < p_2 \Rightarrow t_{NT}(Q/p_1) > t_{NT}(Q/p_2)$ и $q_1 > q_2 \Rightarrow t_{NT}(q_1) > t_{NT}(q_2)$. Применение в ВС технологий с иной зависимостью функции $t_{NT}(q)$

¹ В соответствии с п. 3 используемой здесь концепции.

противоречило бы основной цели распараллеливания – достижению требуемой оперативности в реализации пользовательских приложений и требуемой их достоверности, связанной с повышением сложности алгоритмов и/или объемов обрабатываемых данных. Однако, к сожалению, увеличение числа процессоров p , приводящее к пропорциональному уменьшению удельного объема q обмениваемых данных, приводит к уменьшению элементарного (на единичном расстоянии) времени $t_{NT}(q)$ информационного взаимодействия физически смежных процессоров не в тех же, а в зависимых от применяемой в системе NT пропорциях, т.е. с некоторым коэффициентом непропорциональности k . К примеру, при коммутации сообщений функция элементарных задержек от числа процессоров p является гиперболической: $t_{NT}(q) = \alpha + q / \beta$, здесь α – латентность сети, β – ее пропускная способность [6], а $q = Q / p$ – объем обмениваемых данных между двумя информационно смежными процессорами. Из п. 3, в соответствии с которым $\lim q = 0$, и из определения латентности [12] как времени, затрачиваемого на передачу информационного сообщения с нулевым объемом данных, ясно, что ее значение является горизонтальной асимптотой в графике гиперболической функции $t_{NT}(p)$ при заданном объеме Q . Подобная этой зависимость $t_{NT}(p)$, хоть и в разной степени, но справедлива не только для рассмотренного здесь в качестве примера метода коммутации сообщений, но и для других используемых в современных сетях связи ВС сетевых технологий.

Рассмотрим общий случай масштабирования задачи, включающий в себя масштабирование числа p используемых ею процессоров с коэффициентом $k_p \geq 1$ и масштабирование этой задачи «по данным», увеличивающее общий объем обмениваемых данных Q в $m \geq 1$ раз. При этом элементарное время задержки

$$t_{NT} \left(\frac{m \cdot q}{k_p \cdot p} \right),$$

в соответствии с оговоренным выше свойством непропорционального его изменения и в зависимости от диапазона изменения обобщающего эти виды масштабирования коэффициента $k_{mp} = m / k_p$, определим неравенствами:

$$k_{mp} > 1 \Rightarrow t_{NT}(q) < t_{NT}(k_{mp} \cdot q) < k_{mp} \cdot t_{NT}(q),$$

$$(1) \quad k_{mp} = 1 \Rightarrow t_{NT}(q) = t_{NT}(k_{mp} \cdot q),$$

$$k_{mp} < 1 \Rightarrow k_{mp} \cdot t_{NT}(q) < t_{NT}(k_{mp} \cdot q) < t_{NT}(q).$$

Неравенства учитывают непропорциональность изменения времени задержки $t_{NT}(q)$, заданного общим объемом обмениваемых данных Q и числом задействованных в решении задачи процессоров p , от масштабирования как этого объема Q , так и числа процессоров p . Эта непропорциональность обусловлена сетевой технологией, например, для рассмотренной выше NT с коммутацией сообщений – обратной пропорциональностью от числа процессоров p элементарного объема q и присутствием во времени постоянной составляющей – латентности α . Введем в неравенства (1) коэффициент

$$k_{NT} = k_{mp} \cdot t_{NT}(q) / t_{NT}(k_{mp} \cdot q),$$

учитывающий присущую используемой в системе NT непропорциональность, назовем этот коэффициент технологическим и определим диапазоны его изменения при изменении k_{mp} :

$$k_{mp} > 1 \Rightarrow k_{NT} > 1,$$

$$(2) \quad k_{mp} = 1 \Rightarrow k_{NT} = 1,$$

$$k_{mp} < 1 \Rightarrow k_{NT} < 1.$$

Из данного выше формального определения технологического коэффициента k_{NT} видно, что его величина определена исходным (до масштабирования) элементарным объемом $q = Q / p$, и так как $k_{mp} > t_{NT}(q) / t_{NT}(k_{mp} \cdot q)$ – на эту непропорциональность мы обращали внимание выше, – то k_{NT} находится в прямой (но не в прямо пропорциональной) зависимости от масштабного коэффициента k_{mp} . К примеру, технологический коэффициент уже рассмотренной выше NT с коммутацией сообщений –

$$k_{NT} = \frac{k_{mp} \cdot t_{NT}(q)}{t_{NT}(k_{mp} \cdot q)} = \frac{k_{mp}(\alpha + q / \beta)}{\alpha + k_{mp} \cdot q / \beta} = \frac{k_{mp}(\alpha \beta + q)}{\alpha \beta + k_{mp} \cdot q}.$$

Однако далее мы абстрагируемся от исходных значений p и q и ограничимся использованием определенных выражением (2) качественных оценок k_{NT} и k_{mp} , тем не менее достаточных для

установления взаимозависимости топологической и технологической составляющей архитектуры ВС.

Учитывая вытекающее из п. 8 используемой нами модели неравенство $T_{ND}(p) > w$ и считая обмены максимально совмещенными с вычислениями, определим время фактических задержек разностью $T_{ND}(p) - w$. Модифицируя при этом известный закон Амдала, получим:

$$(3) \quad S_p = \frac{W}{w + (L_S(p) \cdot t_{NT}(q) - w)} = \frac{W}{L_S(p) \cdot t_{NT}(q)}.$$

Тогда максимально допустимое при решении задачи с параметрами W и q расстояние $L_S(p) \geq 1$, обусловленное директивным (предписанным) при заданном числе процессоров p ускорением S_p , определится из

$$(4) \quad L_S(p) = \frac{W}{S_p \cdot t_{NT}(q)}.$$

Отметим, что максимальное для системы с некоторой NT ускорение $S_p = W / t_{NT}(q)$ в решении задачи с объемом Q обмениваемых данных может быть получено при предельно допустимом расстоянии $L_S(p)$ между информационно смежными процессорами, равном единице. Такое могло бы быть реализовано только при использовании в ВС полносвязной топологии – в этом случае информационная топология задачи не имеет значения – или при использовании в ВС топологии, повторяющей информационную топологию задачи с точностью до изоморфизма. Из выражений (3) и (4), учитывая известную для используемой NT зависимость времени задержки t_{NT} от числа процессоров p , несложно определить и нижнее граничное значение этого числа $p_S(L)$ для заданного ускорения S_p при изменении L .

Если для пользователя ВС приоритетным является не ускорение S_p , а эффективность $E_p = S_p / p$ использования задействованных в решении задачи p процессоров, то

$$(5) \quad E_p = \frac{W}{p \cdot L_E(p) \cdot t_{NT}(q)},$$

и предельное расстояние $L_E(p)$ для данной такой эффективности E_p составит

$$(6) \quad L_E(p) = \frac{W}{p \cdot E_p \cdot t_{NT}(q)}.$$

Учитывая, что сохранение прежней эффективности $E_{p+\Delta p} = E_p$ с помощью уменьшения расстояния $L_E(p + \Delta p) < L_E(p)$, компенсирующего увеличение $p + \Delta p$, невозможно из-за относительной несоизмеримости изменения этих величин — $(p + \Delta p) / p > L_E(p) / L_E(p + \Delta p)$, становится понятным, что связанное с потребностью повышения оперативности в решении задач увеличение числа p параллельных ветвей неизбежно ведет к снижению эффективности E_p использования процессоров — это, как отмечено выше, обусловлено непропорциональным к увеличению p уменьшением $t_{NT}(Q/p)$: $\Delta p > 0 \Rightarrow \Rightarrow (p + \Delta p) \cdot t_{NT}(Q/(p + \Delta p)) > p \cdot t_{NT}(Q/p)$ с соответствующим изменением знаменателей в (5) и (6).

Факт снижения эффективности при масштабировании задач подтвержден многочисленными экспериментальными исследованиями (например, в [13]), теоретическое же его обоснование нетрудно увидеть из рассмотрения полносвязной топологии, где $\forall p > 1 L(p) = 1$. Таким образом, определенное из (5), (6) при требуемой эффективности E и при заданном предельном расстоянии L число ветвей $p_E(L)$ параллельной (W, Q) -задачи в системе, использующей сетевую технологию с характерной для нее функцией задержки $t_{NT}(Q/p)$, является верхним пределом распараллеливания.

Учитывая это и то, что нижний предел p_S распараллеливания обусловлен заданным ускорением S , делаем вывод о том, что распараллеливание задачи с требуемыми значениями S и E возможно только если при обусловленном сетевой технологией ВС расстоянии L между информационно смежными в ней процессорами выполняется неравенство $p_S(L) \leq p_E(L)$, а топологически обеспечиваемое при этом L число процессоров p находится между значениями p_S и p_E — $p_S(L) \leq p(L) \leq p_E(L)$.

В дополнение к вышесказанному еще раз обратим внимание на различную природу обусловленности расстояния L и числа процессоров p в (W, Q) -задаче: если предельное расстояние $L(p)$ при заданном p обусловлено функцией $t_{NT}(Q/p)$, т. е. используемой в ВС сетевой технологией, то предельное число $p(L)$ процессоров, которое может быть задействовано при допуске задаче расстояния L , обусловлено только топологически и

зависит от информационной топологии задачи и физической топологии системы.

Отмеченную здесь инвариантность обусловленных директивными значениями E_p и S_p предельных расстояний $L_E(p)$ и $L_S(p)$ к топологии не следует понимать как гарантированную произвольной топологией ВС обеспеченность этих расстояний при заданном p и при прочих равных условиях – равенстве объемов W вычислений и Q обмениваемых в задачах данных, равенстве критериев эффективности реализации этих задач и идентичности используемой в системах NT . Инвариантность к топологии расстояний $L_E(p)$ или $L_S(p)$ состоит в том, что, во-первых, они определены вне зависимости от топологии ВС, и, во-вторых, они определяют общее для любой топологии требование, заключающееся в следующем: успешное (в смысле обеспечения критериев эффективности, предписанных задаче с параметрами W и Q) распараллеливание на p процессоров любой задачи возможно лишь тогда, когда топология системы гарантирует хотя бы одно вложение (W, Q) -задачи, при котором расстояния между информационно смежными процессорами в соответствующем этой задаче подграфе ВС не превышают этих расстояний. Отметим здесь интуитивно понятную прямую зависимость успешности (в данном выше смысле) вложения от величины расстояния $L(p)$: большему допускаемому расстоянию соответствует и большая вероятность наличия в нем подграфов, допускающих удовлетворяющие заданным требованиям вложения. Это объяснимо хотя бы тем, что с увеличением предельно допускаемого расстояния возрастает число удовлетворяющих этому пределу вершин графа, естественно при этом, что кратность возрастания числа таких вершин должна быть выше кратности масштабирования k_p , но этот момент уже определяется топологиями задачи и системы.

Рассмотрим, каким образом технологическая составляющая ВС влияет на масштабирование задачи с коэффициентом $k_{mp} = m / k_p$, где коэффициент $m \geq 1$ соответствует увеличению объема Q в m раз, а $k_p \geq 1$ – увеличению числа задействованных в решении этой задачи процессоров в пределах n ($p \cdot k_p \leq n$), где n – общее число процессоров в системе (порядок графа ВС). Дополнительно к уже использованным ранее обозначениям

ускорения S_p , эффективности E_p и соответствующих этим критериям предельно допустимых при заданном числе процессоров p расстояний $L_E(p)$ и $L_S(p)$ введем обозначения ускорения S_{mp} , эффективности E_{mp} и соответствующих расстояний $L_{S_{mp}}(p)$ и $L_{E_{mp}}(p)$ при масштабированных с коэффициентом m данных и масштабированном с коэффициентом k_p исходном числе процессоров p . В соответствии с п. 2 изложенной в разделе 2 модели объем вычислений W при этом также увеличивается в m раз. Тогда из (6), оставив прежнюю (до масштабирования) эффективность $E_{mp} = E_p$ и используя введенный выше технологический коэффициент $k_{NT} = k_{mp} \cdot t_{NT}(q) / t_{NT}(k_{mp} \cdot q)$, получим

$$(7) \quad L_{E_{mp}}(p) = \frac{m \cdot W}{k_p \cdot p \cdot E_{mp} \cdot t_{NT}(k_{mp} \cdot q)} = \frac{W \cdot k_{NT}}{p \cdot E_p \cdot t_{NT}(q)} = k_{NT} \cdot L_E(p).$$

Если масштабирование числа задействованных в решении задачи процессоров превалирует над масштабированием данных и $k_p \geq m$, то в соответствии с (2) $k_{NT} < 1$, и из (7) – $L_{E_{mp}}(p) < L_E(p)$. Таким образом, сохранение эффективности при масштабировании задачи с кратностью увеличения объема данных m , меньшей кратности увеличения числа процессоров k_p , потребует обеспечения в k_{NT}^{-1} раз меньшего $L_E(p)$ предельно допустимого расстояния на в k_p раз большем множестве информационно смежных процессоров. Понятно, что явная противоречивость этого предъявляет более жесткие требования к топологии системы, каковой в условиях неограниченного (при $k_p \rightarrow \infty$) масштабирования может быть только полносвязная топология.

Если же кратность m масштабирования данных в задаче более или равна кратности масштабирования k_p числа задействованных при ее решении процессоров – $k_{mp} \geq 1$, то в соответствии с (2) $k_{NT} \geq 1$, при этом из (7) следует $L_{E_{mp}}(p) \geq L_E(p)$, и превышение $L_{E_{mp}}(p)$ над $L_E(p)$ будет тем большим, чем больше m превышает k_p . Как видим, здесь, в отличие от предыдущего случая, сохранению прежней эффективности при увеличении числа задействованных в решении задачи процессоров способствует возможность увеличения предельно допустимого расстояния между ними. Естественно, потенциал такого увеличения для единственной задачи определялся бы близостью (в лучшем случае – изоморфностью) топологии системы к топологии раз-

мещаемой в ней задачи. Для множества же решаемых в системе задач потребуется оптимизация топологии ВС под все задачи этого множества с учетом присущих им объемов вычислительных W и обменных Q операций, а также с учетом ожидаемых пределов масштабирования этих задач.

Из изложенного ясно, что сетевая технология играет существенную роль в организации эффективного функционирования ВС. Для каждой из решаемых в системе параллельных (W, Q)-задач, характеризующихся вычислительной сложностью W и общим объемом Q обменных взаимодействий, вне зависимости от их информационных топологий и от физической топологии ВС сетевая технология определяет потребную для достижения директивной эффективности решения задачи функциональную зависимость между числом задействованных процессоров p и предельно допустимым расстоянием $L(p)$ между ними. Отвечающая такой зависимости топология, позволяющая в рамках заданного набора из N решаемых в ВС задач с заданными рангами p_i ($i = 1, \dots, N$) их распараллеливания конфигурировать подсистемы с меньшими $L(p_i)$ расстояниями между информационно смежными процессорами, дает возможность снизить требования к технологической составляющей системы – для осознания этого факта достаточно рассмотреть предельный случай использования полносвязного графа ВС. Поэтому, учитывая, что отнесенное к i -й задаче ранга p_i расстояние $L(p_i)$ является предельно допускаемым в рассматриваемой параллельной системе, топологическую ее составляющую имеет смысл характеризовать адекватностью топологии ВС требованиям, предъявляемым одной или набором из N решаемых в ней задач в обеспечении соответствующих этим задачам предельных расстояний $L(p_i)$.

Таким образом, обусловленное директивной для отдельной задачи эффективностью предельное расстояние может быть обеспечено выбором соответствующей этой задаче топологии. Если же задач, определяющих предназначение системы, некоторое множество, то неизбежно встает вопрос о выборе топологии, оптимизированной под это множество. Однако даже при самой

лучшей оптимизации никакая, кроме полностью связной¹, топология при ее статической организации не способна обеспечить изоморфную вложимость всех топологически разнотипных² задач из набора, заданного ориентацией системы. Поэтому даже без учета предполагаемого в перспективе масштабирования задач необходимость организации опосредованных (транзитных) взаимодействий информационно смежных в задачах процессоров становится неизбежной, и расстояния между ними зависят от степени адекватности топологии задачи топологии системы. Масштабирование системы – увеличение общего числа процессоров n системы – приводит к такому же (в смысле увеличения расстояний) результату в связи со стремлением максимального их использования в решении той или иной задачи из набора (масштабирование задач). Кроме того, связанное с масштабированием задачи наращивание числа вершин ее информационного графа может сопровождаться ростом его степени и снижением вероятности изоморфного его вложения в граф ВС вплоть до нулевой при превышении степенью первого степени второго. Добавим к этому, что увеличение числа p задействованных в решении задачи процессоров неполиномиально увеличивают сложность и время реализации алгоритмов ее изоморфного размещения в системе, и при числе процессоров, большем десяти, это побуждает к замене точных методов приближенными, со свойственным им полярным отношением достоверности и актуальности.

Итак, потенциальные возможности масштабирования задачи, свободные от ограничений, вносимых используемыми в системе сетевой технологией и топологией, определяются общим числом n процессоров и возможностью наращивания их числа. Существенная, в сравнении с вычислительными операциями, продолжительность информационных взаимодействий ограничивает этот потенциал: число p задействованных в решении задачи процессоров не может быть вне диапазона значений,

¹ В полностью связной ВС расстояние между любыми двумя вершинами не зависит от числа процессоров и является единичным.

² Здесь имеется в виду, что, информационно-логическая структура таких задач представлена существенно отличающимися графами.

определенных из (4), (6) и соответствующих директивным значениям ускорения и эффективности. Для узкоспециализированных ВС проблема адекватного вложения единственной задачи может быть решена выбором соответствующей этой задаче топологии. Расширение набора топологически разнотипных задач, определяющих проблемную ориентацию ВС, и потребности обеспечения возможностей эффективного их масштабирования в наращиваемой системе не укладываются в рамки какой-либо одной, кроме полносвязной, топологии. Экономическая нецелесообразность и техническая неосуществимость реализации такой топологии ограничивают возможности эффективного масштабирования как отдельных решаемых в системе задач, так и системы в целом.

В связи с теоретической и физической невозможностью реализации в ВС полносвязной топологии, в работе [7] впервые представлено решение проблемы вложения задач в постановке, основанной на замене отношений смежности вершин графа ВС отношениями лимитированной их достижимости – при этом расстояния между вершинами графа подсистем, образуемых для решения распараллеленных на p ветвей задач, ограничены определяемыми из (4), (6) предельно допускаемыми расстояниями.

4. Вложение параллельных задач с учетом лимитированной достижимости

Как показано в предшествующем разделе, соотношение между объемами вычислений W и информационных взаимодействий Q в параллельной задаче, заданные критерии эффективности ее решения и быстродействие используемой в системе NT определяют взаимозависимость числа p задействованных процессоров и предельно допускаемого при этом p расстояния $L(p)$ между информационно смежными вершинами соответствующего задаче подграфа в графе ВС. Учитывая, что расстояния между вершинами невзвешенного графа $G(V, E)$ определяются числом транзитных участков и могут быть выражены только целыми числами, а также тем, что расстояние между информационно смежными вершинами подграфа не должно превышать

определяемого из (4) и/или (6) предельного при предписанных критериях S_p или E_p и при заданном p значения $L(p)$, предельное расстояние между информационно смежными в задаче процессорами определим целой частью $L(p)^1$ и назовем предельной достижимостью $\hat{\partial}(p)$:

$$(8) \quad \hat{\partial}(p) = \lfloor L(p) \rfloor, \quad 1 \leq \hat{\partial}(p).$$

В соответствии с (4) директивное ускорение S для уменьшенного в сравнении с $L_S(p)$ значения предельной достижимости $\hat{\partial}(p) \leq L(p)$ может быть достигнуто при большей задержке t_{NT} и, соответственно, при уменьшенном в сравнении с p числе процессоров $p_{\hat{\partial}}$. Покажем это на уже использованном примере NT с коммутацией сообщений, обозначив допускаемые при равенстве ускорений $S_p = S_{p_{\hat{\partial}}}$ задержки на соответствующих $L_S(p)$ и $\hat{\partial}(p)$ расстояниях, через $t_{NT}(q)$ и $t_{NT}(q')$. Из

$$t_{NT}(q) = \alpha + \frac{Q}{p \cdot \beta} \text{ и}$$

$$t_{NT}(q') = \alpha + \frac{Q}{p_{\hat{\partial}} \cdot \beta}$$

и из (4) получим минимальные при этих задержках значения числа процессоров:

$$p = \frac{Q \cdot L_S(p) \cdot S_p}{\beta \cdot (W - \alpha \cdot L_S(p) \cdot S_p)} \text{ и}$$

$$p_{\hat{\partial}} = \frac{Q \cdot \hat{\partial}(p) \cdot S_p}{\beta \cdot (W - \alpha \cdot \hat{\partial}(p) \cdot S_p)}.$$

Отсюда нетрудно убедиться в том, что $p_{\hat{\partial}} < p$.

Вершины u и v графа вычислительной системы $G(V, E)$ считаем связанными отношением $\hat{\partial}$ -достижимости² $R_{\hat{\partial}}$ ($\hat{\partial}$ -достижимыми), если расстояние $d(u - v)$ между ними не превышает $\hat{\partial}$: $(u - v) \in R_{\hat{\partial}} \Leftrightarrow d(u - v) \leq \hat{\partial}$. Порожденный множеством вершин $V_{\hat{\partial}} = V$ и матрицей $\hat{\partial}$ -достижимости $R_{\hat{\partial}}$ граф $G_{\hat{\partial}}(V_{\hat{\partial}}, E_{\hat{\partial}})$, $E_{\hat{\partial}} \in R_{\hat{\partial}}$ назван графом $\hat{\partial}$ -достижимости. Очевидно, что превышение степени этого графа над степенью графа $G(V, E)$ тем значительнее, чем больше полученное из (4) или из (6) и из $\hat{\partial}(p) = \lfloor L(p) \rfloor$ значение $\hat{\partial}(p)$. Максимальное по включению подмножество $V_{\hat{\partial}}^* \subseteq V_{\hat{\partial}}$ связанных отношением

¹ Здесь индекс критерия S_p или E_p опущен.

² Здесь и далее зависимость значения достижимости $\hat{\partial}$ от числа процессоров p не указана и предполагается по умолчанию $\hat{\partial} \equiv \hat{\partial}(p)$.

∂ -достижимости R_∂ вершин порождает подграф $G_\partial(V_\partial^*, R_\partial)$, являющийся кликой графа $G_\partial(V_\partial, E_\partial)$, иначе – ∂ -кликкой $K_\partial(G)$ графа $G(V, E)$. Подобно плотности $\varphi(G)$ графа G , определяемой порядком его наибольшей клики [2], порядок наибольшей из максимальных ∂ -клик графа $G(V, E)$ назовем плотностью $\varphi(G_\partial)$ графа ∂ -достижимости, или ∂ -плотностью $\varphi_\partial(G)$ графа $G(V, E)$ – $\varphi(G_\partial) \equiv \varphi_\partial(G)$.

В работе [7] введено понятие ∂ -ограниченной компоненты достижимости (∂ -компоненты) графа. Говоря далее о ∂ -компоненте графа $G(V, E)$, будем иметь в виду данное там определение глобальной компоненты ∂ -достижимости, как подграфа $G^*_{\partial+}(V^*_{\partial+}, E^*_{\partial+})$, порожденного максимальным по включению множеством $V^*_{\partial+} = V^*_{\partial} \cup V^*_+$ вершин, объединяющим подмножество $V^*_{\partial} = \{u \mid \forall v \in V_\partial(u - v) \in R_\partial\}$ попарно ∂ -достижимых вершин с подмножеством $V^*_+ = \{w \notin V^*_{\partial} \mid \exists u, v \in V^*_{\partial} : w \in V_\partial(u - v)\}$ вершин, не являющихся членами подмножества V^*_{∂} , но входящих в состав ∂ -ограниченных путей между его вершинами; здесь $V_\partial(u - v)$ – подмножество вершин, составляющих путь из u в v . Вершины из V^*_{∂} в плане их попарной ∂ -достижимости называем основными; они составляют множество вершин ∂ -клики $K_\partial(G)$ графа $G(V, E)$, и их число (порядок ∂ -клики, порядок глобальной ∂ -компоненты) есть ∂ -плотность $\varphi_\partial(G)$ графа $G(V, E)$ – $\varphi_\partial(G) = |V^*_{\partial}|$. Вершины из V^*_+ – вспомогательные; они обеспечивают ∂ -достижимость основных вершин из V^*_{∂} и, не входя в V^*_{∂} , входят в состав ∂ -путей между ними.

Проблема выявления в графе ВС $G(V, E)$ компоненты ∂ -достижимости (∂ -клики графа G) решена в [7] с помощью проективного описания графа [8]. Суть соответствующего алгоритма состоит в следующем:

1. Строим систему $n = |V|$ ∂ -уровневых проекций $P_\partial(v_i)$ графа $G(V, E)$, определяющих граф ∂ -достижимости $G_\partial(V_\partial, R_\partial)$.

2. В проекциях $P_\partial(v_i)$ графа $G_\partial(V_\partial, R_\partial)$ выявляем подмножества $\bar{V}_\partial(v_i)$ вершин, ∂ -недостижимых из соответствующих курсных вершин v_i : $\bar{V}_\partial(v_i) = V \setminus V_\partial(v_i)$.

3. Используя полученные в п. 2 подмножества $\bar{V}_\partial(v_i)$, строим матрицу \bar{R}_∂ смежности графа ∂ -недостижимости $\bar{G}_\partial(V, \bar{R}_\partial)$ и

выявляем в нем наибольшее множество независимых вершин (МНВ). Входящие состав полученного МНВ вершины образуют искомое подмножество V_{∂}^* вершин, порождающих клику $K_{\partial}(G)$ графа ∂ -достижимости $G_{\partial}(V_{\partial}, R_{\partial})$.

4. Проекции $P_{\partial+}^*$ искомой компоненты ∂ -достижимости $G_{\partial+}^*(V_{\partial+}^*, E_{\partial+}^*)$ получаем последовательным (от $j = \partial$ до $j = 1$) исключением из каждого j -го уровня проекций $\{P_{\partial}(v_i) \mid v_i \in V_{\partial}^*\}$ вершин, не принадлежащих V_{∂}^* и не имеющих продолжения на $(j + 1)$ -м уровне проекции, полученной на предыдущем шаге. Оставшиеся при этом в результирующих проекциях $P_{\partial+}^*$ вершины не принадлежат V_{∂}^* и являются дополнительными.

5. О топологической масштабируемости параллельных задач и вычислительных систем

Уточним используемые в этом разделе обозначения:

p – ранг рассматриваемой задачи (число параллельных ветвей задачи, число задействованных в решении задачи процессоров);

W_p – информационно-логический граф задачи, соответствующий ее распараллеливанию на p параллельных ветвей;

$\partial \equiv \partial(p) = \lfloor L(p) \rfloor$ – обусловленное числом процессоров p и требуемым критерием эффективности S_p или E_p предельно допускаемое в графе ВС значение достижимости;

$\varphi(W_p)$ – плотность распараллеленного на p ветвей информационного графа задачи, порядок наибольшей клики в графе W_p ;

$\varphi_{\partial}(G)$ – ∂ -плотность графа $G(V, E)$, она же – порядок ∂ -клики $K_{\partial}(G)$ графа G , или порядок клики $K(G_{\partial})$ графа ∂ -достижимости $G_{\partial}(V_{\partial}, R_{\partial})$.

5.1. ТОПОЛОГИЧЕСКАЯ МАСШТАБИРУЕМОСТЬ ЗАДАЧИ

Понятно, что, если числу p ветвей (W, Q) -задачи, решаемой в рассматриваемой ВС¹, соответствует предельная достижи-

¹ Здесь и далее, говоря о ВС, по умолчанию предполагаем использование ею сетевой технологии, которая для (W, Q) -задачи ранга p с

мость $\delta(p)$, а порядок p информационного графа W_p вкладываемой подсистемы не выше δ -плотности $\varphi_\delta(G)$ графа G – $\varphi_\delta(G) \geq p$, то вложение графа W_p в эту δ -клику может быть произвольным, даже если граф W_p является полным (при этом $\varphi(W_p) = p$). Попытка наращивания при этом числа процессоров сверх определенного для графа G значения δ -плотности ($p + \Delta p > \varphi_\delta(G)$) неизбежно потребует размещения Δp вершин вне клики $K(G_\delta)$, что увеличит расстояния между информационно смежными в задаче вершинами как минимум на единицу: $\delta(p + \Delta p) - \delta(p) \geq 1$. Так как по определению $\delta(p) = \lfloor L(p) \rfloor$ и $\delta(p + \Delta p) = \lfloor L(p + \Delta p) \rfloor$, то при таком увеличении p предельно-допустимое расстояние $L(p)$ будет превышено – $L(p + \Delta p) - L(p) \geq 1$, что воспрепятствует сохранению директивного критерия эффективности.

В отличие от рассмотренного выше случая информационно-полносвязных задач, вложение в граф δ -достижимости G_δ графа G задач с $\varphi(W_p) < p$ (неполносвязных задач) допускает увеличение числа p процессоров сверх δ -плотности $\varphi_\delta(G) - \varphi(W_p) \leq \varphi_\delta(G) < p$, если в графе G_δ найдется подграф, изоморфный графу W_p задачи. К примеру, если исходная ВС является узкоспециализированной и ориентирована на решение некоторой задачи (или нескольких топологически однотипных задач) с неполносвязным ($p > \varphi(W_p)$) информационным графом W_p , то, в принципе, для этой ВС может быть выбрана топология, повторяющая информационную топологию задачи с точностью до изоморфизма. Если при этом допустимы опосредованные (с $\delta(p) > 1$) взаимодействия информационно смежных процессоров, то топология ВС может быть сведена к одной из известных регулярных топологий, граф δ -достижимости G_δ которой изоморфен графу W_p . Этого можно достичь, например, определив граф G системы суграфом¹ информационного графа W_p задачи, получаемым путем последовательного удаления из W_p ребер до сохраняющего изоморфизм $G_\delta \simeq W_p$ их минимума. Такой подход позволит минимизировать топологию проектируемой ВС, и

требуемыми значениями показателей эффективности ее решения обуславливает предельно допустимую достижимость δ .

¹ Суграф – часть графа, имеющая то же множество вершин, что и сам граф.

он может быть применен и для некоторого набора из N решаемых в системе задач. Соответствующую постановку, состоящую в объединении графов W_{p_i} задач из этого набора ($i = 1, \dots, N$), несложно представить, поэтому останавливаться на ней не имеет смысла.

Итак, связанные с масштабированием параллельной задачи изменения информационной ее топологии существенно влияют на возможности вложения в ВС этой задачи, релевантного требуемому ускорению и эффективности. Понятно, что возможности вложения информационного графа W_p в граф δ -достижимости G_δ при превышении рангом p задачи δ -плотности $\varphi_\delta(G)$ графа ВС коррелированы не только информационной топологией задачи, но и физической топологией системы, и шансы изоморфного вложения задачи ранга p в граф δ -достижимости G_δ возрастают как с уменьшением плотности $\varphi(W_p)$ вкладываемых задач, так и с увеличением δ -плотности $\varphi_\delta(G)$ графа ВС (естественно при этом, что порядок p информационного графа W_p задачи не должен превышать порядка n графа G системы: $p \leq n(G)$). Очевидно, что тезис об увеличении шансов успешного вложения для задач с меньшей плотностью информационного графа актуален для любой системы, топологически адекватной заданному неравенству $\varphi(W_p) \leq \varphi_\delta(G)$ условию. Это позволяет абстрагироваться от используемой в ВС топологии (разумеется, в рамках ее адекватности этому условию) и, таким образом, сосредоточиться только на топологическом аспекте масштабируемости $\mu_\varphi(W_p)$ параллельной задачи, оценивая связанное с наращиванием числа параллельных ветвей относительное изменение возможностей ее вложения лишь качественно:

$$(9) \quad \mu_\varphi(W_p) = 1 - \frac{\varphi(W_p)}{p} = \frac{p - \varphi(W_p)}{p}.$$

Из (9) ясно, что в случаях, например, с неизменной при масштабировании задачи плотностью $\varphi(W_p) = 2$ и с плотностью $\varphi(W_p) = p$ (эти случаи характерны для конвейерных ($W_p \equiv C_p$) и полносвязных ($W_p \equiv K_p$) вычислений) для предельных значений $p = 2$ и $p = \infty$ получим, соответственно, $\mu_\varphi(C_2) = 1/2$, $\mu_\varphi(K_2) = 0$ и $\mu_\varphi(C_\infty) = 1$ и $\mu_\varphi(K_\infty) = 0$. Физический смысл введенной здесь функции $\mu_\varphi(W_p)$ состоит в том, что в первом случае наращивание параллелизма (числа p ее параллельных ветвей) уменьшает

относительную (в отношении к p) плотность информационного графа задачи, соответственно, его вложение в граф δ -достижимости G_δ , $1 \leq \delta \leq d(G)$, облегчается, и масштабируемость $\mu_\varphi(C_p)$ стремится к равному единице максимальному значению. Во втором же случае $\varphi(K_p) = p$, т.е. вне зависимости от ранга p относительная плотность параллельной задачи¹ максимальна и равна единице. Это, как и нулевое значение $\mu_\varphi(K_p)$, говорит о максимальной сложности масштабирования информационно полносвязной параллельной задачи и о том, что вложение в граф δ -достижимости² G_δ задачи K_p возможно, только если плотность $\varphi(G_\delta)$ этого графа не меньше числа ветвей p .

Однако относительная плотность – не единственное, что определяет сложность вложения масштабируемой задачи. К примеру, задачи одного ранга с кольцевой и звездной информационными топологиями обладают одинаковой плотностью, но степень s «кольца» C_p при этом независимо от ранга p задачи постоянна и $s(C_p) = 2$, тогда как наращивание числа p ветвей в задаче со «звездной» ($W_p \equiv Z_p$) топологией увеличивает степень ее информационного графа: $s(Z_p) = p - 1$, и при $p > s(G_\delta) + 1$ одно из необходимых условий наличия изоморфных вложений $s(Z_p) \leq s(G_\delta)$ будет нарушено. Коррелированную степень информационного графа масштабируемость задачи с аналогичным вышеописанному физическим смыслом можно описать тем же выражением, что и в (9), заменив плотность $\varphi(W_p)$ степенью $s(W_p)$:

$$(10) \mu_s(W_p) = 1 - \frac{s(W_p)}{p} = \frac{p - s(W_p)}{p}.$$

Тогда для задач с «кольцевой» и «звездной» топологиями при $p > 2$ получим: $\mu_s(C_p) = 1 - 2p^{-1}$ и $\mu_s(Z_p) = p^{-1}$, что соответствует увеличению $\mu_s(C_p)$ (улучшению возможностей вложения при наращивании «кольца») и уменьшению $\mu_s(Z_p)$ (ухудшению таких возможностей для «звезды»).

Совместное использование введенных выше частных функций масштабируемости параллельной задачи, связанных с изме-

¹ Словосочетание «информационного графа» для краткости опущено.

² Предельная достижимость δ соответствует заданному числу ветвей p .

нением плотности и степени ее информационного графа, позволяет качественно (больше/меньше – лучше/хуже) сопоставлять возможности вложения задач в регулярные топологии вычислительных систем, предельные значения достижимости δ в которых адекватны изменениям p , а плотность $\varphi(G_\delta)$ и степень $s(G_\delta)$ соответствующих графов δ -достижимости удовлетворяют условиям: $\varphi(G_\delta) \geq \varphi(W_p)$ и $s(G_\delta) \geq s(W_p)$. Функцию $\mu(W_p)$ зависимости топологической масштабируемости задачи от числа p параллельных ее ветвей определим произведением функций (9) и (10):
 (11) $\mu(W_p) = \mu_\varphi(W_p) \cdot \mu_s(W_p)$.

Из введенных выше частных функций (9), (10) и из обобщающей их функции (11) видим, что топологическая масштабируемость задачи ухудшается с приближением к единице относительных значений плотности или степени ее информационного графа, т.е. с приближением ее информационной топологии к полностью связной. Понятно также, что при сопоставлении алгоритмов распараллеливания задачи для обеспечения хорошей ее масштабируемости следует отдать предпочтение тем из них, в которых увеличение параллелизма приводит по крайней мере к меньшему росту относительных значений плотности и степени информационного графа.

5.2. ТОПОЛОГИЧЕСКАЯ МАСШТАБИРУЕМОСТЬ СИСТЕМЫ

Понятно, что степень и плотность графа δ -достижимости $G_\delta(V, E_\delta)$, $\delta > 1$, положительно коррелированы степени исходного регулярного¹ графа $G(V, E)$ и величиной полученного из (8) значения достижимости $\delta(p) = \lfloor L(p) \rfloor$: для графов G и H одного порядка ($n(G) = n(H)$)

$$s(G) < s(H) \Rightarrow s(G_\delta) < s(H_\delta), \quad \varphi_\delta(G) < \varphi_\delta(H),$$

$$\delta_1 < \delta_2 \Rightarrow s(G_{\delta_1}) < s(G_{\delta_2}), \quad \varphi(G_{\delta_1}) < \varphi(G_{\delta_2}).$$

Понятно также, что δ -плотность $\varphi_\delta(G)$ графа ВС как порядок $n(K_\delta)$ его максимальной δ -клик, все вершины которой по определению взаимно δ -достижимы, ограничивает сверху параллелизм (число ветвей) только для информационно полностью связных

¹ Напомним, что в соответствии с п. 6 рассматриваемой здесь модели топология ВС представлена регулярным неполностью связным графом.

задач. Для задач же с плотностью информационного графа $\varphi(W_p)$, меньшей числа ветвей p в них, верхние границы могут быть большими δ -плотности графа G ВС, если в графе δ -достижимости G_δ найдется подграф порядка $p > \varphi_\delta(G)$, изоморфный графу W_p . К примеру, число задействованных в гиперкубической ВС процессоров для задачи с кольцевой топологией ограничено сверху порядком гиперкуба вне зависимости от допускаемой при ее решении достижимости δ .

Итак, пределы распараллеливания решаемых в системе задач в зависимости от допускаемой для каждой из них достижимости δ находятся в диапазонах от $\varphi_\delta(G)$ до $n(G)$, поэтому задачи можно классифицировать соответственно принадлежности этим диапазонам, идентифицируемым значениями $\delta \geq 1$. Понятно при этом, что при использовании в системе достаточно быстродействующей NT допускаемая некоторыми задачами предельная достижимость δ не ограничивается диаметром $d(G)$ графа ВС, и $\delta \geq d(G) \Rightarrow \varphi_\delta(G) = n(G_\delta)$.

Рассмотрим, к примеру, решение двух из множества принадлежащих δ -классу задач на системах, топологии которых не привязаны к информационным топологиям этих задач. Пусть топологии систем заданы графами G_1 и G_2 одного порядка $n(G_1) = n(G_2)$, значения их δ -плотности соответствуют оговариваемому δ -классу, но различны: $\varphi_\delta(G_2) > \varphi_\delta(G_1)$. Наделим рассматриваемые здесь задачи W_1 и W_2 значениями топологической масштабируемости μ_1 и μ_2 , близкими к граничным и равными, соответственно, единичному и нулевому значениям. Следует ожидать при этом, что предел распараллеливания первой, наименее топологически сложной для вложения задачи для обеих систем будет близок к одному и тому же значению – порядку n их графов. Для второй задачи, как наиболее топологически сложной для вложения, можно ожидать, что предельное число задействованных в каждой системе процессоров будет близким к порядку наибольшей δ -клики графа – значениям δ -плотности $\varphi_\delta(G_1)$ и $\varphi_\delta(G_2)$. Из этого ясно, что системы с меньшей δ -плотностью графов обладают меньшими возможностями распараллеливания прежде всего топологически более сложных

задач, обладающих меньшей топологической масштабируемостью $\mu(W_p)$.

Понятно тогда, что при решении набора задач усредненное изменение пределов их распараллеливания будет более выигрышным в системах с большими значениями δ -плотности их графов. Поэтому вполне оправданным будет использование δ -плотности в качественной (больше/меньше – лучше/хуже) характеристике графов ВС и в сопоставительном их анализе не только для информационно полносвязных задач, но и для задач, не требующих информационной полносвязности. Естественным при этом будет нормирование значений δ -плотностей сопоставляемых графов ВС порядком этих графов — отношением $\varphi_\delta(G)/n$. Нормированное таким образом значение не только характеризует в абсолютном выражении потенциал системы в распараллеливании информационно полносвязных задач с $\mu(W_p) = 0$, но и позволяет осуществлять качественную оценку потенциала для классифицированных по достижимости $\delta > 1$ задач с $\mu(W_p) > 0$. Эту функцию изменения δ -плотности графа ВС от его порядка n $\mu_\delta(G_n) = \varphi_\delta(G_n)/n$ назовем функцией топологической масштабируемости системы.

При этом, как мы уже говорили выше, если рассматриваемая система допускает для некоторых из решаемых на ней задач $\delta \geq d(G)$, то предел распараллеливания этих задач определяется числом процессоров в ней — $\varphi_\delta(G) = n(G_\delta)$, понятно, что масштабируемость таких задач в этой системе максимальна и равна единице — $\mu_\delta(G_n) = 1$. Отсюда, кстати, вытекает желательность выбора сетевой технологии масштабируемой системы с учетом того, чтобы быстродействие NT было достаточным для эффективного использования увеличенного в результате предполагаемого в процессе эксплуатации масштабирования системы числа ее процессоров: например, для ключевых задач, определяющих функциональное назначение системы, обусловленная сетевой технологией достижимость δ должна быть близкой к текущему диаметру $d(G)$ графа ВС или превышать его.

Итак, учитывая, что предельные значения достижимостей δ в системе обусловлены предписанными ей задачами и используемой сетевой технологией, топологическую масштабируемость

системы, как характеристику изменения потенциала в распараллеливании предписанных ей задач при изменении порядка n графа G_n системы можно характеризовать семейством функций $\mu_{\hat{c}}(G_n)$ относительно к масштабу системы изменения \hat{c} -плотности графа G_n , описывающего топологию исследуемой ВС, при заданных значениях достижимости $\hat{c} > 1$.

6. Заключение

Все параллельные системы по мере их появления и наращивания подвергаются исследованиям их быстродействия и эффективности на различных классах и наборах задач и данных. Понятно, что различия в технической, технологической, топологической, прикладной и прочих архитектурных составляющих системы придают полученным при этом результатам некую эксклюзивность и могут быть распространены на другие системы и задачи лишь с некоторыми ограничениями. Обобщение полученного таким образом и опубликованного к настоящему времени огромного фактологического материала убеждает в качественной идентичности локального влияния этих компонент на показатели масштабируемых систем, но не дает цельной картины взаимной их обусловленности, достаточной для получения требуемых показателей вновь проектируемых и уже эксплуатируемых систем. В настоящей работе предпринята попытка восполнить имеющийся в этом отношении пробел, прежде всего, в оценке влияния топологии на масштабируемость параллельных систем и решаемых на них задач.

С этой целью во втором разделе статьи предложена разделенная на две составляющие модель параллельных вычислений: первая отнесена к параллельным приложениям и приписывает им свойства неограниченной распараллеливаемости, вторая отнесена к вычислительной системе, ограничения параллелизма которой обусловлены недостаточным быстродействием ее коммуникационной среды. В связи с тем, что задержки информационных взаимодействий ветвей зависят от присущих каждой задаче объемов вычислительных и обменных операций, естественной является характеристика задач соответствующими этим объемам параметрами.

Использование предложенной модели позволило сконцентрировать внимание на технологической и топологической составляющих коммуникационной среды и исследовать формальную их взаимозависимость в совокупном влиянии на пределы распараллеливания в заданной ВС задач, соответствующие предписанным критериям эффективности их решения. В третьем разделе статьи получены выражения, связывающие с используемой в системе сетевой технологией предельно допускаемые между информационно смежными процессорами расстояния $L(p)$, минимально необходимое и максимально допускаемое число параллельных ветвей, релевантные заданному ускорению и эффективности. Это позволяет оценить любой из перечисленных выше параметров при заданных остальных.

В связи с тем, что расстояния между вершинами невзвешенного графа определяются числом транзитных участков и могут быть выражены только целыми числами и не должны превышать при этом упомянутых выше предельно допускаемых расстояний между информационно смежными в задаче вершинами графа ВС, введено понятие предельной достижимости $\delta(p)$, определенной целой частью $L(p)$. В четвертом разделе рассматривается проблема вложения в граф ВС параллельных задач с учетом лимитированной достижимости. Даны понятия графа δ -достижимости, δ -компоненты графа, представляющей собой клику графа δ -достижимости и соответствующей порядку последней, δ -плотности графа ВС, приведен алгоритм выявления в графе ВС такой клики.

В пятом разделе введены взаимно абстрагированные формализованные показатели и функции топологической масштабируемости параллельных задач и топологической масштабируемости систем. Свойство взаимной абстрагированности, с одной стороны, позволяет, не привязываясь к конкретной среде реализации задачи, выбрать наименее топологически сложный для ее вложения алгоритм, а с другой – выбирать для унифицированных систем топологии, при прочих равных условиях обладающие наибольшими возможностями успешного вложения произвольных задач. Это не исключает, однако, использование данных функций для специализированных систем.

Введенные показатели топологической масштабируемости задач и систем основаны на использовании плотности информационного графа задачи и δ -плотности графа ВС; допускаемая предложенной в работе моделью нерегулярность топологии задачи учтена дополнением соответствующего показателя степени изменяющегося при ее масштабировании информационного графа. Отметим, что использование в составе показателей масштабируемости задач и вычислительных систем инвариантов соответствующих графов (степени, плотности и δ -плотности) определяет инвариантность этих показателей к способам вложения задач в системы, характеризуя, таким образом, потенциал масштабируемости, который в той или иной степени достижим при использовании этих способов.

Результаты работы будут полезными не только при анализе уже действующих систем и реализуемых на них параллельных алгоритмов, но и при создании новых систем или алгоритмов с учетом предполагаемого масштабирования тех и других.

Литература

1. АБРАМОВ С.М., ЛИЛИТКО Е.П. *Состояние и перспективы развития вычислительных систем сверхвысокой производительности // Информационные технологии и вычислительные системы.* – 2013. – №2. – С. 6–22.
2. БАДЕХА И.А., РОЛДУГИН П.В. *О плотности графов, в которых каждое ребро входит хотя бы в две максимальные клики // Дискрет. матем.* – 2013. – Т. 25, №3. – С. 7–21.
3. БУРОВА И.Г., ДЕМЬЯНОВИЧ Ю.К. *Алгоритмы параллельных вычислений и программирование: Курс лекций.* – СПб: Изд-во С.-Пб. ун-та, 2007. – 206 с.
4. ВОЕВОДИН В.В., ВОЕВОДИН В.В. *Параллельные вычисления.* – Санкт-Петербург. 2002. – 599 с.
5. ВОЕВОДИН В. *Суперкомпьютеры и парадоксы неэффективности // Открытые системы. СУБД.* – 2009. – №10. – С. 17–20.
6. ГЕРГЕЛЬ В.П. *Высокопроизводительные вычисления для многопроцессорных многоядерных систем: [учебник для ву-*

- зов]. – Нижний Новгород: Издательство Нижегородского государственного университета, 2010. – 539 с.
7. МЕЛЕНТЬЕВ В.А. *Вложение подсистем, лимитирующих длину и число путей между вершинами графа вычислительной системы* // Управление большими системами. – 2014. – №47. – С. 212–246.
 8. МЕЛЕНТЬЕВ В.А. *Формальные основы скобочных образов в теории графов* // Труды II Международной конференции «Параллельные вычисления и задачи управления». РАСО'2004 Москва. 2004. – С. 694–706.
 9. ПОЖИЛОВ И.А. *Оценка задержки и пропускной способности сети с топологией «многомерный тор» при наличии отказавших каналов связи* // Научный сервис в сети интернет: многообразие суперкомпьютерных миров. Труды Международной суперкомпьютерной конференции. – С. 211–217.
 10. РАППОПОРТ А.М. *Метрические характеристики графов сетей коммуникаций* // Труды ИСА РАН. – 2005. – Т. 14. – С. 141–147.
 11. СЛЕПУХИН А.Ф. *Перспективы развития аппаратных технологий и их применение в суперкомпьютерах экзафлопного уровня* // Тезисы докладов Четвертого Московского суперкомпьютерного форума (Москва, 23 октября 2013 г.) / [Под. ред. Волкова Д.В. Москва. Россия. Октябрь 2013. – С. 7.
 12. ШПАКОВСКИЙ Г.И., СТЕЦЮРЕНКО В.И., ВЕРХОТУРОВ А.Е. И ДР. *Применение технологии MPI в Грид*. – Минск: Белорусский государственный университет, 2008. – 137 с.
 13. GRAMA A.Y., GUPTA A., KUMAR V. *Isoefficiency: Measuring the Scalability of Parallel Algorithms and Architectures* // Parallel & Distributed Technology: Systems & Applications, IEEE. – August 1993. – Vol. 1. No.3. – P. 12–21.

ON TOPOLOGICAL SCALABILITY OF COMPUTING SYSTEMS

Victor Melentiev, Rzhanov Institute of Semiconductor Physics Siberian Branch of RAS, Novosibirsk, Cand. Sc., senior research associate (melva@isp.nsc.ru).

Abstract: Most results on parallel task and system scaling met in the literature are guided by the specific technological and topological implementation. Unique properties of a problem originated from the class of the task solved, from technical, technological and topological implementation of the system, firstly, prevent using these results directly in the analysis and synthesis of the other parallel systems and tasks, and, secondly, hide the general pattern of mutual conditionality of the given parameters of system with demanded performance metrics. On the basis of the model suggested in this paper we try to fill the existing gap with the main goal to assess the impact of the topology of a parallel system on scalability of a system or a task.

Keywords: topology, system network technology, scalability of parallel computing systems and tasks.

Статья представлена к публикации членом редакционной коллегии М.В. Губко

*Поступила в редакцию 06.09.2015.
Опубликована 30.11.2015.*

УДК 519.83+519.86
ББК 519

РЕШЕНИЕ ЗАДАЧИ СТИМУЛИРОВАНИЯ ИННОВАЦИЙ В ЭЛЕКТРОЭНЕРГЕТИКЕ.¹ ЧАСТЬ 1. ОПТИМИЗАЦИОННЫЕ МОДЕЛИ

**Антоненко А. В.², Лошкарев И. В.³,
Панков В. С.⁴, Угольницкий Г. А.⁵**

(Южный федеральный университет, Ростов-на-Дону)

Рассматриваются оптимизационные модели согласования интересов в электроэнергетике. Представлена программная реализация моделей. На основе полученных результатов сформированы рекомендации и сделаны выводы о целесообразности ввода счетчиков электроэнергии нового поколения.

Ключевые слова: оптимизационные модели, электроэнергетический сектор, математическое моделирование, программная реализация.

1. Введение

Использование оптимизационных и теоретико-игровых моделей организационного управления представляется перспективным инструментом повышения эффективности электроэнергетической отрасли. В работе анализируется ряд моделей мотивационного управления в области стимулирования инноваций в коммерческом учёте электроэнергии. Анализ

¹ Работа поддержана РФФИ, проект №15-01-00432-а.

² Андрей Валерьевич Антоненко, кандидат технических наук, ассистент (andrei80586@yandex.ru).

³ Илья Витальевич Лошкарев, ассистент (loshkarev.i@gmail.com).

⁴ Владимир Сергеевич Панков, магистр прикладной математики и информатики (моб. тел. +79888987246).

⁵ Геннадий Анатольевич Угольницкий, доктор физико-математических наук, профессор (ougoln@mail.ru).

литературных источников позволяет сделать следующие выводы [1]:

1. При управлении распределенным потреблением электроэнергии в условиях децентрализованного рынка достаточно активно используются математические модели и компьютерные системы поддержки решений на их основе.

2. Среди наиболее широко применяемых моделей можно выделить:

– модели прогноза различных параметров функционирования электроэнергетических систем и рынков электроэнергии (цены, нагрузки, потребление) [14];

– модели влияния различных факторов и подготовки данных (эконометрические, динамические) [7, 8];

– модели рынка и его отдельных участников, в том числе регуляторов и распределяющих компаний (экономико-математические, имитационные модели) [3];

– модели оценки инвестиций [16];

– модели и системы поддержки решений [2, 4–6, 9–13, 18–20];

– модели управления спросом [15, 16].

3. Большинство описанных моделей опирается на данные коммерческого учета потребления электроэнергии, для получения которых необходимо устанавливать интеллектуальные счетчики.

4. Наиболее важными с точки зрения решения практических задач, стоящих перед электроэнергетическими компаниями на конкурентном рынке, представляются задачи управления спросом. Именно инновации в управлении спросом могут дать наибольший эффект в краткосрочной и долгосрочной перспективе.

5. Комплексное использование моделей и методов управления спросом позволяет не только решать задачи оперативного управления электроэнергетическими системами, но и обеспечить их долгосрочное устойчивое развитие, под которым понимается выполнение базовых социальных, экономических, экологических и технологических требований при условии согласования интересов всех участников рынка электроэнергии. Постановка и решение задач управления

устойчивым развитием электроэнергетической отрасли представляется магистральным направлением фундаментальных и прикладных научных исследований в электроэнергетике.

В ролях ведущего и ведомого в моделях иерархического управления могут выступать различные агенты, например: регулятор рынка – энергетическая компания, энергетическая компания – клиент, крупная оптовая компания – розничные продавцы электроэнергии, компания – филиалы и т.п.

Для решения практических задач управления спросом на электроэнергию представляется целесообразным использовать модели и механизмы управления организационными системами с учетом концепции управления устойчивым развитием. На первый план здесь выходят модели мотивационного управления, описывающие экономические механизмы стимулирования активных агентов. Изучение таких моделей осуществляется средствами теории контрактов [12], информационной теории иерархических систем [5], теории активных систем [2] (сейчас – теории управления организационными системами [9]).

Цель работы – создание программного комплекса для исследования достаточно простых математических моделей, которые могут быть использованы для решения практических задач мотивационного управления в электроэнергетике, и обсуждение проблем их идентификации.

Практическая значимость работы состоит в создании инструмента для анализа и идентификации параметров моделей стимулирования инноваций в электроэнергетике. В данной статье исследованы первые две модели (оптимизационные) из представленного ниже списка (теоретико-игровые модели рассматриваются во второй части статьи):

- классическая модель стимулирования с экспоненциальным видом функций дохода центра и затрат агента;
- классическая модель стимулирования с показательным видом функций дохода центра и затрат агента;
- теоретико-игровая модель с экспоненциальным видом функций дохода игроков;

- теоретико-игровая модель с показательным видом функций дохода игроков;
- теоретико-игровая модель Гермейера – Вателя с экспоненциальным видом функций дохода игроков;
- теоретико-игровая модель Гермейера – Вателя с показательным видом функций дохода игроков.

В составе программного комплекса все представленные модели численно решены:

- классическая модель стимулирования со степенным видом функций дохода центра и затрат агента;
- классическая модель стимулирования с экспоненциальным видом функций дохода центра и затрат агента;
- классическая модель стимулирования с показательным видом функций дохода центра и затрат агента;
- общий случай теоретико-игровой модели со степенным видом функций дохода игроков;
- общий случай теоретико-игровой модели с экспоненциальным видом функций дохода игроков;
- общий случай теоретико-игровой модели с показательным видом функций дохода игроков;
- частный случай теоретико-игровой модели со степенным видом функций дохода игроков;
- частный случай теоретико-игровой модели с экспоненциальным видом функций дохода игроков;
- частный случай теоретико-игровой модели с показательным видом функций дохода игроков;
- теоретико-игровая модель Гермейера – Вателя со степенным видом функций дохода игроков;
- теоретико-игровая модель Гермейера – Вателя с экспоненциальным видом функций дохода игроков;
- теоретико-игровая модель Гермейера – Вателя с показательным видом функций дохода игроков;
- теоретико-игровая иерархическая модель.

2. Модели стимулирования инноваций в электроэнергетике

В качестве примера инновации рассматривается установка компьютеризованных электросистем «Аргус», обеспечивающих автоматизацию учета потребления электроэнергии и ряд функций управления нагрузкой. Применение данного оборудования позволяет увеличить эффективность использования установленной электрической мощности объектов до 30%, повысить уровень пожарной безопасности объектов, снизить затраты на эксплуатацию систем электроснабжения объектов до 15% и увеличить их срок эксплуатации. Максимальная эффективность применения оборудования достигается при строительстве новых и реконструкции старых объектов. Оборудование имеет европейский и российский сертификаты соответствия, его применение одобрено и рекомендовано НП «Центр энергосбережения и инновационных технологий» Ростовской области и ОАО «Донэнерго»

Однако остается открытым вопрос о субъекте финансирования установки компьютеризованных электросистем (интеллектуальных счетчиков) «Аргус» для различных категорий потребителей. Для его решения целесообразно использовать различные классы моделей стимулирования.

2.1. КЛАССИЧЕСКАЯ МОДЕЛЬ СТИМУЛИРОВАНИЯ

В классической модели стимулирования [4] инноваций, центр компенсирует агенту затраты на совершение действий, приносящих доход центру. Введём следующие обозначения:

u – доля установленных счетчиков от максимального числа (100%-й обеспеченности);

$c(u)$ – затраты агента на установку счетчиков (выпуклая функция);

$s(u)$ – компенсация центром затрат агента.

Модель имеет вид-

$$(1) F(u) = H(u) - s(u) \rightarrow \max, s(u) \in S(u);$$

$$(2) f(u) = s(u) - c(u) \rightarrow \max, 0 \leq u \leq 1.$$

Решение модели:

$$(3) \quad s^*(u) = \begin{cases} c(u) + \delta, & u = u^*, \\ 0, & u \neq u^*; \end{cases}$$

где δ – мотивирующая надбавка, а

$$F(u^*) = \max_{0 \leq u \leq 1} F(u) = \max_{0 \leq u \leq 1} [H(u) - c(u)].$$

Рассмотрим различные виды функций дохода центра ($H(u)$) и затрат агента ($c(u)$):

1. Степенные функции дохода центра и затрат агента:

$$(4) \quad H(u) = au^\alpha, \quad 0 \leq \alpha \leq 1;$$

$$(5) \quad c(u) = bu^\beta, \quad \beta \geq 1.$$

Здесь a – доход центра при 100%-м охвате территории счетчиками; b – затраты агента на 100%-ю обеспеченность счетчиками; α , β – модельные параметры. Найдём аналитическое решение данной модели:

$$(6) \quad F(u) = H(u) - c(u) = au^\alpha - bu^\beta,$$

тогда (14) является точкой максимума, так как справедливо (8):

$$(7) \quad 0 = \frac{dF}{du} = \alpha au^{\alpha-1} - \beta bu^{\beta-1} \Rightarrow u^* = \left(\frac{\alpha a}{\beta b} \right)^{\frac{1}{\beta-\alpha}},$$

$$(8) \quad \frac{d^2 F(u^*)}{du^2} = \alpha(\alpha-1)u^{\alpha-2} - \beta(\beta-1)u^{\beta-2} < 0.$$

Оптимальная доля установленных счётчиков составит

$$u^* = \left(\frac{\alpha a}{\beta b} \right)^{\frac{1}{\beta-\alpha}}.$$

2. Экспоненциальные функции дохода центра и затрат агента:

$$(9) \quad H(u) = a(1 - e^{-\alpha u}), \quad \alpha > 0;$$

$$(10) \quad c(u) = b(e^{\beta u} - 1), \quad \beta > 0.$$

Здесь a – доход центра при 100%-м охвате территории счетчиками; b – затраты агента на 100%-ю обеспеченность счетчиками; α , β – модельные параметры. Найдём аналитическое решение данной модели:

$$(11) F(u) = H(u) - c(u) = a(1 - e^{-au}) - b(e^{\beta u} - 1),$$

тогда (12) – точка максимума, так как справедливо (13):

$$(12) 0 = \frac{dF}{du} = aae^{-au} - \beta be^{\beta u} \Rightarrow u^* = \frac{\ln\left(\frac{aa}{b\beta}\right)}{\alpha + \beta},$$

$$(13) \frac{d^2F(u^*)}{du^2} = -\alpha^2 ae^{-au} - \beta^2 be^{\beta u} < 0.$$

Оптимальная доля установленных счётчиков составит

$$u^* = \frac{\ln\left(\frac{aa}{b\beta}\right)}{\alpha + \beta}.$$

3. Показательные функции дохода центра и затрат агента:

$$(14) H(u) = a(1 - \alpha^u), \quad 0 < \alpha < 1;$$

$$(15) c(u) = b(\beta^u - 1), \quad \beta > 1.$$

Здесь a – доход центра при 100%-м охвате территории счетчиками; b – затраты агента на 100%-ю обеспеченность счетчиками; α , β – модельные параметры. Найдём аналитическое решение данной модели:

$$(16) F(u) = H(u) - c(u) = a(1 - \alpha^u) - b(\beta^u - 1),$$

тогда (17) – точка максимума, так как справедливо (18):

$$(17) 0 = \frac{dF}{du} = a \ln(\alpha) \alpha^u - b \ln(\beta) \beta^u \Rightarrow u^* = -\frac{\ln\left(-\frac{a \ln(\alpha)}{b \ln(\beta)}\right)}{\ln(\alpha) - \ln(\beta)},$$

$$(18) \frac{d^2F(u^*)}{du^2} = -a(\ln(\alpha))^2 \alpha^u - b(\ln(\beta))^2 \beta^u < 0.$$

Оптимальная доля установленных счётчиков составит

$$u^* = -\frac{\ln\left(-\frac{a \ln(\alpha)}{b \ln(\beta)}\right)}{\ln(\alpha) - \ln(\beta)}.$$

Несмотря на существование явных аналитических решений модели, была осуществлена ее программная реализация для анализа и содержательной интерпретации зависимости решений от параметров. Подробное обсуждение этой идеи проводится в заключении к работе.

3. Программная реализация модели

Для реализации программ был выбран язык программирования C#. Среда разработки Microsoft Visual Studio 2008 на платформе .NET.

В C# входит много полезных особенностей – простота, объектная ориентированность, типовая защищенность, «сборка мусора», поддержка совместимости версий и многое другое. Ввиду достаточно удобного объектно-ориентированного дизайна, C# является хорошим выбором для быстрого конструирования различных компонентов.

Visual Studio включает в себя редактор исходного кода с поддержкой технологии IntelliSense и возможностью простейшего рефакторинга кода. Встроенный отладчик может работать и как отладчик уровня исходного кода, и как отладчик машинного уровня. Остальные встраиваемые инструменты включают в себя редактор форм для упрощения создания графического интерфейса приложения, веб-редактор, дизайнер классов и дизайнер схемы базы данных. Visual Studio позволяет создавать и подключать сторонние дополнения (плагины) для расширения функциональности практически на каждом уровне.

Для создания пользовательского интерфейса были использованы следующие средства языка C#: Form, Button, GroupBox, Label, Panel, RadioButton, TextBox, TabControl, PictureBox, DataGridView.

Структура программного комплекса показана на рис. 1.

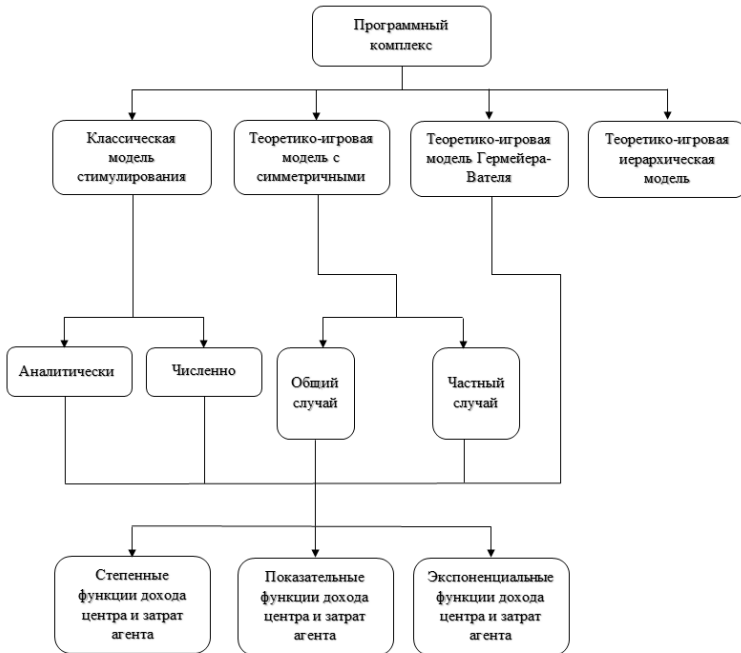


Рис. 1. Структура программного комплекса

4. Выполнение расчётов и анализ результатов

Результаты расчетов представлены на рис. 2–5. В процессе работы были исследованы модели, значения параметров α и β в которых изменялись в интервале от 0 до 10000; с учетом ограниченности числа графиков в основном показаны случаи интервалов $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 10$, где результаты наиболее наглядны.

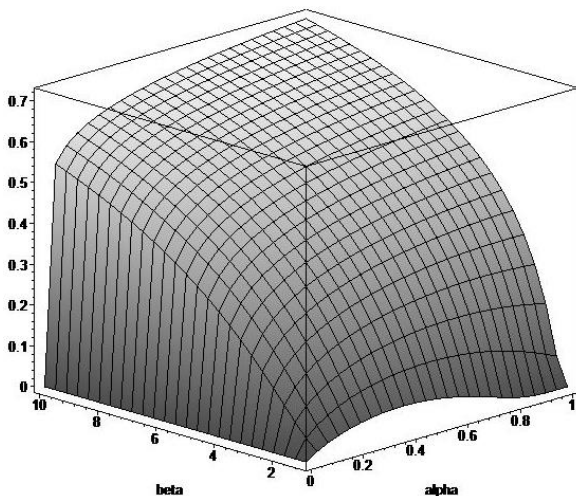


Рис. 2. Вид функции объема инноваций: степенная;
Доход (a) – 10, затраты (b) – 20, α изменяется от 0 до 1,
 β от 0 до 10

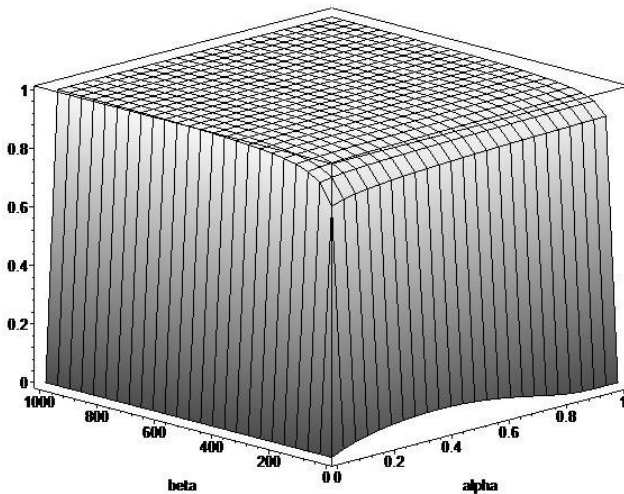


Рис. 3. Вид функции объема инноваций: степенная;
Доход (a) – 10, затраты (b) – 20, α изменяется от 0 до 1,
 β от 0 до 1000

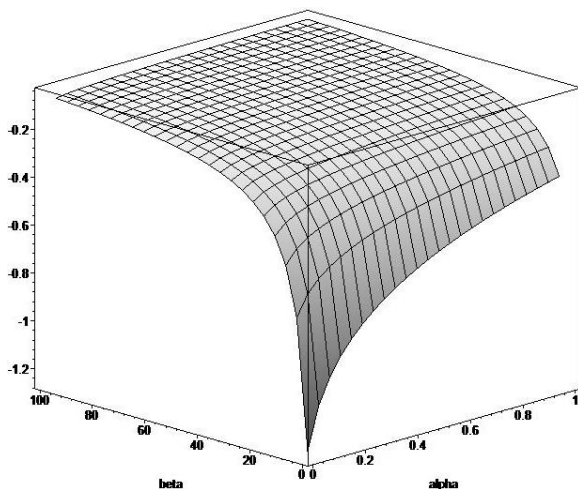


Рис. 4. Вид функции объема инноваций: экспоненциальный;
Доход (a) – 10, затраты (b) – 20, α изменяется от 0 до 1,
 β от 0 до 100

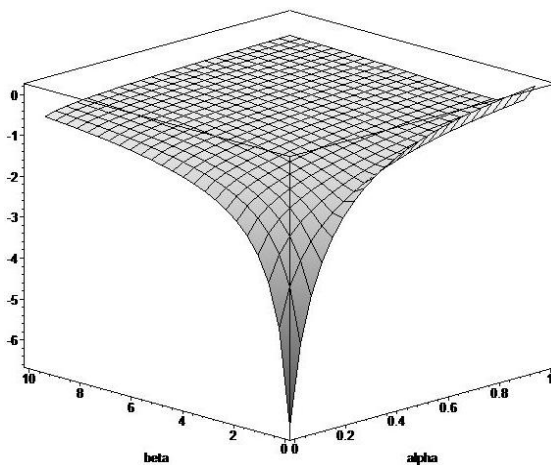


Рис. 5. Вид функции объема инноваций: экспоненциальный;
Доход (a) – 10, затраты (b) – 20, α изменяется от 0 до 1,
 β от 0 до 10

Экономически параметр α можно трактовать как «коэффициент отдачи» (КО) от инноваций – это внешний параметр, влияние центра на который отсутствует. С ростом КО увеличивается возможная прибыль центра. Более интересен параметр β . При малых его значениях (менее 10) наибольшее влияние оказывает КО, однако с ростом параметра β объем инноваций и, как следствие, прибыль центра уже практически не зависят от КО, что видно на рисунках. Поэтому параметр β можно условно назвать «коэффициентом массовости» (КМ). Иными словами, в случае массовой установки счетчиков нового поколения уровень отдачи для центра растет. Из графиков видно, что максимум объема инноваций при достаточно больших значениях КМ не превышает значения $u = 1$. Можно сказать, что при больших (более 100) значениях КМ влияние КО нивелируется и центр может рассчитывать на максимальное значение прибыли от установки счетчиков.

После анализа примеров для данной модели был получен набор выводов и рекомендаций.

1. *Степенные функции дохода центра и затрат агента.* Затраты агента составляют половину дохода центра: $a = 10$, $b = 20$; $0 \leq \alpha \leq 1$, шаг по $\alpha = 0,1$; $1 \leq \beta \leq 10$, шаг по $\beta = 1$.

Итог: максимум u^* достигается при использовании максимально возможных значений КО и КМ.

2. *Степенные функции дохода центра и затрат агента.* Доход центра вдвое больше затрат агента: $a = 20$, $b = 10$; $0 \leq \alpha \leq 1$, шаг по $\alpha = 0,1$; $1 \leq \beta \leq 10$, шаг по $\beta = 1$.

Итог: максимум u^* достигается при использовании значений параметров таких, что $2\alpha \geq \beta$, иначе наиболее выгодно использовать максимальные возможные значения КО и КМ.

3. *Экспоненциальные функции дохода центра и затрат агента.* Затраты агента равны доходу центра: $a = b = 1$; $0,2 \leq \alpha \leq 2$, шаг по $\alpha = 0,2$; $0,2 \leq \beta \leq 2$, шаг по $\beta = 2$.

Итог: максимум u^* достигается при использовании минимального значения параметра β , при параметре α таком, что: $\alpha > \frac{\beta}{2}$ (в общем случае $\alpha > \frac{\beta}{k}$, где $k = b/a$).

4. Показательные функции дохода центра и затрат агента.

Затраты агента равны доходу центра: $a = b = 1$; $0,1 \leq \alpha \leq 0,9$, шаг по $\alpha = 0,1$; $1 \leq \beta \leq 10$, шаг по $\beta = 1$.

Итог: максимум u^* достигается при использовании значений параметра β , близких к 1 справа и при минимально возможных значениях КО.

5. Показательные функции дохода центра и затрат агента.

Затраты агента составляют половину дохода центра: $a = 10$, $b = 20$; $0,1 \leq \alpha \leq 0,9$, шаг по $\alpha = 0,1$; $1 \leq \beta \leq 10$, шаг по $\beta = 1$.

Итог: максимум u^* достигается при использовании значений КО, близких к 1 справа, и при минимально возможных значениях КО.

6. Показательные функции дохода центра и затрат агента.

Доход центра вдвое больше затрат агента: $a = 20$, $b = 10$; $0,1 \leq \alpha \leq 0,9$, шаг по $\alpha = 0,1$; $1 \leq \beta \leq 10$, шаг по $\beta = 1$.

Итог: максимум u^* достигается при использовании значений КО, близких к 1 справа, и при минимально возможных значениях КО.

5. Заключение

В статье исследованы следующие модели:

- классическая модель стимулирования с экспоненциальным видом функций дохода центра и затрат агента;
- классическая модель стимулирования с показательным видом функций дохода центра и затрат агента.

В составе программного комплекса реализовано решение следующих моделей:

- классическая модель стимулирования со степенным видом функций дохода центра и затрат агента;
- классическая модель стимулирования с экспоненциальным видом функций дохода центра и затрат агента;
- классическая модель стимулирования с показательным видом функций дохода центра и затрат агента.

Представлены тестовые примеры, проведён анализ решения. На основе результатов работы программного комплекса можно сделать следующие выводы.

Для классической модели стимулирования были рассмотрены три вида функций.

Для степенного вида функций дохода центра и затрат агента в случае, когда $a \leq b$ (доход центра не превышает затрат агента), максимум целевой функции достигается при использовании максимальных значений КО и КМ. Если же $a > b$, то в целом тенденция роста решения при росте КО и КМ сохраняется, но при этом в случае, когда $\alpha \geq \beta / 2$, решение гарантированно принимает максимальное значение.

В случае экспоненциального вида функций дохода центра и затрат агента влияние КО и КМ на решение зависит от значения параметров a и b . Если $a = b$, то максимум решения достигается при использовании минимального значения КМ и при параметре α таком, что $\alpha > \beta$. Если $a < b$, то максимум решения достигается при использовании минимального значения КМ и при параметре α таком, что $\alpha > \beta / k$, где $k = b / a$. Если $a > b$, то максимум решения достигается при использовании минимального значения КМ и при параметре α таком, что $\alpha > k\beta$, где $k = b / a$.

В случае показательного вида функций дохода центра и затрат агента влияние КО и КМ на решение не зависит от значения параметров a и b . При любом их соотношении максимум решения достигается при значении КМ таком, что $\beta = 1 + \varepsilon$, где ε – сколь угодно малое число, и при использовании минимального значения КО.

В случае модели со степенным видом функций решение модели растёт по параметрам a_i , α_i и убывает по параметрам b_i .

В случае модели с показательным видом функций решение модели зависит от конкретного сочетания параметров a_i и β .

В случае модели с экспоненциальным видом функций решение модели убывает по параметрам β и b_i , растёт по параметрам a_i , а по параметрам α_i сначала растёт, а затем убывает.

Предложена интерпретация модельных параметров α , β как коэффициентов «отдачи» и «массовости» соответственно. С ростом КО увеличивается возможная прибыль центра. При малых значениях КМ основную роль по-прежнему играет КО, а

при их увеличении на объем инноваций и соответствующую прибыль влияет уже КМ.

В целом показано, что для различных функций доходов и затрат и сочетания их числовых параметров оптимальными оказываются различные доли внедрения интеллектуальных счетчиков электроэнергии.

Рассмотренные классы моделей с различными видами функций описывают качественно различные ситуации. Используя данные о влиянии модельных параметров на решение, можно провести идентификацию представленных параметров с помощью экспертных оценок.

Литература

1. АНТОНЕНКО А.В., УГОЛЬНИЦКИЙ Г.А. *Модели мотивационного управления в электроэнергетике и проблемы их идентификации* // Управление большими системами. – 2014. – №47. – С. 92–124.
2. БУРКОВ В.Н., НОВИКОВ Д.А. *Теория активных систем: состояние и перспективы*. – М.: Синтег, 1999. – 128 с.
3. БУРКОВ В.Н., ГУБКО М.В., НОВИКОВ Д.А. *Организационные механизмы управления в электроэнергетике* // Управление развитием крупномасштабных систем. – М.: Изд-во физ.-мат. лит-ры, 2012. – С. 261–278.
4. БУРКОВ В.Н., КОРГИН Н.А., НОВИКОВ Д.А. *Введение в теорию управления организационными системами*. – М.: Книжный дом «ЛИБРОКОМ», 2009. – 264 с.
5. ГЕРМЕЙЕР Ю.Б., ВАТЕЛЬ И.А. *Игры с иерархическим вектором интересов* // Известия АН СССР. Техническая кибернетика. – 1974. – №3. – С. 54–69.
6. ГОРЕЛИК В.А., ГОРЕЛОВ М.А., КОНОНЕНКО А.Ф. *Анализ конфликтных ситуаций в системах управления*. – М.: Радио и связь, 1991. – 288 с.
7. ДОУГЕРТИ К. *Введение в эконометрику*. – М.: Инфра-М, 1999. – 465 с.
8. ЛЬЮНГ Л. *Идентификация систем. Теория для пользователя*. – М.: Наука, 1991. – 432 с.

9. НОВИКОВ Д.А. *Теория управления организационными системами*. – М.: МПСИ, 2005. – 584 с.
10. УГОЛЬНИЦКИЙ Г.А. *Иерархическое управление устойчивым развитием // Общественные науки и современность*. – 2010. – 336 с.
11. УГОЛЬНИЦКИЙ Г.А. *Устойчивое развитие организаций*. – М.: ФИЗМАТЛИТ, 2011. – 320 с.
12. УГОЛЬНИЦКИЙ Г.А. *Модели конфликтов*. – М.: ФИЗМАТЛИТ, 2012. – 320 с.
13. УГОЛЬНИЦКИЙ Г.А. *Математическое моделирование сложных систем // Научное наследие Ю.А.Жданова и современные проблемы моделирования сложных социосистем (на материалах Юга России)*. – Ростов-на-Дону: Изд-во СКНЦ ВШ ЮФУ, 2012. – С. 32–41.
14. BAYON L., SUAREZ P., MATIAS J.M. et al. *Influence of forecasting electricity prices in the optimization of complex hydrothermal systems // Journal of Computational and Applied Mathematics*. – 2009. – No. 232. – P. 262–274.
15. FOLEY A.M., GALLACHOIR B.P., HUR J. et al. *A strategic review of electricity systems markets // Energy*. – 2010. – No. 35. – P. 4533–4530.
16. HASANI M., HOSSEINI S.H. *Dynamic assessment of capacity investment in electricity market considering complementary capacity mechanisms // Energy*. – 2011. – No. 36. – P. 277–293.
17. LAFFONT J.-J., MARTIMORT D. *The Theory of Incentives. The Principal-Agent Model*. – Princeton, 2002. – 440 p.
18. MIRZA F.M., BERGLAND O. *The impact of daylight saving time on electricity consumption: Evidence from southern Norway and Sweden // Energy Policy*. – 2011. – No. 39. – P. 5008–5025.
19. SHI L., ZENG M., Li L. *A Novel Electricity Marketing Model Integrating Intelligent Disaster-Recovery System // Systems Engineering Procedia*. – 2012. – No. 4. – P. 133–142.
20. STADLER M., KRAUSE W., SONNENSCHNEIN M. et al. *Modelling and evaluation of control schemes for enhancing load shift of electricity demand for cooling devices // Environmental Modelling and Software*. – 2009. – No. 24. – P. 285–295.

**SOLVING PROBLEM OF INNOVATION FUNDING
INCENTIVE IN ELECTRICITY POWER INDUSTRY. PART
1.OPTIMIZATION MODELS**

Andrey Antonenko, Southern Federal University, Rostov-on-Don,
Ph.D., Assistant (andrei80586@yandex.ru)

Ilya Loshkarev, Southern Federal University, Rostov-on-Don,
Assistant (loshkarev.i@gmail.com)

Vladimir Pankov, Rostov-on-Don, M.Sc. (+79888987246)

Gennady Ugolnitsky, Southern Federal University, Rostov-on-Don,
Doctor of Physical and Mathematical Sciences, Professor
(ougoln@mail.ru)

Abstract: We study optimization models of interest coordination in the electric power industry. Algorithms are implemented in applied software. Analysis of computer simulations provides motivation for deployment of the new generation of electricity meters.

Keywords: optimization models, electric power industry, mathematical modeling, software implementation.

*Статья представлена к публикации
членом редакционной коллегии В.И. Зоркальцевым*

*Поступила в редакцию 15.01.2015.
Опубликована 30.11.2015.*

УДК 519.816

ББК 78.34

МОДЕЛИ ПРИНЯТИЯ КОЛЛЕКТИВНЫХ РЕШЕНИЙ В ПРОИЗВОДСТВЕННЫХ СИСТЕМАХ¹

Вожаков А. В.²

(ОАО «Мотовилихинские заводы», Пермь)

Гитман М. Б.³, Столбов В. Ю.⁴

*(Пермский национальный исследовательский
политехнический университет, Пермь)*

Рассматриваются четыре модели принятия коллективного решения путем сравнения альтернатив и выбора наиболее предпочтительной группой экспертов. В качестве организационной формы работы экспертов рассматривается комитет. Приводится демонстрационный пример применения моделей при выборе наилучшего производственного плана промышленного предприятия.

Ключевые слова: модели принятия коллективного решения, множество альтернатив, частные показатели качества, предпочтения экспертов.

1. Введение

Качество принимаемых менеджерами решений в значительной степени определяет эффективность функционирования

¹ Работа выполнена при финансовой поддержке Минобрнауки РФ (договор № 02.G25.31.0068 от 23.05.2013г. в составе мероприятия по реализации постановления Правительства РФ № 218).

² Артем Викторович Вожаков, кандидат технических наук (vozhaikov@yandex.ru).

³ Михаил Борисович Гитман, доктор физико-математических наук, профессор (gmb@matmod.pstu.ac.ru).

⁴ Валерий Юрьевич Столбов, доктор технических наук, профессор (valeriy.stolbov@gmail.com).

любой организации, в том числе производственной системы. Повысить качество управленческих решений позволяет механизм коллективного принятия решений [1, 6], который является непосредственной реализацией такого принципа TQM (Total Quality Management – тотальное управление качеством), как вовлечение сотрудников в управление предприятием с целью раскрытия и использования их творческого потенциала [8]. Реализация этого механизма требует разработки соответствующих моделей, позволяющих учитывать многокритериальность выбора, а также предпочтения экспертов на заданном множестве допустимых альтернатив [1, 10].

2. Постановка задачи

Рассматривается задача выбора наилучшего решения из заданного множества альтернатив, возникающая при некоторой сложившейся ситуации на предприятии и требующая быстрого решения с учетом системного анализа ситуации и возможных последствий принятия управленческого решения. Решение принимается в рамках небольшой группы экспертов, которые и представляют собой коллективное ЛПР.

Требуется предложить наилучшую организацию работы ЛПР, а также возможные модели принятия коллективного решения, позволяющие учитывать различные предпочтения и квалификацию экспертов, а также неоднозначность выбора при наличии нескольких показателей качества получаемого решения.

При решении данной задачи вводятся следующие гипотезы:

– ЛПР представляет собой двухуровневую структуру, состоящую из рабочей группы экспертов (ГЛПР) и ответственного лица (ОЛПР), принимающего или не принимающего окончательное решение;

– ГЛПР представляет собой совокупность лиц, представляющих соответствующую структуру (например, для предприятия – это отдел снабжения, производственный отдел, отдел сбыта и т.п.), важность мнения которых может быть как равнозначной, так и различной по каждой ситуационной задаче и каждому показателю качества принимаемого решения;

– частные показатели качества также могут иметь различную важность для принятия коллективного решения.

Для организации процесса принятия решений можно сформировать одну из трех организационных структур: комитет, иерархию или полиархию [6, 10]. В комитете рассматриваемый проект отдается на ознакомление всем менеджерам. По результатам ознакомления проводится голосование и проект принимается, если за него проголосовало больше определенной доли менеджеров. С нашей точки зрения, такая организация оправдана, если квалификация всего менеджерского звена приблизительно одинакова. При иерархии менеджеры выстроены в цепочку и знакомятся с проектом последовательно. Проект окончательно отклоняется, если его отклоняет хотя бы один менеджер в заданной цепочке иерархии, и направляется на рассмотрение к следующему менеджеру в случае его рассмотрения предыдущим. Считаем, что такая организация принятия решений может быть рекомендована, если квалификация всех менеджеров, принимающих решения, очень высока, так как отрицательное мнение любого из них приводит к отклонению всего рассматриваемого проекта. В полиархии проект направляется одному из менеджеров с равной вероятностью и принимается окончательно, если менеджер его принимает, и отклоняется в противном случае. По существу, речь идет о том, что случайно выбранный менеджер является лицом, принимающим решение от всей группы менеджеров. Очевидно, что недостатком этого подхода является зависимость окончательного решения от случайного выбора лица, принимающего решение, независимо от квалификации последнего. Отметим, что рассмотренные организационные формы можно комбинировать, строя из них более сложную организационную структуру принятия решений. Например, можно рассматривать иерархию, каждый элемент которой представляет собой комитет. Также можно рассматривать иерархию из полиархий или полиархию из иерархий.

Для рассматриваемой задачи принятия коллективных решений на промышленном предприятии удобно использовать комитет в рамках ситуационного центра [4]. Это связано с тем, что другие формы организации принятия коллективного решения требуют больших затрат времени на согласование компромисс-

ного решения, а в рамках текущего производства это не всегда оправданно.

3. Математическая постановка задачи

Пусть X – конечное множество альтернатив $x_j \in X$, $j = 1, \dots, n$. Под альтернативой будем понимать допустимый вариант решения ситуационной задачи с набором числовых значений частных показателей качества данного решения. Обозначим численное значение k -го частного показателя для j -й альтернативы J_k^j , $j = 1, \dots, n$; $k = 1, \dots, r$, где r – количество показателей. Отметим, что эти значения для производственных показателей качества всегда будут неотрицательными (например, время выполнения заказа, вероятность поломки оборудования, затраты и т.п.). Поэтому можно считать, что наилучшее решение для некоторого k -го частного показателя достигается при его *максимальном* значении. При этом на предприятии должен быть создан банк моделей, позволяющих строить допустимое множество альтернатив для возможных ситуаций с определением значений частных показателей каждой качества альтернативы. Например, необходимый набор моделей управления предприятием на этапе планирования производства приведен в работе [9].

Считается, что ГЛПР состоит из m экспертов и известны \succ^i , $i = 1, \dots, m$, – предпочтения каждого эксперта на множестве альтернатив X . Под предпочтением i -го эксперта будем понимать набор данных о ранжировании всех рассматриваемых альтернатив. При этом эксперт может высказывать свое мнение как по каждой альтернативе в целом, так и по частным показателям качества рассматриваемой альтернативы.

Требуется найти коллективное решение ситуационной задачи, под которым понимается некое групповое предпочтение \succ на множестве X , полученное на основе обработки частных предпочтений всех участников ГЛПР. При этом формой организации принятия решений ГЛПР считается комитет.

Рассмотрим модели принятия коллективных решений в рамках комитета, которые, на наш взгляд, наиболее приемлемы для производственных систем.

4. Модели принятия коллективных решений

Первоначально рассмотрим две модели, часто применяемые для принятия коллективных решений в социальных и экономических системах [7].

Модель *Кондорсе*. Это одна из простейших и наиболее часто встречающихся на практике моделей. Сущность данной модели состоит в следующем.

На основании полученных от экспертов ранжировок для каждой пары альтернатив x_j, x_k подсчитывается число экспертов $l(x_j, x_k)$, считающих альтернативу x_j более предпочтительной, чем x_k . Если $l(x_j, x_k) > l(x_k, x_j)$, то альтернатива x_j признается предпочтительнее x_k . Наилучшей альтернативой объявляется альтернатива x_j , если выполняется условие

$$(1) \quad l(x_j, x_k) \geq l(x_k, x_j) \text{ для всех } j, k = 1, \dots, n; \quad j \neq k.$$

Основным недостатком данной модели является то, что она не всегда обеспечивает достаточность при определении наилучшей альтернативы с помощью непосредственного подсчета голосов экспертов по правилу большинства (нет единственности решения). В этом случае необходимо использовать более сложные модели.

Модель *Борда*. В последнее время эта модель получила широкое распространение. Согласно данной модели результаты голосования экспертов выражаются в виде числа баллов, набранных каждой альтернативой. Для этого вводится функция Борда $b(x_j)$ следующим образом:

$$(2) \quad b(x_j) = \sum_{k=1}^n [l(x_j, x_k) - l(x_k, x_j)],$$

где $l(x_j, x_k)$ – число экспертов, для которых x_j предпочтительней x_k .

Лучшей считается альтернатива, набравшая максимальное количество баллов, рассчитанное по формуле (2).

Можно предложить еще один способ расчета баллов для различных альтернатив (обобщенное правило Борда), приводящий к тому же результату. За первое место при попарном сравнении альтернатив присуждается n баллов, за второе – $n-1$ баллов и т.д. После этого подсчитывается сумма баллов для каждой альтернативы. Лучшей считается та альтернатива, которая набирает большую сумму баллов. Преимуществом данной модели является то, что всегда существует наилучшая альтернатива. Все описанные выше модели не учитывают многокритериальность выбора при сравнении альтернатив. Другими словами, не учитываются предпочтения экспертов в рамках каждой альтернативы, определяемой набором частных критериев оптимальности. Кроме того, не учитывается *важность* (квалификация) экспертов при многокритериальной оценке каждой альтернативы.

Попытаемся избежать этих недостатков с помощью построения *многокритериальных* моделей выбора с учетом важности частных показателей качества альтернативы. При этом комплексное оценивание качества каждой альтернативы может производиться различными способами, например с помощью линейных и нелинейных (матричных) сверток [1, 3], а также нечетких множеств [3, 5]. Рассмотрим модель коллективного выбора на основе комплексного оценивания альтернатив в виде линейной свертки частных показателей (линейная многокритериальная модель)

Линейная многокритериальная модель. Пусть количество допустимых альтернатив равно n , число экспертов – m . Количество частных показателей качества, характеризующих каждую из возможных альтернатив, равно r . Для построения модели введем следующие обозначения:

$\mu_k^j, j = 1, \dots, m, k = 1, \dots, r$, – оценка *квалификации* j -го эксперта по k -му показателю качества альтернативы;

λ_j^i – предпочтения каждого эксперта $\succ^j, j = 1, \dots, m$, на множестве альтернатив X, i – номер соответствующей альтернативы;

$\nu_k, k = 1, \dots, r$, – важность (значимость) k -го частного показателя качества альтернативы.

Численно все значения предпочтений и важностей задаются в диапазонах от 0 до 1, начиная с наихудшего (наименее значимого) до наилучшего (наиболее значимого).

Как отмечено выше, численное значение k -го частного показателя для i -й альтернативы обозначено J_k^i , $i = 1, \dots, n$; $k = 1, \dots, r$. Будем считать, что *наилучшее* решение для некоторого k -го частного показателя достигается при его *максимальном* значении. Если же *наилучшее* решение достигается при его *минимальном* значении, то в исследуемом случае при положительности значений частных показателей всегда от *максимума* можно перейти к *минимуму* путем несложных математических преобразований.

Теперь для i -й альтернативы введем функцию *предпочтительности* $pr(x_i)$ следующим образом:

$$(3) \quad pr(x_i) = \sum_{j=1}^m \lambda_j^i \cdot \left(\sum_{k=1}^r \mu_k^j \cdot v_k \cdot J_k^i \right), \quad i = 1, \dots, n.$$

Отметим, что если k -й частный показатель для i -й альтернативы J_k^i необходимо минимизировать, $i = 1, \dots, n$; $k = 1, \dots, r$, то переопределим этот показатель следующим образом:

$$J_k^i = (1 - J_k^i / \max_{i,k} J_k^i).$$

Если считать, что квалификация эксперта не связана с экономическим смыслом частных показателей качества альтернативы, то выбор наилучшей альтернативы не будет зависеть от квалификации экспертов и функция *предпочтительности* может быть переписана в виде

$$(4) \quad pr(x_i) = \sum_{j=1}^m \lambda_j^i \cdot \left(\sum_{k=1}^r v_k \cdot J_k^i \right), \quad i = 1, \dots, n.$$

В качестве наилучшей альтернативы x_i выбирается та альтернатива x_i , для которой выполняется условие

$$(5) \quad pr(x_i) > pr(x_l), \quad \forall i, l = 1, \dots, n; \quad i \neq l.$$

Другими словами, лучшей считается та альтернатива, для которой значение функции *предпочтительности* принимает максимальное значение.

К недостаткам предложенной линейной модели можно отнести использование гипотезы линейности при вычислении

функции предпочтительности. Существует несколько подходов, позволяющих отказаться от данной гипотезы. Например, в работе [2] предлагается рассматривать задачу многокритериального выбора наилучшей альтернативы как задачу «аксиоматического построения результирующего ранжирования» при условии, что значения каждого из введенных частных критериев имеют только три градации (например, «плохо», «удовлетворительно» и «хорошо»), а используемое для выбора наилучшей альтернативы правило – это некоторое введенное пороговое правило. Предложенная модель хорошо зарекомендовала себя в задачах выбора в социальных системах, характеризующихся именно трёхранговым ранжированием объектов оценивания. Однако в производственных системах часто объекты исследования имеют количественную оценку, которая не всегда может быть обоснованно сведена к трёхранговому ранжированию. Поэтому рассмотрим еще один способ ухода от линейной гипотезы, который заключается в переходе к нечеткой постановке путем введения обобщенного показателя качества в виде специального нечеткого множества J^r , элементами носителя которого являются значения частных показателей J_k^i , а важности этих показателей, равные v_k , $k = 1, \dots, r$, определяют соответствующие функции принадлежности. При этом для сравнения альтернатив по значениям обобщенного критерия необходимо ввести специальный индекс ранжирования. Необходимость введения такого индекса обусловлена тем, что, во-первых, различные показатели качества объектов производственной системы могут иметь различную физическую природу и, как следствие, различную размерность, которая нивелируется введением безразмерной величины C_i (i – номер соответствующего показателя), а, во-вторых, этот индекс учитывает только те показатели, которые являются определяющими при обобщенной оценке качества. Подробно процедура введения такого индекса ранжирования рассмотрена в работах [5, 8, 9].

Введем J^r в виде

$$J^r = \bigcup_{k=1}^r v_k / J_k,$$

где v_k , $k = 1, \dots, r$, – экспертная оценка значимости k -го частного

показателя. Теперь, используя четкую функцию от нечеткого аргумента $H(J^{r1}, J^{r2})$ (специальный индекс ранжирования [5, 8]), можно записать:

$$(6) \quad H(J^{r1}, J^{r2}) = \text{sign } C_i,$$

где

$$C_i = \frac{v_i^{r1} \cdot J_i^{r1} - v_i^{r2} J_i^{r2}}{d_i},$$

здесь i соответствует

$$\max_i |v_i^{r1} \cdot J_i^{r1} - v_i^{r2} J_i^{r2} / d_i|$$

$$d_i = \max(J_k^{r1}, J_k^{r2}), \quad k = 1, \dots, r.$$

При этом если значение i единственно и если $\text{sign } C_i = +1$, то $J^{r1} > J^{r2}$; если $\text{sign } C_i = -1$, то $J^{r1} < J^{r2}$.

Если значение i не единственно, то определим число l , равное количеству одинаковых по модулю экстремумов, и вычислим вспомогательный параметр

$$\gamma = \sum_{i=1}^l \text{sign } C_i.$$

При этом если $\gamma = 0$, то $J^{r1} = J^{r2}$; если $\gamma < 0$, то $J^{r1} < J^{r2}$; если $\gamma > 0$, то $J^{r1} > J^{r2}$.

Выбор наилучшей альтернативы производится путем попарного сравнения альтернатив с помощью предложенного индекса ранжирования (6). Альтернатива, которая при полном переборе парных сравнений доминирует большее количество раз, считается наиболее предпочтительной.

Рассмотрим пример, демонстрирующий возможности применения рассмотренных моделей принятия коллективного решения.

5. Демонстрационный пример

Рассмотрим пример применения предложенных моделей при выборе наилучшего варианта операционного плана производства при заданном главном календарном плане производства (ГКПП) [5, 8].

Считается, что качество составления операционного плана производства оценивается по трем базовым показателям:

1. *Комфортность производства* (оценивается от 0 до 1). Календарный план производства должен быть скомпонован таким образом, чтобы операции одного вида запускались в производство как можно большими партиями, минимизируя при этом потери на переналадку оборудования.

2. *Равномерность производства* (от 0 до 1). Нагрузка на рабочие центры должна быть распределена таким образом, чтобы загрузка рабочих центров в течение планового периода изменялась постепенно, без резких перепадов.

3. *Минимальность срока изготовления* (от 0 до 1). Цикл производства каждой позиции ГКПП от момента запуска производства до выпуска должен быть минимальным.

Пусть в результате расчета найдено 3 допустимых варианта операционного плана производства (3 альтернативы), эффективность каждого из которых может характеризоваться различными значениями заданных показателей качества [8] (см. таблицу 1).

Таблица 1. Варианты планов производства

| Наименование альтернативы | Достигнутые значения показателей | | |
|-------------------------------|----------------------------------|---------------|-------------------|
| | Комфортность | Равномерность | Срок изготовления |
| Точно вовремя (x_1) | 0,62 | 0,44 | 0,89 |
| Максимальные партии (x_2) | 0,76 | 0,53 | 0,56 |
| Ритмичность (x_3) | 0,55 | 0,87 | 0,61 |

Отметим, что альтернатива x_1 соответствует тому варианту плана производства, при котором большинство изделий производится точно в срок, установленный ГКПП; альтернатива x_2 обеспечивает «комфортность» производства за счет увеличения объемов партий и уменьшения времени переналадок оборудования; альтернатива x_3 позволяет обеспечить более равномерную

загрузку оборудования при производстве запланированной продукции.

Для принятия коллективного решения сформирован комитет, который состоит из ГЛПР, включающего 10 экспертов из различных подразделений предприятия, которые оценивают данные варианты путем их сравнения, а также ОЛПР, возглавляющего этот комитет.

Количество возможных вариантов сравнения в данном случае равно 6. Результаты оценки экспертов приведены в таблице 2.

Таблица 2. Предпочтения экспертов на множестве альтернатив

| Вариант сравнения | Число экспертов, проголосовавших за данный вариант | Предпочтения экспертов |
|-------------------|--|---------------------------|
| 1 | 4 | $x_1 \succ x_2 \succ x_3$ |
| 2 | 2 | $x_1 \succ x_3 \succ x_2$ |
| 3 | 2 | $x_2 \succ x_1 \succ x_3$ |
| 4 | - | $x_2 \succ x_3 \succ x_1$ |
| 5 | 1 | $x_3 \succ x_1 \succ x_2$ |
| 6 | 1 | $x_3 \succ x_2 \succ x_1$ |

Первоначально произведем выбор наилучшей альтернативы с помощью модели *Кондорсе*. Количество экспертов, предпочитающих альтернативу x_1 по сравнению с x_2 , в данном случае равно $4 + 2 + 1 = 7$, альтернативу x_2 по сравнению с x_1 – $2 + 1 = 3$. Значит, альтернатива x_1 предпочтительней x_2 , т.е. $x_1 \succ x_2$. Аналогично сравниваем попарно альтернативы x_1 и x_3 , а также x_2 и x_3 . Для альтернатив x_1 и x_3 получим $4 + 2 + 2 = 8$ и $1 + 1 = 2$, т.е. $x_1 \succ x_3$. При сравнении x_2 и x_3 количество экспертов равно $4 + 2 = 6$ и 2 , т.е. $x_2 \succ x_3$. Переходя от индивидуальных предпочтений к коллективному решению, получим, что

$x_1 \succ x_2 \succ x_3$, т.е. в данном примере лучшей альтернативой является альтернатива x_1 .

Произведем выбор наилучшей альтернативы по модели Борда. Для этого подсчитаем число баллов, набранных каждой альтернативой по формуле (2):

$$b(x_1) = 6 \cdot 3 + 3 \cdot 2 + 1 \cdot 1 = 25 \text{ баллов};$$

$$b(x_2) = 2 \cdot 3 + 5 \cdot 2 + 3 \cdot 1 = 19 \text{ баллов};$$

$$b(x_3) = 2 \cdot 3 + 2 \cdot 2 + 6 \cdot 1 = 16 \text{ баллов}.$$

Значит, по модели Борда лучшим коллективным решением также является альтернатива x_1 .

Отметим, что в рассмотренном примере коллективное решение по двум моделям совпадает, что в данном конкретном случае говорит о лучшей обоснованности выбранного решения при принятых условиях. При этом выбор альтернативы x_1 большинством экспертов объясняется тем, что при заданных показателях качества сроки выполнения заказов для рассматриваемого предприятия играют основную роль. В общем случае ранжирование альтернатив по многим показателям является достаточно сложной задачей для экспертов. Поэтому желательно предоставлять экспертам дополнительную информацию о важности принятых показателей качества возможной альтернативы, характеризующих текущую политику предприятия на рынке продукции.

С учетом вышесказанного для выбора наилучшего коллективного решения с учетом важности частных показателей качества альтернативы применим линейную и нелинейную многокритериальные модели, рассмотренные выше.

Пусть для частных показателей качества альтернативы заданы оценки важности, установленные ОЛПР (таблица 3).

Таблица 3. Важности показателей качества альтернатив

| № | Показатели | Важность, v_k |
|---|-------------------|-----------------|
| 1 | Комфортность | 0,5 |
| 2 | Равномерность | 0,4 |
| 3 | Срок изготовления | 1,0 |

Первоначально произведем выбор наилучшего коллективного решения по многокритериальной линейной модели в предположении, что квалификация экспертов не связана с экономическим смыслом частных показателей качества альтернативы. Тогда предпочтения экспертов в модели задаются формулами (4), а выбор осуществляется по формуле (5).

Оценим линейные свертки альтернатив, исходя из достигнутых значений частных показателей и их важностей (таблица 4).

Таблица 4. Значения показателей качества альтернатив

| № | Наименование альтернативы | Достигнутые значения показателей | | | $S^i = \sum_{k=1}^r v_k \cdot J_k^i$ |
|-------|---------------------------|----------------------------------|---------|-------------|--------------------------------------|
| | | Комф. | Равном. | Срок изгот. | |
| x_1 | Точно вовремя | 0,62 | 0,44 | 0,89 | 1,38 |
| x_2 | Максимальные партии | 0,76 | 0,53 | 0,56 | 1,15 |
| x_3 | Ритмичность | 0,55 | 0,87 | 0,61 | 1,23 |

В таблице 5 приведены предпочтения каждого эксперта в виде значений λ_j^i , изменяющихся от 0 до 1, а также многокритериальные оценки альтернатив, рассчитанные для каждого эксперта. Подчеркнем, что в рамках данной модели варианты предпочтений экспертов формируются автоматически без их участия (последний столбец таблицы 5).

Таблица 5. Предпочтения экспертов

| № | Эксперт | Альтернативы | | | | | | Предпочтения экспертов |
|---|-------------|---------------|-------------------|---------------|-------------------|---------------|-------------------|---------------------------|
| | | x_1 | | x_2 | | x_3 | | |
| | | λ_j^1 | $\lambda_j^1 S^1$ | λ_j^2 | $\lambda_j^2 S^2$ | λ_j^3 | $\lambda_j^3 S^3$ | |
| 1 | «Продажа» 1 | 0,60 | 0,82 | 0,20 | 0,23 | 0,20 | 0,25 | $x_1 \succ x_3 \succ x_2$ |

| № | Эксперт | Альтернативы | | | | | | Предпочтения экспертов |
|----|------------------|---------------|------------------|---------------|------------------|---------------|------------------|---------------------------|
| | | x_1 | | x_2 | | x_3 | | |
| | | λ_j^1 | λ_j^{1S} | λ_j^2 | λ_j^{2S} | λ_j^3 | λ_j^{3S} | |
| 2 | «Продажа» 2 | 0,50 | 0,68 | 0,30 | 0,32 | 0,20 | 0,24 | $x_1 \succ x_2 \succ x_3$ |
| 3 | «Экономика» 1 | 0,34 | 0,48 | 0,33 | 0,35 | 0,33 | 0,40 | $x_1 \succ x_3 \succ x_2$ |
| 4 | «Экономика» 2 | 0,40 | 0,55 | 0,40 | 0,45 | 0,20 | 0,25 | $x_1 \succ x_2 \succ x_3$ |
| 5 | «Производство» 1 | 0,00 | 0,00 | 0,50 | 0,58 | 0,50 | 0,63 | $x_3 \succ x_2 \succ x_1$ |
| 6 | «Производство» 2 | 0,30 | 0,41 | 0,40 | 0,46 | 0,30 | 0,38 | $x_2 \succ x_1 \succ x_3$ |
| 7 | «Оборудование» | 0,25 | 0,34 | 0,50 | 0,58 | 0,25 | 0,30 | $x_2 \succ x_1 \succ x_3$ |
| 8 | «Персонал» | 0,25 | 0,35 | 0,25 | 0,30 | 0,50 | 0,63 | $x_3 \succ x_1 \succ x_2$ |
| 9 | «Финансы» 1 | 0,50 | 0,70 | 0,30 | 0,33 | 0,20 | 0,23 | $x_1 \succ x_2 \succ x_3$ |
| 10 | «Финансы» 2 | 0,60 | 0,85 | 0,25 | 0,30 | 0,15 | 0,20 | $x_1 \succ x_2 \succ x_3$ |

Следует отметить, что мнения экспертов, выраженное вариантами их предпочтений, совпало с результатами голосования по первым двум моделям. При этом предпочтение по 1-му из возможных вариантов выразили эксперты 2, 4, 9 и 10; по 2-му – 1 и 3; по 3-му – 6 и 7; по 5-му – 8; по 6-му – 5.

Значения функций *предпочтительности* альтернатив распределены следующим образом:

$$pr(x_1) = 5,18; pr(x_2) = 3,9; r(x_3) = 3,51.$$

Значит, по данной модели лучшим коллективным решением является альтернатива x_1 .

Отметим, что полученное коллективное решение совпадает с решениями, полученными по первым двум моделям. Однако преимуществом последней модели является упрощение работы экспертов при многокритериальном сравнении альтернатив. Им требуется только дать численные оценки каждой альтернативе, а конечное предпочтение формируется с помощью самой модели.

Как было отмечено выше, к недостаткам предложенной модели можно отнести гипотезу линейности при формировании комплексной оценки качества альтернативы. Поэтому рассмотрим применение четвертой модели, основанной на нечеткой постановке.

Результаты попарного сравнения частных показателей качества J_k^i с учетом их важности v_k , приведены в таблице 6.

Таблица 6. Результаты попарного сравнения альтернатив

| № | Пары альтернатив | | C_1 | C_2 | C_3 | Результат сравнения |
|---|------------------|----------|--------|--------------|---------------|---------------------|
| | J^{r1} | J^{r2} | | | | |
| 1 | x_1 | x_2 | -0,092 | -0,06 | 0,370 | $x_1 \succ x_2$ |
| 2 | x_2 | x_3 | 0,138 | -0,15 | -0,081 | $x_3 \succ x_2$ |
| 3 | x_3 | x_1 | -0,056 | 0,2 | -0,314 | $x_1 \succ x_3$ |

Полужирным шрифтом в таблице 6 отмечены максимальные по модулю элементы сравнения рассматриваемых альтернатив, входящих в специальные индексы ранжирования (6). В результате парного сравнения (последний столбец таблицы 6) можно сделать вывод о том, что лучшим коллективным решением по данной модели является альтернатива x_1 .

Отметим, что последняя из рассмотренных моделей позволяет провести комплексное оценивание альтернатив, при котором роль экспертов заключается не в оценивании каждой аль-

тернативы, а в установлении коэффициентов важности введенных частных показателей качества. В приведенном примере эти коэффициенты совпали с коэффициентами, установленными ОЛПР в рамках 3-й модели. Поэтому результаты коллективного решения по этим моделям совпали. Очевидно, что при изменении экспертной оценки важности частных показателей качества предпочтительная альтернатива может измениться.

6. Заключение

Рассмотрены четыре модели принятия коллективных решений, которые могут быть использованы для решения ситуационных задач, возникающих в производственных системах. Показано, что учет дополнительной исходной информации и применение современного математического аппарата комплексного оценивания позволяет повысить объективность принимаемого коллективного решения за счет снижения роли экспертов, а также повысить оперативность принятия управленческих решений в рамках промышленного предприятия.

Литература

1. АЛЕСКЕРОВ Ф.Т., ХАБИНА Э.Л., ШВАРЦ Д.А. *Бинарные отношения, графы и коллективные решения*. – М.: Физматлит, 2012. – 341 с.
2. АЛЕСКЕРОВ Ф.Т., ЮЗБАШЕВ Д.А., ЯКУБА В.И. *Пороговое агрегирование трехградационных ранжировок // Автоматика и телемеханика*. – 2007. – №1. – С. 147–152.
3. АНДРОННИКОВА Н.Г., ЛЕОНТЬЕВ С.В., НОВИКОВ Д.А. *Механизмы нечеткой активной экспертизы // Автоматика и телемеханика*. – 2002. – №8. – С. 128–135.
4. ВОЖАКОВ А.В., ГИТМАН М.Б., СТОЛБОВ В.Ю. *Ситуационный центр как инструмент интеллектуализации системы управления производством // Интеллектуальные системы в производстве*. – 2013. – №2(22). – С. 45–49.
5. ВОЖАКОВ А.В., ГИТМАН М.Б., ФЕДОСЕЕВ С.А. *Комплексное оценивание при выборе оптимального плана производства на тактическом уровне с учетом нечетких крите-*

- риев и ограничений // Управление большими системами. – 2010. – №30. – С. 164–179.*
6. ГУБКО М.В. *Математические модели оптимизации иерархических структур.* – М.: ЛЕНАНД, 2006. – 264 с.
 7. РЫКОВ А.С. *Модели и методы системного анализа: принятие решений и оптимизация.* – М.: МИСИС, 2005. – 352 с.
 8. ФЕДОСЕЕВ С.А., ГИТМАН М.Б., СТОЛБОВ В.Ю. и др. *Управление качеством продукции на современных промышленных предприятиях: монография.* – Пермь: Изд-во Перм. нац. исслед. политехн. ун-та, 2011. – 229 с.
 9. ФЕДОСЕЕВ С.А., ГИТМАН М.Б., СТОЛБОВ В.Ю. *Математические модели управления качеством продукции на этапе планирования производства // Проблемы управления.* – 2011. – №4. – С. 60–67.
 10. SAN R.K., STIGLITZ J.E. *The Committees, Hierarchies and Polyarchies // The Economic Journal.* – 1988. – Vol. 98. – No. 391. – P. 451–470.

MODELS OF COLLECTIVE DECISION-MAKING IN PRODUCTION

Artem Vozhakov, Motovilikhinskiye Zavody, Perm, *Cand.Sc.*,
(vozhakov@ya.ru)

Mikhail Gitman, Perm National Research Polytechnic University,
Perm, Doctor of Science, professor (mgiman@netzero.net)

Yaleriy Stolbov, Perm National Research Polytechnic University,
Perm, Doctor of Science, professor (valeriy.stolbov@gmail.com)

Abstract: Four models of collective decision-making are considered. The models are based on the comparison of alternatives and on the expert group selection. A committee is considered to be an organizational form of expert work. The models are compared in an illustrative example of the best production plan selection in an industrial company.

Keywords: models of adoption of the collective decision, set of alternatives, private indicators of quality, preference of experts.

*Статья представлена к публикации
членом редакционной коллегии Ф.Т. Алескеровым*

*Поступила в редакцию 21.04.2015.
Опубликована 30.11.2015.*

УДК 517.977.5

ББК 01.01.01

ОПТИМАЛЬНЫЕ МЕХАНИЗМЫ ПЛАНИРОВАНИЯ ПОЗАКАЗНОГО ПРОИЗВОДСТВА ПО ФИНАНСОВЫМ И ВРЕМЕННЫМ КРИТЕРИЯМ

Гераськин М. И.¹, Егорова В. В.²

*(Национальный исследовательский университет
«Самарский государственный аэрокосмический
университет имени С.П. Королева», Самара)*

Рассматривается проблема планирования оптимальных заказов ресурсов производственных фирм в отраслях промежуточной переработки, производящих товары конечного потребления и спрос на продукцию которых определяется ограниченным кругом покупателей. На основе анализа факторов, предопределяющих позаказную систему планирования фирм на рынках олигопсонии, разработаны модели оптимизации заказов производственных ресурсов фирмы по критериям прибыли, а также производственного и операционного циклов при монотонно возрастающих трендах параметров издержек, выручки и денежного потока с учетом ценовых и технологических ограничений. Сформированы аналитические механизмы оптимального планирования заказов, на основе которых для фирм подшипниковой промышленности проведено моделирование области компромисса заказов, оптимальных по различным критериям.

Ключевые слова: модель оптимизации, оптимальный механизм планирования, производственный цикл, операционный цикл, прибыль, многокритериальная область компромисса.

¹ Михаил Иванович Гераськин, доктор экономических наук, профессор (Самара, Московское шоссе, д. 34, (846) 267-44-96, innovation@ssau.ru)

² Виктория Викторовна Егорова, ассистент (Самара, Московское шоссе, д.34, (846) 267-44-96, v_v_egorova@inbox.ru)

1. Введение

Организация промышленного производства в отраслях промежуточной переработки (машиностроение и металлообработка, в том числе подшипниковая промышленность, производство строительных материалов, электротехническая промышленность и др.) в современной мировой, а также российской экономике базируется на позаказной системе планирования, отражающей олигопсоническую структуру рынка, преобладание которой обусловлено действием следующих факторов. Аккумуляция капитала и концентрация производства в рамках крупных корпораций, доминирующих во второй половине XX века, способствовали обострению ценовой и объемной конкуренции, что, в свою очередь, предопределило перенасыщение рынков конечного потребления и постепенное превращение экономики предложения в экономику спроса, в которой сбыт товаров имеет преобладающее значение над производством. Конкуренция продавцов, в отличие от конкуренции производителей, на современном этапе выражается в наращивании диверсификации ассортимента однотипных товаров за счет расширения многообразия их качеств, что неизбежно приводит к сокращению объемов серий. В различных отраслях, производящих товары конечного потребления (автомобилестроение, производство бытовой техники и электроники, жилищное строительство и др.), систематически обновляются модельные ряды при снижении объемов партий, что обуславливает рост номенклатуры и сокращение объемов заказов в перерабатывающих отраслях.

Позаказная модель планирования основана на стратегической зависимости производителя от заказчика и предопределяет возникновение следующих системных проблем, осложняющих развитие производственных фирм. Колебания спроса сезонного или циклического характера побуждают производителя накапливать значительные запасы материальных ресурсов, что приводит к иммобилизации финансовых ресурсов и упущенной прибыли. Динамичное варьирование ассортимента заказов обуславливает непрогнозируемые потребности в специфических ресурсах, производители которых в связи с этим устанавливают

повышенные цены, вследствие чего производственная фирма попадает в стратегическую зависимость также и от поставщиков ресурсов. В результате в условиях позаказного мелкосерийного полиассортиментного производства прибыль фирмы перестает быть основным критерием эффективности; на передний план выходит фактор времени, выражающийся в длительности бизнес-цикла и позволяющий оценить реальную (с учетом недополученной прибыли вследствие иммобилизации в остатках ресурсов), а не номинальную прибыль.

Наряду с отмеченными проблемами фирмы перерабатывающих отраслей, ориентированные на так называемый производный, т.е. обусловленный потребностями конечного потребления, спрос, получают определенное преимущество. Развиваются механизмы стабилизации системы внутрифирменного планирования относительно периодических повторяющихся (сезонно или циклически) подъемов и спадов спроса на конечные продукты. В результате влияние колебаний спроса при постоянных структуре заказчиков и общетехнических требованиях к качеству продукции на длительном ретроспективном временном горизонте приводит к формированию устойчивых трендов важнейших параметров состояния фирм под действием адаптационных механизмов в подсистемах снабжения, производства и сбыта за счет координации с циклическим и сезонным компонентами спроса. В итоге колебательный процесс снабжения синхронизируется с экзогенно заданной динамикой процесса сбыта, что позволяет использовать инструментарий производственных функций для построения моделей взаимосвязей ресурсов с результатами производства (выпуском, выручкой и прибылью).

Таким образом, для большинства современных промышленных фирм сектора промежуточной переработки базовая задача оптимального планирования ассортимента при экзогенной функции спроса и эндогенной функции издержек трансформируется в задачу оптимального планирования заказов ресурсов при экзогенной производственной функции выручки (и издержек) от ресурсов, экзогенных функциях цен ресурсов, и ограничениях по объемам заказов покупателей. При этом критериями оптимальности, наряду с прибылью, являются периоды

иммобилизации финансовых ресурсов в фазах производства (производственные циклы) и реализации продукции (операционные циклы), отдельный учет которых играет весомую роль для фирм, имеющих стабильную высокую дебиторскую задолженность и для которых средства в расчетах выступают как особо значимый ресурс.

2. Анализ оптимальных механизмов производственного планирования

В современных исследованиях задача оптимального планирования, как правило, формулировалась для одного из подпроцессов бизнес-процесса промышленной фирмы (заготовления, производства или сбыта) при ограничениях на параметры управления другими подпроцессами; в качестве критерия оптимальности рассматривалась прибыль в форме линейной функции ресурсов, причем игнорировались временные показатели бизнес-процесса.

Однопроцессные модели формулировались в виде задач объемно-календарного оперативного планирования производственной программы [1, 14, 15, 18]. Моделирование этапов заготовления и выпуска исследовалось на основе линейных статических моделей прибыли с учетом ограничений на складские остатки [20]; анализировалось влияние функций спроса покупателей на заготовительные расходы [10]; изучалась взаимосвязь между остатками ресурсов и заказами на продукцию с учетом ограничений на производственные мощности [16, 21], а также влияние различных критериев (издержек, прибыли и выручки) на выбор объемов заказов ресурсов; оптимизировалась отпускная цена производителя с учетом расходов на заготовление и хранение ресурсов [5, 6]. Моделирование этапов заготовления и производства рассматривалось в [7, 22] в форме моделей терминального управления по критерию минимума отклонений производственных расходов от плановых показателей производственной программы; оптимизировалась себестоимость продукции по стадиям технологического процесса производства [8, 13].

Комплексное моделирование процессов заготовления, производства и сбыта также в рамках линейных моделей прибыли проводилось путем оптимизации производственной программы при ограничениях на поставки сырья, производственные и складские мощности [17]. Многопроцессные модели оптимизации формулировались в виде задач с несколькими критериями: рассматривалось календарное планирование с нечеткими ограничениями по критериям прибыли и затрат [3]; оптимизировался ассортимент выпускаемой продукции по критериям объема продаж, прибыли, себестоимости и трудоемкости [23]; оптимизировались проекты с точки зрения критериев времени и стоимости выполнения [9].

Таким образом, современные механизмы оптимального производственного планирования не учитывают, во-первых, наличия устойчивых взаимосвязей между показателями различных подпроцессов бизнес-процесса фирмы, в том числе в виде нелинейных зависимостей; во-вторых, существенности временного фактора в виде производственных и операционных циклов иммобилизации ресурсов наряду с финансовым фактором (прибылью) для выбора целей оптимизации. Поэтому актуальной представляется задача разработки оптимальных механизмов планирования заказов ресурсов для реализации заданной производственной программы с позиций комплекса временных и финансовых критериев оптимальности.

3. Модели оптимального планирования позаказного производства

Рассматривается задача оптимального планирования заказов производственных ресурсов в бизнес-процессе фирмы по критериям, характеризующим временную и финансовую эффективность:

$$(1) \min F_1(u),$$

$$(2) \min F_2(u),$$

$$(3) \max F_3(u),$$

где F_1 , F_2 – длительности производственного и операционного циклов фирмы, F_3 – прибыль фирмы. Матрица параметров

управления включает в себя объемы заказов (предъявлений требований к поставке или оплате) ресурсов на различных стадиях производственного процесса:

$$(4) \quad u = \{u_i^j, i = 1, \dots, 4, j = 1, \dots, J_i\}.$$

В качестве ресурсов исследуются такие виды (обозначены индексом i), как материалы, незавершенное производство (НЗП), готовая продукция, средства в расчетах, номенклатура которых обозначена индексом j .

Критерии эффективности определяются по формулам

$$(5) \quad F_1(u) = \frac{T}{C(u)} \left(u_{\Sigma}^0 + \frac{u_1 + P(u) - C(u)}{2} \right), \quad u_{\Sigma}^0 = \sum_{i=1}^3 u_i^0,$$

$$(6) \quad F_2(u) = F_1(u) + \frac{T}{R(u)} \left(u_4^0 + \frac{R(u) - g(u)}{2} \right),$$

$$(7) \quad F_3(u) = R(u) - C(u),$$

где T – продолжительность периода (в днях); u_i^0 , $i = 1, \dots, 4$ – остатки ресурсов (материалов, НЗП, готовой продукции, средств в расчетах соответственно) на начало периода; R , C – выручка и себестоимость продаж за период; P – нематериальные производственные расходы за период (расходы на оплату труда, социальные отчисления, амортизация и прочие расходы); g – денежный поток от покупателей за период.

Поскольку выражения циклов (5), (6) включают в себя в качестве аргумента в явном виде объем заказов материальных ресурсов u_1 , представим функции нематериальных расходов, издержек и выручки также в зависимости от этого аргумента, что позволит свести задачу многомерной оптимизации (1)–(3) к оптимизации функций одной переменной u_1 . Для этого *предположим наличие устойчивых взаимосвязанных трендов* динамики объемов заказов на продукцию фирмы и объемов заказов ресурсов на различных стадиях производства в виде функций материалоемкости производства $P(u_1), P'_{u_1}(u_1) > 0$, материалоемкости выпуска продукции $C(u_1), C'_{u_1}(u_1) > 0$, материалоемкости отгруженной продукции $R(u_1), R'_{u_1}(u_1) > 0$, материалоемкости

денежного потока $g(u_1), g'_{u_1}(u_1) > 0$. Рассмотрим следующие модели этих трендов в виде степенных функций связи параметров управления и состояния фирмы:

$$(8) P(u_1) = B_p u_1^{\beta_p}, \quad B_p > 0, \quad 0 < \beta_p < 2,$$

$$(9) C(u_1) = B_c u_1^{\beta_c}, \quad B_c > 0, \quad 0 < \beta_c < 2,$$

$$(10) R(u_1) = B_R u_1^{\beta_R}, \quad B_R > 0, \quad 0 < \beta_R < 2,$$

$$(11) g(u_1) = B_g u_1^{\beta_g}, \quad B_g > 0, \quad 0 < \beta_g < 2,$$

где $B_c, B_p, B_R, B_g, \beta_c, \beta_p, \beta_R, \beta_g$ – коэффициенты регрессионных моделей, ограничение $0 < \beta < 2$ наложено в связи с реальным характером эффекта расширения масштаба [2]. Отметим, что функции (8)–(11) могут не отражать непосредственные регрессионные связи с аргументом u_1 , а формируются путем выявления наиболее коррелируемых из компонентов матрицы (4), а затем последовательного получения регрессий, как будет показано при моделировании.

Также предположим наличие убывающих трендов цен ресурсов, закупаемых фирмой, $z_{j_1} = z_{j_1}(u_{1j_1}), j_1 = 1, \dots, J_1, z'_{j_1 u} < 0$, моделируемых в виде степенных функций:

$$(12) z_{j_1}(u_{1j_1}) = A_{z_{j_1}} u_{1j_1}^{\alpha_{z_{j_1}}}, \quad \alpha_{z_{j_1}} < 0, \quad j_1 = 1, \dots, J_1,$$

где $A_{z_{j_1}}, \alpha_{z_{j_1}}$ – коэффициенты регрессионных моделей цен ресурсов.

Учитываются следующие ограничения на управление. Ограничение по нормативу потребности ресурсов в зависимости от объема заказов покупателей определяется исходя из норм расхода материалов, а также наличных остатков готовой продукции и НЗП:

$$(13) \begin{aligned} u_1^{\max}(N) &\geq u_1 \geq u_1^{\min}(N), \\ u_{1j_3}^{\max(\min)} &= \sum_{j_1=1}^{J_1} \left[z_{j_1} \sum_{j_3=1}^{J_3} m_{j_1 j_3} \left(N_{j_3}^{\max(\min)} - u_{3j_3}^0 - k_{j_3} \cdot u_{2j_3}^0 \right) \right], \end{aligned}$$

где $m_{j_1 j_3}$ – массовый норматив расхода j_1 -го ресурса на производство единицы продукции j_3 -го типа; $N_{j_3}^{\max(\min)}$ – диапазон

колебаний спроса на продукцию j_3 -го типа; k_{j_3} – коэффициент выхода готового изделия j_3 -го типа из НЗП; z_{j_1} – закупочная цена j_1 -го ресурса (за единицу массы).

Ограничение по предельному уровню заготовительных издержек

$$(14) u_1 \geq u_1^{\min}(z)$$

обусловлено ростом цены ресурсов при снижении u_1 по тренду (12).

Сформулируем задачи оптимизации заказов ресурсов по временным и финансовым критериям:

$$(15) u_{1(1)}^* = \arg \min_{u_1 \in U} F_1(u_1),$$

$$(16) u_{1(2)}^* = \arg \min_{u_1 \in U} F_2(u_1),$$

$$(17) u_{1(3)}^* = \arg \max_{u_1 \in U} F_3(u_1),$$

с учетом ограничений

$$(18) U = \left\{ u_1 \in R^+ \mid u_1^{\max}(N) \geq u_1 \geq u_1^{\min}(N), u_1 \geq u_1^{\min}(z) \right\},$$

где U – область допустимых значений u_1 .

4. Механизмы оптимального планирования заказов ресурсов

4.1. МЕХАНИЗМ МИНИМИЗАЦИИ ИЗДЕРЖЕК

В подразделе решена вспомогательная задача оценки нижней границы заказов ресурсов для ограничения (14). Определим граничное условие в ограничении (14) исходя из условия минимума суммарных производственных расходов

$$(19) P_{\Sigma}(u) = \sum_{j_1=1}^{J_1} z_{j_1}(u_{1_{j_1}})u_{1_{j_1}} + P(u_1).$$

Утверждение 1. Минимум суммарных производственных расходов (19) при трендах вида (8), (12) достигается для неотрицательного аргумента $u_{1_{j_1}}^{\min} \geq 0$ при условиях

$$(20) \quad u_{1j_1}^{\min}(z) = \left(\frac{B_p \beta_p}{-(\alpha_{zj_1} + 1) A_{zj_1}} \right)^{\frac{1}{\alpha_{zj_1}}} \left(\sum_{j_1=1}^{J_1} u_{1j_1}^{\min}(z) \right)^{\frac{\beta_p - 1}{\alpha_{zj_1}}},$$

$$j_1 = 1, \dots, J_1,$$

$$\beta_p \leq 1 \cap \alpha_{zj_1} < -1 \cap$$

$$(21) \quad \cap A_{zj_1} (\alpha_{zj_1} + 1) \alpha_{zj_1} u_{1j_1}^{\alpha_{zj_1} - 1} > B_p \beta_p |\beta_p - 1| \left(\sum_{j_1=1}^{J_1} u_{1j_1} \right)^{\beta_p - 2}.$$

Выполнение достаточного условия минимума суммарных производственных расходов (21) соответствует высокоэластичным кривым цен ресурсов ($\alpha_{zj_1} < -1$). Как правило, на практике реализуется случай низкоэластичной кривой цены ($|\alpha_{zj_1}| < 1$), следовательно, ограничение (14) не играет роли, поскольку минимум производственных расходов достигается при $u_{1j_1}^{\min} < 0$. Отметим, что условие (21) выполняется при типичном для производственных фирм замедленном росте нематериальных расходов по увеличению заказов ресурсов ($\beta_p \leq 1$).

В предположении, что фирма использует однотипные материальные производственные ресурсы, можно приближенно заменить фактические значения коэффициентов функций цен и объемов заказов ресурсов на средние значения (обозначены $\bar{\alpha}_z, \bar{A}_z, \bar{u}_1^{\min}$); тогда уравнение (20) имеет аналитическое решение

$$(22) \quad \bar{u}_1^{\min}(z) = \left(-\frac{B_p \beta_p J_1^{\beta_p - 1}}{(\bar{\alpha}_z + 1) \bar{A}_z} \right)^{\frac{1}{\bar{\alpha}_z - \beta_p - 1}}.$$

Таким образом, нижняя граница заказов ресурсов действует при низкоэластичной кривой цены и определяется по (22).

Далее сформируем механизмы оптимального планирования заказов ресурсов по критериям (1)–(3) без учета ограничений (18), обозначив оптимальные заказы $u_{1F1}^*, u_{1F2}^*, u_{1F3}^*$. Предварительно отметим, что совместный анализ трендов (8), (9) показывает, что темпы роста нематериальных расходов и издержек должны удовлетворять соотношению

$$(23) \beta_p < \beta_c + 1,$$

поскольку в противном случае ($\beta_p > \beta_c + 1$) рост нематериальных расходов должен значительно опережать рост себестоимости, что в невозможно промышленности, для которой характерна высокая материалоемкость. Аналогично, анализ трендов (10), (11) приводит к следующему практическому ограничению на соотношение темпов роста выручки и денежного потока от покупателей

$$(24) \beta_g < \beta_R + 1,$$

поскольку обратное соотношение ($\beta_g > \beta_R + 1$) означает опережающий рост денежного потока относительно роста объема отгруженной продукции, что невозможно в долгосрочной тенденции.

4.2. МЕХАНИЗМ МИНИМИЗАЦИИ ПРОИЗВОДСТВЕННОГО ЦИКЛА

В подразделе определим механизм оптимального планирования заказов ресурсов по критерию минимума производственного цикла.

Утверждение 2. Минимум длительности производственного цикла (5) при трендах (8), (9) с учетом (23) достигается для неотрицательного аргумента $u_{1F1}^* \geq 0$, удовлетворяющего условиям

$$(25) -2u_{\Sigma}^0 \beta_c + (1 - \beta_c)u_{1F1}^* + B_p(\beta_p - \beta_c)u_{1F1}^{*bp} = 0,$$

$$(26) u_{1F1}^* \in U_1 = \left\{ \begin{array}{l} \beta_c > \beta_p \cup \beta_c < \beta_p \cap \beta_c + 1 > \beta_p \cap \\ \cap \varphi(\beta_c, \beta_p, u_{\Sigma}^0, u_1) > 0 \end{array} \right\},$$

где введено обозначение

$$\begin{aligned} \varphi(\beta_c, \beta_p, u_{\Sigma}^0, u_1) &= \beta_c [2(\beta_c + 1)u_{\Sigma}^0 + \beta_c - 1] u_1 + \\ & B_p |\beta_c - \beta_p| (\beta_c + 1 - \beta_p) u_1^{bp}. \end{aligned}$$

В некоторых частных, но характерных случаях, уравнение (25) можно решить аналитически в виде следующих механизмов оптимизации заказов ресурсов по критерию длительно-

сти производственного цикла, области применимости которых определены с учетом $u_{1F1}^* \geq 0$:

$$(27) u_{1F1}^* = \left(\frac{2u_{\Sigma}^0 \beta_C}{B_P (\beta_P - \beta_C)} \right)^{1/\beta_P} \quad \forall \beta_C \approx 1, \text{ при } \beta_P > \beta_C,$$

$$(28) u_{1F1}^* = \frac{2u_{\Sigma}^0 \beta_C}{1 - \beta_C} \quad \forall \beta_C \approx \beta_P, \text{ при } \beta_C < 1,$$

$$(29) u_{1F1}^* = \frac{2u_{\Sigma}^0 \beta_C}{1 - \beta_C + B_P (\beta_P - \beta_C)} \quad \forall \beta_P \approx 1, \text{ при } \beta_C < \frac{1 + B_P \beta_P}{1 + B_P},$$

$$(30) u_{1F1}^* = \frac{2u_{\Sigma}^0 \beta_C - B_P (\beta_P - \beta_C)}{1 - \beta_C} \quad \forall \beta_P \ll 1, \text{ при } \beta_C < 1.$$

Механизм (27), область применимости которого соответствует опережающему росту нематериальных расходов над общими издержками ($\beta_P > \beta_C$), получен из условия прямой пропорциональности себестоимости продукции и заказов ресурсов ($\beta_C \approx 1$), что отражает полное использование материальных ресурсов в производстве; при такой системе планирования не происходит накопления запасов и не возникает дефицит ресурсов. Механизм (28), применимый в случае замедленного роста издержек по сравнению с заказами ресурсов ($\beta_C < 1$), отражает совпадение темпа роста нематериальных производственных расходов с темпом роста себестоимости ($\beta_P \approx \beta_C$) и реализуется в условиях жестких технологических нормативов, обеспечивающих постоянное соотношение общих издержек и нематериальных расходов, что типично при системе планирования нормативного планирования (стандарт-костинг). Механизм (29), применимый, как правило, в случае $\beta_C < 1$, поскольку обычно $\beta_P < 1$, соответствует концепции нормативного планирования, как и вариант (28), но при условии установления постоянного соотношения между заказами материальных ресурсов и уровнем нематериальных расходов ($\beta_P \approx 1$). Наконец, вариант (30) типичен для фирм, производственный процесс которых отличается высокой материалоемкостью (металлургия, машиностроение и металлообработка, промышленность строительных материалов), вследствие чего темп роста нематериальных расходов значи-

тельно ниже темпа роста затрат материальных ресурсов ($\beta_P \ll 1$); область применимости механизма (30) также соответствует $\beta_C < 1$. Таким образом, аналитические механизмы (27)–(30) применимы для наиболее характерного способа построения производственного процесса современных промышленных фирм, при котором за счет технологических эффектов темп роста издержек ниже темпа роста материальных ресурсов.

Анализ механизмов (27)–(30) показывает, что значение u_{1F1}^* по порядку величины близко к u_Σ^0 , а поскольку $u_\Sigma^0 \gg u_1^0$, то оптимум цикла имеет место при значениях, существенно превышающих средние остатки материальных ресурсов.

Отметим, что с учетом соотношения порядка величин остатков запасов u_Σ^0 , объема заказов u_{1F1}^* и коэффициентов регрессий β_C, β_P выполняется следующее соотношение

$$\beta_C [2(\beta_C + 1)u_\Sigma^0 + \beta_C - 1] u_1 \gg B_P |\beta_C - \beta_P| (\beta_C + 1 - \beta_P) u_1^{\beta_P},$$

поэтому достаточное условие (26) выполняется практически для всех $u_{1F1}^* \geq 0$, следовательно

$$(31) \varphi(\beta_C, \beta_P, u_\Sigma^0, u_1) > 0 \quad \forall u_{1F1}^* \geq 0.$$

Условие (31) используем в дальнейшем для анализа достаточных условий минимума операционного цикла и максимума прибыли.

Полученные приближенные механизмы оптимального планирования заказов ресурсов по критерию минимума производственного цикла (27)–(30) определяют решение задачи (1) в практически важных ситуациях.

4.3. МЕХАНИЗМ МИНИМИЗАЦИИ ОПЕРАЦИОННОГО ЦИКЛА

В подразделе определим механизм оптимального планирования заказов ресурсов по критерию минимума операционного цикла.

Утверждение 3. Минимум длительности операционного цикла (6) при трендах (8)–(11) с учетом (23), (24) достигается

для неотрицательного аргумента $u_{1F2}^* \geq 0$, удовлетворяющего условиям

$$(32) \quad \begin{aligned} & -\beta_c \frac{u_\Sigma^0}{B_c} + \frac{1-\beta_c}{2B_c} u_{1F2}^* + \frac{B_p}{2B_c} (\beta_p - \beta_c) u_{1F2}^{*\beta_p} - \\ & -\beta_R \frac{u_4^0}{B_R} u_{1F2}^{*\beta_R + \beta_c} - \frac{B_g}{2B_R} (\beta_g - \beta_R) u_{1F2}^{*\beta_g - \beta_R + \beta_c} = 0, \end{aligned}$$

$$(33) \quad \begin{aligned} & u_{1F2}^* \in U_2 = \\ & \left\{ \begin{aligned} & (\beta_g > \beta_R \cap \beta_g < \beta_R + 1) \forall u_{1F2}^*, \\ & \beta_g \leq \beta_R \cap \varphi(\beta_c, \beta_p, u_\Sigma^0, u_{1F2}^*) + \beta_R (\beta_R + 1) \frac{u_4^0}{B_R} u_{1F2}^{*\beta_R - 2} + \\ & + \frac{B_g}{2B_R} (\beta_g - \beta_R) (\beta_R + 1 - \beta_g) u_{1F2}^{*\beta_g - \beta_R - 2} > 0. \end{aligned} \right. \end{aligned}$$

В характерных частных случаях уравнение (32) можно решить аналитически в виде следующих механизмов оптимизации заказов ресурсов по критерию длительности операционного цикла:

$$(34) \quad u_{1F2}^* = \frac{2(B_R \beta_c u_\Sigma^0 + B_C \beta_R u_4^0)}{B_g B_C (\beta_g - \beta_R) - B_R (1 - \beta_C) - B_p B_R (\beta_p - \beta_C)}$$

$$\forall \beta_p \approx 1 \cap \beta_c \approx 1 \cap \beta_R \approx \beta_C \cap \beta_R \approx \beta_g,$$

$$(35) \quad u_{1F2}^* = \frac{2(B_R \beta_c u_\Sigma^0 + B_C \beta_R u_4^0) - B_p B_R (\beta_p - \beta_C)}{B_R (1 - \beta_C) - B_g B_C (\beta_g - \beta_R)} \forall$$

$$\beta_p \ll 1 \cap \beta_c \approx \beta_R \cap \beta_g \approx 1.$$

Области применимости механизмов (34), (35), определим из условия $u_{1F2}^* \geq 0$ с учетом того, что числители этих выражений неотрицательны (для формулы (35) это следует из условия $\beta_p < \beta_C$) в виде ограничений положительности знаменателей:

$$(34a) \quad B_g B_C (\beta_g - \beta_R) - B_R (1 - \beta_C) - B_p B_R (\beta_p - \beta_C) > 0,$$

$$(35a) \quad B_R (1 - \beta_C) - B_g B_C (\beta_g - \beta_R) > 0.$$

Ограничение (35а) выполняется в случае замедленного роста издержек ($\beta_C < 1$) и опережающего роста выручки над ростом денежного потока ($\beta_R > \beta_g$), а также в случае $\beta_R < \beta_g$, но при условии $B_R(1 - \beta_C) > B_g B_C |\beta_g - \beta_R|$. Ограничение (34а) выполняется при противоположных соотношениях темповых параметров: при $\beta_C > 1$ для случая $\beta_R < \beta_g$, а при $\beta_C < 1$ – в случае $B_P B_R (\beta_P - \beta_C) > B_g B_C (\beta_g - \beta_R) + B_R(1 - \beta_C)$.

Вариант (34) аналогичен рассмотренной выше комбинации случаев (27), (28) и характеризуется строгим соответствием объемов заказов ресурсов программе выпуска ($\beta_C \approx 1$) при соблюдении нормативов нематериальных расходов ($\beta_P \approx \beta_C \approx 1$) по отношению к темпу изменения общих издержек, а также нормативов производства продукции в соответствии с темпами роста спроса ($\beta_R \approx \beta_C$); но, наряду с этим, бизнес-процесс фирмы отличается строгой платежно-расчетной дисциплиной покупателей, когда темп роста отгруженной продукции не существенно отличен от темпа роста денежного потока $\beta_R \approx \beta_g$. Вариант (35) соответствует производству с высокой материалоемкостью (как и в случае (30), $\beta_P \ll 1$), в котором контролируется норматив выпуска продукции в соответствии с темпами роста спроса ($\beta_R \approx \beta_C$), т.е. не происходит значительного накопления товарных остатков ($\beta_R < \beta_C$), а также отсутствует выраженная тенденция роста отпускных цен товаров, опережающего рост цен ресурсов ($\beta_R > \beta_C$); кроме того, система снабжения фирмы скорординирована с динамикой денежного потока ($\beta_g \approx 1$), что характерно для фирм, поставщики которых отпускаю ресурсы на основе предварительной оплаты.

Поскольку из (34), (35) следует, что u_{1F2}^* по порядку величины близко к $u_{\Sigma}^0 + u_4^0$, то оптимум операционного цикла достигается при объеме заказов ресурсов, значительно превышающем средние остатки ресурсов.

Достаточное условие минимума длительности операционного цикла (33) соблюдается в случае опережающего роста денежного потока по сравнению с ростом выручки ($\beta_R < \beta_g < \beta_R + 1$), а также если темп роста денежного потока не превышает темп роста отгруженной продукции ($\beta_g \leq \beta_R$) при

дополнительном ограничении, которое, введя упрощающие предположения, экономически описанные выше, $\beta_P \approx \beta_C \approx 1$, $\beta_R \approx \beta_g \approx 1$, представим в виде неравенства

$$\beta_C(\beta_C - 1) \left[2u_{\Sigma}^0 + u_{1F2}^* \right] + B_P(\beta_P - \beta_C)(\beta_P - \beta_C - 1)u_{1F2}^* + \\ + \beta_R(\beta_R + 1) \frac{u_4^0}{B_R} u_{1F2}^{*-3} - \frac{B_g}{2B_R} (\beta_g - \beta_R)(\beta_g - \beta_R - 1)u_{1F2}^{*-2} > 0,$$

отсюда, пренебрегая несущественными слагаемыми (полагая $u_4^0 \cong u_{1F2}^*$, $u_{1F2}^{*-2} \ll u_{1F2}^*$), получим

$$u_{1F2}^* > - \frac{\beta_C(\beta_C - 1)2u_{\Sigma}^0}{\beta_C(\beta_C - 1) + B_P(\beta_P - \beta_C)(\beta_P - \beta_C - 1)},$$

которое при тех же условиях, что и достаточное условие минимума производственного цикла (26) ($\beta_C > 1$, $\beta_P < \beta_C$) выполняется для неотрицательных аргументов $u_{1F2}^* \geq 0$.

Таким образом, механизмы оптимального планирования заказов ресурсов по критерию минимума операционного цикла (34), (35) определяют решение задачи (2) в практически важных ситуациях.

4.3. МЕХАНИЗМЫ МАКСИМИЗАЦИИ ПРИБЫЛИ И КОМПЛЕКСНОЙ ОПТИМИЗАЦИИ

В подразделе определим механизм оптимального планирования заказов ресурсов по критерию максимума прибыли.

Утверждение 4. Максимум прибыли (7) при трендах (9), (10) достигается при неотрицательном значении аргумента $u_{1F3}^* \geq 0$, удовлетворяющим условиям

$$(36) \quad u_{1F3}^* = \left(\frac{B_R \beta_R}{B_C \beta_C} \right)^{\frac{\beta_R - 1}{\beta_C - 1}},$$

$$(37) u_{1F3}^* \in U_3 = \left\{ \begin{array}{l} \beta_C > 1 \cap \left\{ \left(\beta_R > 1 \cap \frac{B_R(\beta_R - 1)\beta_R}{B_C(\beta_C - 1)\beta_C} < u_{1F3}^{*\frac{\beta_C - 2}{\beta_R - 2}} \right) \cup \right. \\ \cup (\beta_R < 1 \vee \cap u_{1F3}^* > 0) \left. \right\}, \\ \beta_C < 1 \cap \left\{ \left[\beta_R < 1 \cap \frac{B_R(\beta_R - 1)\beta_R}{B_C(\beta_C - 1)\beta_C} > u_{1F3}^{*\frac{\beta_C - 2}{\beta_R - 2}} \right] \cup \right. \\ \left. \cup [\beta_C < 1 \cap \beta_R > 1] \right\}. \end{array} \right.$$

Анализ механизма оптимизации прибыли фирмы показывает, что оптимум прибыли u_{1F3}^* возрастает с увеличением коэффициентов регрессии выручки B_R , β_R и снижается с ростом коэффициентов регрессии издержек B_C , β_C . Сравнение этих тенденций с механизмами оптимизации циклов фирмы приводит к следующим выводам: из механизмов (27)–(30) следует, что с увеличением темпа регрессии издержек β_C оптимум производственного цикла растет, а параметры B_C , B_R , β_R на этот оптимум не влияют, т.е. критерии прибыли и производственного цикла противоречивы по β_C ; механизмы (34), (35) показывают, что оптимум операционного цикла опережающим образом растет с увеличением темпа регрессии издержек β_C даже на фоне роста β_R (поскольку $u_{\Sigma}^0 \gg u_4^0$), поэтому критерии прибыли и операционного цикла противоречивы по β_C . Следовательно, *изменение эндогенной функции издержек фирмы оказывает противоречивое влияние на финансовый и временные критерии бизнес-процесса.*

Поскольку в механизме (36) коэффициенты регрессий B_R , B_C и β_R , β_C есть величины соответственно одного порядка, то максимизирующее прибыль значение u_{1F3}^* должно быть не велико.

Достаточное условие максимума прибыли для случая $\beta_C < 1$, при котором анализировались достаточные условия минимума циклов, приводит к выводу, что максимум прибыли достигается практически при любом неотрицательном значении

аргумента $u_{1F3}^* \geq 0$, поскольку из (37) следует, что должно выполняться условие замедленного по сравнению с нарастанием заказов темпа роста выручки ($\beta_R < 1$), а также неравенство

$$u_{1F3}^* > \frac{\beta_C - 2}{\beta_R - 2} \frac{B_R (\beta_R - 1) \beta_R}{B_C (\beta_C - 1) \beta_C},$$

в котором

$$\frac{\beta_C - 2}{\beta_R - 2} \approx 1, \frac{B_R (\beta_R - 1) \beta_R}{B_C (\beta_C - 1) \beta_C} \approx 1,$$

так как соответствующие коэффициенты регрессий являются величинами одного порядка. В случае $\beta_C < 1$ значение u_{1F3}^* , соответствующее условию (37), должно быть порядка единицы, что для реальных фирм равносильно закрытию, поэтому данный случай в дальнейшем не рассматривается.

Утверждение 5. Механизмы

$$(38) \quad u_{1(k)}^* = \max \left\{ \min \left\{ u_{1Fk}^*, u_1^{\min}(N), u_1^{\min}(z) \right\}, u_1^{\max}(N) \right\}, \quad k = 1, 2, 3,$$

являются решениями задач оптимизации (15)–(17) при ограничении (18), если

$$(39) \quad U_k \cap U \neq \emptyset, \quad k = 1, 2, 3.$$

Механизмы (38) определяют единственное решение из области компромисса для противоречивых, в общем случае, критериев (15)–(17).

5. Моделирование оптимальных механизмов

Сформируем модели (8)–(11) на основе ретроспективной информации о квартальной динамике технико-экономических показателей ООО «Завод приборных подшипников» [12] за период 2011–2013 гг.

На основе динамических рядов параметров управления $u_i(t)$, $i = 1, \dots, 3$, и параметров состояния $P(t)$, $C(t)$, $R(t)$, $g(t)$ (где t – номер квартала) рассчитаем коэффициенты парной корреляции

(таблица 1), анализ которых приводит к следующим выводам. Имеется выраженная (по шкале Чеддока [11] коэффициент корреляции более 0,7) связь между параметрами управления u_2 и u_1 , u_3 и u_2 , подтверждающая выдвинутую гипотезу о наличии устойчивых трендов изменения объемов заказов на последовательных стадиях бизнес-процесса фирмы и позволяющая сформировать задачи одномерной оптимизации (15)–(18), поскольку видна также тесная корреляционная связь параметров состояния P , C , R , g и параметра u_3 .

Поэтому сформируем регрессии (8)–(11) на основе следующих вспомогательных функций:

$$(40) \quad u_3(u_2) = A_{u_3} u_2^{\alpha_{u_3}}, \quad u_2(u_1) = A_{u_2} u_1^{\alpha_{u_2}}, \quad P(u_3) = A_P u_3^{\alpha_P}, \\ C(u_3) = A_C u_3^{\alpha_C}, \quad g(u_4) = A_g u_4^{\alpha_g}, \quad R(u_3) = A_{u_4} u_3^{\alpha_{u_4}}.$$

Определение коэффициентов регрессий (40) методом наименьших квадратов позволило получить функции, адекватно описывающие фактические временные ряды (коэффициент детерминации свыше 0,9). На основе (40) получим регрессии (8)–(12), преобразовав их к годовым показателям с учетом

$$B_{T(k)} = B_{t(k)} \cdot 4^{1-\beta_{t(k)}}$$

(где $k = P, C, R, g$, индекс T обозначает годовое значение):

$$(41) \quad P(u_1) = 8585 \cdot u_1^{0,3}, \quad C(u_1) = 45 \cdot u_1^{0,76}, \\ g(u_1) = 144 \cdot u_1^{0,7}, \quad R(u_1) = 151 \cdot u_1^{0,7}.$$

Анализ динамических рядов цен и объемов поставок основных материалов, используемых фирмой, подтвердил правомерность гипотезы (12) об убывающих ценовых трендах и позволил сформировать однофакторные регрессии в виде:

$$(42) \quad z_1(u_1^1) = 205 \cdot u_1^{-0,3}, \quad z_2(u_1^2) = 105 \cdot u_1^{-0,2}.$$

На рис. 1–3 представлены фактические значения цен материалов и параметров состояния бизнес-процесса фирмы, а также регрессионные модели (41), (42), имеющие характерную нелинейную динамику. На рис. 3 видно соответствие регрессионных моделей фактической динамике закупочных цен металлов.

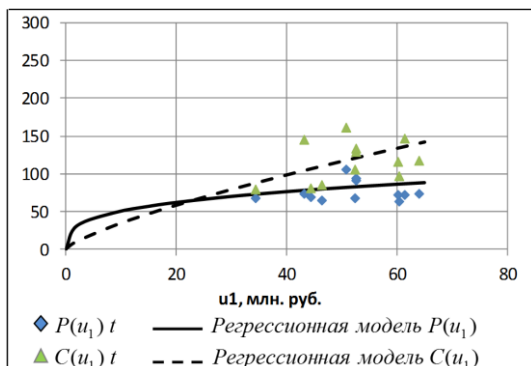


Рис. 1. Динамика изменения производственных расходов и себестоимости в 2011–2013 гг., млн руб.

Таблица 1. Матрица коэффициентов корреляции параметров управления и состояния

| Показатель, тыс. руб. | u_1 | u_2 | u_3 | $R(u_4)$ | C | P | g |
|-----------------------|-------|-------|-------|----------|------|------|------|
| u_1 | | 0,80 | 0,35 | 0,51 | 0,35 | 0,06 | 0,51 |
| u_2 | | | 0,71 | 0,68 | 0,62 | 0,30 | 0,61 |
| u_3 | | | | 0,89 | 0,93 | 0,77 | 0,83 |
| $R(u_4)$ | | | | | 0,93 | 0,69 | 0,94 |
| C | | | | | | 0,71 | 0,89 |
| P | | | | | | | 0,61 |
| g | | | | | | | |

В таблице 2 представлены результаты анализа точности приближенных механизмов (27)–(29), по которым вычислены значения u_{1F1}^* (млн руб.) и относительные отклонения Δ (%) от точных значений по формуле (25). Погрешность формул (27)–(29) убывает либо не меняется с увеличением начальных остатков ресурсов u_{Σ}^0 (млн руб.) и коэффициента регрессии издержек β_C , не превышая 18,73% на границах диапазона практически реализуемых значений. В данной таблице не представлены результаты анализа точности механизма (30) в связи с

отсутствием положительных значений u_{1F1}^* при заданных коэффициентах регрессий.

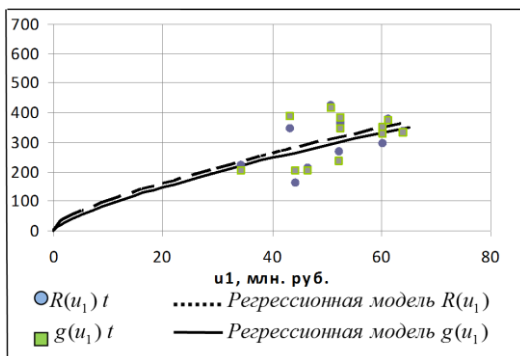


Рис. 2. Динамика изменения выручки от реализации и денежного потока в 2011–2013 гг., млн руб.

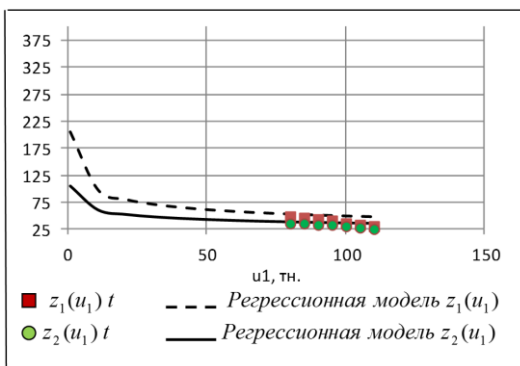


Рис. 3. Динамика изменения закупочных цен металлов ООО «ЗПП» в 2011–2013 гг., т

Таблица 2. Анализ оптимальных механизмов модели (15)

| u_{Σ}^0 | $\beta_p = 1, \beta_c = 0,93$ | | | $\beta_p = 1, \beta_c = 0,95$ | | |
|----------------|-------------------------------|-----------------|--------------|-------------------------------|-----------------|--------------|
| | $u_{1F1}^*(25)$ | $u_{1F1}^*(27)$ | $\Delta, \%$ | $u_{1F1}^*(25)$ | $u_{1F1}^*(27)$ | $\Delta, \%$ |
| 100 000 | 91 626 | 94 898 | 3,57% | 131 034 | 135 714 | 3,57% |

| u_{Σ}^0 | $\beta_P = 1, \beta_C = 0,93$ | | | $\beta_P = 1, \beta_C = 0,95$ | | |
|----------------|----------------------------------|------------------|--------------|---------------------------------|------------------|--------------|
| | $u_{1F1}^* (25)$ | $u_{1F1}^* (27)$ | $\Delta, \%$ | $u_{1F1}^* (25)$ | $u_{1F1}^* (27)$ | $\Delta, \%$ |
| 200 000 | 183 251 | 189 796 | 3,57% | 262 069 | 271 429 | 3,57% |
| 300 000 | 274 877 | 284 694 | 3,57% | 393 103 | 407 143 | 3,57% |
| | | | | | | |
| u_{Σ}^0 | $\beta_P = 0,45, \beta_C = 0,35$ | | | $\beta_P = 0,4, \beta_C = 0,3$ | | |
| | $u_{1F1}^* (25)$ | $u_{1F1}^* (28)$ | $\Delta, \%$ | $u_{1F1}^* (25)$ | $u_{1F1}^* (28)$ | $\Delta, \%$ |
| 100 000 | 106 903 | 107 692 | 0,74% | 85 339 | 85 714 | 0,44% |
| 200 000 | 214 305 | 215 385 | 0,50% | 170 933 | 171 429 | 0,29% |
| 300 000 | 321 781 | 323 077 | 0,40% | 256 560 | 257 143 | 0,23% |
| | | | | | | |
| u_{Σ}^0 | $\beta_P = 0,88, \beta_C = 0,7$ | | | $\beta_P = 0,9, \beta_C = 0,75$ | | |
| | $u_{1F1}^* (25)$ | $u_{1F1}^* (29)$ | $\Delta, \%$ | $u_{1F1}^* (25)$ | $u_{1F1}^* (29)$ | $\Delta, \%$ |
| 100 000 | 88 329 | 104 869 | 18,73 % | 94 605 | 101 124 | 6,89% |
| 200 000 | 190 220 | 209 738 | 10,26 % | 201 581 | 202 247 | 0,33% |
| 300 000 | 297 739 | 314 607 | 5,67% | 313 657 | 303 371 | -3,28% |

В таблице 3 представлены результаты анализа точности приближенных механизмов (34)–(35), по которым вычислены значения u_{1F1}^* (млн руб.) и относительные отклонения Δ (%) от точных значений по формуле (32).

Погрешность формул (34)–(35) убывает с увеличением начальных остатков ресурсов u_{Σ}^0 (млн руб.), остатков дебиторской задолженности u_4^0 и коэффициенты регрессии издержек

β_C , не превышая 22,39% на границах диапазона практически реализуемых значений.

Таблица 3. Анализ оптимальных механизмов модели (16)

| u_{Σ}^0 | u_4^0 | $\beta_P = 0,96, \beta_C = 0,83, \beta_R = 0,78, \beta_g = 0,83$ | | | $\beta_P = 1, \beta_C = 0,92, \beta_R = 0,9, \beta_g = 0,9$ | | |
|----------------|---------|--|-----------------|--------------|---|-----------------|--------------|
| | | $u_{1F1}^*(32)$ | $u_{1F1}^*(34)$ | $\Delta, \%$ | $u_{1F1}^*(32)$ | $u_{1F1}^*(34)$ | $\Delta, \%$ |
| 100 000 | 50 000 | 100 092 | 83 230 | – 16,85% | 94 241 | 91 184 | –3,24% |
| 200 000 | 100 000 | 204 944 | 166 461 | – 18,78% | 188 900 | 182 368 | –3,46% |
| 300 000 | 200 000 | 334 820 | 260 156 | – 22,30% | 299 048 | 285 426 | –4,56% |
| | | | | | | | |
| u_{Σ}^0 | u_4^0 | $\beta_P = 0,1, \beta_C = 1,25, \beta_R = 0,9, \beta_g = 0,8$ | | | $\beta_P = 0,05, \beta_C = 1,1, \beta_R = 0,99, \beta_g = 0,91$ | | |
| | | $u_{1F1}^*(32)$ | $u_{1F1}^*(35)$ | $\Delta, \%$ | $u_{1F1}^*(32)$ | $u_{1F1}^*(35)$ | $\Delta, \%$ |
| 100 000 | 50 000 | 106 527 | 130 029 | 22,06% | 110 116 | 106 710 | –3,09% |
| 200 000 | 100 000 | 237 021 | 260 042 | 9,71% | 222 815 | 213 407 | –4,22% |
| 300 000 | 200 000 | 519 236 | 402 960 | – 22,39% | 378 115 | 333 022 | – 11,93% |

На рис. 4 представлены кривая операционного цикла $F_2(u_1)$, рассчитанная по формуле (6) с учетом регрессий (41), и кривая прибыли $F_3(u_1)$ – по формуле (7), график $F_1(u_1)$ не показан в силу его близости к графику $F_2(u_1)$ для данных параметров бизнес-процесса фирмы; также отмечено ограничение (13) в виде

$u_1^{\min}(N)$, а ограничение (14) не показано, поскольку на рынке ресурсов выявлены низкоэластичные ценовые кривые (42), при которых в виду условия (21) данное ограничение не играет роли; показаны фактические значения критериев F_2 , F_3 , реализовавшиеся в 2013 г. Анализ рис. 4 показывает, что поскольку для исследуемой фирмы $u_{1F3}^* < u_{1F2}^*$, то областью компромиссных по критериям F_2 , F_3 объемов заказов ресурсов u_1 является диапазон $u_1 \in [u_1^{\min}(N), u_{1F3}^*]$, отмеченный отрезком EG кривой прибыли; в противном случае область компромисса выражалась бы диапазоном $u_1 \in [u_1^{\min}(N), u_{1F2}^*]$ (отрезок DH кривой цикла). Отметим, что в практически значимом диапазоне значений u_1 критерии F_2 , F_3 не противоречивы: прибыль уменьшается с ростом u_1 только при $u_1 > u_{1F3}^*$, а цикл при этом продолжает сокращаться, что отражает низкий темп роста заготовительных издержек фирмы.

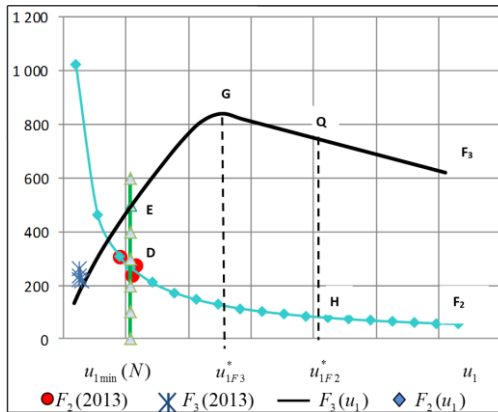


Рис. 4. Расчетные и фактические (при объемах заказов в 2013 г.) значения операционного цикла (дни), ограничение по нормативу потребности ресурсов в зависимости от объема заказов покупателей (млн. руб.), прибыль (млн. руб.)

В соответствии с механизмом (38) решением задачи (15)–(18) является значение $u_1 = u_1^{\min}(N)$, при котором значение

прибыли выше, а продолжительность цикла ниже, чем фактические данные 2013 г. Общий экономический эффект при иммобилизации прибыли в запасах может составить 5,8 млн. руб.

6. Заключение

Разработана модель оптимизации заказов производственных ресурсов фирмы по критериям прибыли и бизнес-циклов, адекватная в условиях действия устойчивых нелинейных трендов издержек, выручки, денежного потока и цен ресурсов в зависимости от объема их закупок. Определены оптимальные механизмы планирования заказов ресурсов по каждому критерию, а также механизм решения задачи оптимизации по комплексу критериев. Анализ оптимальных механизмов приводит к следующим выводам.

В практически значимых случаях сопоставимых темпов роста нематериальных расходов и себестоимости, а также темпов роста денежного потока и выручки, существуют значения объемов заказов материалов, минимизирующие длительность бизнес-циклов фирмы. Аргументы оптимумов бизнес-циклов достаточно велики по сравнению со средним уровнем остатков материальных ресурсов, поэтому в реальном диапазоне изменения объема заказов ресурсов длительность циклов убывает с ростом этого показателя. Следовательно, с точки зрения временных критериев оптимальности для фирмы целесообразно наращивать объем заказов ресурсов. Также для практически реализуемых соотношений темпов роста издержек и выручки существует объем заказов ресурсов, максимизирующий прибыль фирмы, причем аргумент оптимума прибыли, как правило, существенно меньше аргументов оптимумов циклов. Поэтому с позиции финансового критерия фирма также заинтересована в повышении объемов заказов ресурсов, а в целом по комплексу временных и финансовых критериев допустимый диапазон изменения объемов заказов ограничен наименьшим из аргументов оптимумов прибыли и циклов. Следовательно, состояние противоречивости критериев наступает только при объемах заказов, превышающих аргумент оптимума прибыли, когда встает необходимость анализа множества Парето. В остальных

случаях выбор оптимального по комплексу критериев объема заказов ресурсов определяется на основе сравнительного анализа оптимумов критериев и ограничений по уровню заказов покупателей и по ценовым функциям поставщиков.

7. Приложение

Обоснование формул (5), (6): формулы вытекают из следующих соотношений и преобразований: базовые формулы расчета производственного и операционного циклов [4]

$$(П1) F_1(u) = \frac{T}{C(u)} \sum_{i=1}^3 \bar{u}_i, \quad F_2(u) = F_1(u) + \frac{T}{R(u)} \bar{u}_4,$$

где \bar{u}_i – средние остатки ресурсов за период, рассчитанные по формулам.

$$(П2) \bar{u}_i = \frac{u_i^0 + u_i^T}{2}, i = 1, \dots, 4.$$

С учетом соотношений материального, производственного, товарного и платежного балансов [19]

$$(П3) u_1 = u_1^T - u_1^0 + u_2, u_2 = u_2^T - u_2^0 + u_3 - P, \\ u_3 = u_3^T - u_3^0 + C, u_4 = u_4^T - u_4^0 + g,$$

средние остатки ресурсов определяются в виде

$$(П4) \bar{u}_1 = u_1^0 + \frac{u_1 - u_2}{2}, \bar{u}_2 = u_2^0 + \frac{u_2 + P - u_3}{2}, \\ \bar{u}_3 = u_3^0 + \frac{u_3 - C}{2}, \bar{u}_4 = u_4^0 + \frac{u_4 - g}{2},$$

где $u_i u_i^T$ – остатки ресурсов на конец периода T . Полагаем средства в расчетах с дебиторами численно равными выручке фирмы за отгруженную продукцию

$$(П5) R(u) = u_4.$$

Подставив (П2), (П4), (П5) в (П1), получим (4), (5). ■

Доказательство утверждения 1. Подставим (8), (12) в (19):

$$(П6) P_\Sigma = \sum_{j_1=1}^{J_1} A_{\zeta_{j_1}} u_1^{\alpha_{\zeta_{j_1}}+1} + B_P \left(\sum_{j_1=1}^{J_1} u_{1j_1} \right)^{\beta_P},$$

дифференцируя (П6), запишем необходимое условие минимума (19):

$$P'_{\Sigma u_{j_1}} = A_{z_{j_1}} (\alpha_{z_{j_1}} + 1) u_1^{\alpha_{z_{j_1}}} + B_p \beta_p \left(\sum_{j_1=1}^{J_1} u_{1j_1} \right)^{\beta_p - 1} = 0, j_1 = 1, \dots, J_1,$$

откуда получим соотношение (20). Отметим, что неотрицательное решение уравнения (20) соответствует ограничению (П7) $\alpha_{z_{j_1}} < -1$.

Достаточное условие минимума

$$(П8) \quad P''_{\Sigma u_{j_1}} = A_{z_{j_1}} \alpha_{z_{j_1}} (\alpha_{z_{j_1}} + 1) u_1^{\alpha_{z_{j_1}} - 1} + B_p \beta_p (\beta_p - 1) \left(\sum_{j_1=1}^{J_1} u_{1j_1} \right)^{\beta_p - 2} > 0,$$

$$j_1 = 1, \dots, J_1,$$

выполняется при

$$\left\{ \begin{array}{l} \beta_p > 1 \cap B_p \beta_p (\beta_p - 1) \left(\sum_{j_1=1}^{J_1} u_{1j_1} \right)^{\beta_p - 2} \geq \\ A_{z_{j_1}} |\alpha_{z_{j_1}} + 1| \alpha_{z_{j_1}} u_{1j_1}^{\alpha_{z_{j_1}} - 1}, \forall \alpha_{z_{j_1}} < 0, \\ \beta_p \leq 1 \cap \alpha_{z_{j_1}} < -1 \cap A_{z_{j_1}} (\alpha_{z_{j_1}} + 1) \alpha_{z_{j_1}} u_{1j_1}^{\alpha_{z_{j_1}} - 1} > \\ B_p \beta_p |\beta_p - 1| \left(\sum_{j_1=1}^{J_1} u_{1j_1} \right)^{\beta_p - 2}, \end{array} \right.$$

однако, учитывая (П7), условие (П8) выполняется только при соотношении (21). ■

Доказательство утверждения 2. Исходя из необходимого условия оптимальности (5)

$$(П9) \quad F'_{1u_1}(u_1) = -\beta_c \frac{T u_{\Sigma}^0}{B_c} u_1^{-\beta_c - 1} + \frac{T}{2B_c} (1 - \beta_c) u_1^{-\beta_c} + \frac{T B_p}{2B_c} (\beta_p - \beta_c) u_1^{\beta_p - \beta_c - 1} = 0,$$

получим уравнение (25) для определения значения u_{1F1}^* . Достаточное условие минимума (5)

$$F'_{u_1}(u_1) = -\beta_c(-\beta_c - 1) \frac{Tu_{\Sigma}^0}{B_c} u_1^{-\beta_c-2} - \frac{T}{2B_c}(1 - \beta_c)\beta_c u_1^{-\beta_c-1} + \\ + \frac{TB_p}{2B_c}(\beta_p - \beta_c)(\beta_p - \beta_c - 1)u_1^{\beta_p-\beta_c-2} > 0$$

преобразуем к виду

$$(П10) \quad \beta_c [2(\beta_c + 1)u_{\Sigma}^0 + \beta_c - 1] u_1 + \\ + B_p(\beta_c - \beta_p)(\beta_c + 1 - \beta_p)u_1^{\beta_p} > 0.$$

Поскольку для реальных фирм по порядку величин $u_{\Sigma}^0 \gg \beta_c$, то $2(\beta_c + 1)u_{\Sigma}^0 + \beta_c \gg 1$; поэтому на знак (П10) влияет только соотношение β_c, β_p . С учетом диапазонов изменения β_c, β_p , определенных (8), (9), условие (П10) выполняется при

$$(П11) \quad \begin{cases} \beta_c > \beta_p \cup [\beta_c < \beta_p \cap \beta_c + 1 < \beta_p] \quad \forall u_{1F1}^* \geq 0, \\ \beta_c < \beta_p \cap \beta_c + 1 > \beta_p \cap \varphi(\beta_c, \beta_p, u_{\Sigma}^0, u_1) > 0, \end{cases}$$

где

$$\varphi(\beta_c, \beta_p, u_{\Sigma}^0, u_1) = \beta_c [2(\beta_c + 1)u_{\Sigma}^0 + \beta_c - 1] u_1 + \\ + B_p |\beta_c - \beta_p| (\beta_c + 1 - \beta_p) u_1^{\beta_p}.$$

С учетом (23) условие (П11) имеет вид (26). ■

Доказательство утверждения 3. Запишем необходимое условие оптимума (6):

$$F'_{2u_1}(u_1) = -\beta_c \frac{Tu_{\Sigma}^0}{B_c} u_1^{-\beta_c-1} + \frac{T}{2B_c}(1 - \beta_c)u_1^{-\beta_c} + \frac{TB_p}{2B_c}(\beta_p - \beta_c)u_1^{\beta_p-\beta_c-1} - \\ - \beta_R \frac{Tu_4^0}{B_R} u_1^{-\beta_R-1} - \frac{TB_g}{2B_R}(\beta_g - \beta_R)u_1^{\beta_g-\beta_R-1} = 0,$$

преобразуя которое

$$-\beta_c \frac{u_{\Sigma}^0}{B_c} u_{1F2}^{*-\beta_c-1} + \frac{1 - \beta_c}{2B_c} u_{1F2}^{*-\beta_c} + \frac{B_p}{2B_c}(\beta_p - \beta_c)u_{1F2}^{*\beta_p-\beta_c-1} - \\ - \beta_R \frac{u_4^0}{B_R} u_{1F2}^{*-\beta_R-1} - \frac{B_g}{2B_R}(\beta_g - \beta_R)u_{1F2}^{*\beta_g-\beta_R-1} = 0,$$

после деления на $u_{1F2}^{*-\beta_C-1}$, получим уравнение (32) для определения u_{1F2}^* . Достаточное условие минимума (6)

$$F_{2u_1}^{//}(u_1) = \beta_C(\beta_C + 1) \frac{u_{\Sigma}^0}{B_c} u_1^{-\beta_C-2} + \frac{\beta_c(\beta_C - 1)}{2B_c} u_1^{-\beta_C-1} + \\ + \frac{B_p}{2B_c} (\beta_p - \beta_c)(\beta_p - \beta_C - 1) u_1^{\beta_p - \beta_C - 2} + \\ + \beta_R(\beta_R + 1) \frac{u_4^0}{B_R} u_1^{-\beta_R-2} - \frac{B_g}{2B_R} (\beta_g - \beta_R)(\beta_g - \beta_R - 1) u_1^{\beta_g - \beta_R - 2} > 0$$

преобразуем к виду

$$(П12) \quad \varphi(\beta_C, \beta_p, u_{\Sigma}^0, u_{1F2}^*) + \beta_R(\beta_R + 1) \frac{u_4^0}{B_R} u_{1F2}^{*-\beta_R-2} + \\ + \frac{B_g}{2B_R} (\beta_g - \beta_R)(\beta_R + 1 - \beta_g) u_{1F2}^{*\beta_g - \beta_R - 2} > 0.$$

Будем искать минимум операционного цикла на том же множестве параметров управления, что и минимум производственного цикла, т.е. предположим выполнение достаточного условия (26):

$$(П13) \quad \varphi(\beta_C, \beta_p, u_{\Sigma}^0, u_{1F2}^*) > 0.$$

В случае (П13) неравенство (П12) выполняется в следующих диапазонах значений темповых параметров бизнес-процесса фирмы и u_{1F2}^* :

$$(П14) \quad \left\{ \begin{array}{l} (\beta_g > \beta_R \cap \beta_g > \beta_R + 1) \cap \varphi(\beta_C, \beta_p, u_{\Sigma}^0, u_{1F2}^*) + \beta_R(\beta_R + 1) \frac{u_4^0}{B_R} u_{1F2}^{*-\beta_R-2} + \\ + \frac{B_g}{2B_R} (\beta_g - \beta_R)(\beta_R + 1 - \beta_g) u_{1F2}^{*\beta_g - \beta_R - 2} \geq 0, \\ (\beta_g > \beta_R \cap \beta_g < \beta_R + 1) \forall u_{1F2}^*, \\ \beta_g \leq \beta_R \cap \varphi(\beta_C, \beta_p, u_{\Sigma}^0, u_{1F2}^*) + \beta_R(\beta_R + 1) \frac{u_4^0}{B_R} u_{1F2}^{*-\beta_R-2} + \\ + \frac{B_g}{2B_R} (\beta_g - \beta_R)(\beta_R + 1 - \beta_g) u_{1F2}^{*\beta_g - \beta_R - 2} > 0. \end{array} \right.$$

С учетом темпового соотношения (24) в (П14) реализуется второй и третий вариант в виде (33). ■

Доказательство утверждения 4. Для выражения прибыли

$$(П15) \quad F_3(u_1) = B_R u_1^{\beta_R} - B_C u_1^{\beta_C},$$

полученного подстановкой (9), (10) в (7), запишем необходимое условие оптимума

$$(П16) \quad F_{3u_1}'(u_1) = B_R \beta_R u_1^{\beta_R-1} - B_C \beta_C u_1^{\beta_C-1} = 0,$$

откуда выразим (36). Достаточное условие максимума (П15)

$$F_{3u_1}''(u_1) = B_R (\beta_R - 1) \beta_R u_1^{\beta_R-2} - B_C (\beta_C - 1) \beta_C u_1^{\beta_C-2} < 0,$$

выполняется при условиях (37). ■

Доказательство утверждения 5. Запишем функцию Лагранжа для задач (15)–(17) при ограничении (18):

$$(П17) \quad L_k = F_k(u_1) + \lambda_{1k} (u_1^{\min}(N) - u_1) + \lambda_{2k} (u_1^{\max}(N) - u_1) + \\ + \lambda_{3k} (u_1^{\min}(z) - u_1), k = 1, 2, 3,$$

дифференцируя которую получим систему необходимых условий оптимальности:

$$(П18) \quad \frac{\partial L_k}{\partial u_1} = \frac{\partial F_k(u_1)}{\partial u_1} - \lambda_{1k} - \lambda_{2k} - \lambda_{3k} = 0,$$

$$(П19) \quad \frac{\partial L_k}{\partial \lambda_{1k}} = u_1^{\min}(N) - u_1 \leq 0, \quad \frac{\partial L_k}{\partial \lambda_{2k}} = u_1^{\max}(N) - u_1 \geq 0,$$

$$\frac{\partial L_k}{\partial \lambda_{3k}} = u_1^{\min}(z) - u_1 \leq 0, k = 1, 2, 3.$$

При выполнении (П19) как строгих неравенств решением системы (П18), (П19) будет вектор множителей Лагранжа $\lambda_{1k} = \lambda_{2k} = \lambda_{3k} = 0$, следовательно, из (П18) получаем необходимые условия оптимальности $\partial F_k / \partial u_1 = 0$, $k = 1, 2, 3$, записанные в формах (25), (32), (36) для определения оптимумов критериев без ограничений u_{1Fk}^* , $k = 1, 2, 3$, т.е. при этом $u_{1(k)}^* = u_{1Fk}^*$, $k = 1, 2, 3$. В случае выполнения какого либо из условий (П19) как строгих равенств решением системы (П18), (П19) будет параметр управления, удовлетворяющий условиям

$$(П20) \quad u_1 = u_1^{\min}(N) \cup u_1 = u_1^{\max}(N) \cup u_1 = u_1^{\min}(z),$$

которые формально запишем в виде (38). Достаточные условия экстремумов в задачах (15)–(17) при ограничении (18) выполняются, если достаточные условия (26), (33), (37) определяют диапазоны $u_{1Fk}^* \in U$. ■

Литература

1. БРАЖНИКОВ М.А. *Моделирование календарных планов сборочных процессов в условиях машиностроительного производства* // Вестник Самарского государственного технического университета. Серия физико-математические науки. – 2004. – №26. – С. 165–173.
2. ВАЙНЕР ДЖ. *Кривые затрат и кривые предложения* // Вехи экономической мысли. Теория фирмы. – Спб: Экономическая школа. – 2000. – №27. – С. 94–135.
3. ВОЖАКОВ А.В., ГИТМАН М.Б. *Модель календарного планирования с нечеткими ограничениями* // Вестник МГТУ им. Г.И. Носова. – 2008. – №4. – С. 79–82.
4. ГЕРАСЬКИН М.И., ЕГОРОВА В.В. *Статическая оптимизация производственных циклов на предприятиях подшипниковой промышленности* // Вестник Самарского государственного экономического университета. – 2014. – №11(121). – С. 53–60.
5. ГОРЛАЧ Б.А., САВЕЛЬЕВ Г.Л. *Прогнозирование и оптимизация процесса поставок в условиях колебания спроса* // Вестник Самарского государственного аэрокосмического университета им. академика С.П. Королева (Национального исследовательского университета). – 2011. – №4. – С. 48–57.
6. ГОРЛАЧ Б.А., ЧУЙКОВА Ю.С. *Прогнозирование объемов продаж в модели управления запасами* // Вестник Самарского государственного аэрокосмического университета им. академика С.П. Королева (Национального исследовательского университета). – 2008. – №8. – С. 129–133.
7. ГРИГОРЬЕВ В.П., КАЛЮТА В.Н., КИСЕЛЕВ К.А. *Модель оптимального распределения ресурсов в производство* // Известия Томского политехнического университета. – 2005. – Т.308, №5. – С. 179–181.

8. ЗУБКОВА Н.В. *Применение экономико-математических моделей при формировании затрат машиностроительного предприятия на стадии планирования* // Вектор науки ТГУ. – 2010. – №2(12). – С. 166–169.
9. КОНОНЕНКО В.Н., МИРОНЕНКО И.Л. *Анализ финансового состояния предприятия (экспресс-оценка)* // Менеджмент в России и за рубежом. – 2008. – №34. – С. 26–37.
10. КУЗНЕЦОВ Л.А., ЧЕРНЫХ М.В. *Новый подход к решению задачи планирования производственной деятельности организации* // Управление предприятием. – 2005. – №1. – С. 66–76.
11. КУПРИЕНКО Н.В. *Статистика. Методы анализа распределений. Выборочное наблюдение.* – СПб.: Изд-во Политехн. ун-та, 2009. – 138 с.
12. *Материалы сайта ООО «Завод приборных подшипников»* [Электронный ресурс] – URL: <http://www.mbf-samara.ru>. (дата обращения 20.11.2015.)
13. МУЛКИДЖАНЫН В.С. *Совершенствование процесса управления оборотными средствами промышленного предприятия: организационно-методический аспект*: автореф. дис. канд. экон. наук – Ростов н/Д, 2011. – 26 с.
14. ПАНЮКОВ А.В. *Подходы к формированию производственной программы для предприятий с дискретным механосборочным типом производства* // Вестник Пермского университета. Серия экономика. – 2011. – №4(11). – С. 74–83.
15. ПИЖУРИН А.А., МУРАЩЕНКО Д.Д. *Оптимизационная математическая модель задачи оперативного планирования и управления лесопильно-деревообрабатывающим производством в условиях рыночной экономики* // Лесной вестник. – 2008 – №3. – С. 32–35.
16. РАДИОНОВ Р.А. *Нормирование и управление запасами и оборотными средствами предприятия в условиях рыночной экономики* // Вестник машиностроения. – 2004. – №9. – С. 69–75.
17. РОСТОВА Е.П., ВЕРХОВЕЦ О.А. *Постановка задачи линейного программирования для распределения средств по управлению рисками промышленного предприятия* // Вест-

- ник Омского университета. Серия: Экономика. – 2013. – №2. – С. 116–119.
18. РУЖАНСКАЯ Н.В. *Методика оптимизации запасов торговой организации: модели и возможности применения* // Корпоративное управление и инновационное развитие Севера. Вестник научно-исследовательского центра корпоративного права, управления и венчурного инвестирования Сыктывкарского гос. ун-та. – 2008. – №3. – С. 107–115.
 19. СОКОЛОВ Я.В. *Основы теории бухгалтерского учета.* – М.: Финансы и статистика, 2005. – 496 с.
 20. ТОЛЫСБАЕВ Б.С. *Экономико-математическое моделирование процесса поставки сельскохозяйственной продукции* // Вестник ОГУ. – 2006. – №5. – С. 85–88.
 21. ТКАЧ В.Р. *Экономико-математические модели оперативного планирования камнеобрабатывающего производства* // Материалы симпозиума «Неделя горняка – 2001». – М.: МГТУ, 2001.
 22. ФЕДОРИН В.Ю. *Проблема рынка природного камня в России* // Камень и бизнес. – 2000. – №2. – С. 76–80.
 23. ШОРИКОВ А.Ф. *Динамическая оптимизация комплексного управления технологическими процессами на предприятии* // Известия Уральского государственного экономического, университета. – 2007. – №1(18). – С. 254–266.

OPTIMAL PLANNING MECHANISMS FOR ASSEMBLE-TO-ORDER MANUFACTURING: TIME AND MONEY CRITERIA

Michail Geraskin, Samara State Aerospace University, Samara, Doctor of Economics, professor ((846)267-44-96, innovation@ssau.ru).

Victoriia Egorova, Samara State Aerospace University, assistant of the Department of mathematical methods in Economics ((846)267-44-96, v_v_egorova@inbox.ru).

Abstract: The problem is studied of raw material inventory optimization in companies, which manufacture consumer goods for market of limited volume. We analyze factors leading to the use of the assemble-to-order planning policy in firms, which compete on oligopsony markets, and suggest optimization models of procurement order planning either to maximize profit or to minimize the operation/manufacturing cycle under the monotonic increase of costs, revenue, and cash flow. The model takes into account price and technology constraints. We design analytical mechanisms of optimal order planning for bearing industry and model the region of compromise for orders being extremal with respect to different criteria.

Key words: production cycle, work in process, finished products, optimization, supply and demand, the ratio of production costs.

*Статья представлена к публикации
членом редакционной коллегии В.В. Клочковым*

*Поступила в редакцию 18.08.2015.
Опубликована 30.11.2015.*

УДК 519.876.2
ББК 32.81

МОДЕЛЬ БЕЗОПАСНОСТИ ГОСУДАРСТВА

Шумов В. В.¹

(Отделение погранологии

Международной академии информатизации, Москва)

Рассмотрена модель безопасности государства как социально-политической системы, отражающая дихотомию ценностей развития и ценностей сохранения. Модель учитывает основные факторы, влияющие на исторические процессы: географию, демографию и технологии.

Ключевые слова: безопасность государства, математическая модель, критерий безопасности.

1. Введение

Безопасность («отсутствие опасности; сохранность, надежность» [16]) есть процесс и результат развития общества, государства, их взаимодействия с другими этносами и культурами; это присущее людям чувство упорядоченности и правильности их жизни [1]. Безопасность является потребностью, общественным благом и ценностью. Ценности по своей природе дихотомичны, двойственны (добро – зло, любовь – ненависть, жизнь – смерть, развитие – сохранение и т.д.): тотальное господство одних нравственных ориентиров ведет к неутолимой жажде других [35]. Главное условие и источник развития человека и общества – способность справляться с конфликтностью ценностей.

Три ключевых фактора влияют на человеческую историю: география, демография и технология [38, 39]. Географический фактор относительно постоянен. Он задает сцену, на которой разворачиваются исторические события, и формирует культурно-

¹ Владислав Вячеславович Шумов, кандидат технических наук, доцент (vshum59@yandex.ru).

хозяйственные типы (этносy – группы людей, обладающих общей моделью поведения, позволяющей им отделить себя от других [15]). Суперэтносy (культурно-исторические типы, локальные цивилизации по Н.Я. Данилевскому, О. Шпенглеру, А. Тойнби, С. Хантингтону), в отличие от государств (стран), существуют долгое время – как правило, более тысячелетия.

В силу разнообразия локальных цивилизаций критерии их выделения нечетки [56]. Относительно недавно (XVIII век) в гуманитарных науках и социальной практике появились термины «нация», «этнос». Споры о правильных идентификационных критериях нации и этноса продолжаются в научной среде по настоящее время, однако на практике для целей государственного управления и планирования с XVIII–XIX вв. проводятся регулярные переписи населения [17], в опросные листы которых часто включаются вопросы о родном языке, вероисповедании, национальности (народности, этнической группе).

Существует множество содержательных концепций, объясняющих влияние технологического фактора: концепция технологических укладов Д.С. Львова и С.Ю. Глазьева [32], концепция типов организационной культуры В.А. Никитина [40, 41], концепция волн Э. Тоффлера [53] и др. Для классификации государств по уровню развития технологий, в частности, используется Глобальный индекс инноваций (The Global Innovation Index) как взвешенная сумма оценок двух групп показателей [63]:

- располагаемые ресурсы и условия для проведения инноваций (институты, человеческий капитал и исследования, инфраструктура, состояние и степень развития внутреннего рынка и бизнеса);
- достигнутые практические результаты осуществления инноваций.

В теории безопасности и геополитике традиционно используются вербальные конструкции, концептуальные модели и экспертные оценки. В настоящее время применяется и математическое моделирование, ориентированное на получение количественных оценок, характеризующих геополитические

статусы и потенциалы государств [2]. Винокуровым Г.Н. и др. предложена следующая модель для расчета геополитического потенциала $G(t)$ государства в момент времени t [10]:

$$(1) \quad G(t) = 0,5(1 + X_M^{0,43}) X_T^{0,11} X_D^{0,19} X_E^{0,27},$$

где X_i ($i = T, D, E, M$) – доли государства в общемировых показателях в территориальной, демографической, экономической и военной сферах соответственно (часть площади территории страны от площади всех стран, доля ее населения от населения Земли и т.д.). Значения показателей степени в уравнении (1) рассчитывались методом наименьших квадратов. При этом данные для долей X_i брались из соответствующей статистики (применительно к ведущим странам), а левая часть уравнения (геопотенциал) – из сравнительных оценок могущества государства по модели корпорации RAND [10].

В моделях высокого уровня, где оперируют агрегированными характеристиками населения и территории, военный потенциал избыточен, поскольку он полностью определяется демографическим и экономическим фактором. В частности, чрезмерно высокий процент мобилизации в годы войны (или военных расходов в мирное время) приводит к расстройству экономики и других сфер государства [58].

Из модели (1) следует, что расширение государства за счет присоединения новых территорий, создания блоков и союзов всегда повышает его потенциал, что логически ведет к обоснованию создания единого мирового правительства, при котором безопасность индивидов и народов будет максимальной. Названного недостатка лишены политико-экономические модели А. Алесины и Э. Сполаоре [59]: чем больше государство, тем ниже удельные издержки производства общественных благ (включая, безопасность) и прочих услуг, которые это государство предоставляет гражданам. Поэтому крупные государства обладают преимуществом большого внутреннего рынка. Но если население большой страны слишком разнородно, то ее гражданам трудно прийти к согласию, какие именно общественные блага должно предоставлять государство, каким должен быть размер

государства и, следовательно, уровень налогообложения. Тогда возникают тенденции к дезинтеграции страны.

При построении математических макромоделей, описывающих динамику социально-экономического развития [49], часто используется производственная функция Кобба–Дугласа, определяющая зависимость объема производства Y от создающих его факторов [61]:

$$(2) \quad Y = AL^\alpha K^{1-\alpha},$$

где A – технический прогресс или совокупная производительность факторов; L – рабочая сила; K – капитал; α – коэффициент эластичности по труду. Применительно к обрабатывающей промышленности в начале прошлого века получены следующие значения коэффициента эластичности по труду: $\alpha = 0,65 \div 0,75$. В частности, Дж. Комлос и С. Нефедов [65] при анализе земледельческих обществ используют модель Кобба–Дугласа для расчета объема собранного урожая $P(t)$ в момент времени t :

$$P(t) = T(t)^{1/3} N(t)^{2/3},$$

где $T(t)$ – текущее состояние технологии; $N(t)$ – численность населения государства в момент времени t .

В настоящей работе предпринята попытка сформулировать критерий безопасности, учитывающий (пусть и в максимально агрегированном виде) существенные факторы и аспекты, влияющие на безопасность государства (региона, союза государств).

2. Модель безопасности государства

2.1. КРИТЕРИЙ БЕЗОПАСНОСТИ

Безопасность как индивидуума, так и некоторой общности людей может быть определена через дуализм (дихотомию) ценностей развития (самореализации, суверенитета, свободы) и ценностей сохранения, удобства, комфорта. В логике высказываний дихотомии пары ценностей (их непреодолимой связности) можно поставить в соответствие бинарную связку конъюнкции & («и», логическое умножение). В математике

дихотомии соответствует мультипликативная функция. Следовательно, функция u_i безопасности i -го государства равна

$$(3) \quad u_i = w_i q_i, \quad w_i \geq 0, \quad 0 \leq q_i \leq 1,$$

где w_i – функция суверенности (развития) i -го государства; q_i – функция сохранения¹ i -го государства.

В геополитике функции суверенности соответствует функция $G(t)$ геопотенциала страны (1). Функция сохранения основана на учете гетерогенностей (социально-экономических, этнических и других).

2.2. ФУНКЦИЯ СУВЕРЕННОСТИ

Следуя сложившейся традиции [2, 10, 65], определим базовый суверенитет w_{bi} i -го государства с использованием степенной производственной функции [61]:

$$(4) \quad w_{bi} = \left(\frac{z_i}{z_{\max}} \right)^{\omega} \left(\frac{s_i}{s_{\max}} \right)^{1-\omega}, \quad 0 \leq \omega \leq 1,$$

где z_i – численность населения i -й страны; s_i – ее площадь; s_{\max} – площадь крупнейшей страны (России); z_{\max} – численность населения Китая; ω – параметр важности демографического фактора (параметр эластичности по демографическому фактору).

Ориентируясь на модели Кобба–Дугласа ($\alpha = 0,65 \div 0,75$), Винокура и др. ($\alpha = 0,19 / (0,19 + 0,11) = 0,63$), Дж. Комлоса и С. Нефедова ($\alpha = 0,67$), можно положить $\omega = 0,67$.

В выражении (4) мы используем две геополитические константы: *численность населения Китая* (примерно с 1600 г. Китай является самой многочисленной страной мира) и *размер территории России* (по площади территории Российская империя – СССР занимала третье место после Британской и Монгольской империй). Территория современной России почти в два раза больше территорий таких стран, как Канада, США и Китай.

¹ В качестве показателя надежности (сохранности) традиционно используются показатели вероятностного или индикаторного типа со значениями на отрезке $[0, 1]$.

Функцию суверенности с учетом технологического фактора определим в виде

$$(5) \quad w_i = \lambda_i w_{bi}, \quad \lambda_i \geq 1,$$

где λ_i – параметр, отражающий технологический фактор; w_{bi} – базовая суверенность, вычисляемая по формуле (4).

Поскольку Глобальный индекс инноваций ГИ [63] принимает значения от 0 до 1, выражение для параметра λ_i может быть следующим:

$$(6) \quad \lambda_i = (1 + I_i)^\chi, \quad 0 < I_i \leq 1, \quad \chi > 0,$$

где I_i – индекс технологического развития (инноваций) i -й страны; χ – статистический параметр, отражающий содержание рассматриваемого технологического уклада (степень технологического фактора).

2.3. ФУНКЦИЯ СОХРАНЕНИЯ

При исследовании социальных, экономических, физических и иных явлений часто используется распределение Парето [64], являющееся одной из разновидностей обратного степенного распределения. Функция F и плотность распределения f случайной величины $x > x_0 > 0$, подчиняющейся закону Парето, имеют вид

$$F(\alpha, x_0, x) = 1 - \left(\frac{x_0}{x}\right)^\alpha, \quad f(\alpha, x_0, x) = \frac{\alpha}{x_0} \left(\frac{x_0}{x}\right)^{1+\alpha}, \quad \alpha > 0,$$

или после перенормировки $x/x_0 \Rightarrow y$ для нормированной величины y : $F(\alpha, y) = 1 - y^{-\alpha}$, где x_0 является параметром сдвига, а α – параметром масштаба. Подобные зависимости называют распределениями с «тяжелыми хвостами» (fat heavy tails) – их нельзя «обрезать», т.е. нельзя пренебрегать редкими событиями.

Содержательные интерпретации явлений, описываемых распределением Парето, сформулированы Р. Кохом [26]:

- значимых факторов немного, а факторов тривиальных множество, – лишь единичные действия приводят к важным результатам;
- большинство удачных событий обусловлено действием небольшого числа конструктивных сил, большинство

неприятностей связано с действием небольшого числа деструктивных сил;

- большая часть действий (групповых или индивидуальных) не дает ничего реального для достижения результата.

Для распределения Парето существуют только моменты порядка, меньшего чем параметр α . В частности, математическое ожидание случайной величины x , подчиненной закону Парето, существует только при $\alpha > 1$ и равно

$$Ex = \frac{\alpha}{\alpha - 1} x_0,$$

откуда можно получить выражение для вычисления параметра масштаба (при известных Ex и x_0):

$$\alpha = \frac{Ex}{Ex - x_0}.$$

В моделях с дискретным временем при расчете численности населения $Y(t + 1)$ аграрного общества в момент времени $t + 1$ используется логистическое отображение [38]:

$$Y(t + 1) = \frac{rY(t)}{1 + (r - 1)\frac{Y(t)}{K}},$$

где r – коэффициент естественного прироста населения в благоприятных условиях, а K – емкость экологической ниши, т.е. максимально возможная численность населения при имеющихся продовольственных ресурсах. Дж. Майнард Смит и М. Слаткин [69] предложили модель, в которой член $Y(t)/K$ заменен степенной функцией (распределением Парето):

$$\left(\frac{Y(t)}{K}\right)^n,$$

где $n > 0$ – показатель компенсации, позволяющий учесть тот факт, что в человеческом обществе голод приводит не только к высокой смертности, но также к восстаниям и войнам, резко увеличивающим коэффициент смертности.

Распределение Парето обладает свойством самоподобия, т.е. распределение значений, превышающих величину $z_0 \geq x_0$, также характеризуется распределением Парето:

$$f(\alpha, z_0, x) = \frac{f(\alpha, x_0, x)}{1 - F(\alpha, x_0, z_0)} = \frac{\alpha}{z_0} \left(\frac{z_0}{x} \right)^{1+\alpha}.$$

Использование свойства самоподобия и распределения Парето позволяет с единых позиций описывать сложные иерархические системы безопасности, где на нижних уровнях (социальная группа, населенный пункт, район) различия обусловлены возрастными, психологическими и иными характеристиками людей. На средних уровнях (субъект федерации, государство) часть различий игнорируется и рассматриваются преимущественно социально-экономические и этнические характеристики, а на высших уровнях (блоки, союзы государств) – цивилизационные характеристики.

Пусть z_i есть численность населения i -го региона (субъекта федерации, административно-экономического района, страны, союза), z_{ij} – численность j -го этноса в нем, причем

$$(7) \quad z_i = \sum_{j=1}^{n_i} z_{ij}, \quad \zeta_i = \max_{j=1, \dots, n_i} z_{ij},$$

где n_i – количество этносов в i -м регионе; ζ_i – численность регионообразующего этноса. Допустим, что социальные взаимодействия осуществляются только по линии «регионообразующий этнос – другой этнос», а взаимодействиями между нергионообразующими этносами можно пренебречь. Регионообразующий этнос i характеризуется параметром $\delta_i > 0$ притяжения, отражающим способность этноса к объединению других этносов в единую общественную единицу. Каждый этнос i -го региона характеризуется параметром разнородности $\mu_{ij} \geq 1$ с регионообразующим этносом (для регионообразующего этноса положим $\mu_{ij} = 1$).

Используя распределение Парето, определим *функцию сохранения* i -го региона:

$$(8) \quad q_i = \left(\frac{\zeta_i}{z_i} \right)^{\delta_i \mu_i}, \quad \mu_i = \frac{1}{z_i} \sum_{j=1}^{n_i} \mu_{ij} z_{ij}.$$

Мы имеем функцию с двумя параметрами: ζ_i – параметр сдвига и $\delta_i \mu_i$ – параметр формы. Чем выше однородность населения (доля регионообразующего этноса), тем больше

значение функции сохранения. Для разнородного (полиэтнического) региона комфортность проживания в нем определяется значением параметра формы: чем меньше величина произведения $\delta_i \mu_i$, тем комфортнее проживание и тем выше значение функции сохранения. Параметр μ_i отражает существующие различия, а параметр δ_i – способности и возможности по их смягчению. Иными словами, параметры μ_i , δ_i отражают дихотомию ценностей конкуренции – кооперации.

Таким образом, функция (3) безопасности i -го государства приобретает вид

$$(9) \quad u_i = (1 + I_i)^{\chi} \left(\frac{z_i}{z_{\max}} \right)^{\omega} \left(\frac{s_i}{s_{\max}} \right)^{1-\omega} \left(\frac{\zeta_i}{z_i} \right)^{\delta_i \mu_i}.$$

Далее рассмотрим вопросы оценки параметров модели.

3. Оценка параметров функции суверенности

3.1. ОЦЕНКА ИНДЕКСА СОЦИАЛЬНЫХ ТЕХНОЛОГИЙ

А.Н. Зубец полагает, что важнейшими характеристиками качества жизни людей являются [20, 21]: ВВП на душу населения, уровень развития образования, уровень урбанизации. Многими исследователями [42] отмечается устойчивая положительная корреляционная зависимость между уровнем урбанизации и показателями качества жизни (уровень образования и продолжительность жизни). С инновациями тесно связана традиция – это «новации, пережившие смену трех или более поколений, т.е. предложенные 75–100 лет назад» [50, с. 62]. Такие социальные институты, как конкуренция и кооперация, формируются в первую очередь в семье и существенно зависят от темпов роста населения в стране.

На основании исторических данных [5, 31, 37, 45, 68, 71], характеризующих уровни социально-экономического развития ряда государств, выполнен расчет индекса социальных технологий i -й страны по формуле

$$(10) \quad I_i = (V_i / V_{\max} + D_i / D_{\max} + N_i / N_{\max}) / 3,$$

$$V_{\max} = \max_i V_i, \quad D_{\max} = \max_i D_i, \quad N_{\max} = \max_i N_i,$$

где V_i – ВВП на душу населения i -й страны [68]; D_i – доля городского населения в i -й стране; N_i – прирост населения i -й страны в среднем за 75–100 лет (в расчетах принято 80 лет); $V_{\max}, D_{\max}, N_{\max}$ – максимальные значения указанных показателей.

Таким образом, индекс социальных технологий определен как совокупность трех показателей: 1) *уровень урбанизации* (результат исторического развития, прошлое); 2) *уровень валового внутреннего продукта (ВВП) на душу населения* (настоящее) и 3) *темпы роста численности населения* (будущее). Здесь мы видим связь структуры показателей со свойством человеческого мышления: мышление – процесс, который как бы связывает прошлое, настоящее и будущее, становится над временем, устанавливая связь причин (прошлого), следствий (будущего) и условий реализации причинно-следственных отношений (настоящего) [44].

Отметим, что коэффициент корреляции между значениями индексов I_i и ГИ для сорока крупнейших стран мира равен 0,75 (достаточно высокое значение для социально-политических моделей).

Темпы роста населения в существенной степени определяют эффективность государства. По оценкам М. Ливи Баччи [31] заселение Северной Америки в основном осуществлено силами британских и испано-португальских иммигрантов, с менее значительным участием голландцев, немцев и французов. Британская и Российская империи создавались организованными группами людей, на первых этапах зачастую без какой-либо серьезной поддержки со стороны центральной власти [18, 19].

В силу простоты модели (10) и доступности данных по населению, ВВП и урбанизации, можно вычислить значения индекса I_i за достаточно большой исторический период (рис. 1).

Представленные на рисунке данные не противоречат историческим сведениям о роли государств в мировой истории и в целом отражают динамику смены мировых лидеров [3, 4, 9, 43, 57]. В частности, на территории современного Китая в XVI веке возникла империя Цин, достигшая своего внешнего могущества к 1790 году (население – 380 млн чел., территория – 14,7 млн

км²). Отставание в уровне социально-технологического развития от геополитических соперников (империи Цин от Японии, России и других стран, Британской империи от США) через некоторое время привело к крушению этих империй.

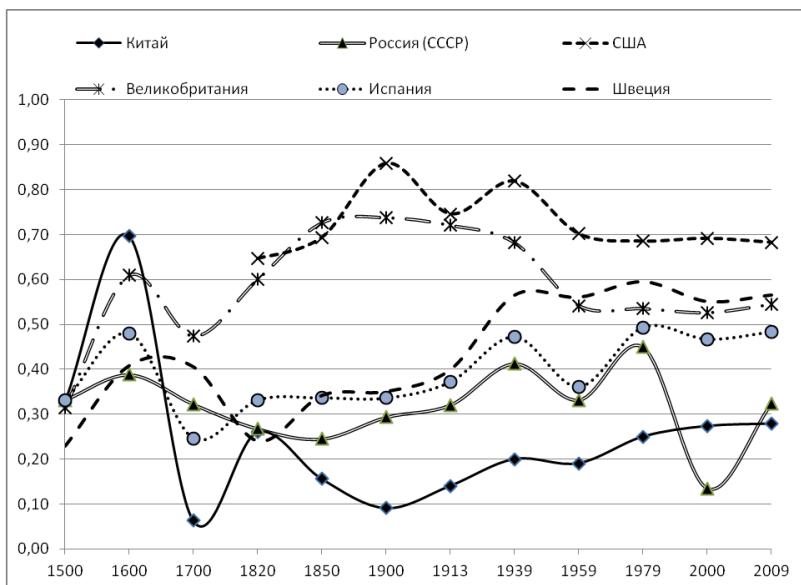


Рис. 1. Значения индекса социальных технологий

Следует отметить, что используемые в расчетах данные по ВВП (А. Мэддисон, [67, 68]) и уровню урбанизации для стран Восточной Европы, Азии и Африки до 1800 г., возможно, занижены. Альтернативные измерения межстранового неравенства по состоянию на 1800 г. позволяют сделать вывод: весьма вероятно, что средний доход в Японии, Китае, некоторых регионах Юго-Восточной Азии был сравним (или даже выше) с доходами в Западной Европе даже в конце XVIII века [29]. Не исключено, что урбанизация в основных регионах Азии и Среднего Востока в XVIII веке также была не ниже (а иногда и выше) уровня Европы [29].

Высокий уровень социальных технологий создает предпосылки для *фундаментальных* открытий (это открытия, «расширяющие экологическую нишу этноса и способствующие увеличению его численности» – достижения в области производства пищи, новое оружие, транспортные средства и др. [38, с. 111]). Эффект фундаментальных открытий таков, что они дают народу-первооткрывателю решающее преимущество перед другими народами. Фундаментальное открытие и вызванная им военная или технологическая революция порождают миграционную и диффузионную волну, постепенно образующую культурный или цивилизационный круг – область распространения данного культурного комплекса [33, 38, 73].

Рассмотрим пример фундаментального открытия [38]. В 1629 г. в Швеции была создана легкая полковая пушка, стреляющая картечью и обеспечивающая пехоте постоянную огневую поддержку. Шведский король Густав Адольф впервые в Европе ввел всеобщую воинскую повинность, создав регулярную армию, ввел поземельный налог, частично монополизировал торговлю. Данный комплекс мер позволил в небольшой по численности стране создать 80-тысячную армию, разгромить немецкую армию, поставить на грань гибели Речь Посполитую и стать хозяевами Центральной Европы. К 1700 г. значение индекса технологий России было незначительно ниже значения индекса Швеции, что позволило России в кратчайшие сроки внедрить у себя шведские новации и некоторое время спустя разгромить шведскую армию.

Промышленная революция, возникшая в Англии в последней трети XVIII века, за короткое историческое время охватила большинство стран Европы и Северной Америки. В связи с низким индексом технологий и другими причинами (внутренние неурядицы и т.д.) страны Азии, Африки, Латинской Америки не смогли перенять новые технологии и оказались в длительной зависимости от Европы и США.

3.2. ОЦЕНКА СТЕПЕНИ ТЕХНОЛОГИЧЕСКОГО ФАКТОРА

Рост социальных технологий в некоторой стране служит образцом для подражания среди других народов. Страна, достигшая определенного уровня развития, притягивает другие

государства, объединяет их в формальные и неформальные союзы. Крепость и силу этих союзов можно оценить по результатам военных действий, поскольку в ходе войны «все политические и социальные учреждения подвергаются проверке и испытанию «огнем и мечом». Сила и слабость учреждений и порядков любого народа определяется исходом войны и последствиями ее» [30].

В таблице 1 представлены основные демографические показатели накануне первой мировой войны и военные потери Великобритании, Франции, их колоний и доминионов [36, 55].

В ходе боевых действий Франция потеряла погибшими и умершими военными служащими 3,2% населения, тогда как в ее колониях доля погибших составила 0,44% (в семь раз меньше).

Таблица 1. Основные характеристики Великобритании и Франции, их колоний и доминионов накануне первой мировой войны, военные потери

| Страна | Население, млн чел. | Погибло и умерло военными служащими, млн чел. | Доля погибших от численности населения, % | Отношение к доли метрополии, Δ_{ij} |
|---------------------|---------------------|---|---|--|
| Франция | 41,463 | 1,327 | 3,20 | 1 |
| Французские колонии | 52,7 | 0,23 | 0,44 | 0,14 |
| Великобритания | 45,426 | 0,715 | 1,57 | 1 |
| Канада | 7,852 | 0,061 | 0,78 | 0,49 |
| Австралия | 4,821 | 0,06 | 1,24 | 0,79 |
| Индия | 303,7 | 0,054 | 0,02 | 0,01 |
| Новая Зеландия | 1,122 | 0,016 | 1,43 | 0,91 |
| Южная Африка | 6,153 | 0,007 | 0,11 | 0,07 |

Для оценки степени технологического фактора χ воспользуемся выражением

$$(11) \sum_{j \in J_i} \Delta_{ij} z_j = (1 + I_i)^\chi z_i,$$

где J_i – множество стран, являющихся союзниками (доминионами) i -й страны, включая союзообразующее

государство (метрополию); Δ_{ij} – отношение доли погибших в j -й стране к доли погибшим в метрополии (i -й стране).

Содержательно левая часть выражения (11) есть численность «эффективного населения» i -й страны, учитывающая вклад союзников в решение жизненно важных дел метрополии (союзобразующего государства).

При значении индекса $I_i = 0,41$ социальных технологий Франции в 1913 г. получим:

$$48,65 = 41,463(1 + I_i)^{\chi}, \quad \chi = \ln(1,17)/\ln(1,41) = 0,47.$$

Для Великобритании в 1913 г. индекс равен $I_i = 0,71$. Получим по формуле (11):

$$\chi = \ln(1,28)/\ln(1,71) = 0,46.$$

В таблице 2 представлены основные демографические показатели Германии и ее сателлитов накануне второй мировой войны [13, 52, 55, 70], а также количество погибших военнoслужаших.

Таблица 2. Основные характеристики Германии и ее сателлитов накануне второй мировой войны, военные потери

| Страна | Население, млн чел. | Погибло военнoслужаших, млн чел. | Доля погибших от численности населения, % | Отношение к доли Германии, Δ_{ij} |
|--------------|---------------------|----------------------------------|---|--|
| Германия | 69,31 | 6,5 | 9,38 | 1,00 |
| Австрия | 7,0 | 0,23 | 3,29 | 0,35 |
| Чехословакия | 15,5 | 0,15 | 0,97 | 0,10 |
| Италия | 43,2 | 0,33 | 0,76 | 0,08 |
| Венгрия | 9,2 | 0,14 | 1,52 | 0,16 |
| Румыния | 19,9 | 0,2 | 1,01 | 0,11 |
| Финляндия | 3,8 | 0,082 | 2,16 | 0,23 |

Выполнив по формуле (11) расчеты, получим:

$$\chi = \ln(1,17)/\ln(1,61) = 0,34.$$

В ходе войны в Ираке войска США и их союзников с 2003 г. по 2010 г. потеряли 4704 чел. (США – 4386 чел. или 93%, Великобритания – 179 чел., другие – 139 чел.) [28]. Получим следующую оценку степени технологического фактора: $\chi = 0,13$.

Таким образом, мы получили *нижнюю оценку* степени технологического фактора $\chi \approx 0,13 \div 0,5$, отражающую возможности государства по эффективному удержанию контролируемых территорий и населения и созданные для этой цели институты. Минимальное значение степени (0,13) характерно для ситуаций, когда преследуются преимущественно экономические интересы. Если речь идет о военной безопасности, реализации социально-политических целей, то следует использовать значение для нижней оценки степени, равное примерно 0,5.

4. Оценка параметров функции сохранения

4.1. ОЦЕНКА ПАРАМЕТРА РАЗНОРОДНОСТИ

Известны два подхода к трактовке этнических (культурных) гетерогенностей [22]. При первом подходе измеряется фрагментированность общества (раздробленность страны по некоторому признаку). При втором оценивается поляризованность, учитывающая степень сходства между группами (чем меньше сходства, тем острее антагонизм). Для измерения этнической гетерогенности (разнородности) используются следующие основные индексы [22]:

- индекс этнолингвистической фрагментированности (ИЭЛФ);
- индекс Гринберга (GI);
- индекс второстепенной гетерогенности (PH);
- индекс поляризованности Эстебана – Рэя (ER);
- индекс поляризованности Рейнал – Кэрл (RQ);
- индекс поляризованности Вольфсона;
- индекс поляризованности Алескерова – Голубенко.

Различение между группами в политических сообществах производится по одной или нескольким объективным характеристикам: языку, религии, расовой, этнической или племенной принадлежности. Для оценки дистанций между группами используются экспертные оценки, лингвистическое дерево, лексикостатистический анализ языков, генетические расстояния, опросные данные [22].

Перечисленные выше индексы фиксируют различия между группами, но не дают ответа на вопрос, как выявленные разнородности транслируются в действия.

Для этносов существующего государства (союза) параметр разнородности между государствообразующим этносом (национальностью) i и этносом j определим по формуле

$$(12) \mu_{ij} = \frac{Pr_G}{Pr_j},$$

где Pr_j – процент потерь от численности для j -й национальности (этноса); Pr_G – процент потерь от численности государствообразующего этноса.

Вычислим значения параметра разнородности между русским этносом и другими этносами на примере вклада различных национальностей в победу СССР в Великой Отечественной войне. Для чего воспользуемся результатами переписи населения СССР 1939 г. и данными по безвозвратным потерям (по национальностям) [11, 47].

В Великой Отечественной войне русские потеряли 5,78% (безвозвратные потери), татары – 4,35%, евреи – 4,71%, казахи – 4,05%, узбеки – 2,43% и т.д. Соответственно, значения параметра разнородности равны: 1; 1,3; 1,2; 1,4; 2,4.

Наряду с этническими различиями иногда следует учитывать влияние на разнородность населения и социальных различий, в том числе порождаемых демографическими циклами. Основы концепции демографических циклов были заложены Т. Мальтусом [34] и получили дальнейшее развитие в работах Дж. Кейнса [23], Ф. Броделя [6], И. Валлерстайна [8], П.В. Турчина [54], А.В. Коротаева, А.С. Малкова, Д.А. Халтуриной [25], С.А. Нефедова [38, 39] и др.

4.2. ОЦЕНКА ПАРАМЕТРА ПРИТЯЖЕНИЯ

Наряду с исследователями, безопасность и комфортность проживания в тех или иных странах и регионах оценивают тысячи и миллионы обычных граждан, голосуя за безопасность «ногами», мигрируя в другие страны и регионы. Для характеристики миграции обычно используется коэффициент миграционного прироста (разница между прибывшими и

убывшими) – отношение миграционного прироста к среднегодовой численности или на 10 000 чел. населения.

Пусть M_j есть коэффициент миграции в j -м регионе (может принимать как положительные, так и отрицательные значения). Приведем его значения к отрезку $[0, 1]$:

$$(13) \quad m_j = \frac{M_j - M_{\min}}{M_{\max} - M_{\min}},$$

где M_{\max} (M_{\min}) – максимальное (минимальное) значение коэффициента миграции по всем регионам. Если приведенное значение m_j коэффициента близко к нулю, то j -й регион неблагоприятен. При $m_j \rightarrow 1$ в регион стремится переехать максимальное количество граждан.

Э.Г. Равенштейн обосновал, что главными причинами миграции являются экономические; мигранты на длинные расстояния мигрируют в крупные центры промышленности и торговли; жители городов менее подвижны, чем жители сельской местности [72]. Выбор мигрантами региона (страны) производится и с учетом фактора безопасности проживания [66]. Исходя из законов миграции [66, 72] сделаем предположение, что мигранты выбирают регион назначения исходя из его социально-экономических показателей и комфортности (безопасности) проживания в нем.

Определим функцию привлекательности для мигрантов j -го региона:

$$(14) \quad U_{R_j}(\delta_i) = K_{E_j} \left(\frac{\zeta_{R_j}}{z_{R_j}} \right)^{\delta_i \mu_{ij}},$$

где $0 < K_{E_j} < 1$ – уровень социально-экономического развития региона; z_{R_j} – численность населения j -го региона; ζ_{R_j} – численность самой многочисленной национальности в j -м регионе; μ_{ij} – параметр разнородности; δ_i – параметр притяжения русского этноса.

В данном выражении параметр μ_{ij} отражает существующие в обществе и государстве этнические разнородности, параметр K_{E_j} показывает относительные различия между регионами преимущественно экономического характера. Соответственно, параметр δ_i характеризует возможности основного этноса,

социальных и государственных институтов по формированию единого культурного, политического, экономического пространства.

Используя социально-экономические показатели регионов России за 2009–2011 гг. [46, 51] оценим параметр притяжения δ_i русского этноса методом наименьших квадратов:

$$(15) \delta_i = \text{ArgMin} \sum_{j=1}^{K_R} (U_{R_j}(\delta_i) - m_j)^2,$$

где K_R – количество регионов; m_j – коэффициент миграционного прироста в j -м регионе, приведенный к отрезку $[0, 1]$ (параметр миграции).

В каждом государстве имеются свои особенности ведения статистики по регионам, поэтому набор используемых показателей для каждой страны обычно различен. Для России определим социально-экономический критерий j -го региона (субъекта федерации) в виде

$$(16) K_{Ej} = \frac{1}{4} \left(\frac{K_{1j}}{K_{1\max}} + \frac{K_{2j}}{K_{2\max}} + \frac{K_{3j}}{K_{3\max}} + K_{4j} + K_{5j} \right),$$

где K_{ij} – значение i -го показателя для j -го региона; $K_{i\max}$ – максимальное значение i -го показателя по всем регионам; $i = 1$ – валовой региональный продукт на душу населения; $i = 2$ – отношение среднемесячной зарплаты к величине прожиточного минимума; $i = 3$ – количество легковых автомобилей на тысячу человек; $i = 4$ – климатический фактор; $i = 5$ – доля городского населения.

Пяти основным климатическим регионам современной России [48] поставлен в соответствие весовой коэффициент K_{4j} – комфортность проживания, значение которого зависит от существующих технологических возможностей и социальной политики государства (таблица 3).

Таблица 3. Комфортность проживания в регионах России

| Климатический регион | Комфортность проживания |
|----------------------|-------------------------|
| Ia («особый») | 0,2 |
| Iб | 0,4 |
| II | 0,6 |

| | |
|----------------------|-------------------------|
| Климатический регион | Комфортность проживания |
| III | 0,8 |
| IV | 1 |

Получим для России $\delta_i = 0,15$. При различных наборах социально-экономических факторов, характеризующих уровень развития региона, оценка параметра находится в интервале $\delta_i = 0,1 \div 0,5$.

В таблице 4 представлены результаты вычислений для некоторых регионов России по состоянию на 2013 г. Значение параметра миграции, равное 0, соответствует минимальному значению коэффициента миграции (происходит существенный отток населения из региона). Значение параметра, равное 1, соответствует максимальному притоку населения в регион.

Таблица 4. Количественные характеристики регионов России

| № | Субъект Федерации | Параметр миграции | Доля осн. этноса | Критерий соц-эк. разв. | Функция сохранения | Функция привлекательности |
|-------------------------------|----------------------|-------------------|------------------|------------------------|--------------------|---------------------------|
| Центральный федеральный округ | | | | | | |
| 1 | Белгородская область | 0,68 | 0,944 | 0,55 | 0,992 | 0,547 |
| 2 | Брянская область | 0,45 | 0,967 | 0,59 | 0,995 | 0,590 |
| 3 | Владимирская область | 0,54 | 0,956 | 0,54 | 0,993 | 0,536 |
| 4 | Воронежская область | 0,68 | 0,955 | 0,47 | 0,993 | 0,470 |
| 5 | Ивановская область | 0,55 | 0,956 | 0,59 | 0,993 | 0,589 |
| 6 | Калужская область | 0,61 | 0,931 | 0,54 | 0,990 | 0,529 |
| 7 | Костромская область | 0,42 | 0,966 | 0,47 | 0,995 | 0,471 |
| 8 | Курская область | 0,51 | 0,965 | 0,47 | 0,995 | 0,468 |
| 9 | Липецкая область | 0,53 | 0,963 | 0,46 | 0,994 | 0,455 |
| 10 | Московская область | 1,00 | 0,929 | 0,48 | 0,989 | 0,476 |
| 11 | Орловская область | 0,48 | 0,960 | 0,47 | 0,994 | 0,468 |
| 12 | Рязанская область | 0,57 | 0,951 | 0,42 | 0,993 | 0,418 |
| 13 | Смоленская область | 0,54 | 0,947 | 0,48 | 0,992 | 0,472 |
| 14 | Тамбовская область | 0,51 | 0,970 | 0,50 | 0,996 | 0,500 |
| 15 | Тверская область | 0,53 | 0,935 | 0,43 | 0,990 | 0,430 |
| 16 | Тульская область | 0,59 | 0,953 | 0,47 | 0,993 | 0,466 |
| 17 | Ярославская область | 0,57 | 0,960 | 0,62 | 0,994 | 0,612 |

| № | Субъект Федерации | Параметр миграции | Доля осн. этноса | Критерий соц-эк. разв. | Функция сохранения | Функция привлекательности |
|-------------------------------------|-------------------------------------|-------------------|------------------|------------------------|--------------------|---------------------------|
| 18 | г. Москва | 0,80 | 0,917 | 0,59 | 0,987 | 0,582 |
| Северо-Кавказский федеральный округ | | | | | | |
| 19 | Республика Дагестан | 0,49 | 0,294 | 0,58 | 0,444 | 0,257 |
| 20 | Республика Ингушетия | 0,40 | 0,941 | 0,58 | 0,894 | 0,518 |
| 21 | Кабардино-Балкарская Республика | 0,36 | 0,572 | 0,58 | 0,750 | 0,438 |
| 22 | Карачаево-Черкесская Республика | 0,50 | 0,410 | 0,54 | 0,519 | 0,281 |
| 23 | Республика Северная Осетия – Алания | 0,36 | 0,651 | 0,58 | 0,887 | 0,514 |
| 24 | Чеченская Республика | 0,40 | 0,953 | 0,55 | 0,914 | 0,503 |
| 25 | Ставропольский край | 0,60 | 0,809 | 0,52 | 0,969 | 0,508 |

Из таблицы видно, что минимальное значение функции сохранения имеет Республика Дагестан с долей основного этноса 0,29. В этой республике продолжительное время наблюдаются террористические акты; несмотря на ряд серьезных усилий, все еще не удалось снизить уровень межэтнических и поликультурных противоречий, являющихся питательной средой для внутренних конфликтов, экстремизма и терроризма.

Используя данные переписей населения и другие статистические данные [12, 14, 24, 60, 62, 74, 75, 76], выполнены расчеты по оценке значений параметра сохранения для пяти государств (таблица 5).

Таблица 5. Значения параметра сохранения

| | | | | |
|---------|---------|----------|-----------|---------|
| Россия | США | Германия | Казахстан | Украина |
| 0,1÷0,5 | 0,3÷0,8 | 0,2÷2,4 | 0,8÷1,3 | 0,8÷1,4 |

Малые значения параметра сохранения отражают высокие возможности государствообразующего этноса по формированию устойчивых социально-политических

институтов. Поскольку во всех землях Германии немцы составляют свыше 50% от численности населения, то значение параметра сохранения для Германии имеет большой разброс, что свидетельствует о необходимости проведения дополнительных исследований. Эти исследования можно, например, провести в форме ежегодных вычислений значения параметра по актуальным статистическим данным на интервале в несколько лет.

В силу различий национальных статистик государств результаты вычислений значений функции привлекательности для регионов разных государств несравнимы между собой. Полученные значения функции можно использовать только для сравнительного анализа регионов внутри одной страны.

5. Оценка безопасности России

В подавляющем большинстве жители Российской империи стали фиксироваться с 1719 г., когда на смену подворным переписям пришли ревизии. В таблице 6 представлены результаты исследований С.И. Брука и В.М. Кабузана [7] по численности населения Российской империи (в границах XIX в.).

Таблица 6. Численность населения Российской империи

| Вид и год учета | Всего населения, млн чел. | Численность русских, млн чел. | Доля русских, % | Средний коэффициент прироста населения, % | Средний коэффициент прироста русских, % |
|------------------------|---------------------------|-------------------------------|-----------------|---|---|
| Подворный учет 1678 г. | 20,0 | 8,12 | 40,6 | | |
| I ревизия 1719 г. | 27,18 | 11,128 | 40,9 | 0,75 | 0,77 |
| IV ревизия 1782 г. | 40,261 | 18,082 | 44,9 | 0,63 | 0,77 |
| V ревизия 1795 г. | 46,587 | 20,118 | 43,2 | 1,13 | 0,82 |
| VIII ревизия 1833 г. | 66,731 | 28,644 | 42,9 | 0,95 | 0,93 |
| X ревизия 1858 г. | 80,499 | 34,821 | 43,3 | 0,75 | 0,78 |
| Перепись 1897 г. | 128,203 | 55,765 | 43,5 | 1,20 | 1,21 |
| Перепись 1916–1917 гг. | 171,75 | 76,676 | 44,6 | 1,47 | 1,60 |
| 1678–1917 гг. | | | | 0,90 | 0,94 |

Средний коэффициент r прироста населения вычислен по формуле

$$(17) r = \exp\left(\frac{\ln P_n - \ln P_0}{n}\right) - 1,$$

где P_n – численность населения по текущей переписи; P_0 – численность населения по предыдущей переписи; n – количество лет между указанными переписями.

Отмена крепостного права и первые успехи медицины способствовали повышению темпов естественного прироста населения в конце XIX – начале XX в. в основном за счет снижения смертности. Причем в исконных великорусских регионах (Центрально-промышленном, Озерном и Северном) естественный прирост был существенно ниже за счет огромных размеров отходничества: сотни тысяч людей, главным образом мужчин производительных возрастов, надолго уходили из родных мест на заработки. Одновременно русское население из регионов своего исконного обитания распространялось на все части империи [7].

В таблице 7 представлены основные результаты переписей населения после 1917 г.

Таблица 7. Численность населения СССР (Российской Федерации)

| Год переписи | Всего населения, млн чел. | Численность русских, млн чел. | Доля русских, % | Средний коэффициент прироста населения, % | Средний коэффициент прироста русских, % |
|--------------|---------------------------|-------------------------------|-----------------|---|---|
| 1926 г. | 146,638 | 77,791 | 53 | -1,6 | 0,1 |
| 1939 г. | 170,557 | 99,592 | 58,4 | 1,2 | 1,9 |
| 1959 г. | 208,827 | 114,114 | 54,6 | 1,0 | 0,7 |
| 1970 г. | 241,720 | 129,015 | 53,4 | 1,3 | 1,1 |
| 1979 г. | 262,087 | 137,397 | 52,4 | 0,9 | 0,7 |
| 1989 г. | 285,740 | 145,155 | 50,8 | 0,9 | 0,6 |
| 2002 г. | 145,167 | 115,9 | 79,8 | -5,1 | -1,7 |
| 2010 г. | 142,857 | 111,017 | 77,7 | -0,2 | -0,5 |

На рис. 2 показаны результаты расчета функции безопасности (и ее компонентов) России за период с 1500 г. по 2009 г.

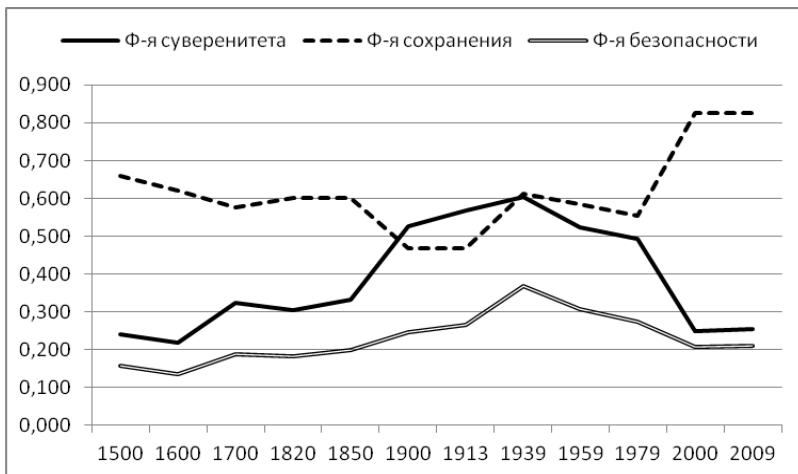


Рис. 2. Компоненты функции безопасности России

Расчеты основаны на данных, представленных в таблице 8. Отметим, что состав народов на территории Российской империи – СССР практически не поменялся.

Таблица 8. Статистические данные, характеризующие безопасность России

| Год | Площадь, млн кв. км | Население, млн чел./ макс. знач. | Доля русских, % | ВВП на душу населения/ макс. знач. | Урбанизация / макс. знач. | Индекс инноваций | Параметры $\mu\delta$ |
|------|---------------------|----------------------------------|-----------------|------------------------------------|---------------------------|------------------|-----------------------|
| 1500 | 2,4 | 17/110 | 50 | 0,5 | -/0,21 | 0,33 | 0,6 |
| 1600 | 5,7 | 21/160 | 45 | 0,55 | -/0,24 | 0,39 | 0,6 |
| 1700 | 13,5 | 26/165 | 40 | 0,62/1,4 | 0,13/0,37 | 0,32 | 0,6 |
| 1820 | 16 | 55/381 | 43 | 0,69/2,11 | 0,09/0,29 | 0,27 | 0,6 |
| 1850 | 20 | 69/412 | 43 | 0,88/1,85 | 0,09/0,5 | 0,24 | 0,6 |
| 1900 | 23 | 129/400 | 43 | 1,19/4,49 | 0,13/0,6 | 0,29 | 0,9 |
| 1913 | 22,4 | 156/437 | 43 | 1,49/5,3 | 0,15/0,78 | 0,32 | 0,9 |

| Год | Площадь, млн кв. км | Население, млн чел./ макс. знач. | Доля рус- ских, % | ВВП на душу населения/ макс. знач. | Урба- низация / макс. знач. | Индекс инно- ваций | Параме- тры $\mu\delta$ |
|------|---------------------------|--|----------------------------|--|--------------------------------------|--------------------------|----------------------------|
| 1939 | 21 | 192/516 | 58 | 2,24/6,56 | 0,32/0,78 | 0,41 | 0,9 |
| 1959 | 22,4 | 210/666 | 55 | 3,67/11,86 | 0,54/0,92 | 0,33 | 0,9 |
| 1979 | 22,4 | 264/969 | 52 | 6,48/18,79 | 0,69/0,94 | 0,45 | 0,9 |
| 2000 | 17,1 | 146/1263 | 81 | 5,29/28,47 | 0,70/0,97 | 0,13 | 0,9 |
| 2009 | 17,1 | 143/1331 | 81 | 8,97/30,9 | 0,73/0,98 | 0,32 | 0,9 |

Результатом двух модернизаций, начавшихся в 1917 г. (Большая русская революция) и в 1991 г. (распад СССР) явилось увеличение доли русских в общей численности населения. К 1926 г. доля русских увеличилась с 45% до 53%, к 2002 г. – с 51% до 80%. С точки зрения демографической истории России нынешняя ситуация является исключительной. Вероятно, идущие ныне интеграционные процессы приведут к созданию нового восточноевропейского (евроазиатского) союза.

Выполненные расчеты дают основания предположить, что важнейшей предпосылкой возможного распада (трансформации) государства (союза государств) является снижение значений функции сохранения ниже $0,5 \div 0,6$.

6. Заключение

Таким образом, нами рассмотрена модель безопасности государства, учитывающая основные факторы, влияющие на исторические процессы (по С. Нефедову и Л. Гумилеву): география, демография и технологии.

Выполнена оценка основных параметров модели:

1. Индекс социальных технологий, отражающий качество социальных институтов и уровень социально-экономического развития государства. Индекс вычисляется как среднее значение трех нормированных показателей: ВВП на душу населения, доля городского населения и темпы роста численности населения.

2. Степень социально-технологического фактора, отражающая способность государства по привлечению

союзников (сателлитов) для решения жизненно важных задач (отражение военной агрессии и т.д.).

3. Параметр разнородности, отражающий существующие социально-этнические гетерогенности в обществе. Параметр оценен для основных национальностей СССР (России) с учетом доли военных потерь в годы Великой Отечественной войны.

4. Параметр притяжения, характеризующий возможности государствообразующего этноса устранять существующие различия и сплачивать другие этносы в единую систему. параметр оценен на основе статистических данных, характеризующих социально-экономическое развитие государства.

Актуальной научной задачей является более точный учет параметров, характеризующих социальные и этнические различия и взаимодействия социальных групп (этносов), в том числе с использованием теории графов.

Имея базовую модель безопасности государства, можно рассматривать ее расширения, связанные с учетом дополнительных факторов: модели динамики безопасности, модели безопасности приграничных регионов (районов) и др.

Литература

1. АБЕРКРОМБИ Н., ХИЛЛ С., ТЕРНЕР Б. *Социологический словарь*. 2-е изд., перераб. и доп. / Пер. с англ. И.Г. Ясавеева, под ред. С.А. Ерофеева. – М.: Экономика, 2004. – 620 с.
2. АКАЕВ А.А., МАЛКОВ С.Ю. *Геополитическая динамика: возможности логико-математического моделирования* // Геополитика и безопасность. – 2009. – №4(8). – С. 39–55.
3. БАЛАЦКИЙ Е. *О возможной смене глобального лидера мировой экономики* // Общество и экономика. – 2014. – №1. С. 3–21.
4. БАЛАЦКИЙ Е.В. *Предпосылки глобальной геополитической инверсии* // TERRA ECONOMICUS. – 2014. – Том 12, №3. – С. 15–18.

5. БИЧУРИН Н.Я. *Статистическое описание Китайской империи* (в двух частях). – М.: Восточный Дом, 2002. – 464 с.
6. БРОДЕЛЬ Ф. *Материальная цивилизация, экономика и капитализм XV-XVIII вв.* Том 1. Структуры повседневности: возможное и невозможное. – М.: Прогресс, 1986. – 623 с.
7. БРУК С.И., КАБУЗАН В.М. Динамика численности и расселения русского этноса (1678–1917 гг.). [Электронная публикация]. – URL: <http://www.politpros.com/journal/read/?ID=431> (дата обращения: 01.09.2015).
8. ВАЛЛЕРСТАЙН И. *Анализ мировых систем и ситуация в современном мире* / Пер. с англ. П.М. Кудюкина. Под общ. ред. Б.Ю. Кагарлицкого. – СПб.: Изд-во «Университетская книга», 2001. – 416 с.
9. ВАЛЛЕРСТАЙН И. *После либерализма*: Пер. с англ. / Под ред. Б.Ю. Каргалицкого. – М.: Едиториал УРСС, 2003. – 256 с.
10. ВИНОКУРОВ Г.Н., КОНЯХИН Б.А., ПОДКОРЫТОВ Ю.А. *Геополитический статус Китая как фактор российской политики ядерного сдерживания Соединенных Штатов // Стратегическая стабильность.* – 2008. – №2. – С. 49–53.
11. *Всесоюзная перепись населения 1939 года: основные итоги.* – М.: Наука, 1992. – 254 с.
12. *Всеукраинская перепись населения, 2001.* [Электронная публикация]. – URL: <http://2001.ukrcensus.gov.ua/rus/> (дата обращения: 05.05.2015).
13. ГОЛУБЧИК М.М. *Политическая география мира: Учебное пособие.* – Смоленск: Изд-во СГУ, 1998. – 312 с.
14. *Государственная служба статистики Украины.* [Электронный ресурс]. – URL: <http://www.ukrstat.gov.ua/> (дата обращения: 05.05.2015).
15. ГУМИЛЕВ Л.Н. *Этногенез и биосфера Земли* / Под ред. В.С. Жекулина. – 2-е изд. испр. и доп. – Л.: Изд-во ЛГУ, 1989. – 496 с.
16. ДАЛЬ В.И. *Толковый словарь живого великорусского языка:* в 4 т. – СПб., 1863–1866 гг.

17. ЕЛИСЕЕВА И.И., ПОПОВА И.Н. *Начало международного признания российской государственной статистики* // Вопросы статистики. – 2013. – №8. – С. 80–85.
18. ЕРОФЕЕВ Н.А. *Империя создавалась так. Английский колониализм в XVIII веке.* – М.: Наука, 1964. – 117 с.
19. ЕФИМОВ А.В. *Из истории великих русских географических открытий.* – М.: Учпедгиз, 1949. – 150 с.
20. ЗУБЕЦ А.Н. *Истоки и история экономического роста.* – М.: Изд-во «Экономика», 2014. – 463 с.
21. ЗУБЕЦ А.Н. *Количественные оценки в истории (инструменты для клиометрии)* // Финансовый университет при Правительстве РФ. – 2014. – [Электронная публикация]. – URL: <http://www.fa.ru> (дата обращения: 05.03.2015).
22. КАМАЛОВА Р.У. *Этническая гетерогенность: основные понятия и проблемы измерения* // «Полития». – 2013. – №4(71). – С. 127–149.
23. КЕЙНС Дж.М. *Экономические последствия Версальского договора.* – М.: ГИЗ, 1922. – XIV, 15 – 165 с.
24. *Киевский международный институт социологии.* [Электронный ресурс]. – URL: <http://www.kiis.com.ua/> (дата обращения: 05.05.2015).
25. КОРОТАЕВ А.В., МАЛКОВ А.С., ХАЛТУРИНА Д.А. *Законы истории. Математическое моделирование исторических макропроцессов. Демография, экономика, войны* / Отв. ред. Н.Н. Крадин. – М.: КомКнига, 2005. – 344 с.
26. КОХ Р. *Принцип 20/80: секреты достижения больших результатов при затрате меньших усилий.* – М.: Попурри, 2002. – 352 с.
27. КУЗНЕЦОВ Д.В. *Проблемы Ближнего Востока и общественное мнение: в 2-х частях. Часть II: Иракский кризис.* – Благовещенск: Изд-во БГПУ, 2009. – 440 с.
28. КУЗНЕЦОВ Д.В. *Использование военной силы во внешней политике США: Учебное пособие.* – Благовещенск: Изд-во БГПУ, 2010. – 430 с.
29. ЛАТОВ Ю.В. *Генезис глобального неравенства* // Историко-экономические исследования. – 2009. – Т. 10, №3. – С. 120–132.

30. ЛЕНИН В.И. *«Сожаление» и «стыд»* // Полное собрание сочинений, 5-е изд-е. – Т.20. – М.: Политиздат, 1973. – С. 245.
31. ЛИВИ БАЧЧИ М. *Демографическая история Европы* / Пер. с итал. А. Миролубовой. – СПб.: Александрия, 2010. – 310 с.
32. ЛЬВОВ Д.С., ГЛАЗЬЕВ С.Е. *Теоретические и прикладные аспекты управления научно-техническим прогрессом* // Экономика и математические методы. – 1987. – Т. 23, Вып. 5. – С. 793–804.
33. ЛУРЬЕ С.В. *Историческая этнология*. 2-е изд. – М.: Аспект Пресс, 1998. – 448 с.
34. МАЛЬТУС Т. *Опыт о законе народонаселения*. Пятое издание (1817 г.) // Антология экономической классики. – М.: «Эконов», «Ключ», 1993. – 486 с.
35. МАСЛОУ А. *По направлению к психологии бытия*. – М.: Изд-во: ЭКСМО-Пресс, 2002. – 272 с.
36. *Мировая война в цифрах*. – М., Л.: Гос. воен. издат., 1934. – 128 с.
37. *Народонаселение стран мира: Справочник* / Под ред. Б.Ц. Урланиса и В.А. Борисова. – 3-е изд., перераб. и доп. – М.: Финансы и статистика, 1983. – 447 с.
38. НЕФЕДОВ С.А. *Концепция демографических циклов*. – Екатеринбург: Издательство УГГУ, 2007. – 141 с.
39. НЕФЕДОВ С.А. *Факторный анализ исторического процесса. История Востока*. – М.: «Территория будущего», 2008. – 752 с.
40. НИКИТИН В.А. *Организационные типы современной культуры*: Автореферат дисс. д-ра культурологии. – Тольятти, 1998. – 49 с.
41. НОВИКОВ А.М., НОВИКОВ Д.А. *Методология*. – М.: СИНТЕГ, 2007. – 668 с.
42. *Обзорный доклад о модернизации в мире и Китае (2001–2010)* / Пер. с англ. под общ. ред. Н.И. Лапина / Предисл. Н.И. Лапин, Г.А. Тосунян. – М.: Издательство «Весь Мир», 2011. – 256 с.
43. ПОПОВ В.В. *Почему Запад разбогател раньше, чем другие страны, и почему Китай сегодня догоняет Запад? Новый*

- ответ на старый вопрос // Журнал Новой Экономической Ассоциации.* – 2012. – №3(15). – С. 35–64.
44. *Психология: Учебник для гуманитарных вузов / Под ред. В.Н. Дружинина.* – СПб., М., Харьков, Минск: Изд-во Питер, 2001. – 656 с.
 45. РАШИН А.Г. *Население России за 100 лет (1811–1913 гг.).* – М.: Государственное статистическое издательство, 1956. – 352 с.
 46. *Регионы России. Социально-экономические показатели. 2012: Стат. сб.* – М.: Росстат, 2012. – 990 с.
 47. *Россия и СССР в войнах XX века. Потери вооруженных сил. Статистическое исследование / Под общ. ред. Г.Ф. Кривошеева.* – М.: ОЛМА-ПРЕСС, 2001. – 608 с.
 48. *Руководство по гигиенической оценке факторов рабочей среды и трудового процесса. Критерии и классификация условий труда.* Руководство Р 2.2.2006-05. Утв. Главным государственным санитарным врачом РФ 29.07.2005 г.
 49. САДОВНИЧИЙ В.А., АКАЕВ А.А., КОРОТАЕВ А.В. и др. *Моделирование и прогнозирование мировой динамики.* – М.: ИСПИ РАН, 2012. – 359 с.
 50. СОКОЛОВ А.В. *Общая теория социальной коммуникации: Учебное пособие.* – СПб.: Изд-во Михайлова В.А., 2002 г. – 461 с.
 51. *Социально-демографический портрет России: По итогам Всероссийской переписи населения 2010 года / Федер. служба гос. статистики.* – М.: ИИЦ «Статистика России», 2012. – 183 с.
 52. ТИППЕЛЬСКИРХ К., КЕССЕЛЬРИНГ А. и др. *Итоги Второй мировой войны. Выводы побежденных.* Пер. с нем. – СПб.: Полигон; М.: АСТ, 1998. – 640 с.
 53. ТОФФЛЕР Э. *Третья волна.* – М.: АСТ, 2010. – 784 с.
 54. ТУРЧИН П.В. *Историческая динамика: На пути к теоретической истории.* – М.: УРСС, 2007. – 368 с.
 55. УРЛАНИС Б.Ц. *Войны и народонаселение Европы.* – М.: Издательство социально-экономической литературы, 1960. – 568 с.

56. ФРОЛОВ Э.Д. *Проблема цивилизаций в историческом процессе* // Вестник Санкт-Петербургского университета. Серия 2: История. – 2006. – №2. – С. 96–100.
57. ХОКСВОРТ Дж., ТИВАРИ А. *Мир в 2050 году. Ускорение процесса изменения баланса экономических сил в мире: проблемы и возможности.* – ПрайвотерхаусКуперс LLP, 2011. – [Электронная публикация]. – URL: http://www.pwc.ru/ru_RU/ru/globalisation/assets/World-in-2050-ru.pdf (дата обращения: 06.10.2014).
58. ШТЕМЕНКО С.М. *Генеральный штаб в годы войны.* 2-е изд. Лит. запись Г.А. Сомова. – М.: Воениздат, 1989. – 560 с.
59. ALESINA A., SPOLAORE E. *War, Peace and the Size of Countries* // Journal of Public Economics. – 2005. – №89(7). – P. 1333–1354.
60. *Bureau of Economic.* U.S. Department of Commerce. [Электронная публикация]. – URL: <http://www.bea.gov/> (дата обращения: 06.08.15).
61. COBB C.W., DOUGLAS P.H. *A Theory of Production* // Amer. Econ. Rev. Suppl. – 1928. – Vol. 18, March. – P. 139–165.
62. *Demographics the United States.* – [Электронная публикация]. – URL: http://en.wikipedia.org/wiki/Demographics_of_the_United_States (дата обращения: 05.05.2015).
63. *Global Innovation Index 2014 Edition.* – [Электронная публикация]. – URL: <http://www.globalinnovationindex.org/> (дата обращения: 06.08.2014).
64. GUERRIERO V. *Power Law Distribution: Method of Multi-scale Inferential Statistics* // Journal of Modern Mathematics Frontier. – 2012. – Vol. 1, №1. – P. 21–28.
65. KOMLOS J., NEFEDOV S. *A Compact Macromodel of Pre-Industrial Population Growth* // Historical Methods. – 2002. – Vol. 35, №2. – P. 92–93.
66. LEE E.S. *A Theory of Migration* // Demography. – 1966. – Vol. 3, No. 1. – P. 47–57.
67. MADDISON A. *The World Economy: A Millennial Perspective.* – Paris: OECD, 2001. – 667 p.
68. *Maddison Project.* – [Электронный ресурс]. – URL: <http://www.ggdcc.net/maddison/maddison-project/home.htm> (дата обращения: 05.03.2015).

69. MAYNARD SMITH J., SLATKIN M. *The stability of the predator-prey systems* // Ecology. – 1973. – Vol. 53. – P. 384–391.
70. MITCHELL B.R. *European historical statistics 1750-1970*. – New York; London: Columbia Univ., 1978. – 446 p.
71. *Population History*. [Электронный ресурс]. – URL: <http://www.tacitus.nu/historical-atlas/population/> (дата обращения: 05.03.2015).
72. RAVENSTEIN E.G. *The Laws of Migration* // Journal of the Statistical Society of London. – 1885. – Vol. 48, No. 2. – P. 167–235.
73. ROBERTS M. *Essays in Swedish History*. – Minneapolis: Univ. of Minnesota Press, 1967. – 358 p.
74. *Statistische Ämter des Bundes und der Länder*. – [Электронный ресурс]. – URL: <http://www.statistik-portal.de/> (дата обращения: 05.05.2015).
75. *U.S. Census Bureau Statistical Abstract of the United States: 2011*. – [Электронный ресурс]. – URL: <http://2010.census.gov/2010census/index.php> (дата обращения: 05.03.2015).
76. *Willkommen bei der Zensusdatenbank des Zensus 2011*. – [Электронный ресурс]. – URL: <https://ergebnisse.zensus2011.de/> (дата обращения: 05.05.2015).

MODEL OF SAFETY OF THE STATE

Vladislav Shumov, International Informatizational Academy, Moscow, Cand.Sc., senior lecturer (vshum59@yandex.ru).

Abstract: In this paper we consider the model of safety of the state that reflects the dichotomy between development and conservation values. The model takes into account the main factors affecting the historical processes: geography, demography and technology.

Keywords: state safety, the mathematical model, the criterion of safety.

*Статья представлена к публикации
членом редакционной коллегии Д.А. Новиковым*

*Поступила в редакцию 15.08.2014.
Опубликована 30.11.2015.*

УДК 519.854.2

ББК 22.1

СОСТАВЛЕНИЕ ОПТИМАЛЬНОГО РАСПИСАНИЯ ДВИЖЕНИЯ Поездов МЕЖДУ ДВУМЯ СТАНЦИЯМИ, СОЕДИНЕННЫМИ ОДНОПУТНОЙ ЖЕЛЕЗНОЙ ДОРОГОЙ С РАЗЪЕЗДОМ¹

Лазарев А. А.²

*(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, ФГБОУ ВПО «Московский
государственный университет им. М.В. Ломоносова»,
ФГАОУ ВПО «Национальный исследовательский университет
«Высшая школа экономики», ФГБОУ ВПО «Московский
физико-технический институт (государственный
университет)», Москва)*

Тарасов И. А.³

*(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, ФГБОУ ВПО «Московский
государственный университет им. М.В. Ломоносова», Москва)*

Рассматривается проблема составления оптимального расписания движения поездов между двумя станциями, соединенными однопутной железной дорогой с разъездом. Подобные задачи являются предметом интенсивных исследований из-за практической значимости. Разработан точный алгоритм решения задачи минимизации времени окончания перевозок для случая одновременного поступления поездов.

¹ Работа выполнена при финансовой поддержке РФФИ: №13-01-12108, №13-08-13190, №15-07-03141, №15-07-07489, №16-31-00354; DAAD A/1400328.

Авторы признательны профессору Сиднейского технологического университета Я. Зиндеру за ценную критику и комментарии.

² Александр Алексеевич Лазарев, доктор физико-математических наук, профессор, (jobmath@mail.ru, г. Москва, ул. Профсоюзная, д. 65, тел. (495) 334-87-51).

³ Илья Алексеевич Тарасов, студент (ilya_tarasof@mail.ru).

Ключевые слова: теория расписаний, комбинаторная оптимизация, транспортные задачи, алгоритм.

Введение

Задача составления оптимального расписания на однопутных участках актуальна как для пассажирских, так и для грузовых поездов, так как такие участки составляют значительную часть любой железнодорожной сети. В основном железнодорожные линии мира однопутные; общая длина двухпутных и многопутных дорог составляет около 180 тыс. км (примерно 13% мировой сети), в том числе многопутных – около 10 тыс. км (менее 1%). Только в немногих странах Западной Европы двухпутные и многопутные линии образуют значительную часть железнодорожной сети: в Великобритании – 74% (в том числе многопутные линии – 10%), в Бельгии – 60%, в Нидерландах – 48%, во Франции, ФРГ, Швейцарии – 40-44%, в Италии – 30%, в Австрии – 25%, в других европейских странах – 10-15%. В США 42 тыс. км двухпутных и многопутных линий (треть мирового протяжения), но они составляют лишь 12% общей длины всех железнодорожных линий; в Японии и Индии двухпутные и многопутные линии составляют 18-20%, в Канаде – 6% общей железнодорожной сети. Максимальные по длине многопутные участки принадлежат США (более 100 км); в других странах длина таких участков не превышает нескольких десятков километров; обычно многопутные участки расположены на подходах к крупнейшим городам. В России протяженность путей общего пользования составляет 86 тыс. км (всего 120 тыс. км), из них двухколейных и многоколейных 43,8 тыс. км [1]. Для повышения пропускной способности однопутных участков используются разьезды.

Рассматриваемая задача является одной из типичных задач управления транспортными потоками на железной дороге, т.е. задачей построения оптимального расписания движения состава на участке с жестким ограничением на пропускную способность путей (так называемая задача об «узких местах», или «bottleneck»).

1. Общее описание проблемы

Рассмотрим постановку задачи поиска оптимального расписания движения поездов. Две станции соединены однопутной железной дорогой. Имеется два множества поездов, N_1 и N_2 . Поезда из множества N_1 следуют со станции 1 на станцию 2, поезда из N_2 следуют в обратном направлении со станции 2 на станцию 1. Между станциями находится разъезд для пропуска встречных поездов. В разъезде есть главный путь для безостановочного пропуска поездов и один дополнительный путь (или несколько) для пропуска встречных поездов. Время прохождения разъез-

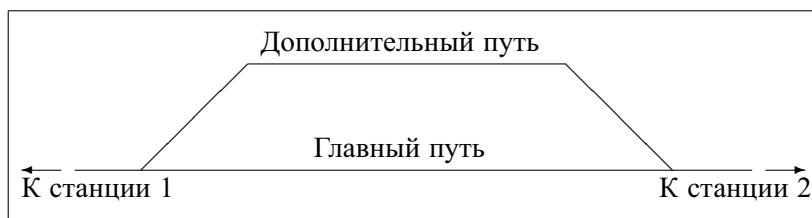


Рис. 1. Схема разъезда

да поездами много меньше времени, за которое поезда проходят путь между станциями 1 и 2. Необходимо составить расписание, т.е. для каждого поезда из N_1 и N_2 указать:

- время отправления;
- время стоянки в разъезде (если оно есть).

Будем рассматривать задачу с минимизацией целевой функции времени окончания перевозок C_{max} («makespan») при одновременном поступлении поездов на станции.

В соответствии с трехпозиционной системой обозначений, принятой в теории расписаний [6], для данной задачи можно ввести краткое обозначение $STRSP2-siding||C_{max}$, где STRSP2 является сокращением от «Single Track Railway Scheduling Problem with two stations» (такое обозначение было предложено в [5]).

2. Математическая постановка задачи

Исходными данным для задачи в общем случае являются параметры железной дороги и поездов:

- на пути между станциями находится один разъезд, вмещающий один поезд при пропуске встречных поездов;
- скорость движения поездов одинакова и постоянна;
- интервал безопасности для любых одновременно находящихся в движении поездов $\tau > 0$ (минимальное время между отправлением двух поездов с одной станции и между прибытием к разъезду двух поездов с разных станций);
- время прохождения поездами участков пути между станциями слева и справа от разъезда – p_1 и p_2 , без потери общности будем считать $p_1 \geq p_2 > 0$;
- число поездов n_1 множества N_1 и n_2 множества N_2 .

Необходимо составить оптимальное расписание движения поездов, т.е. указать:

- время отправления поезда номер i с s -й станции S_i^s , $i = \overline{1, n_s}$, $s \in \{1, 2\}$;
- время стоянки поезда i со станции s в разъезде k_i^s , $i = \overline{1, n_s}$, $s \in \{1, 2\}$.

Однопутная железная дорога разделена разъездом на два отрезка. Поезда из множества N_1 проходят их в порядке $1 \rightarrow 2$, поезда из N_2 в порядке $2 \rightarrow 1$. Если на отрезке находится поезд из N_1 , то ни один из поездов из N_2 не может находиться на данном отрезке, и наоборот.

Обозначим время старта поезда номер i с s -й станции как S_i^s , а время прибытия поезда номер i с s -й станции – C_i^s , тогда:

$$(1) \quad C_i^s = S_i^s + p_1 + k_i^s + p_2, \quad i = \overline{1, n_s}, \quad s \in \{1, 2\}.$$

Целевая функция – время окончания перевозок, которое определяется по формуле

$$(2) \quad C_{max} = \max_{i=1, n_s, s \in \{1, 2\}} \{C_i^s\}.$$

3. Существующие методы решения

Обзор публикаций по моделям и методам железнодорожного планирования, в том числе по моделям с однопутными железными дорогами, был рассмотрен Ласби с соавторами в 2011 г. [10]. Кроме планирования движения на однопутных дорогах, в работе описаны другие типы задач – распределение по платформам, организация движения в сетях различного типа. В 2011 г. Хэррод [7] классифицировал модели по типу расписания (периодическое или аperiodическое) и по способу представления железнодорожной сети (явно или неявно).

Задачи планирования движения поездов на однопутной железной дороге из-за практической значимости и сложной математической природы являются предметом интенсивных исследований. Одна из первых работ по данной проблеме – работа Шпигеля, опубликованная в 1973 г. [14]. Шпигель рассмотрел задачу планирования движения поездов по однокорейной железной дороге с возможностью обгона на станциях. Он первым отметил сходство между задачей планирования движения поездов и задачей теории расписаний для нескольких приборов, рассматривая участки пути как «приборы», а поезда – как «работы».

Хиггинс с соавторами в 1996 г. описали метод ветвей и границ для однопутных линий железных дорог, похожий на метод Шпигеля [8]. Брэннлунд с соавторами в 1998 г. рассмотрели модель однокорейной дороги, на сегментах которой установлены ограничения на пропускную способность, а поезда движутся с постоянной скоростью [2]. Кэри и Локвуд свели задачу планирования движения для небольшой сети железных дорог к задаче целочисленного программирования с бинарными переменными [4]. Для решения авторы предложили эвристический метод, который последовательно упорядочивает поезда. В отличие от постановки

Шпигеля, в задаче были учтены порядковые ограничения.

В работе Оливейры [12] представлен литературный обзор по задачам планирования движения на однопутных линиях и метод программирования в ограничениях для их решения. В публикации китайских ученых [15] исследуется задача составления оптимального расписания на однопутной железной дороге с минимизацией целевой функции общего времени прохождения пути. В этой работе ставится обобщенная задача, учитывающая пропускную способность станций как ограниченных ресурсов, и предлагается метод ветвей и границ для получения допустимых расписаний с гарантированной точностью.

Исследование задачи составления оптимального расписания поездов с равными и постоянными скоростями на однопутной линии между двумя станциями было представлено в [5]. В работе задача сводится к одной из уже достаточно изученных задач в теории расписаний – к задаче одного прибора. Поезда рассмотрены как «работы», а путь – как «прибор». Между обслуживанием требований разного типа (т.е. поездов с разных станций) должна проводиться «настройка» прибора. В модели присутствуют множества требований разного типа с равными временами выполнения, поэтому при таком представлении задачи возможно применить уже разработанные методы («batching»), описанные в работах Монма и Поттс [11], Кравченко и Вернера [9]. Задача составления оптимального расписания поездов на однопутной железной дороге может быть также представлена как задача цеха с заданным отношением предшествования между операциями («job-shop scheduling») [3]. В работе Сотскова и Голами [13] для случая с несколькими станциями задача представлена как задача цеха и для нее предложен эвристический алгоритм.

Можно сделать вывод о том, что задачи составления оптимального расписания на однопутной железной дороге при исследовании в основном представляются как уже изученные задачи теории расписаний, для которых разработаны методы решения: динамическое программирование, метод ветвей и границ, «batching», эвристические алгоритмы. Для модели, исследуемой

в данной работе, еще не было предложено точных алгоритмов решения. При определенных условиях полученные в работе алгоритмы можно сравнить с существующими эвристическими алгоритмами для более обобщенных моделей, например для сети из нескольких станций.

4. Исследование проблемы

4.1. СВОЙСТВА РАСПИСАНИЙ МОДЕЛИ

Определение 1. Назовем поезд i со станции s ($i = \overline{1, n_s}$, $s \in \{1, 2\}$) активным в момент времени t , если в этот момент он либо движется между станциями с постоянной скоростью, либо находится в разъезде, т.е. справедливо

$$(3) \quad S_i^s \leq t < C_i^s.$$

Расписание должно соответствовать ограничениям модели. Сформулируем некоторые свойства расписания для данной модели.

- Одновременно на каждом отрезке пути (справа или слева от разъезда) не могут находиться два активных поезда с разных станций. Это следует из модели – пропуск встречных поездов происходит только в разъезде. На одном сегменте два встречных поезда не могут разойтись.
- Одновременно на всем пути могут находиться A активных поездов со станции 1 и B активных поездов со станции 2, при условии, что выполняется неравенство $\min\{A, B\} \leq 1$.
- Для нашей модели будем предполагать, что

$$0 < \tau \leq \min\{p_1, p_2\}.$$

- Расстояние между активными поездами не может быть меньше расстояния, которое поезд проходит за минимальный интервал безопасности – время $\tau > 0$. Он вводится из-за значительной длины грузового поезда, которую необходимо учитывать (поезда не могут двигаться вплотную один

за другим), и для разрешения экстренных ситуаций при непредвиденной остановке поезда.

Определение 2. Будем называть расписания с описанными выше свойствами допустимым расписанием.

4.2. СВОЙСТВА ОПТИМАЛЬНОГО РАСПИСАНИЯ

Определим необходимые свойства оптимального расписания для данной модели.

- Расписание должно быть допустимым.
- При оптимальном расписании нет искусственных перерывов в движении поездов. Для любой регулярной функции в случае добавления перерыва в движении целевая функция только возрастает. Это означает, что при оптимальном расписании в любой момент t при $t \in [0, C_{max})$ число активных поездов больше нуля.
- При оптимальном расписании при остановке в разъезде поезд пропускает как минимум один встречный состав. Остановка в разъезде без пропуска встречных поездов не имеет смысла и не уменьшит значение целевой функции.
- При оптимальном расписании поезд, пропускающий в разъезде один или несколько встречных поездов, покидает разъезд сразу после того как последний из пропускаемых поездов пройдет его.

Определение 3. Множество расписаний, удовлетворяющих необходимым свойствам оптимального расписания, назовем регулярными расписаниями.

Лемма 1. При любом регулярном расписании время завершения перевозок не больше чем $(p_1 + p_2)n$, где $n = n_1 + n_2$.

Доказательство. Рассмотрим случай, когда $n_1 = n_2 = n/2$ или значения n_1 и n_2 отличаются на 1. В таком случае возможно, что все поезда будут стартовать в момент прибытия поезда с другой станции (и в нулевой момент – условие для первого поезда),

т.е. поочередно с разных станций. Очевидно, что при расписании с максимальным временем завершения перевозок поезда с разных станций будут начинать движение по очереди (см. рис. 2).

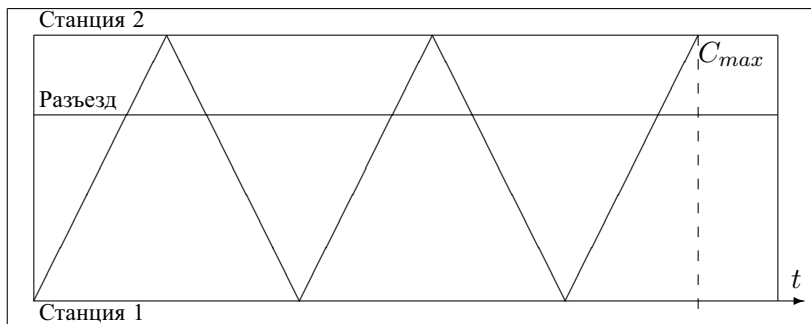


Рис. 2. График движения при регулярном расписании с максимальным временем окончания перевозок для $n_1 = 3$ и $n_2 = 2$

Тогда время завершения перевозок будет равно

$$(4) \quad C_{max} = (n_1 + n_2)(p_1 + p_2) = n(p_1 + p_2).$$

В случае, если $n_1 \neq n_2$, при регулярном расписании с максимальным временем окончания перевозок часть поездов будут двигаться с минимальным интервалом (см. рис. 3).

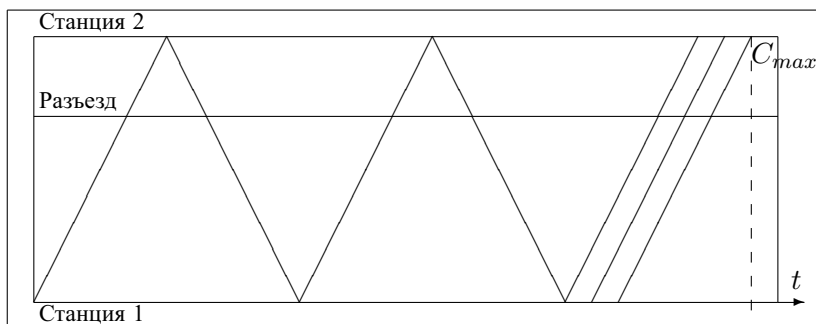


Рис. 3. График движения при регулярном расписании с максимальным временем окончания перевозок для $n_1 = 5$ и $n_2 = 2$

В таком случае C_{max} будет равен

$$(5) \quad C_{max} = (p_1 + p_2)n^* + \tau n^{**},$$

где n^* – число поездов, идущих поочередно с разных станций, n^{**} – число поездов, стартующих с одной станции через минимальный интервал, $n^* + n^{**} = n$. По условию $\tau \leq \min\{p_1, p_2\}$, т.е. при любых n справедливо $C_{max} \leq (p_1 + p_2)n$.

Последовательный старт с одной станции нескольких поездов через минимальный интервал, так же как и использование разъезда, только уменьшит C_{max} . Оценка сверху для C_{max} достигается, например, в случае, который представлен на рис. 3.

Отметим, что при любом регулярном расписании с произвольными n_1 и n_2 поездами на станциях в разъезде делают остановку максимум $\min\{n_1, n_2\}$ поездов. При регулярном расписании все поезда, делающие остановку в разъезде, пропускают как минимум один встречный состав. Это значит, что необходимо составить пары из поездов с разных станций (пропускающий и проходящий). Число таких пар может быть не больше $\min\{n_1, n_2\}$.

4.3. ВЫВОД ОПТИМАЛЬНОГО РЕШЕНИЯ

Из определения регулярного расписания и допустимого расписания можно сделать вывод, что при регулярном расписании количество ситуаций с различным числом активных поездов с разных станций конечно. Необходимо оценить количество регулярных расписаний.

При регулярном расписании не определен порядок, в котором поезда покидают станции. Обозначим множества поездов со станций N_1 и N_2 как

$$\{a_1, a_2, \dots, a_{n_1}\}$$

и

$$\{b_1, b_2, \dots, b_{n_2}\}.$$

Если не использован разъезд, то количество возможных регулярных расписаний равно числу перестановок порядка $n = n_1 + n_2$.

Число перестановок множества из $n = n_1 + n_2$ элементов (в произвольном порядке) определяется формулой

$$(6) \quad m = \frac{(n_1 + n_2)!}{n_1!n_2!}.$$

Как следует из определения 2, $\min\{A, B\} \leq 1$, поэтому рассмотрим четыре возможных случая:

- активны A поездов со станции 1 и нет ни одного активного поезда со станции 2;
- активны B поездов со станции 2 и нет ни одного активного поезда со станции 1;
- активны A поездов со станции 1 и один поезд со станции 2;
- активны B поездов со станции 2 и один поезд со станции 1.

В последних двух вариантах единственный со своей станции поезд пропускает встречные поезда с другой станции.

Любое регулярное расписание представляет собой набор из последовательностей сегментов, соответствующих одной из этих четырех ситуаций. Для нахождения решения разберем подзадачи, оптимальные решения которых включают в себя каждый из 4 сегментов. Затем опишем правила их комбинации и построения порядка решения. Далее необходимо выяснить, сколько поездов и с каких станций сделают остановку в разъезде при оптимальном расписании.

Описанные далее подзадачи по постановке отличаются от исходной задачи только числом поездов на станциях в начальный момент.

4.3.1. ПОДЗАДАЧА $A(M)$

На станции 1 находятся m поездов, на станции 2 нет ни одного поезда.

Очевидно, что в оптимальном решении поезда со станции 1 идут подряд с начального момента $t = 0$ с минимальным интервалом τ .

$$(7) \quad C_{\max}[A(m)] = C_m^1 = p_1 + p_2 + (m - 1)\tau.$$

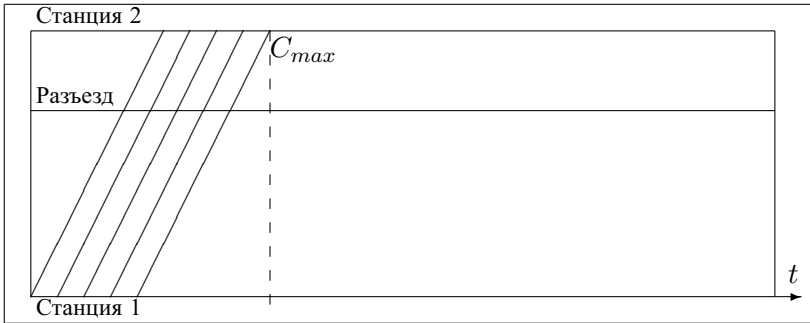


Рис. 4. График движения поездов при оптимальном расписании для подзадачи $A(m)$ при $m = 5$

4.3.2. ПОДЗАДАЧА $B(M)$

Подзадача $B(m)$ аналогична подзадаче $A(m)$: на второй станции m поездов, а на первой станции нет ни одного поезда. В оптимальном решении поезда со станции 2 идут подряд с начального момента $t = 0$ с минимальным интервалом τ .

$$(8) \quad C_{max}[B(m)] = C_m^2 = p_1 + p_2 + (m - 1)\tau.$$

Очевидно, что задачу минимизации времени окончания перевозок при одновременном поступлении поездов с параметрами $n_1 = l$, $n_2 = m$ можно разделить на две подзадачи $A(l)$, $B(m)$, выполняемые в произвольном порядке, однако при этом не будет использоваться разъезд.

4.3.3. ПОДЗАДАЧА $E1(M)$

Количество поездов на станции 2 равно $n_2 = m \geq 1$, а на станции 1 находится 1 поезд, т. е. $n_1 = 1$. Обозначим данную подзадачу как $E1(m)$.

Решение данной задачи не так очевидно, как в подзадачах $A(m)$ и $B(m)$. Если перебрать все возможные варианты регулярных расписаний, то можно определить общую структуру оптимального расписания для данной задачи.

Возможны три варианта:

- в разъезде не останавливается ни один из поездов;
- в разъезде останавливается поезд со станции 2;
- в разъезде останавливается поезд со станции 1.

Для данной задачи нет других вариантов (в разъезд заходит только один поезд), так как $\min\{n_1, n_2\} = 1$, т.е. в разъезде будет останавливаться для пропуска не более чем один поезд.

Рассмотрим первый вариант. Обозначим множества поездов со станций N_1 и N_2 как $\{a_1\}$ и $\{b_1, b_2, \dots, b_{n_2}\}$. Существует $n_2 + 1$ способов построения порядка движения поездов:

$$\{a_1, b_1, b_2, \dots, b_{n_2}\};$$

$$\{b_1, a_1, b_2, \dots, b_{n_2}\};$$

...

$$\{b_1, b_2, \dots, b_{n_2}, a_1\}.$$

Для случаев $\{a_1, b_1, b_2, \dots, b_{n_2}\}$ и $\{b_1, b_2, \dots, b_{n_2}, a_1\}$ имеем

$$(9) \quad C_{max} = 2(p_1 + p_2) + (n_2 - 1)\tau.$$

В остальных случаях

$$(10) \quad C_{max} = 3(p_1 + p_2) + (n_2 - 1)\tau.$$

Во втором варианте поезд со станции 2 пропускает в разъезде поезд со станции 1. Существует n_2 способов построения порядка движения: в разъезд заходит первый из поездов, идущих со станции 2, второй, ..., n_2 -й. Если в разъезд заходит первый из поездов, идущих со станции 2, то при условии $\tau \leq p_1 - p_2$ поезд со станции 1 начинает движение в начальный момент (см. рис. 5)

$$(11) \quad C_{max} = 2(p_1 + p_2) + (n_2 - 2)\tau.$$

При $\tau \geq p_1 - p_2$ поезд со станции 1 начинает движение в момент $t = p_2 + \tau - p_1$ (см. рис. 6)

$$(12) \quad C_{max} = (2p_2 + \tau) + (p_1 + p_2) + (n_2 - 2)\tau.$$

Если в разъезд заходит второй из поездов, идущих со станции 2

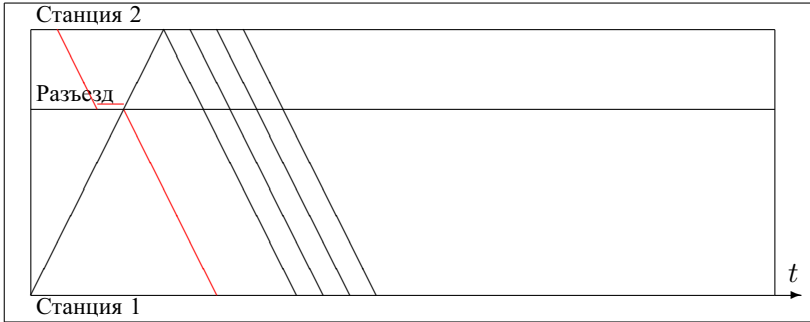


Рис. 5. График движения при втором варианте структуры расписания для подзадачи E1(5) при $\tau \leq p_1 - p_2$, в разъезд заходит первый поезд со станции 2

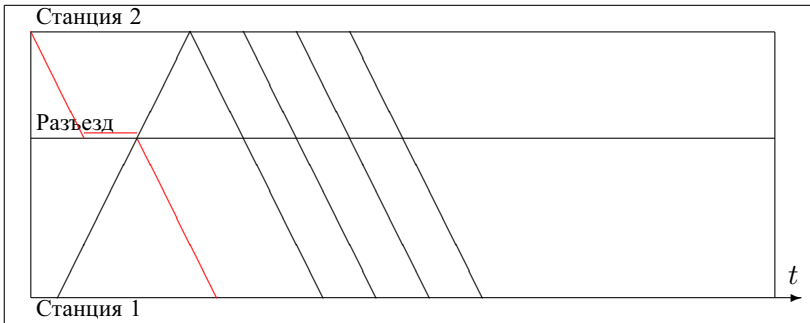


Рис. 6. График движения при втором варианте структуры расписания для подзадачи E1(5) при $\tau \geq p_1 - p_2$, в разъезд заходит первый поезд со станции 2

(см. рис. 7), то

$$(13) \quad C_{max} = 3(p_1 + p_2) + (n_2 - 2)\tau.$$

Данная формула также справедлива, если в разъезд заходит третий, четвертый, ..., $n_2 - 1$ из поездов, идущих со станции 2.

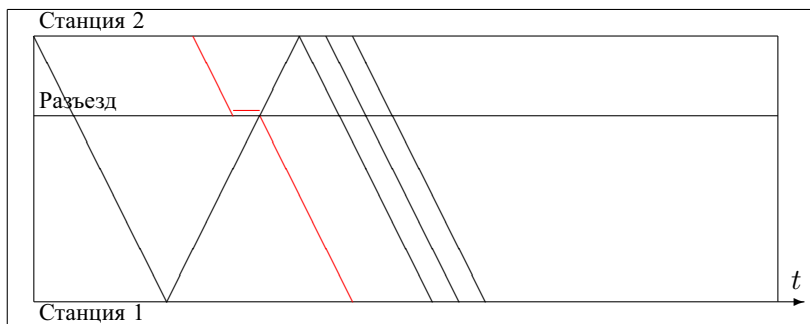


Рис. 7. График движения при втором варианте структуры расписания для подзадачи E1(5), в разъезд заходит второй поезд со станции 2

Если в разъезд заходит последний из поездов, идущих со станции 2 (под номером n_2), то

$$(14) \quad C_{max} = p_1 + p_2 + (n_2 - 2)\tau + 2p_1.$$

В третьем варианте поезд со станции 1 пропускает в разъезде поезда со станции 2 (один или несколько). Можно выделить три случая:

- поезд со станции 1 начинает движение, когда ни один из поездов со станции 2 еще не прибыл на станцию 1 (см. рис. 8);
- поезд со станции 1 начинает движение, когда несколько поездов со станции 2 уже прибыли на станцию 1, и пропускает все оставшиеся поезда со станции 2 в разъезде;
- поезд со станции 1 начинает движение, когда несколько поездов со станции 2 уже прибыли на станцию 1, и пропускает не все оставшиеся поезда со станции 2 в разъезде.

В первом случае поезд со станции 1 может пропустить не все из n_2 поездов со станции 2 в разъезде (см. рис. 8). Отметим, что

в таком случае для любого числа пропускаемых поездов (от 1 до $n_2 - 1$) формула будет одинаковой:

$$(15) \quad C_{max} = 2(p_1 + p_2) + (n_2 - 1)\tau.$$

Если поезд со станции 1 пропускает все (см рис. 9) поезда со станции 2 в первом варианте, то

$$(16) \quad C_{max} = 2p_1 + n_2\tau.$$

Во втором случае поезд со станции 1 начинает движение, когда

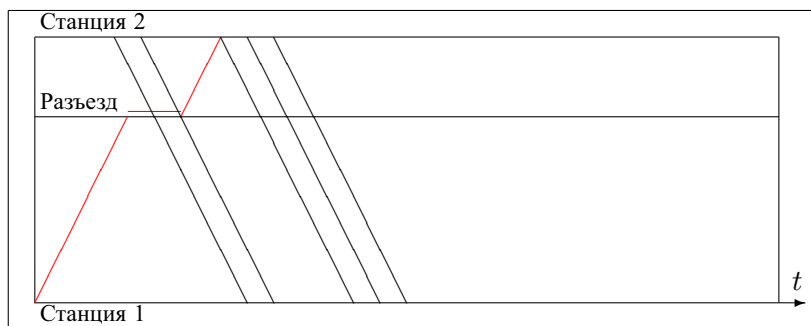


Рис. 8. График движения при третьем варианте структуры расписания для подзадачи E1(5), поезд со станции 1 начинает движение, когда ни один из поездов со станции 2 еще не прибыл на станцию 1

несколько поездов со станции 2 уже прибыли на станцию 1, и пропускает все оставшиеся поезда со станции 2 в разъезде. При такой структуре расписания для любого допустимого числа пропускаемых поездов справедлива формула

$$(17) \quad C_{max} = p_1 + p_2 + (n_2 - 2)\tau + 2p_1 + \tau.$$

В третьем случае поезд со станции 1 начинает движение, когда несколько поездов со станции 2 уже прибыли на станцию 1, и пропускает не все оставшиеся поезда со станции 2 в разъезде. Для любого допустимого числа пропускаемых поездов имеем

$$(18) \quad C_{max} = p_1 + p_2 + (n_2 - 2)\tau + 2(p_1 + p_2).$$

Сравнив полученные формулы (9)–(18), можно найти среди них решение – вариант с минимальным значением C_{max} . При оптимальном расписании для подзадачи E1(m) (см. рис. 9) поезд со

станции 1 начинает движение в нулевой момент времени, заходит в разъезд и пропускает все встречные поезда со станции 2, целевая функция определяется формулой

$$(19) \quad C_{max}[E1(m)] = 2p_1 + m\tau.$$

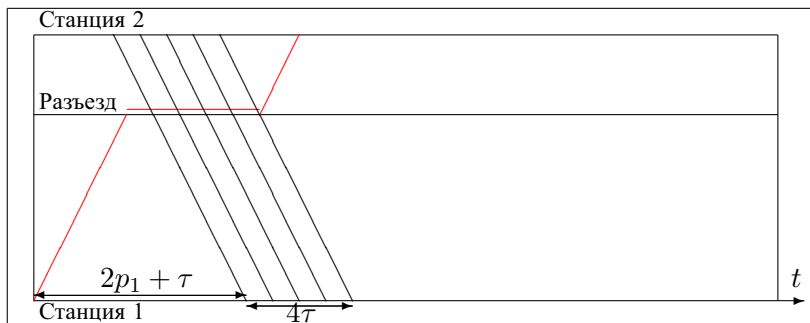


Рис. 9. График движения поездов при оптимальном расписании для подзадачи $E1(5)$

4.3.4. ПОДЗАДАЧА $E2(M)$

Необходимо решить задачу минимизации времени окончания перевозок при одновременном поступлении поездов с одним разъездом, вмещающим один поезд. Количество поездов на станции 1 в начальный момент равно $n_1 = m \geq 1$, на станции 2 находится один поезд. Время прохождения поездом отрезков слева и справа от разъезда равно p_1 и p_2 , $p_1 \geq p_2$. Обозначим данную задачу как подзадачу $E2(m)$.

Общую структуру оптимального расписания возможно найти, рассмотрев все варианты регулярных расписаний для данной подзадачи. Как и для подзадачи $E1(m)$, при оптимальном расписании для данной подзадачи один поезд со станции 2 пропускает в разъезде все встречные поезда со станции 1. При этом возможны две ситуации, когда первым начинает движение поезд со станции 2 или поезд со станции 1. Поезд со станции 2 должен пробыть в разъезде как минимум время τ до того, как первый

поезд со станции 1 пройдет разъезд. Допустимый момент начала движения первого поезда со станции 1 определяется из неравенства

$$(20) \quad S_1^1 + p_1 \geq p_2 + \tau,$$

т.е. момент начала движения первого поезда со станции 1 равен

$$(21) \quad S_1^1 = \max\{0, p_1 - (p_2 + \tau)\}.$$

Если $\tau \leq p_1 - p_2$, то при оптимальном расписании $S_1^1 = 0$ (см. рис. 10). Время окончания перевозок для подзадачи $E2(m)$ определяется формулой

$$(22) \quad C_{max}[E2(m)]^1 = 2p_1 + m\tau.$$

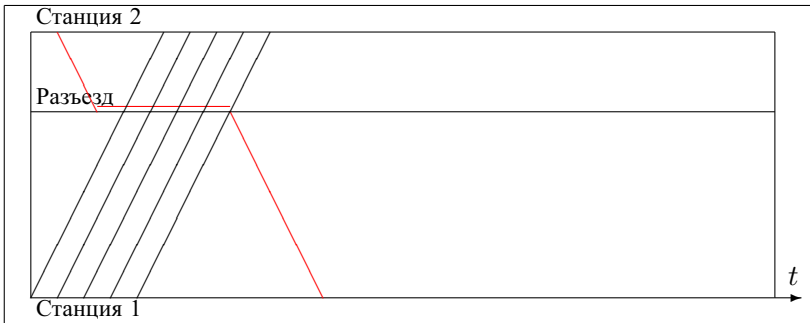


Рис. 10. Оптимальное решение подзадачи $E2(5)$ при $\tau \leq p_1 - p_2$

Если $\tau \geq p_1 - p_2$, то при оптимальном расписании $S_1^1 = p_1 - p_2 + \tau$ (см. рис. 11). Время окончания перевозок для подзадачи $E2(m)$ определяется формулой

$$(23) \quad C_{max}[E2(m)]^2 = p_1 + m\tau + p_2.$$

Значения целевой функции берутся как разница между началом движения первого поезда и прибытием последнего поезда на станцию назначения.

4.3.5. ПРАВИЛА КОМБИНАЦИИ ПОДЗАДАЧ

Лемма 2. Исходная задача минимизации времени окончания перевозок при одновременном поступлении n_1 и n_2 поездов и одним разъездом может быть представлена как комбинация:

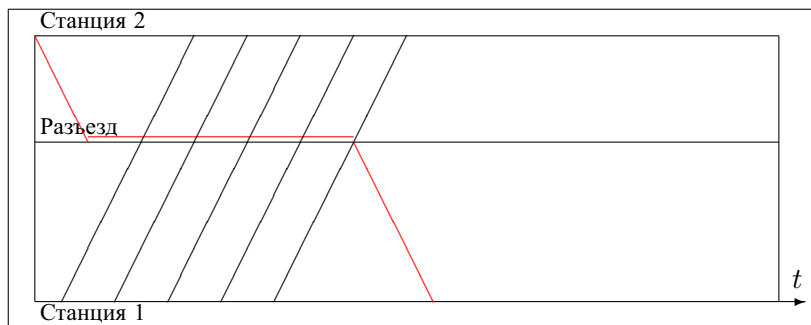


Рис. 11. Оптимальное решение подзадачи E2(5) при $\tau \geq p_1 - p_2$

- k_A подзадач $A(m_i^A)$, $i = 1, \dots, k_A$;
- k_B подзадач $B(m_i^B)$, $i = 1, \dots, k_B$;
- k_{E1} подзадач $E1(m_i^{E1})$, $i = 1, \dots, k_{E1}$;
- k_{E2} подзадач $E2(m_i^{E2})$, $i = 1, \dots, k_{E2}$;

при условии, что

$$(24) \quad n_1 = \sum_{i=1}^{k_A} m_i^A + k_{E1} + \sum_{i=1}^{k_{E2}} m_i^{E2};$$

$$(25) \quad n_2 = \sum_{i=1}^{k_B} m_i^B + k_{E2} + \sum_{i=1}^{k_{E1}} m_i^{E1}.$$

Доказательство. Комбинации подзадач, решаемых в определенном порядке, позволяют получить все возможные регулярные расписания, в том числе оптимальное решение исходной задачи. Это можно объяснить тем, что при регулярном расписании поезда могут идти либо без остановок в одну сторону с минимальным интервалом, что соответствует решениям подзадач $A(m_i^A)$, $i = 1, \dots, k_A$, и $B(m_i^B)$, $i = 1, \dots, k_B$, либо поезд со станции 1 (или 2) пропускает встречные m_i^{E1} (m_i^{E2}) поездов, что соответствует решениям подзадач $E1(m_i^{E1})$, $i = 1, \dots, k_{E1}$, и $E2(m_i^{E2})$, $i = 1, \dots, k_{E2}$. Условие в лемме связывает параметры подзадач m и параметры исходной задачи. В исходной задаче на станции 1 находятся n_1 поездов, на станции 2 – n_2 поездов.

- После решения k_A подзадач $A(m_i^A)$, $i = 1, \dots, k_A$, $\sum_{i=1}^{k_A} m_i^A$ поездов со станции 1 достигнут станции 2.
- После решения k_B подзадач $B(m_i^B)$, $i = 1, \dots, k_B$, $\sum_{i=1}^{k_B} m_i^B$ поездов со станции 2 достигнут станции 1.
- После решения k_{E1} подзадач $E1(m_i^{E1})$, $i = 1, \dots, k_{E1}$, k_{E1} поездов со станции 1 достигнут станции 2, а $\sum_{i=1}^{k_{E1}} m_i^{E1}$ поездов со станции 2 достигнут станции 1.
- После решения k_{E2} подзадач $E2(m_i^{E2})$, $i = 1, \dots, k_{E2}$, k_{E2} поездов со станции 2 достигнут станции 1, а $\sum_{i=1}^{k_{E2}} m_i^{E2}$ поездов со станции 1 достигнут станции 2.

При выполнении условия в лемме комбинация подзадач и исходная задача согласованы по постановке – число поездов на станциях в начальный момент совпадает.

Исходная задача минимизации времени окончания перевозок при одновременном поступлении поездов и одним разъездом может быть разделена на комбинацию подзадач A , B , $E1$ и $E2$, решаемых в определенном порядке. Решение очередной подзадачи начинается в тот момент, когда все поезда предыдущей задачи либо достигли станций назначения, либо уже в пути. Для одной и той же задачи можно указать несколько вариантов комбинации подзадач. Возможные расписания можно представить в виде упорядоченного набора подзадач, в общем виде без конкретных значений аргументов обозначим его как

$$\{A, B, A, A, E1, E2, B, B, E1, \dots\}.$$

Порядок выполнения подзадач влияет на значение функции C_{max} , так как у данных подзадач существуют моменты $t < C_{max}$, в которые возможно начать решение следующей по порядку подзадачи. Для нахождения оптимального порядка необходимо сравнить значения функций C_{max} для выполняемых подряд подзадач,

и выделить правила оптимального порядка выполнения подзадач A , B , $E1$ и $E2$.

Лемма 3. Для получения оптимального расписания с минимальным значением C_{max} порядок решаемых подряд подзадач $A(l)$ и $B(m)$ не важен. Для двух возможных вариантов $\{A(l), B(m)\}$ или $\{B(m), A(l)\}$ целевая функция C_{max} будет одинаковой.

Доказательство. Для пары подзадач $A(l)$ и $B(m)$ решение следующей подзадачи всегда начинается в момент окончания решения предыдущей, а целевая функция C_{max} равна сумме C_{max} каждой отдельной подзадачи. Если подзадача $A(l)$ решается первой (см. рис. 12), т.е. начиная с момента $t^1 = 0$, то решение подзадачи $B(m)$ начнется в момент $t^2 = C_{max}[A(l)]$. Целевая функция при таком порядке:

$$C_{max} = C_{max}[A(l)] + C_{max}[B(m)] = p_1 + p_2 + (l - 1)\tau + p_1 + p_2 + (m - 1)\tau.$$

При обратном порядке (см. рис. 13):

$$C_{max} = C_{max}[B(m)] + C_{max}[A(l)] = p_1 + p_2 + (m - 1)\tau + p_1 + p_2 + (l - 1)\tau.$$

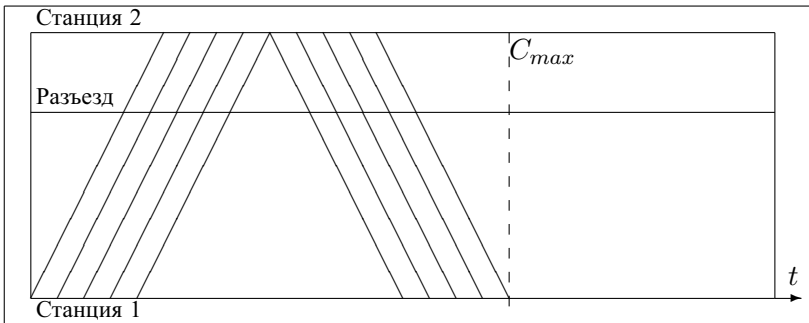


Рис. 12. График движения при последовательном выполнении подзадач $\{A(l), B(m)\}$ при $l = 5$ и $m = 5$

Лемма 4. Для получения оптимального расписания с минимальным значением C_{max} решаемые подряд подзадачи $A(l)$ и $A(m)$ сводятся к подзадаче $A(l + m)$.

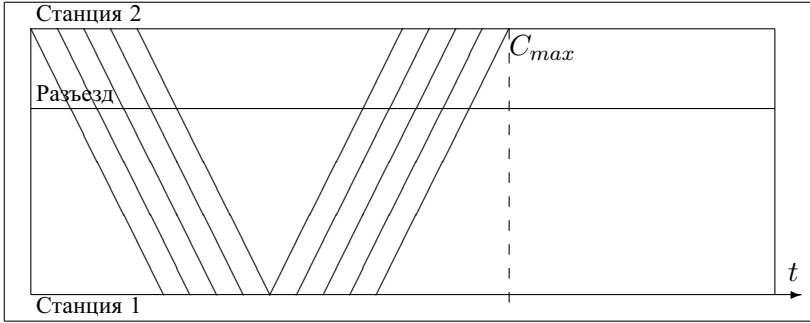


Рис. 13. График движения при последовательном выполнении подзадач $\{B(m), A(l)\}$ при $l = 5$ и $m = 5$

Доказательство. Для доказательства леммы достаточно показать, что если решать подзадачи $A(l)$ и $A(m)$ подряд (см. рис. 14), то время окончания перевозок будет равно сумме C_{max} каждой из подзадач, т.е. $C_{max}[A(l)] + C_{max}[A(m)]$. Подзадача $A(l + m)$ (см. рис. 15) по постановке эквивалентна комбинации этих подзадач, но имеет $C_{max}[A(l + m)] < C_{max}[A(l)] + C_{max}[A(m)]$:

- $C_{max}[A(l + m)] = p_1 + p_2 + (l + m - 1)\tau$;
- $C_{max}[A(l)] + C_{max}[A(m)] = 2p_1 + 2p_2 + (l + m - 2)\tau$.

Лемма 5. Для получения оптимального расписания с минимальным значением C_{max} решаемые подряд подзадачи $B(l)$ и $B(m)$ сводятся к подзадаче $B(l + m)$.

Доказательство. Лемма аналогично доказывается из неравенства

$$C_{max}[B(l + m)] < C_{max}[B(l)] + C_{max}[B(m)]:$$

- $C_{max}[B(l + m)] = p_1 + p_2 + (l + m - 1)\tau$;
- $C_{max}[B(l)] + C_{max}[B(m)] = 2p_1 + 2p_2 + (l + m - 2)\tau$.

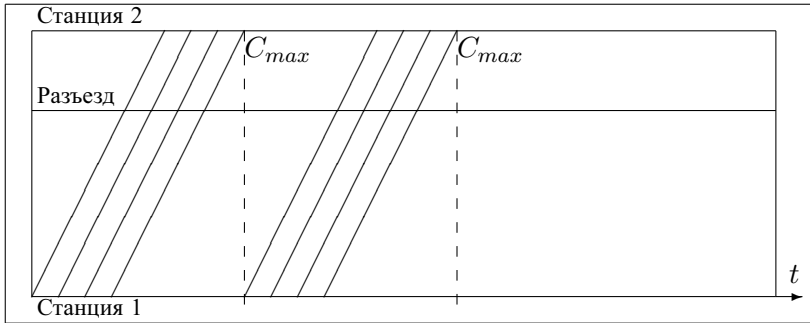


Рис. 14. График движения при последовательном выполнении подзадач $\{A(l), A(m)\}$ при $l = 4$ и $m = 4$

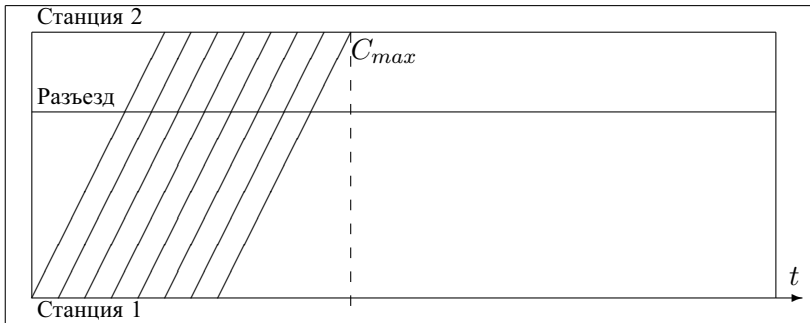


Рис. 15. График движения при выполнении подзадачи $A(l + m)$ при $l = 4$ и $m = 4$

Лемма 6. При оптимальном расписании с минимальным значением C_{max} неважен порядок решаемых подряд пар подзадач $E1(l)$ и $E1(m)$, $E2(l)$ и $E2(m)$. Целевая функция C_{max} будет одинаковой в каждой паре возможных вариантов порядка выполнения:

- $\{E1(l), E1(m)\}$ и $\{E1(m), E1(l)\}$;
- $\{E2(l), E2(m)\}$ и $\{E2(m), E2(l)\}$.

Доказательство. Порядок выполнения подзадач $E2(l)$ или $E1(m)$ с любыми l и m не имеет значения, так как для любой из

двух этих подзадач нет моментов $t < C_{max}$, в которые можно начать выполнение следующей подзадачи $E2(l)$ или $E1(m)$. Чтобы это доказать, рассмотрим алгоритм движения поездов в данных подзадачах. В каждой из них в начальный момент со станции 1 начинает движение поезд, а в момент окончания перевозок на станцию 1 приходит последний поезд со станции 2. На рис. 16 показан график движения при оптимальном решении для подзадачи $E1$ и отмечены моменты завершения предыдущей подзадачи и начала следующей подзадачи типа $E1$.

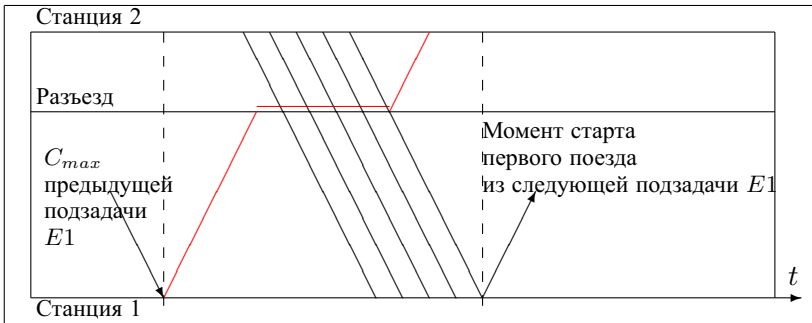


Рис. 16. График движения при оптимальном расписании для подзадачи $E1(5)$

Итак, для каждой пары вариантов справедливы выражения:

$$C_{max}[E1(l), E1(m)] = C_{max}[E1(m), E1(l)],$$

$$C_{max}[E2(l), E2(m)] = C_{max}[E2(m), E2(l)].$$

Лемма 7. Если $\tau \geq p_1 - p_2$, то при оптимальном расписании с минимальным значением C_{max} решаемые подряд подзадачи $E1(l)$ и $E2(m)$ должны выполняться в порядке $\{E1(l), E2(m)\}$. Если $\tau \leq p_1 - p_2$, то порядок последовательного выполнения подзадач $E1(l)$ и $E2(m)$ не влияет на C_{max} , т.е. оба варианта $\{E1(l), E2(m)\}$ и $\{E2(m), E1(l)\}$ имеют одинаковые значения целевой функции.

Доказательство. Если $\tau \geq p_1 - p_2$, то при оптимальном расписании для подзадачи $E2(l)$ поезд со станции 2 начинает движение раньше, чем первый поезд со станции 1 ($S_1^1 = p_2 + \tau - p_1$). Предположим, что подзадачи выполняются в порядке $\{E2(m), E1(l)\}$. Выполнение подзадачи начинается в момент прибытия поезда со станции 2, т.е. в момент завершения выполнения подзадачи $E2(m)$. Общее время окончания перевозок равно сумме C_{max} отдельных задач $E1(l)$ и $E2(m)$.

(26)

$C_{max}[\{E2(m), E1(l)\}] = p_1 + p_2 + m\tau + 2p_1 + l\tau = 3p_1 + p_2 + (l+m)\tau$. Если подзадачи выполняются в порядке $\{E1(l), E2(m)\}$, тогда выполнение подзадачи $E2(m)$ начинается в момент прибытия поезда со станции 1 из подзадачи $E1(l)$, т.е. раньше момента завершения выполнения подзадачи $E1(l)$. Общее время окончания перевозок равно времени прибытия последнего поезда со станции 2 в подзадаче $E2(m)$:

$$(27) \quad C_{max}[\{E1(l), E2(m)\}] =$$

$$p_1 + l\tau + p_2 + p_2 + m\tau + p_1 = 2p_1 + 2p_2 + (l+m)\tau.$$

Так как $p_1 \geq p_2$, то для любых l и m

$$(28) \quad C_{max}[\{E1(l), E2(m)\}] < C_{max}[\{E2(m), E1(l)\}].$$

При выполнении условия $\tau \leq p_1 - p_2$ время окончания перевозок будет равно сумме C_{max} отдельных задач $E1(l)$ и $E2(m)$ при любом порядке. В предыдущих леммах были описаны правила построения оптимального порядка последовательного выполнения пар подзадач типа A и B , $E1$ и $E2$. Теперь определим оптимальный порядок последовательного выполнения пар подзадач A , B и $E1$, $E2$.

Лемма 8. Для получения оптимального расписания с минимальным значением C_{max} решаемые подряд подзадачи $\{A(k), E2(m)\}$ и $\{E2(m), A(k)\}$ сводятся к подзадаче $E2(k+m)$.

Доказательство. Комбинация подзадач $E2(m)$ и $A(k)$ – задача с поступлением $k+m$ поездов на станцию 1 и одного поезда на станцию 2, т.е. по постановке это подзадача $E2(k+m)$, для которой уже был определен вид оптимального решения.

Справедливо $C_{max}[E2(k + m)] < C_{max}[\{A(k), E2(m)\}]$ (см. рис. 17 и рис. 18):

- $C_{max}[E2(k + m)] = 2p_1 + (k + m)\tau$;
- $C_{max}[\{A(k), E2(m)\}] = p_1 + p_2 + (k - 1)\tau + p_2 + m\tau + p_1 = 2(p_1 + p_2) + (k + m - 1)\tau$.

Также верно $C_{max}[E2(k + m)] < C_{max}[\{E2(m), A(k)\}]$:

- $C_{max}[E2(k + m)] = 2p_1 + (k + m)\tau$;
- $C_{max}[\{E2(m), A(k)\}] = 2p_1 + m\tau + p_1 + p_2 + (k - 1)\tau = 2(p_1 + p_2) + (k + m - 1)\tau$, если $\tau \leq p_1 - p_2$;
- $C_{max}[\{E2(m), A(k)\}] = p_1 + p_2 + m\tau + p_1 + p_2 + (k - 1)\tau = 2(p_1 + p_2) + (k + m - 1)\tau$, если $\tau \geq p_1 - p_2$.

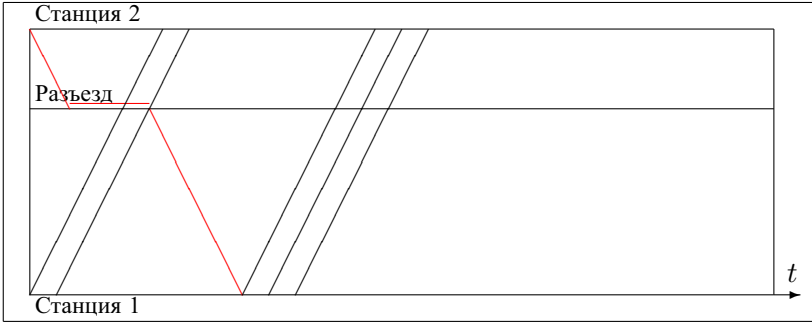


Рис. 17. График движения при последовательном выполнении подзадач $\{E2(m), A(k)\}$ при $m = 2$ и $k = 3$

Лемма 9. Для получения оптимального расписания с минимальным значением C_{max} решаемые подряд подзадачи $\{B(k), E1(m)\}$ и $\{E1(m), B(k)\}$ необходимо свести к подзадаче $E1(m + k)$.

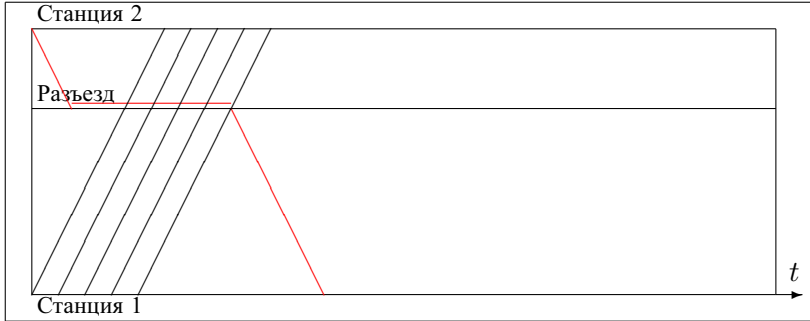


Рис. 18. График движения при выполнении подзадачи $E2(m+k)$ при $m=2$ и $k=3$

Доказательство. Комбинация подзадач $B(k)$ и $E1(m)$ – задача с поступлением $k+m$ поездов на станцию 2 и одного поезда на станцию 1, т.е. по постановке это подзадача $E1(k+m)$, для которой уже было определено решение. Справедливо неравенство $C_{max}[E1(k+m)] < C_{max}[\{B(k), E1(m)\}]$:

- $C_{max}[E1(k+m)] = 2p_1 + (k+m)\tau$;
- $C_{max}[\{B(k), E1(m)\}] = p_1 + p_2 + (k-1)\tau + 2p_1 + m\tau = 3p_1 + p_2 + (m+k-1)\tau$.

Также верно $C_{max}[E1(k+m)] < C_{max}[\{E1(m), B(k)\}]$ (см. рис. 20 и рис. 19):

- $C_{max}[E1(k+m)] = 2p_1 + (k+m)\tau$;
- $C_{max}[\{E1(m), B(k)\}] = p_1 + m\tau + p_2 + p_1 + p_2 + (k-1)\tau = 2(p_1 + p_2) + (m+k-1)\tau$.

Лемма 10. Для получения оптимального расписания с минимальным значением C_{max} решаемые подряд подзадачи $\{E1(m), A(k)\}$ должны выполняться в порядке $\{A(k), E1(m)\}$.

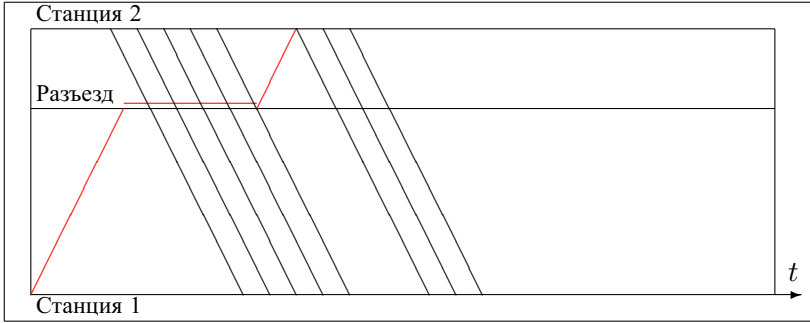


Рис. 19. График движения при последовательном выполнении подзадач $\{E1(m), B(k)\}$ при $m = 5$ и $k = 3$

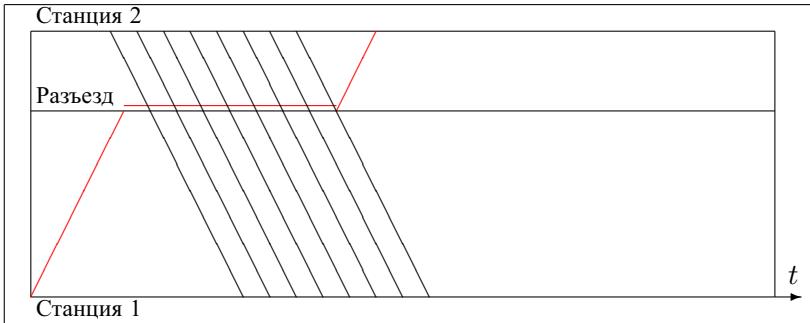


Рис. 20. График движения для последовательного выполнения подзадач $E1(m + k)$ при $m = 5$ и $k = 3$

Доказательство. Пусть подзадачи выполняются в порядке $\{E1(m), A(k)\}$, начиная с $t^1 = 0$ (см. рис. 21). Подзадача $A(k)$ может выполняться только начиная с момента завершения $E1(m)$, когда последний из m поездов достигнет станции 1, т.е. в момент $C_{max}[E1(m)]$. Следовательно, время выполнения двух подзадач в указанном порядке будет равно сумме C_{max} для каждой из подзадач:

$$(29) \quad C_{max}[\{E1(m), A(k)\}] = 3p_1 + p_2 + m\tau + (k - 1)\tau.$$

Если выполнять подзадачи в порядке $\{A(k), E1(m)\}$ (см. рис. 22), начиная с $t^1 = 0$, то для любых $m \geq 1$ и $k \geq 1$ по-

лучаем C_{max} меньше, чем при порядке $\{E1(m), A(k)\}$:

$$(30) \quad C_{max}[\{A(k), E1(m)\}] = 2(p_1 + p_2) + (k - 1)\tau + m\tau.$$

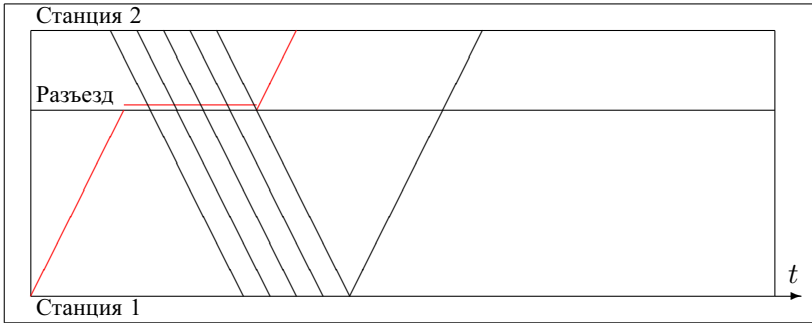


Рис. 21. График движения при выполнении подзадач в порядке $\{E1(m), A(k)\}$ при $m = 5$ и $k = 1$

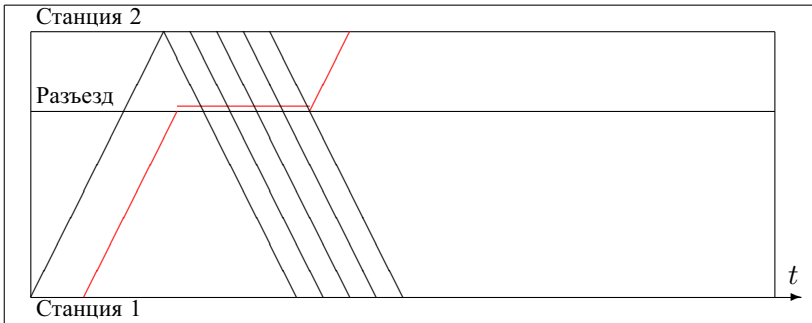


Рис. 22. График движения при последовательном выполнении подзадач $\{A(k), E1(m)\}$ при $m = 5$ и $k = 1$

Лемма 11. При $\tau \leq p_1 - p_2$ для получения оптимального расписания с минимальным значением C_{max} решаемые подряд подзадачи $B(k)$ и $E2(m)$ решаются в порядке $\{E2(m), B(k)\}$. При $\tau \geq p_1 - p_2$ порядок выполнения подзадач $B(k)$ и $E2(m)$ не влияет на C_{max} .

Доказательство. Имеем $\tau \leq p_1 - p_2$, а подзадача $E2(m)$ выполняется первой (см. рис. 23), начиная с момента $t^1 = 0$. Подзадача $B(k)$ может начать выполняться в момент прибытия поезда m на станцию 2, т.е. в момент $t^2 = C_m^1 = p_1 + p_2 + (m-1)\tau$. Отметим, что $t^2 < C_{max}[E2(m)]$. Время окончания выполнения подзадач в порядке $\{E2(m), B(k)\}$:

$$(31) \quad C_{max}\{E2(m), B(k)\} = t^1 + C_{max}[B(k)] = 2p_1 + 2p_2 + (m-1)\tau + (k-1)\tau.$$

Если выполнять подзадачи в порядке $\{B(k), E2(m)\}$, начиная с момента $t^1 = 0$, то выполнение подзадачи $B(k)$ закончится в момент

$$(32) \quad C_{max}[B(k)] = p_1 + p_2 + (k-1)\tau.$$

Время окончания перевозок подзадачи $E2(m)$:

$$(33) \quad C_{max}\{B(k), E2(m)\} = 3p_1 + p_2 + (k-1)\tau + (m-1)\tau.$$

При условии $p_1 \geq p_2$ для получения оптимального расписания

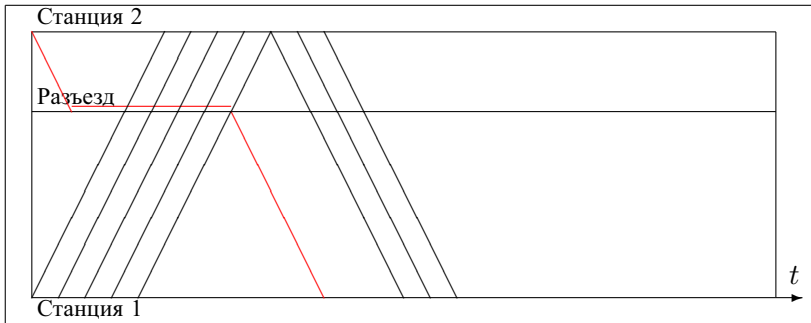


Рис. 23. График движения при последовательном выполнении подзадач $\{E2(m), B(k)\}$ при $m = 5$ и $k = 3$

подзадачи необходимо выполнять в порядке $\{E2(m), B(k)\}$ (см. рис. 24). На рис. 24 стрелками обозначены самый ранний и самый поздний из возможных вариантов момента начала движения поезда со станции 2, который делает остановку в разъезде.

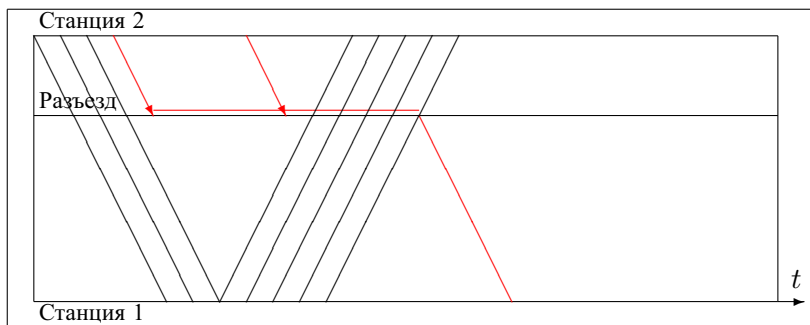


Рис. 24. График движения при последовательном выполнении подзадач $\{B(k), E2(m)\}$ при $m = 5$ и $k = 3$

4.3.6. ПОИСК ОПТИМАЛЬНОЙ КОМБИНАЦИИ

Итак, были получены правила комбинации для любых пар подзадач $A(m^A)$, $B(m^B)$, $E1(m^{E1})$ и $E2(m^{E2})$.

Теорема 1. Исходная задача может быть представлена как комбинация подзадач вида $\{A, B, A, A, E1, E2, B, B, E1, \dots\}$ с соответствующими аргументами. В соответствии с леммами 3–11 для получения оптимального расписания комбинация будет в зависимости от количества k_{E1} подзадач $E1(m_{E1})$ и k_{E2} подзадач $E2(m_{E2})$ изменена.

- При $k_{E1} = 0$, $k_{E2} \geq 1$ задача будет сведена к комбинации подзадачи $B(n_2 - k_{E2})$ и k_{E2} подзадач $E2(m_i^{E2})$, $i = 1, \dots, k_{E2}$, выполняемых в порядке: $\{E2(m_1^{E2}), \dots, E2(m_{k_{E2}}^{E2}), B(n_2 - k_{E2})\}$. Порядок выполнения k_{E2} подзадач $E2$ не важен.
- При $k_{E1} \geq 1$, $k_{E2} = 0$ задача будет сведена к комбинации подзадачи $A(n_1 - k_{E1})$ и k_{E1} подзадач $E1(m_i^{E1})$, $i = 1, \dots, k_{E1}$, выполняемых в порядке: $\{A(n_1 - k_{E1}), E1(m_1^{E1}), \dots, E1(m_{k_{E1}}^{E1})\}$. Подзадачи $E1$ могут выполняться в произвольном порядке.
- При $k_{E1} \geq 1$, $k_{E2} \geq 1$ задача будет сведена к комбинации k_{E1} подзадач $E1(m_i^{E1})$, $i = 1, \dots, k_{E1}$, и k_{E2} подзадач

$E2(m_j^{E2}), j = 1, \dots, k_{E2}$, выполняемых в порядке:
 $\{E1(m_1^{E1}), \dots, E1(m_{k_{E1}}^{E1}), E2(m_1^{E2}), \dots, E2(m_{k_{E2}}^{E2})\}$.

Доказательство. Можно доказывать от противного: сделать предположение, что существует некоторая комбинация другого вида, для которой справедливы все леммы. Докажем, что в соответствии с леммами комбинация любой структуры будет сведена к перечисленным вариантам при любых $\tau \leq \min\{p_1, p_2\}$.

В первом случае, при $k_{E1} = 0, k_{E2} \geq 1$, все комбинации составлены из k_{E2} подзадач $E2$ и набора задач A и B между ними:

$$\{E2, A, B, A, \dots, E2, \dots, E2, \dots\}.$$

Согласно лемме 3, порядок решаемых подряд подзадач A и B не важен. В комбинации все наборы подзадач A и B можно упорядочить так, чтобы первыми шли все A , а затем все подзадачи B . Согласно леммам 4 и 5 все идущие подряд подзадачи типа A должны быть сведены к одной подзадаче типа A , а все подзадачи типа B – к одной подзадаче B . Новая комбинация будет состоять из k_{E2} подзадач $E2$ и пар подзадач A и B между ними:

$$\{E2, A, B, E2, A, B, E2, \dots\}.$$

Согласно лемме 8 все пары подзадач $\{E2, A\}$ и $\{A, E2\}$ сводятся к подзадаче $E2$:

$$\{E2, B, E2, B, E2, \dots\}.$$

Из леммы 11 все пары подзадач $\{B, E2\}$ выполняются в порядке $\{E2, B\}$:

$$\{E2, E2, B, B, E2, \dots\}.$$

В соответствии с леммой 5 все решаемые подряд подзадачи B сводятся к одной подзадаче B ,

$$\{E2, \dots, E2, E2, B\}.$$

По лемме 6 порядок выполнения подзадач $E2$ не важен.

Во втором случае, при $k_{E1} \geq 1$, $k_{E2} = 0$, все комбинации составлены из k_{E1} подзадач $E1$ и набора подзадач A и B между ними:

$$\{E1, A, B, A, E1, B, B, E1, \dots, E1, \dots\}.$$

Так же как и в первом случае, в соответствии с леммами, все наборы подзадач A и B разделяются и сводятся к парам A и B :

$$\{E1, A, B, E1, B, E1, \dots, E1, \dots\}.$$

По лемме 9 решаемые подряд подзадачи $\{E1, B\}$ и $\{B, E1\}$ сводятся к подзадаче $E1$, по лемме 10 все пары подзадач $\{E1, A\}$ выполняются в порядке $\{A, E1\}$:

$$\{E1, A, E1, E1, \dots, E1, \dots\};$$

$$\{A, E1, E1, E1, \dots, E1, \dots\}.$$

По лемме 6 порядок выполнения подзадач $E1$ не важен.

В третьем случае, при $k_{E1} \geq 1$, $k_{E2} \geq 1$, все комбинации составлены из k_{E1} подзадач $E1$, k_{E2} подзадач $E2$ и набора подзадач A и B между ними:

$$\{A, B, E2, E1, B, B, E1, \dots, E2, \dots, E2\}.$$

В соответствии с леммами 3, 4 и 5 все наборы подзадач A и B разделяются и сводятся к парам A и B . По леммам 10 и 11 пары подзадач A, B и $E1, E2$ меняют порядок. Согласно леммам 8 и 9 пары подзадач A, B и $E1, E2$ сводятся к подзадачам $E1$ и $E2$. Полученное расписание состоит только из подзадач $E1$ и $E2$, выполняемых в соответствии с леммой 7 в порядке

$$\{E1, \dots, E1, E2, \dots, E2\}.$$

В теореме описаны оптимальные комбинации для различного числа подзадач $E1$ и $E2$, т.е. разного числа поездов, пропускающих в разъезде встречный поезд (или поезда). Чтобы узнать

оптимальное число подзадач $E1$ и $E2$, необходимо получить формулу C_{max} в зависимости от k_{E1} и k_{E2} .

Если $k^{E1} = 0$ и $k^{E2} \geq 1$, то оптимальная комбинация имеет вид

$$\{E2(m_1^{E2}), \dots, E2(m_{k^{E2}}^{E2}), B(n_2 - k^{E2})\}.$$

Время окончания перевозок для данной комбинации можно представить как

$$(34) \quad C_{max} = C_{max}[\{E2(m_1^{E2}), \dots, E2(m_{k^{E2}-1}^{E2})\}] + C_{max}[\{E2(m_{k^{E2}}^{E2}), B(n_2 - k^{E2})\}].$$

Обозначим комбинацию $\{E2(m_1^{E2}), \dots, E2(m_{k^{E2}-1}^{E2})\}$ как π_1 , и $\{E2(m_{k^{E2}}^{E2}), B(n_2 - k^{E2})\}$ как π_2 . Пусть $\tau \leq p_1 - p_2$, тогда $C_{max}[E2(m)] = 2p_1 + (m - 1)\tau$. Из леммы 6 следует, что порядок последовательного выполнения подзадач $E2$ не влияет на целевую функцию данной комбинации, поэтому

$$(35) \quad C_{max}[\pi_1] = \sum_{i=1}^{k^{E2}-1} C_{max}[E2(m_i^{E2})] = (2p_1 - \tau)(k^{E2} - 1) + \tau \left(\sum_{i=1}^{k^{E2}-1} m_i^{E2} \right).$$

Из леммы 2 получаем

$$(36) \quad \sum_{i=1}^{k^{E2}-1} m_i^{E2} = n_1 - k^{E1} - m_1^{E2} = n_1 - m_{k^{E2}}^{E2},$$

и

$$(37) \quad C_{max}[\pi_1] = (2p_1 - \tau)(k^{E2} - 1) + \tau(n_1 - m_{k^{E2}}^{E2}).$$

Для комбинации вида $\{E2(m_{k^{E2}}^{E2}), B(n_2 - k^{E2})\}$ уже была описана формула C_{max} в лемме 11:

$$(38) \quad C_{max}[\pi_2] = 2(p_1 + p_2) + (m_{k^{E2}}^{E2} - 1)\tau + (n_2 - k^{E2} - 1)\tau.$$

Из формул (37) и (38) получаем выражения для C_{max} . При $\tau \leq p_1 - p_2$ у комбинации $\{E2(m_1^{E2}), \dots, E2(m_{k^{E2}}^{E2}), B(n_2 - k^{E2})\}$, возможны два варианта. При $n_2 > k^{E2}$

$$(39) \quad C_{max} = 2(p_1 + p_2) + (n_1 + n_2 - 3)\tau + 2(p_1 - \tau)(k^{E2} - 1),$$

при $n_2 = k^{E2}$ комбинация, обозначенная как π_2 , имеет вид $\{E2(m_{k^{E2}}^{E2})\}$ и

$$(40) \quad C_{max} = (2p_1 - \tau)k^{E2} + \tau n_1 = (2p_1 - \tau)n_2 + \tau n_1.$$

Пусть $\tau \geq p_1 - p_2$, тогда $C_{max}[E2(m)] = p_1 + p_2 + m\tau$. Если $n_2 > k^{E2}$, то

$$(41) \quad C_{max} = k^{E2}(2p_2 - \tau) + (n_1 + n_2 - 1)\tau + p_1 + p_2.$$

Если $n_2 = k^{E2}$, то π_2 имеет вид $\{E2(m_{k^{E2}}^{E2})\}$ и

$$(42) \quad C_{max} = k^{E2}2p_2 + n_1\tau + p_1 - p_2 = n_2p_2 + n_1\tau + p_1 - p_2.$$

Были получены формулы (39) и (41) для данной комбинации при любых $\tau \leq \min\{p_1, p_2\}$.

Если $k^{E1} \geq 1$ и $k^{E2} = 0$, то оптимальная комбинация имеет вид

$$\{A(n_1 - k^{E1}), E1(m_1^{E1}), \dots, E1(m_{k^{E1}}^{E1})\}.$$

Целевую функцию данной комбинации можно представить формулой

$$(43) \quad C_{max} = C_{max}[\{A(n_1 - k^{E1}), E1(m_1^{E1})\}] + C_{max}[\{E1(m_2^{E1}), \dots, E1(m_{k^{E1}}^{E1})\}].$$

Определим комбинацию $\{A(n_1 - k^{E1}), E1(m_1^{E1})\}$ как π_1 , и $\{E1(m_2^{E1}), \dots, E1(m_{k^{E1}}^{E1})\}$ как π_2 . Для p_1 и p_2 получаем формулы целевой функции комбинации π_1

$$(44) \quad C_{max}[\pi_1] = 2(p_1 + p_2) + (n_1 - k^{E1} - 1)\tau + (m_1^{E1} - 1)\tau$$

и π_2

$$(45) \quad C_{max}[\pi_2] = \sum_{i=2}^{k^{E1}} C_{max}[E1(m_i^{E1})] = 2p_1(k^{E1} - 1) + \tau \sum_{i=2}^{k^{E1}} m_i^{E1}.$$

Из леммы 2 получаем

$$(46) \quad \sum_{i=2}^{k^{E1}} m_i^{E1} = n_2 - k^{E2} - m_1^{E1} = n_2 - m_1^{E1}$$

и

$$(47) \quad C_{max}[\pi_2] = 2p_1(k^{E1} - 1) + (n_2 - m_1^{E1})\tau.$$

Если $n_1 > k^{E1}$, то

$$(48) \quad C_{max} = (k^{E1} - 1)(2p_1 - \tau) + (n_1 + n_2 - 3)\tau + 2(p_1 + p_2).$$

Если $n_1 = k^{E1}$, то

$$(49) \quad C_{max} = 2p_1k^{E1} + \tau n_2 = 2p_1n_1 + \tau n_2.$$

Если $k^{E1} \geq 1$ и $k^{E2} \geq 1$, то оптимальная комбинация имеет вид

$$\{E1(m_1^{E1}), \dots, E1(m_{k^{E1}}^{E1}), E2(m_1^{E2}), \dots, E2(m_{k^{E2}}^{E2})\}.$$

Целевую функцию данной комбинации можно представить формулой

$$(50) \quad C_{max} = C_{max}[\{E1(m_1^{E1}), \dots, E1(m_{k^{E1}-1}^{E1})\}] + C_{max}[\{E1(m_{k^{E1}}^{E1}), E2(m_1^{E2}), \dots, E2(m_{k^{E2}}^{E2})\}].$$

Обозначим комбинацию $\{E1(m_1^{E1}), \dots, E1(m_{k^{E1}-1}^{E1})\}$ как π_1 , и $\{E1(m_{k^{E1}}^{E1}), E2(m_1^{E2}), \dots, E2(m_{k^{E2}}^{E2})\}$ как π_2 . Для π_2 справедлива формула

$$(51) \quad C_{max}[\pi_1] = 2p_1(k^{E1} - 1) + \tau \sum_{i=1}^{k^{E1}-1} m_i^{E1}.$$

Из леммы 2 получаем

$$(52) \quad \sum_{i=1}^{k^{E1}-1} m_i^{E1} = n_2 - k^{E2} - m_{k^{E1}}^{E1}$$

и находим формулы целевых функций комбинаций π_1 и π_2 :

$$(53) \quad C_{max}[\pi_1] = 2p_1(k^{E1} - 1) + \tau(n_2 - k^{E2} - m_{k^{E1}}^{E1});$$

$$(54) \quad C_{max}[\pi_2] = 2p_1 + \tau m_{k^{E1}}^{E1} + 2p_2 k^{E2} + (n_1 - k^{E1})\tau.$$

Складываем формулы (53) и (54), получаем

$$(55) \quad C_{max} = \tau(n_1 + n_2) + k^{E1}(2p_1 - \tau) + k^{E2}(2p_2 - \tau).$$

Были найдены оптимальные комбинации для любого количества поездов, заходящих в разъезд (т.е. k^{E1} и k^{E2}). Для каждой комбинации целевая функция зависит от k^{E1} и (или) k^{E2} , и увеличивается с ростом k^{E1} и k^{E2} . Целевая функция в каждой комбинации минимальна, при условии что:

- $k^{E1} = 0$, $k^{E2} = 1$, $n_2 \geq 2$ и оптимальная комбинация имеет вид $\{E2(n_1), B(n_2 - 1)\}$.

Целевая функция:

$$C_{max} = 3p_2 + (n_1 + n_2 - 2)\tau + p_1;$$

- $k^{E2} = 0, k^{E1} = 1, n_1 \geq 2$ и оптимальная комбинация имеет вид $\{A(n_1 - 1), E1(n_2)\}$.

Целевая функция:

$$C_{max} = 2(p_1 + p_2) + (n_1 + n_2 - 3)\tau;$$

- $k^{E1} = 1, k^{E2} = 1, n_1 \geq 2, n_2 \geq 2$, оптимальная комбинация имеет вид $\{E2(n_1 - 1), E1(n_2 - 1)\}$.

Целевая функция:

$$C_{max} = 2(p_1 + p_2) + (n_1 + n_2 - 2)\tau.$$

Если $\tau \geq p_1 - p_2$, то комбинация $\{A(n_1 - 1), E1(n_2)\}$ имеет наименьшее значение целевой функции. Если $\tau \leq p_1 - p_2$, то оптимальной комбинацией является $\{E2(n_1), B(n_2 - 1)\}$.

5. Вывод

Можно сделать вывод, что в зависимости от соотношения параметров τ, p_1 и p_2 оптимальное расписание однозначно определяется комбинацией подзадач, получаемой без перебора. При оптимальном расписании для задачи минимизации времени завершения перевозок с 1 разездом и $n_1 \geq 2, n_2 \geq 2$, в разезде будет делать остановку только один поезд. Есть два варианта – в разезде останавливается поезд со станции 1 или со станции 2:

- при $\tau \leq p_1 - p_2$ оптимальная комбинация $\{E2(n_1), B(n_2 - 1)\}$ с $C_{max} = 3p_2 + (n_1 + n_2 - 2)\tau + p_1$;
- при $\tau \geq p_1 - p_2$ оптимальная комбинация $\{A(n_1 - 1), E1(n_2)\}$ с $C_{max} = 2(p_1 + p_2) + (n_1 + n_2 - 3)\tau$.

Для задачи минимизации времени окончания перевозок при одновременном поступлении поездов с одним разездом был получен точный алгоритм без перебора. Разработанный метод представления расписаний в виде комбинаций подзадач может быть применен для описания свойств модели и исследования задач с другими целевыми функциями.

В дальнейшем планируется:

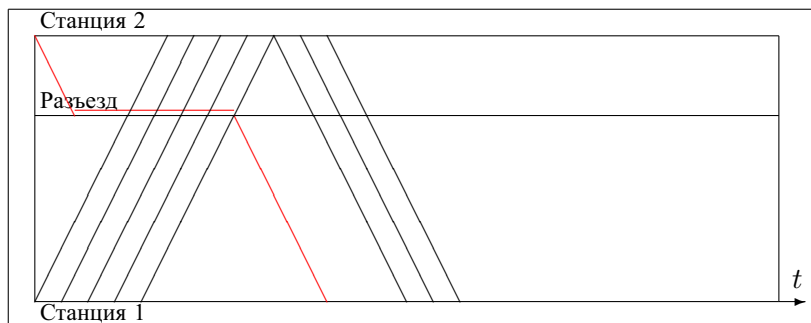


Рис. 25. График движения при оптимальном расписании для задачи с $n_1 = 5$ и $n_2 = 4$ при $\tau \leq p_1 - p_2$

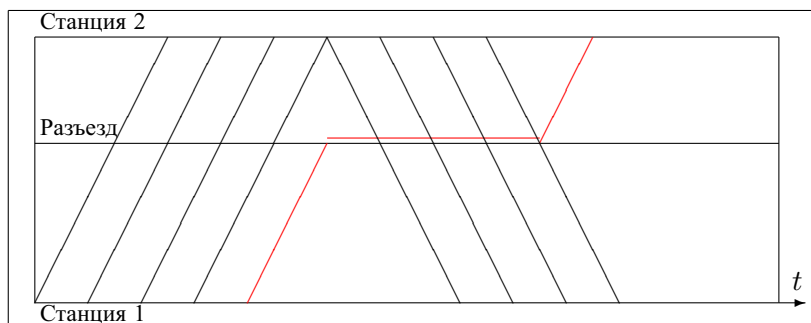


Рис. 26. График движения при оптимальном расписании для задачи с $n_1 = 5$ и $n_2 = 4$ при $\tau \geq p_1 - p_2$

- разработка алгоритма решения задачи минимизации максимально временного смещения при заданных директивных сроках с одним разъездом;
- исследование задачи минимизации времени окончания перевозок при заданных моментах поступления поездов на станции с одним разъездом.

Литература

1. *Basic indicator of Transport in Russia.* – Rosstat, 2010. – 95 p.

2. BRANNLUND U., LINDBERG P.O., NOU A., NILSSON J.E. *Railway timetabling using lagrangian relaxation* // *Transportation Science*. – 1998. – Vol. 32, No. 4. – P. 358–369.
3. BRUCKER P. *Scheduling Algorithms*. – Springer-Verlag New York, Inc., 2001. – 365 p.
4. CAREY M., LOCKWOOD D. *A model, algorithms and strategy for train pathing* // *The Journal of Operational Research Society*. – 1995. – Vol. 8, No. 46. – P. 988–1005.
5. GAFAROV E.R., DOLGUI A., LAZAREV A.A. *Two-station single-track railway scheduling problem with trains of equal speed* // *Computers and Industrial Engineering*. – 2015. (In Press, Accepted Manuscript).
6. GRAHAM R.L., LAWLER E.L., LENSTRA J.K., RINNOOY KAN A.H.G. *Optimization and approximation in deterministic sequencing and scheduling: a survey* // *Annals of Discrete Mathematics*. – 1979. – Vol. 5. – P. 287–326.
7. HARROD S.S. *A tutorial on fundamental model structures for railway timetable optimization* // *Surveys in Operations Research and Management Science*. – 2012. – Vol. 17, No. 2. – P. 85–96.
8. HIGGINS A., KOZAN E., FERREIRA L. *Optimal scheduling of trains on a single line track* // *Transportation Research Part B: Methodological*. – 1996. – Vol. 30, No. 2. – P. 147–161.
9. KRAVCHENKO S.A., WERNER F. *Parallel machine problems with equal processing times: a survey* // *Journal of Scheduling*. – 2011. – Vol. 14, No. 5. – P. 435–444.
10. LUSBY R.M., LARSEN J., EHRGOTT M., RYAN D. *Railway track allocation: Models and methods* // *OR Spectr.* – 2011. – Vol. 33, No. 4. – P. 843–883.
11. MONMA C.L., POTTS C.N. *On the complexity of scheduling with batch setup times* // *Operations Research*. – 1989. – Vol. 37, No. 5. – P. 798–804.
12. OLIVEIRA E.S. *Solving Single-Track Railway Scheduling*

- Problem Using Constraint Programming.* – PhD thesis, University of Leeds, 2001. – 129 p.
13. SOTSKOV Y., GHOLAMI O. *Shifting bottleneck algorithm for train scheduling in a single-track railway* // Proc. 14th IFAC Symposium on Information Control Problems. Part 1. – 2012. – P. 87–92.
 14. SZPIGEL B. *Optimal train scheduling on a single line railway* // Oper Res. – 1973. – P. 344–351.
 15. XUESONG Z., ZHONG M. *Single-track train timetabling with guaranteed optimality: Branch-and-bound algorithms with enhanced lower bounds* // Transportation Research Part B: Methodological. – 2007. – Vol. 41, No. 3. – P. 320–341.

SCHEDULING PROBLEM FOR TWO-STATION SINGLE TRACK RAILWAY WITH SIDINGS

Alexander Lazarev, Institute of Control Sciences of RAS, Moscow State University, Higher School of Economics, Moscow Institute of Physics and Technology State University, Moscow, Doctor of Science, professor (Moscow, Profsoyuznaya st., 65, (495)334-87-51, *jobmath@mail.ru*).

Ilya Tarasov, Institute of Control Sciences of RAS, Moscow State University, Moscow, student (*ilya_tarasof@mail.ru*).

Abstract: The paper is concerned with the problem of scheduling trains traveling between two stations, which are connected by a single railway track with one siding. The presented algorithm constructs an optimal schedule.

Keywords: scheduling theory, combinatorial optimization, transport problems, algorithm.

*Статья представлена к публикации
членом редакционной коллегии Е.Н. Хоботовым*

*Поступила в редакцию 07.07.2015.
Дата опубликования 30.11.2015.*

УДК 004.896
ББК 30.1

РЕКОНФИГУРАЦИЯ ПРОСТРАНСТВЕННОГО ПОЛОЖЕНИЯ РОЯ РОБОТОВ

Ватаманюк И. В.¹, Панина Г. Ю.², Ронжин А. Л.³
(ФГБУН Санкт-петербургский институт
информатики и автоматизации РАН)

Предложен алгоритм управления и навигации роя автономных гомогенных мобильных роботов, задействованных в формировании заданной пространственной выпуклой поверхности. Расчет траекторий роботов при реконфигурации роя проводится с учетом минимизации временных и энергетических затрат (времени вычислений и суммарной длины траекторий), а также контроля коллизий. Разработанный алгоритм имеет квадратичную вычислительную сложность. В ходе экспериментов по моделированию реконфигурации роя от 10 до 10000 роботов была проведена оценка числа коллизий, возникающих в случае прямолинейного движения роботов к своим целевым точкам, при предположении, что роботы стартуют одновременно.

Ключевые слова: рой роботов, топологическая робототехника, имитация поверхностей, групповое управление, беспилотные аппараты, робототехнические комплексы.

¹ Ирина Валерьевна Ватаманюк, аспирант СПИИРАН (тел.: (812) 328-70-81, vatananiuk@iias.spb.su).

² Гаянэ Юрьевна Панина, доктор физико-математических наук, профессор, ведущий научный сотрудник лаборатории автономных робототехнических систем СПИИРАН (тел.: (812) 328-70-81, gaiane-panina@rambler.ru).

³ Андрей Леонидович Ронжин, доктор технических наук, профессор, заведующий лабораторией автономных робототехнических систем СПИИРАН (тел.: (812) 328-70-81, ronzhin@iias.spb.su).

1. Введение

Системы автономных мобильных роботов способны решать большой класс гражданских и военных задач, включая тушение пожаров, спасательные операции, аграрные и другие задачи в различных средах [1, 10, 13, 16]. Поэтому задачи реконфигурации роя роботов являются актуальными и достаточно широко изучаются и внедряются на практике. Используемые подходы основаны на различных принципах и, соответственно, базируются на различном математическом аппарате. Мы отсылаем читателя к работе [6], где представлен подробный обзор базовых алгоритмов и используемого математического аппарата. Как правило, в существующих работах учитываются взаимодействие роботов с соседями (граф взаимодействия агентов), а расчетные траектории движений представляют собой численные решения соответствующих дифференциальных уравнений. В отличие от традиционных работ, в данной статье мы предлагаем иной подход, существенно упрощающий вычислительную сложность задачи, но наносящий ущерб робастности алгоритма. Возможность его применения обусловлена спецификой конкретной задачи. Прежде чем описать его, рассмотрим существующие подходы и принципы решения схожих задач.

Применение многоагентных децентрализованных способов взаимодействия групп роботов позволяет создавать робототехнические системы, которые являются гибкими к реконфигурации, устойчивыми к противодействующим факторам и обладающими низкой себестоимостью [2, 7, 8, 9, 11, 12, 18].

В работе [11] изучается проблема разделения роя мобильных роботов на сбалансированные подгруппы, предлагаются алгоритмы контроля положения и ориентации групп роботов при выстраивании определенной пространственной структуры.

Работа [14] близка по постановке задачи к данной статье: совокупность роботов должна образовать некоторую пространственную фигуру. Авторы предлагают декомпозировать имитируемую фигуру на части и разделить рой на несколько подмножеств (групп) роботов, отвечающих за формирование своей части фигуры. После этого шага проблема управления роботами решается на двух уровнях: 1) взаимодействие роботов внутри

группы; 2) взаимодействие между группами роботов. При этом авторы используют стратегию «ведущий (лидер) – ведомый» и, в зависимости от сложности формируемой фигуры, вводят несколько уровней доминирования для ведомых групп роботов.

В работе [15] используется некоторый естественный аналог градиентного спуска. Каждый робот «видит» своих ближайших соседей и при реконфигурации движется в направлении, оптимизирующем требуемые целевые расстояния между роботами.

В работе [18] предложен новый топологический метод формирования пространственной фигуры из множества роботов на основе многоуровневой архитектуры, причем построение производится послойно с определенным расположением роботов на каждом уровне внутри всей формы.

Имеется ряд работ по самоорганизующимся многоагентным системам [1, 2]. Такие системы решают самые разнообразные задачи и (теоретически) могут выполнить реконфигурацию группы роботов, рассматриваемую в нашей статье. Однако специфика данной задачи позволяет произвести реконфигурацию пространственного расположения роя роботов с меньшими временными затратами на основе предложенного подхода.

Ниже (во втором разделе) представлена формальная постановка задачи реконфигурации роя роботов и перечислены основные этапы предложенного алгоритма. В разделе 3 приводится математическое описание алгоритма расчета траекторий движения роботов при реконфигурации. В разделе 4 приводится описание вспомогательных функций и процедур, используемых при работе алгоритма. Раздел 5 посвящен анализу результатов экспериментальной проверки алгоритма при моделировании реконфигурации роя от 10 до 10000 роботов.

2. Постановка задачи

В данной работе в результате реконфигурации роботам необходимо покрыть с заданной плотностью некоторую выпуклую поверхность. Выпуклость поверхности, а также отсутствие препятствий, придает задаче определенную специфику, в частности, позволяет предложить относительно простой алгоритм и сэкономить вычислительные и технические ресурсы.

Опишем подробнее постановку задачи. Мы имеем дело со следующими входными данными:

1. Имеется n идентичных объектов-роботов (*рой роботов*). Каждый из них представляет собой шар. Заданы координаты их центров.
2. R – радиус каждого из роботов.
3. $MinDist$ – минимальное допустимое расстояние между центрами роботов.
4. S – некоторая выпуклая поверхность. Мы предполагаем, что поверхность задана двумя наборами данных:
 - а) набором вершин $v_i = (x_i, y_i, z_i)$;
 - б) набором треугольных граней, т.е. троек различных индексов (i, j, k) , задающих грани поверхности. Мы считаем, что грани занумерованы от 1 до M . (Таким образом, M – число граней поверхности);
 - в) данные пунктов а) и б) позволяют составить список ребер поверхности, т.е. набор всех ребер треугольных граней без повторений.
5. V – скорость движения роботов. Предполагается, что роботы могут либо двигаться со скоростью V , либо стоять неподвижно.
6. ρ – желаемая плотность размещения роботов на поверхности. А именно, ρ есть желаемое число объектов на квадрате 10×10 .

В этих условиях требуется, избегая коллизий, переместить часть роботов так, чтобы они покрыли поверхность S с плотностью ρ . Точнее говоря, требуется получить следующие выходные параметры:

- 1) конечные координаты центров всех объектов, участвующих в реконфигурации;
- 2) расчетные траектории движения объектов, задействованных в моделировании фигуры;
- 3) оценку вычислительных затрат;
- 4) а также оценку времени перестроения объектов в требуемую фигуру.

При разработке алгоритма учитывались следующие технические условия:

1. Число объектов-роботов достаточно для покрытия поверхности с требуемой плотностью.

2. Плотность и минимальное допустимое расстояние связаны соотношением

$$(1) \frac{\sqrt{\rho}}{MinDist} < \frac{1}{2}.$$

Неформально говоря, это означает, что требуемая плотность не слишком велика.

3. Длина каждого из ребер триангуляции больше $2MinDist$.

4. Угол между соседними гранями поверхности не меньше прямого.

В настоящей статье предлагается алгоритм реконфигурации роя, вычислительная часть которого имеет квадратичную сложность. Время реконфигурации существенно зависит от взаимного расположения роботов и поверхности. Нам представляется, что оно близко к минимально возможному, так как мы выбираем ближайших к цели роботов и предписываем траектории, близкие к прямолинейным. Чем меньше соотношение «радиус робота / перемещение робота», тем ближе траектории к прямолинейным, а значит, объективно оптимальным. Новизна алгоритма состоит в ряде приемов: назначении точек портала и пересечении поверхности траекториями по нормали, что позволяет получить почти прямолинейные траектории.

Теперь рассмотрим детально эти и другие особенности предлагаемого алгоритма реконфигурации роя роботов. Работа алгоритма начинается с проверки корректности входных данных и задания на формируемой поверхности сети точек с требуемой плотностью. Каждой точке сети приписывается некоторый объект-робот, и эта точка сети является *таргетной* (т.е. целевой) для данного робота. На рис. 1 показаны примеры траекторий движения роботов. Во избежание коллизий необходимо ввести следующие корректировки.

– Если роботу (на пути к таргетной точке) приходится пересекать поверхность, то он делает это в *точке портала*. Точки портала образуют некоторую сеть точек на поверхности, задаваемую алгоритмом, и служат для бесконфликтного перехода роботов из внутренней части пространства, ограни-

ченного поверхностью, во внешнюю. Рис. 1 показывает набор таргетных (черных) точек и набор (белых) точек портала.

– Роботы приближаются к поверхности по вектору нормали (см. рис. 1). Этот (новый) прием позволит избежать коллизий с уже прибывшими на свое место роботами. Таким образом, движение каждого робота состоит из (самое большее) четырех прямолинейных участков.

– Однако такое движение все еще может привести к коллизиям – ситуациям, когда расстояние между центрами роботов окажется меньше минимально допустимого $MinDist$. Поэтому некоторым роботам предпишем задержку, т.е. они начинают движение не в момент 0, а в момент T , свой для каждого робота.

На рис. 1 и 2 мы изображаем «плоскую» модель нашей задачи: роботы расположены на плоскости и после реконфигурации должны сформировать плоскую выпуклую фигуру, в данном случае – пятиугольник. Далее рассмотрим более детально предлагаемый алгоритм реконфигурации роя роботов.

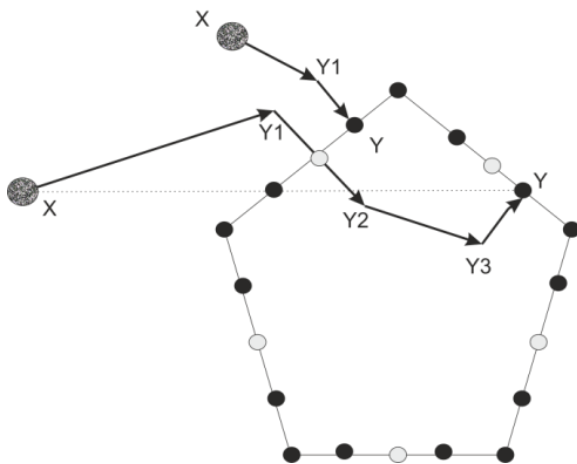


Рис. 1. Пример пересекающей и непересекающей поверхность траекторий. Черным цветом отмечены таргетные точки, белым - точки портала

3. Алгоритм расчета траекторий движения роботов при реконфигурации роя

Разработанный алгоритм состоит из следующих основных этапов: 1) анализ и оснащение формируемой поверхности и расчет координат целевых точек и точек портала; 2) задание соответствия между роботами и целевыми точками; 3) расчет траекторий движения роботов с учетом возможных коллизий и времени начальных задержек каждого робота. Рассмотрим последовательно каждый из указанных этапов.

3.1. АНАЛИЗ И ОСНАЩЕНИЕ ФОРМИРУЕМОЙ ПОВЕРХНОСТИ

На первом этапе алгоритма выполняется расчет координат целевых точек и точек портала. *Целевые точки* – это те точки поверхности, в которые роботы должны прибыть в результате реконфигурации. В общем случае целевые точки, а также их число, зависят от самой поверхности и от заданной плотности. Назначение целевых точек на поверхности производится следующим образом:

1. Вычислим желаемое расстояние между целевыми точками на поверхности:

$$(2) \quad \delta = \frac{10}{\sqrt{\rho}}.$$

Целевые точки условно делятся на три серии.

2. Первая серия целевых точек – вершины поверхности.

3. Вторая серия целевых точек – это точки, лежащие на ребрах; они задаются следующим способом:

Переберем все ребра по одному. Для каждого ребра зададим несколько целевых точек; они расположены равномерно вдоль ребра. А именно:

а) пусть вершины ребра – точки $A = (a_1, a_2, a_3)$, $B = (b_1, b_2, b_3)$.

Вычислим $N = [\text{dist}(A, B) / \delta]$. Здесь и далее квадратные скобки означают взятие целой части;

б) **for** $i = 1$ **to** $N - 1$ **do** Задаем целевую точку

$$T = A + i \frac{(B - A)}{N}.$$

4. Прежде чем задать третью серию целевых точек, зададим точки портала. На каждой грани поверхности будет по одной точке портала. В дальнейшем число точек портала можно увеличить. Но пока придерживаемся следующих действий:

Переберем все грани по одной.

Путь вершины грани F – точки: $A = (a_1, a_2, a_3)$, $B = (b_1, b_2, b_3)$, $C = (c_1, c_2, c_3)$.

а) зададим точку портала грани F :

$$(3) \text{ Portal} = \frac{1}{3}(A + B + C) = \left(\frac{1}{3}(a_1, b_1, c_1), \frac{1}{3}(a_2, b_2, c_2), \frac{1}{3}(a_3, b_3, c_3)\right);$$

б) зададим целевые точки третьего типа, лежащие строго внутри этой грани:

i. вычислим

$$(4) N_1 = \left\lfloor \frac{\text{dist}(A, B)}{\delta} \right\rfloor,$$

$$(5) N_2 = \left\lfloor \frac{\text{dist}(A, C)}{\delta} \right\rfloor;$$

ii. **for** $i = 1$ **to** $N_1 - 1$ **do**

for $j = 1$ **to** $N_2 - i \left\lfloor \frac{N_2}{N_1} \right\rfloor - 1$ **do**

$$\text{Зададим целевую точку } T = A + i \frac{(B - A)}{N_1} + j \frac{(C - A)}{N_2}.$$

Возможно, некоторые из целевых точек расположены недопустимо близко друг к другу или к точке портала. Для их устранения используем следующий порядок действий:

1. Переберем все вершины поверхности и удалим те из целевых точек, которые оказались ближе к данной вершине, чем минимальное допустимое расстояние.

2. Переберем все точки портала и удалим те из (оставшихся) целевых точек, которые оказались ближе к данной точке портала, чем минимальное допустимое расстояние.

3. Переберем все оставшиеся целевые точки второго и третьего типа и удалим те из (оставшихся) целевых точек, которые оказались ближе к данной точке, чем минимальное допустимое расстояние.

После того как точки портала и целевые точки определены, производится оснащение поверхности системой нормалей и

нормальными точками. Для каждой из целевых точек и точек портала зададим по две *нормальные точки* (см. рис. 2). Каждая нормальная точка отстоит от исходной в направлении нормали на расстояние *MinDist*.

При оснащении точек портала и целевых точек третьего типа перебираются по очереди все грани поверхности. Пусть вершина грани *F* – точки $A = (a_1, a_2, a_3)$, $B = (b_1, b_2, b_3)$, $C = (c_1, c_2, c_3)$, далее:

1. Зададим вектор нормали к грани:

$$(6) \vec{n}(F) = \frac{\overline{AB} \times \overline{AC}}{|\overline{AB} \times \overline{AC}|},$$

где \times – векторное произведение.

2. Добьемся нужной ориентации (вектор нормали должен смотреть «наружу»):

а) зададим две точки:

$$(7) N(\text{Portal}) = \text{Portal} + \infty \cdot \vec{n}(F),$$

(при программной реализации знак бесконечности заменяется константой достаточно большой величины);

$$(8) \text{Norm}(\text{Portal}) = \text{Portal} + \text{MinDist} \cdot \vec{n}(F),$$

б) запускаем

$$(9) \text{IntersectS}(\text{Norm}(\text{Portal}), N(\text{Portal}));$$

в) **if** Count \neq 0 **and** Count \neq 2

begin

$$(10) \vec{n}(F) = -\vec{n}(F)$$

end.

3. Зададим нормальные точки для точки портала этой грани правилом

$$(11) \text{Norm}_1(\text{Portal}) = \text{Portal} + \text{MinDist} \cdot \vec{n}(F),$$

$$(12) \text{Norm}_2(\text{Portal}) = \text{Portal} - \text{MinDist} \cdot \vec{n}(F).$$

4. Переберем по очереди все точки третьего типа, принадлежащие данной грани. Для каждой точки *T* зададим нормальные точки правилом

$$(13) \text{Norm}_1(T) = T + \text{MinDist} \cdot \vec{n}(F),$$

$$(14) \text{Norm}_2(T) = T - \text{MinDist} \cdot \vec{n}(F).$$

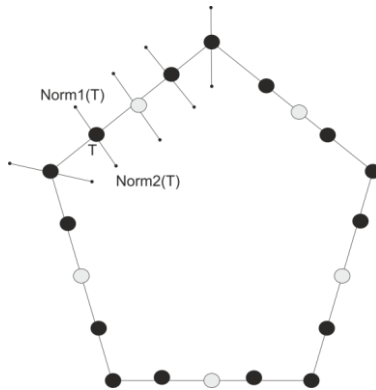


Рис. 2. Задание на поверхности системы нормалей и нормальных точек.

При оснащении таргетных точек второго типа перебираются по очереди все ребра поверхности.

Для каждого ребра E с вершинами AB

1. Находим ABC и ABD – примыкающие к ребру грани.
2. Зададим вектор

$$(15) \vec{n}(E) = \frac{\vec{n}(ABC) + \vec{n}(ABD)}{|\vec{n}(ABC) + \vec{n}(ABD)|}.$$

3. Переберем по очереди все таргетные точки второго типа, принадлежащие ребру AB . Для каждой точки T зададим нормальные точки правилом

$$(16) Norm_1(T) = T + MinDist \cdot \vec{n},$$

$$(17) Norm_2(T) = T - MinDist \cdot \vec{n}.$$

При оснащении таргетных точек первого типа перебираются по очереди все вершины (они же точки первого типа). Для каждой из них:

1. Пусть таргетная точка T – некоторая вершина. Выберем три различные грани F_1, F_2, F_3 , примыкающие к этой вершине.

2. Зададим вектор

$$(18) \vec{n}(T) = \frac{\vec{n}(F_1) + \vec{n}(F_2) + \vec{n}(F_3)}{|\vec{n}(F_1) + \vec{n}(F_2) + \vec{n}(F_3)|}$$

3. Зададим нормальные точки правилом

$$(19) Norm_1(T) = T + MinDist \cdot \vec{n},$$

$$(20) Norm_2(T) = T - MinDist \cdot \vec{n}.$$

3.2. ЗАДАНИЕ СООТВЕТСТВИЯ МЕЖДУ РОБОТАМИ И ТАРГЕТНЫМИ ТОЧКАМИ

Следующим этапом алгоритма является присвоение каждому роботу, задействованному в реконфигурации, своей таргетной точки следующим образом:

1. Упорядочим линейно таргетные точки (т.е. занумеруем их числами от 1 до N).

2. Определим, какой именно объект-робот направится к каждой из таргетных точек. Для этого переберем таргетные точки по порядку номеров, и каждой из них припишем ближайший объект-робот. Точнее говоря, сделаем следующее:

- а) для таргетной точки номер 1 находим ближайший робот, присваиваем ему номер 1, заносим робот в отдельный список *действующих роботов* под номером 1;
- б) для таргетной точки номер 2 находим ближайший *из оставшихся роботов*, присваиваем ему номер 2, заносим его в список действующих роботов под номером 2;
- в) и так далее для всех таргетных точек.

3. Таким образом, у нас есть упорядоченный список действующих роботов. Именно они будут участвовать в реконфигурации роя. Возможно, остались «лишние» роботы, в реконфигурации не участвующие. Мы будем называть их *пассивными роботами*. У каждого действующего робота тем самым имеется:

- а) порядковый номер k ;
- б) тройка координат, задающих начальную позицию его центра;
- в) тройка координат, задающих его таргетную точку. В эту точку должен быть перемещен центр робота в результате реконфигурации.

4. Для каждого робота в дальнейшем определим:

- а) число *NumberPathSegments* прямолинейных участков пути, которое предстоит выполнить роботу. Оно равно двум или четырем (см. рис. 1);
- б) траекторию робота. Она задается перечислением промежуточных точек;
- в) время задержки отправления робота.

3.3. РАСЧЕТ ТРАЕКТОРИЙ ДВИЖЕНИЯ РОБОТОВ

Перед расчетом траекторий выполняется сортировка действующих роботов на два типа. К первому типу относятся те, чья траектория не пересекает поверхность; ко второму типу – остальные роботы. Далее реконфигурация роя производится в следующем порядке.

Действующий робот первого типа с номером k в момент времени $T(k)$ начинает двигаться к нормальной точке своей таргетной точки. Добравшись до нее, он (без задержки) двигается к своей таргетной точке, в которой и останавливается.

Действующий робот второго типа с номером k в момент времени $T(k)$ начинает двигаться к нормальной точке портала, пересекает поверхность по нормали, затем начинает двигаться к нормальной точке своей таргетной точки. Добравшись до нее, он (без задержки) двигается к своей таргетной точке, в которой и останавливается.

Зададим траектории каждого робота следующим образом: Пусть X – стартовая точка робота, Y – таргетная точка.

Запускаем процедуру $IntersectS(X, Norm_1(Y))$ (см. раздел 4)

if $Count = 0$ **then**

begin

$Y_1 = Norm_1(Y)$.

Назначаем роботу X траекторию: $X \rightarrow Y_1 \rightarrow Y$.

Присваиваем значение $NumberPathSegments := 2$.

Завершаем назначение траектории для этого робота.

end

else

Запускаем процедуру $IntersectS(X, Norm_2(Y))$

if $Count \neq 0$ **then**

begin

$Y_1 = Norm_2(Y)$

Назначаем роботу X траекторию: $X \rightarrow Y_1 \rightarrow Y$.

Присваиваем значение $NumberPathSegments := 2$.

Завершаем назначение траектории для этого робота.

end

else

```

    I := FaceNumber
if Count = 1 then
    begin
        Запускаем  $IntersectS(X, Norm_1(Portal(I)))$ 
        if Count = 0 then
            begin
                 $Y_1 = (Portal(I))$ .
                 $Y_2 = Norm_2(Portal(I))$ .
                 $Y_3 = Norm_2(Y)$ .
                Назначаем роботу X траекторию:  $X \rightarrow Y_1 \rightarrow Y_2 \rightarrow Y_3 \rightarrow Y$ .
                Присваиваем значение  $NumberPathSegments(X) = 4$ .
                Завершаем назначение траектории для этого робота.
            end
        else
            begin
                 $Y_1 = (Portal(I))$ .
                 $Y_2 = Norm_1(Portal(I))$ .
                 $Y_3 = Norm_2(Y)$ .
                Назначаем роботу X траекторию:  $X \rightarrow Y_1 \rightarrow Y_2 \rightarrow Y_3 \rightarrow Y$ .
                Присваиваем значение  $NumberPathSegments(X) = 4$ .
                Завершаем назначение траектории для этого робота.
            end
        end.

```

Теперь опишем, как рассчитывается момент времени $T(k)$ начала движения каждого робота. Введем вспомогательную величину – *квант времени*:

$$(21) \tau = \frac{2MinDist}{V}.$$

Это время, за которое робот проходит расстояние $2MinDist$. Вначале все задержки начального движения роботов полагаются равными нулю.

```

For k = 2 to N
begin

```

Тестируем робот с номером k на наличие коллизий с роботами с меньшими номерами. В случае коллизий робот со старшим номером уступает дорогу. А именно:

for $i = 1$ **to** $k - 1$ **do**

begin

if $\text{CollisionTest}(i, k) = \text{TRUE}$ **then**

1. $T(k) = T(k) + \tau$.

2. Запускаем заново тест для k -го робота для младших номеров начиная с $i = 1$.

end

end.

Теперь оценим сложность предложенного алгоритма, исходя из естественного предположения, что общее число роботов (и активных, и пассивных) соразмерно с числом целевых точек (или, что то же самое, с числом активных роботов). Очевидно, сложность каждой из ступеней алгоритма не более чем квадратична (относительно числа целевых точек). Следовательно, мы имеем квадратичный алгоритм, сложность которого не зависит ни от времени потенциального перемещения роботов, ни от расстояний. Деликатным моментом в оценке сложности алгоритма оказывается разрешение коллизий, создающее циклы. Здесь мы пользуемся следующим эвристическим соображением в сочетании с результатами экспериментов (см. раздел 5): поскольку число конфликтующих пар роботов не превышает 2,22% от общего числа пар активных роботов, общее число циклов разрешения коллизий растет не быстрее, чем логарифм от общего числа пар роботов и, следовательно, вносит пренебрежимо малый вклад в общую сложность.

4. Вспомогательные процедуры проверки возможных коллизий роботов

Здесь опишем ряд вспомогательных функций и процедур, использованных в третьем разделе. Первые две из них определяют наличие пересечения отрезка и поверхности; вторые две – тестируют движения роботов на наличие коллизий.

Булева функция *IntersectTriangle* определяет наличие пересечения отрезка и треугольника, лежащих в трехмерном пространстве. На вход подаются отрезок *DE* и треугольник *ABC*.

IntersectTriangle(D, E, A, B, C) определяет, пересекаются ли отрезок и треугольник:

begin

IntersectTriangle = FALSE.

begin

if

$$(22) \text{Sign}(\text{DET}(D-A, B-A, E-A)) = \text{Sign}(\text{DET}(D-B, C-B, E-B)) = \\ \text{Sign}(\text{DET}(D-C, B-C, E-C))$$

and

$$(23) \text{Sign}(\text{DET}(B-A, C-A, D-A)) = -\text{Sign}(\text{DET}(B-A, C-A, E-A))$$

then

IntersectTriangle = TRUE.

end

end.

Процедура *IntersectS* определяет наличие пересечений отрезка с поверхностью *S*. На вход подается отрезок *DE*. Считается, что при этом поверхность *S* задана априори.

Процедура выдает два целых числа:

1. Число *Count(D, E)* говорит, сколько раз отрезок *DE* пересекает поверхность *S*. Оно принимает значения 0, 1, или 2 (в силу выпуклости поверхности).

2. Если пересечение не пусто, то *FaceNumber* - номер грани, содержащей точку пересечения. (Примечание: такая точка может быть не одна. В силу особенности алгоритма нам подойдет любая).

В процессе работы процедура перебирает все грани поверхности и запускает для каждой грани *IntersectTriangle*.

Функция *SmallCollisionTest*. Пусть имеются два робота. Известны положения каждого из них в момент времени t_0 . Каждый из роботов движется от момента времени t_0 до момента времени t_1 с постоянной скоростью (она равна либо нулю, либо V). Известны координаты точки, в направлении которой движется каждый из роботов. Функция *SmallCollisionTest* определяет наличие коллизии на временном промежутке (t_0, t_1) .

Работа функции сводится к поиску минимума квадратной функции от времени (расстояния между роботами) на некотором временном интервале, поэтому конкретные расчеты нами опущены в силу их элементарности.

Процедура *CollisionTest* тестирует траектории движения двух отдельно взятых роботов и работает следующим образом. Имеются два робота. Каждый из них начинает движение в (свой) момент времени по предписанной траектории, состоящей из (самое большее) четырех прямолинейных участков. Булева функция *CollisionTest* определяет наличие коллизии между этими двумя роботами следующим образом:

1. Для обоих роботов рассчитываются временные интервалы, соответствующие прямолинейным участкам траектории.

2. Для всех получившихся пересечений временных интервалов применяется функция *SmallCollisionTest*.

3. Если хотя бы для одного временного интервала определяется наличие коллизии, т.е. функция *SmallCollisionTest* возвращает *TRUE*, то процедура *CollisionTest* возвращает *TRUE*.

5. Компьютерное моделирование: результаты экспериментов

Основная цель проведенных экспериментов состояла в оценке числа коллизий, возникающих в том случае, когда роботы движутся к своим целевым точкам прямолинейно, начиная движение одновременно. При помощи данной оценки были сделаны выводы о сложности предложенного алгоритма. Для проведения экспериментов была смоделирована поверхность в форме граней куба. Эксперименты проводились для следующих значений числа роботов: 10, 50, 100, 500, 1000, 5000, 10000. Были рассмотрены три варианта размещения роботов: при максимально возможной плотности размещения ρ_{max} , при плотности, равной $0,5\rho_{max}$, а также при плотности $0,25\rho_{max}$. Размер ребра куба в каждом случае был рассчитан исходя из плотности размещения роботов на поверхности. Для каждого случая рассматривались два варианта

начального расположения роботов: вокруг центра куба на расстоянии не большем, чем 10 длин ребра куба и вне поверхности на расстоянии, не меньшем 10 длин ребра куба. Были оценены: время расчета траекторий (рис. 3(а)), количество коллизий при максимальной плотности ρ_{max} (рис. 3(б)), а также количество коллизий при различных значениях плотности размещения роботов (рис. 4). Кроме того, на рис. 4 для наглядности представлен график количества всех возможных сочетаний пар роботов.

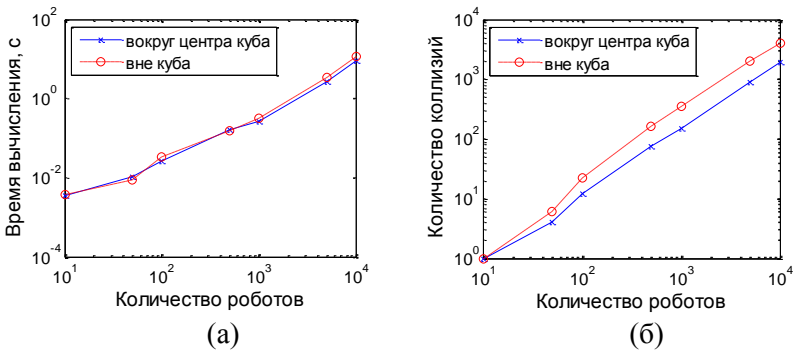


Рис. 3. (а) оценка времени расчета всех траекторий;
(б) оценка количества коллизий при ρ_{max}

Как видно из рис. 3(б), число коллизий при указанных условиях не превышает 2,22% от всех возможных пар роботов. На рис. 3(а) представлены средние значения времени вычисления траекторий. Заметим, что время расчета траекторий растет пропорционально количеству роботов и практически не зависит от начального расположения роботов.

Из рис. 4 видно, что с уменьшением плотности расположения роботов количество коллизий уменьшается пропорционально, вне зависимости от способа начального расположения роботов. Причем число коллизий во всех экспериментах не превышает 2,22% от всех возможных пар роботов.

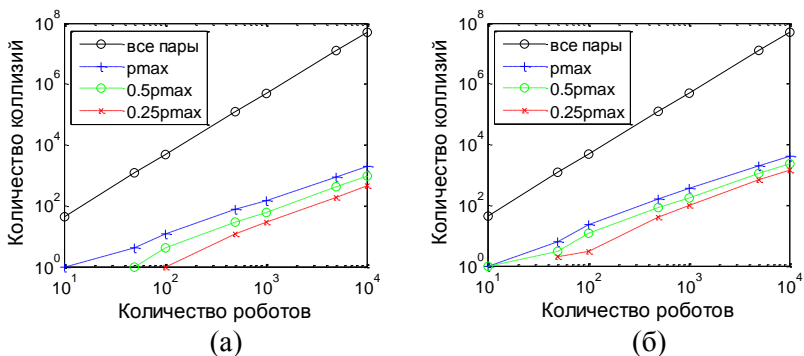


Рис. 4. Оценка количества коллизий при расположении роботов: (а) вокруг центра куба; (б) вне куба.

6. Заключение

В статье предложен алгоритм, обладающий следующими особенностями: 1) траектории движения почти прямолинейны, следовательно, энергозатраты на перемещения роботов практически минимальны; 2) естественно ожидать, что с вычислительной точки зрения алгоритм будет работать быстрее, чем метод градиентного спуска, так как отсутствует многоступенчатый расчет траекторий, неизбежный в последнем (отметим, однако, что экспериментально эти ожидания нами не проверялись); 3) направление движения каждого из роботов меняется самое большее трижды, причем незначительно, что экономит затраты на пространственную переориентацию робота.

Разработанная программная модель позволяет наглядно продемонстрировать движение большого числа малых робототехнических комплексов при формировании пространственных фигур. При моделировании движения роя разработанный алгоритм экспериментально проверен при разном количестве роботов (от 10 до 10000 единиц) с целью оценивания масштабируемости математического обеспечения и объема потребляемых вычислительных ресурсов. Предлагаемые решения управления движением роя роботов могут быть применены в быстро развер-

тываемых имитационных средствах, а также распределенных киберфизических системах [3, 4, 5, 17].

Одной из наиболее актуальных практических задач, где будут использованы разработанные алгоритмы, является разработка методики конфигурации автоматизированных робототехнических средств, позволяющих осуществить транспортировку пострадавшего в определенной позе, оптимальной для его состояния. Другой практической задачей является применение роя гомогенных роботов для формирования поверхности имитируемой пространственной фигуры, позволяющей средствам удаленной регистрации обнаружить заданные классы объектов.

Работа выполнена в рамках проекта программы Президиума РАН I.40П «Актуальные проблемы робототехники».

Литература

1. ГОРОДЕЦКИЙ В.И. *Самоорганизация и многоагентные системы. II. Приложения и технология разработки* // Известия РАН. Теория и системы управления. – 2012. – №3. – С. 55–75.
2. ИВАНОВ Д.Я. *Формирование строя группой беспилотных летательных аппаратов при решении задач мониторинга* // Известия Южного федерального университета. Технические науки. – 2012. – Вып.4, том 129. – С. 219-224.
3. КРЮЧКОВ Б.И., КАРПОВ А.А., УСОВ В.М. *Перспективные подходы к применению сервисных роботов в области пилотируемой космонавтики* // Труды СПИИРАН. – 2014. – Вып.32. – С. 125–151.
4. МИРОНОВ В.И., МИРОНОВ Ю.В., ПРИЩЕПА Ю.В. *О некоторых способах улучшения точностных характеристик координатной привязки изображений, получаемых беспилотными летательными аппаратами* // Труды СПИИРАН. – 2009. – Вып.9. – С. 159–167.
5. РОНЖИН А.Л., ЮСУПОВ Р.М. *Многомодальные интерфейсы автономных мобильных робототехнических комплексов* // Известия Южного федерального университета. Технические науки. – 2015. – №1(162). – С. 195–206.
6. BULLO F., CORTES J., MARTINEZ S. *Distributed Control of Robotic Networks*. – Princeton University Press, 2009. – 314 p.

7. CHEN Y.Q., WANG Z. *Formation control: a review and a new consideration* // Proc. IEEE Int. Conf. Intell. Robots Syst. – 2005. – P. 3181–3186.
8. EFRIMA A., PELEG D. *Distributed algorithms for partitioning a swarm of autonomous mobile robots* // Theoretical Computer Science. – 2009. – Vol.410. – P. 1355–1368.
9. GAZI V. *Swarms aggregation using artificial potentials and sliding mode control* // IEEE Transactions on Robotics. – 2005. – Vol.21(4). – P. 1208–1214.
10. HAGHIGHI R., CHEAH C. *Multi-group coordination control for robot swarms* // Automatica. – 2012. – Vol.48. – P. 2526–2534.
11. HSIEH M.A., KUMAR V., CHAIMOWICZ L. *Decentralized controllers for shape generation with robotic swarms* // Robotica. – 2008. – Vol.26(5). – P. 691–701.
12. HU J., XU J., XIE L. *Cooperative search and exploration in robotic networks* // Unmanned Systems. – 2013. – Vol.01. – P. 121–142.
13. LAWTON J.R., BEARD R.W., YOUNG B.J. *A decentralized approach to formation maneuvers* // IEEE Transactions on Robotics and Automation. – 2003. – Vol.19(6). – P. 933–941.
14. REN W., SORENSEN N. *Distributed coordination architecture for multirobot formation control* // Journal of Robotics and Autonomous Systems. – 2008. – Vol.56(4). – P. 324–333.
15. SHUANG L., DONG S., CHANGAN Z. *Coordinated motion planning for multiple mobile robots along designed paths with formation requirement* // IEEE/ASME Transactions on Mechatronics. – 2011. – Vol.16. – P. 1021–1031.
16. SOUSSELIER T., DREO J., SEVAUX M. *Line formation algorithm in a swarm of reactive robots constrained by underwater environment* // Expert Systems with Applications. – 2015. – [Электронный ресурс] – URL: <http://dx.doi.org/10.1016/j.eswa.2015.02.040>. – (дата обращения 20.11.2015.)
17. STANKEVICH L., SEREBRYAKOV S., IVANOV A. *Data mining techniques for RoboCup soccer agents* // Proc. AIS-ADM'2005. – Vol. 3505 LNAI, 2005. – P. 289–301.

18. YAN X., CHEN J., SUN D. *Multilevel-based topology design and shape control of robot swarm* // Automatica. – 2012. – Vol. 48. – P. 3122–3127.

RECONFIGURATION OF ROBOT SWARM FORMATION

Irina Vatamaniuk, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg, PhD student (St. Petersburg, 14 Line St., 39, (812) 328-70-81, vatamaniuk@ias.spb.su).

Gayane Panina, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg, Doctor of Science, professor, senior researcher (St. Petersburg, 14 Line St., 39, (812) 328-70-81, gaiane-panina@rambler.ru).

Andrey Ronzhin, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg, Doctor of Science, professor, deputy director for research (St. Petersburg, 14 Line St., 39, (812) 328-70-81, ronzhin@ias.spb.su).

Abstract: An algorithm of control and navigation for a swarm of autonomous homogeneous mobile robots involved in the formation of a convex space surface is discussed. Robot trajectories during reconfiguration minimize the cost of time and energy (computation time and the total length of robot paths), taking into account collision control. The developed algorithm has quadratic computational complexity. During the experiments reconfiguration of a swarm with various numbers of robots (from 10 to 10,000 robots) was modeled. The number of collisions was evaluated assuming massed start of all robots.

Keywords: robot swarm, topological robotics, surface imitation, group control, unmanned vehicles, robotic systems.

*Статья представлена к публикации
членом редакционной коллегии Б.Т. Поляком*

Поступила в редакцию 03.06.2015.

Опубликована 30.11.2015.

УДК 021.8 + 025.1
ББК 78.34

МЕТОД АВТОМАТИЧЕСКОГО ПЛАНИРОВАНИЯ СОВОКУПНОСТИ ТРАЕКТОРИЙ ДЛЯ НАВИГАЦИИ БЕСПИЛОТНЫХ ТРАНСПОРТНЫХ СРЕДСТВ

Яковлев К. С.¹, Баскин Е. С.², Андрейчук А. А.³

(Институт системного анализа Федерального исследовательского центра «Информатика и управление» РАН)

Рассматривается задача планирования совокупности альтернативных траекторий беспилотного транспортного средства (БТС). Эта задача является ключевой подзадачей более общей проблемы – построения множества неконфликтных траекторий для коалиции БТС. Предлагается новый метод планирования, косвенно учитывающий ограничения на динамику движения транспортного средства, а также ряд его модификаций. Приводятся результаты экспериментальных исследований. В качестве модельной рассматривается задача планирования траектории маловысотного полета беспилотного летательного аппарата мультироторного типа в городских условиях.

Ключевые слова: беспилотное транспортное средство, беспилотный летательный аппарат, интеллектуальная система управления, навигация, планирование, планирование траектории, альтернативные траектории, построение пути, поиск пути, поиск k кратчайших путей, эвристический поиск, геометрические ограничения, LIAN, MultiLIAN.

¹ Константин Сергеевич Яковлев, кандидат физико-математических наук (yakovlev@isa.ru).

² Егор Сергеевич Баскин, аспирант ИСА РАН (baskin@isa.ru)

³ Антон Андреевич Андрейчук, студент (andreychuk@mail.com).

1. Введение

В настоящее время наблюдается значительное повышение интереса к управлению комплексами взаимодействующих сложных технических систем. Так, в интеллектуальной робототехнике активно развивается направление, связанное с автоматизацией управления коллективов беспилотных транспортных средств (в терминологии искусственного интеллекта – интеллектуальных агентов), в частности – малых беспилотных летательных аппаратов мультироторного типа, таких как Parrot Ar.Drone [14, 31], Mikrokopter [28], AscTec Hummingbird [11]. При этом, когда идет речь об интеллектуальных системах управления сложными техническими объектами [4, 9], подразумевается разделение управления на так называемые уровни управления [8]. Обычно выделяются три уровня управления – стратегический, тактический и реактивный. Навигационные задачи, в том числе рассматриваемая в статье задача планирования траектории, относятся к тактическому уровню. Модель мира на тактическом уровне содержит все необходимые для построения траектории пространственные данные (и обычно является графовой моделью). Координаты целевой области передаются на тактический уровень управления со стратегического [3].

Традиционно в искусственном интеллекте (ИИ) задача планирования траектории рассматривается как задача поиска пути на графе и разделяется на две подзадачи: построения графа и поиска пути на этом графе. Известно достаточно много различных графовых моделей, извлекаемых из пространственных данных, подходящих для решения задачи планирования – см., например, обзор в [5]. Также в ИИ известно большое число алгоритмов поиска пути на графе (в основном – эвристических) – A^* [19], Θ^* [29], R^* [27], JPS [18] и др. Эти алгоритмы подразумевают поиск единственного пути на графе, следовательно, подходят для автоматизации управления одним агентом. Для построения согласованного множества неконфликтных траекторий используются методы, использующие вышеуказанные алгоритмы в качестве отдельных составных частей. Другой важной компонентой являются алгоритмы, реализующие стратегии разрешения конфликтов, которые мож-

но разбить на два класса: централизованные и децентрализованные. При централизованном подходе используется единая пространственная модель местности и информация обо всех агентах коллектива, конфликты устраняются на этапе построения траекторий. Такой подход является весьма ресурсоемким, так как пространство поиска растет экспоненциально с числом вовлеченных агентов [25, 33], и его использование на практике весьма ограничено. При децентрализованном подходе считается, что отдельные агенты могут обладать различными пространственными моделями, согласование которых возможно лишь при соблюдении определенных условий (например, лишь при нахождении агентов на определенном расстоянии друг от друга, необходимом для установления связи), поэтому отдельные агенты строят траектории достижения своих целей независимо друг от друга, а потенциально возникающие конфликты разрешаются на этапе выполнения. К подобным алгоритмам можно отнести, например, FAR [34], WHCA* [32], MAPP [35]. Важной компонентой этих алгоритмов является построение альтернативных траекторий для отдельных агентов. Именно за счет наличия таких траекторий и устраняются конфликты. То есть сначала для каждого агента строится множество альтернативных траекторий, а затем (на этапе разрешения конфликтов) эти траектории согласуются между собой по определенным правилам, и за каждым агентом закрепляется уже вполне определенная траектория из построенного ранее множества. Таким образом, большую важность и актуальность приобретает задача построения совокупности траекторий для отдельного агента, решению которой и посвящена данная работа.

Дальнейшее изложение организовано следующим образом: в разделе 2 описывается рассматриваемая задача и описывается её формальная постановка (задача поиска множества путей на графе особой структуры), в разделе 3 описываются и анализируются существующие методы решения, предлагается новый метод и его модификации, раздел 4 посвящен экспериментальному исследованию предложенных алгоритмов.

2. Постановка задачи

В качестве модельной задачи будет рассматриваться задача планирования совокупности различных траекторий для малого беспилотного летательного аппарата вертикального взлета и посадки мультироторного типа (например, AscTec Hummingbird), осуществляющего маловысотный полет в городских условиях. Полет осуществляется с фиксированной скоростью, в одной плоскости, ниже уровня высотных строений. Каждое строение представляет собой замкнутый многоугольник, координаты вершин которого известны. В качестве источника пространственных данных используется открытая геоинформационная база данных – OpenStreetMaps (OSM) [30].

В работе [7] представлен подход к планированию траектории БЛА в описываемых условиях, опирающийся на следующие положения. Ограничения на динамику движения объекта управления, изначально заданные в виде системы дифференциальных уравнений (модель динамики БЛА), могут быть сведены к геометрическим ограничениям, накладываемым на форму траектории. Последние являются ограничениями на максимальный угол отклонения сегментов траектории, т.е. «считается, что реализуемая (для заданного режима полета) траектория представляет собой последовательность отрезков таких, что угол отклонения между любыми смежными отрезками последовательности не превышает (по модулю) некоторого фиксированного значения α » [7]. Для построения траектории с учетом геометрических ограничений используются графы особой структуры – графы регулярной декомпозиции (ГРД), они же МТ-графы (метрические топологические графы), они же в англоязычной терминологии grids.

Заметим, что указанный подход к учету динамики объекта управления при планировании траектории не является единственным. Так, когда речь идет об объектах, обладающих достаточно простой динамикой движения – например, автомобилях – находят широкое применение методы поиска в пространстве управлений. Граф поиска в этом случае строится итерационно с помощью параметризованного стохастического моделирования, фаза поиска пути на графе совмещается с фазой построения

графа [26]. Для проверки связности вершин графа на каждой итерации алгоритма выполняется моделирование следования по соответствующей траектории. Это весьма ресурсоемкий процесс, временные затраты на осуществление которого напрямую зависят от сложности модели динамики движения объекта управления. Именно поэтому подобный подход обычно применяется для планирования траектории транспортных средств, обладающих достаточно простой моделью движения – в основном для колесных [2, 12], а не для БЛА мультироторного типа.

Существуют также альтернативные способы перехода от ограничений на динамику движения объекта управления к геометрическим ограничениям. Так весьма распространен подход, когда граф, моделирующий окружающую среду, содержит не только информацию о возможных положениях объекта в пространстве, но и о его ориентации (обычно дискретизированной до нескольких десятков значений). При этом вопрос об определении ребёр графа решается эвристически – обычно исследователями определяется набор правил перехода, запрещающих «резко» менять пространственную ориентацию объекта [24]. Поиск пути на подобных графах – гораздо менее трудоемкий процесс, чем упомянутый выше поиск пути в пространстве управлений. Развитием этой идеи и является подход, принятый в данной работе, когда считается, что исполняемая траектория состоит из последовательности прямолинейный секций, таких что угол отклонения между двумя смежными секциями не превышает заданного значения. Впервые алгоритм планирования с учетом ограничения на угол отклонения был представлен в [23] и апробирован в составе системы управления беспилотным катером. К сожалению, предложенный алгоритм не является полным, т.е. не гарантирует отыскание пути (даже если путь существует). Для устранения этого недостатка в [7] был предложен алгоритм LIAN, который является полным для вполне определенного класса задач. Применимость LIAN (и его модификаций) для решения задач планирования траектории маловысотного полета в городских условиях показана в [36].

Итак, задача планирования совокупности траекторий для БЛА мультироторного типа рассматривается в данной работе как задача поиска k путей на графе особой структуры – ГРД/МТ-

графе. При этом (для выдерживания ограничений на динамику движения БЛА) искомые пути должны соответствовать ограничениям на максимальный угол отклонения. Формальная постановка задачи приведена ниже.

Будем считать, что ГРД/МТ-граф – это четверка

$\mathbf{Gr} = \langle A, ADJ, d, los \rangle$, где:

A – множество клеток (вершин), представляющее собой матрицу $A_{m \times n} = \{a_{ij}\}$: $a_{ij} = 0 \vee a_{ij} = 1, \forall i, j: 0 \leq i < m, 0 \leq j < n, m, n \in \mathbf{N}$; клетку будем называть проходимой, если $a_{ij} = 0$; непроходимой, если $a_{ij} = 1$;

$ADJ \subseteq A \times A$ – отношение, задающее смежность на множестве клеток (множество смежных клеток);

$d: A \times A \rightarrow \mathfrak{R}$ – метрика на множестве A (функция задающая способ определения расстояния между клетками);

$los: A^+ \times A^+ \rightarrow \{true, false\}$, где $A^+ = \{a_{ij} \mid a_{ij} \in A, a_{ij} = 0\}$ – функция видимости (от англ. line-of-sight).

Здесь и далее будем периодически использовать запись $a(i; j)$, или просто $(i; j)$, для обозначения клетки (вершины) a_{ij} . Индексы i, j – координаты клеток (вершин). В тех случаях, когда координаты не важны, будем обозначить клетки буквами латинского алфавита: a, b, c и т.д.

Две различные клетки $(i_1; j_1)$ и $(i_2; j_2)$ будем считать смежными, если $|i_1 - i_2| \leq 1 \vee |j_1 - j_2| \leq 1$. Фактически смежным клеткам соответствуют смежные элементы матрицы $A_{m \times n}$.

Будем считать, что функция d определена следующим образом:

$$(1) \quad d(a_{ij}, a_{lk}) = \sqrt{(l - i)^2 + (k - j)^2},$$

т.е. для измерения расстояний используется евклидова метрика.

Нуль-траекторией $nt(a, b)$ между двумя проходимыми клетками a, b будем называть последовательность клеток, построенную по алгоритму Брезенхема [13]. Нуль-траектории соответствует отрезок прямой на плоскости, соединяющий центры соответствующих клеток – см. рис. 1.

Будем также считать, что функция видимости – los – возвращает $true$ только в том случае, если соответствующая нуль-траектория не содержит непроходимых клеток.

Упорядоченную пару клеток будем называть секцией и обозначать $e = \langle a, b \rangle$. Длиной секции будем называть величину: $len(e) = len\langle a, b \rangle = d(a, b)$. Секция $\langle a, b \rangle$ проходима, если $los(a, b) = true$.

Две секции, имеющие хотя бы одну общую клетку $e_1 = \langle a_{ij}, a_{lk} \rangle$, $e_2 = \langle a_{lk}, a_{vw} \rangle$, будем называть смежными. Углом отклонения секции e_2 от $e_1 - \alpha(e_1, e_2)$ — будем называть модуль угла между векторами $\overrightarrow{a_{ij}a_{lk}}$ и $\overrightarrow{a_{lk}a_{vw}}$, координаты которых равны $(l - i, k - j)$ и $(v - l, w - k)$ соответственно (см. рис. 1).

Пусть зафиксированы две клетки графа: начальная — s , и целевая — g . Путем из s в g будем считать последовательность смежных, проходимых секций: $\pi(s, g) = \{e_1, \dots, e_v\}$: $e_1 = \langle s, x \rangle$, $e_v = \langle z, g \rangle$. Аналогично можно определить путь как последовательность клеток. Длина пути — сумма длин соответствующих секций: $len(\pi) = len(e_1) + \dots + len(e_v)$. Величину $\alpha_m(\pi) = \max\{\alpha(e_1, e_2), \alpha(e_2, e_3), \dots, \alpha(e_{v-1}, e_v)\}$ будем называть максимальным углом отклонения пути π .

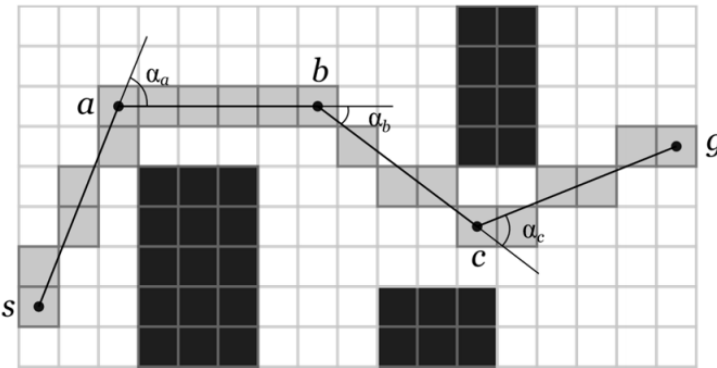


Рис. 1. Задача планирования траектории как задача поиска пути на графе особой структуры. $\pi = \{s, a, b, c, g\}$ — путь, состоящий из проходимых секций (обозначены линиями): $\langle s, a \rangle$, $\langle a, b \rangle$, $\langle b, c \rangle$, $\langle c, g \rangle$. Углы отклонения секций обозначены как α_a , α_b , α_c . Серым цветом обозначены клетки, найденные по алгоритму Брезенхема. Темным цветом обозначены непроходимые клетки

Два пути $\pi_1(s, g)$ и $\pi_2(s, g)$ будем считать различными, если они различаются хотя бы парой секций (или, что то же самое, хотя бы одной вершиной).

Задача планирования совокупности траекторий формально представляет из себя пятерку

$$\mathbf{PTask} = \langle \mathbf{Gr}, s, g, \alpha_m, k \rangle$$

и формулируется следующим образом. Задан ГРД/МТ-граф \mathbf{Gr} и зафиксированы две вершины: начальная – s и целевая – g . Необходимо найти k путей π_1, \dots, π_k , таких что $\alpha_m(\pi_i) \leq \alpha_m \quad \forall \pi_i: i = 1, \dots, k$.

3. Методы решения

3.1. ОБЗОР МЕТОДОВ ПОИСКА К ПУТЕЙ НА ГРАФАХ

Задача планирования совокупности различных траекторий, сформулированная выше как задача поиска нескольких путей на графе специального вида, в современной литературе чаще всего рассматривается как задача поиска k кратчайших путей. При этом рассматриваются пути, которые различаются по крайней мере одной вершиной.

Одним из наиболее известных алгоритмов, решающих задачу поиска k кратчайших путей, является алгоритм EA (Eрstein's algorithm) описанный в работе [16]. Алгоритм начинают работу с того, что строит кратчайший путь из начальной вершины до всех остальных вершин в графе с помощью алгоритма Дейкстры [15]. Временная оценка этой операции $O(m + n \log n)$, где n – количество вершин, m – количество ребер в графе. На втором шаге алгоритм EA строит граф путей $P(G)$. Граф $P(G)$ в неявном виде содержит в себе искомые k кратчайших путей и для их построения необходимо еще раз применить алгоритм Дейкстры, что потребует ещё $O(k \log k)$ времени. В работе [17] был описан более эффективный способ извлечения k кратчайших путей из $P(G)$, имеющий временную сложность $O(k)$. Таким образом, итоговая оценка асимптотической сложности алгоритма EA составляет $O(m + n \log n + k)$. Данная оценка является самой эффективной для алгоритмов поиска k кратчайших путей.

В работе [22] представлен алгоритм REA (Recursive Enumeration Algorithm). На первом шаге, как и алгоритм EA, REA

строит кратчайший путь из начальной вершины до каждой вершины в графе. Однако на втором шаге алгоритм не строит граф $P(G)$. Вместо этого алгоритм REA работает рекурсивно и для поиска каждого последующего пути использует предыдущий путь. Это позволяет ему быстрее отыскивать решения во многих практических задачах, несмотря на то, что асимптотическая оценка сложности REA выше, чем у EA.

В работе [21] описан алгоритм LVEA (lazy version of Erstein's algorithm). Данный алгоритм имеет ту же асимптотическую оценку сложности, что и алгоритм EA, однако на практике работает быстрее. Значительную часть времени алгоритм EA тратит на то, чтобы построить граф $P(G)$. Алгоритм LVEA строит лишь часть этого графа, необходимую для нахождения k кратчайших путей до целевой вершины. В работе [21] было показано, что алгоритм LVEA работает существенно быстрее не только алгоритма EA, но и во многих случаях быстрее алгоритма REA.

Общим недостатком алгоритмов EA, REA и LVEA является первый шаг, который требует построения всех кратчайших путей из начальной вершины (до всех остальных) с помощью алгоритма Дейкстры. Построение этих путей на графах, содержащих несколько десятков тысяч вершин (а именно такие графы обычно рассматриваются в практических задачах, связанных с построением траектории), чрезвычайно ресурсоемко. Для того чтобы снизить вычислительные затраты, был предложен алгоритм K^* [10]. Идея алгоритма K^* заключается в использовании эвристического алгоритма A^* вместо алгоритма Дейкстры и построении кратчайшего пути (с помощью A^*) лишь из начальной вершины в целевую. Во время этого процесса также формируется часть графа $P(G)$. После того как путь найден, с помощью алгоритма Дейкстры из графа $P(G)$ извлекается k кратчайших путей. Если в графе $P(G)$ меньше чем k путей, алгоритм расширяет граф $P(G)$ за счет обхода большего числа вершин в графе G . Операция повторяется до тех пор, пока не будет построено k кратчайших путей. Алгоритм K^* имеет ту же асимптотическую оценку сложности, что и алгоритмы EA и LVEA, но в большинстве практических задач его показатели вычислительной эффективности существенно выше. И чем

больше размер исходного графа G , тем выше разница в этих показателях в пользу алгоритма K^* .

Все рассмотренные алгоритмы применимы на различных типах графов, в том числе и на МТ-графах. Однако эти алгоритмы не применимы в контексте поставленной нами задачи, так как в принципе не могут строить траектории с учетом ограничений на динамику движения БЛА. Это обуславливает необходимость разработки нового метода поиска совокупности путей, применимого на МТ-графах и учитывающего ограничения на максимальный угол отклонения. Такой метод будет описан ниже.

3.2. АЛГОРИТМ MULTILIAN

Предлагаемый алгоритм поиска совокупности путей, удовлетворяющих ограничениям на максимальный угол отклонения (и тем самым косвенно удовлетворяющим ограничениям на динамику движения беспилотного летательного аппарата), является модификацией предложенного ранее алгоритма LIAN, впервые описанного в работе [7]. Алгоритм LIAN основывается на тех же принципах, что и известный в литературе по искусственному интеллекту алгоритм A^* [19], и в упрощенном виде принцип работы алгоритма можно описать следующим образом (подробнее см. в работах [7, 36]).

Алгоритм осуществляет сфокусированный (с помощью эвристики) перебор вершин графа и поддерживает в памяти два списка вершин: список *OPEN*, в котором располагаются нерассмотренные вершины графа – кандидаты на дальнейшую обработку, и список *CLOSE* – множество уже рассмотренных вершин. На каждой итерации алгоритма из списка *OPEN* выбирается вершина $a = \operatorname{argmin}_{a \in \text{OPEN}} \{f(a)\}$, где $f(a) = g(a) + h(a)$, где $g(a)$ – длина кратчайшего пути из s в a , найденного к текущему моменту (эта величина может изменяться в ходе работы алгоритма), $h(a)$ – эвристическая оценка длины пути из a в g ($h(a) = d(a, g)$). Возможно также использование других эвристик – см. например [6]. Далее для выбранной вершины генерируются потомки, т.е. такие вершины a' , для которых выполняются следующие условия:

1. Вершина a' проходима ($a' \in A^+$).

2. Расстояние от вершины-потомка до вершины-родителя приблизительно равно заранее заданной константе Δ , где Δ – входной параметр алгоритма ($d(a, a') \approx \Delta$).
3. Угол между двумя секциями, которые образуют вершина, ее родитель и родитель родителя, не превышает по модулю заданного значения α_m ($\alpha(e_1, e_2) < \alpha_m$, $e_1 = \langle bp(a), a \rangle$, $e_2 = (a, a')$, $bp(a)$ – родитель вершины a).
4. Между a и a' нет препятствий, другими словами – секция $\langle a, a' \rangle$ является проходимой (т.е. $los(a, a') = true$, проверяется с помощью алгоритма Брезенхема).

Сгенерированные таким образом вершины добавляются в список *OPEN*, и процесс повторяется. Если на каком-то шаге окажется, что целевая вершина находится на расстоянии меньшем либо равном Δ от текущей вершины, для целевой вершины удовлетворяется ограничение на угол и между целевой и текущей вершиной отсутствуют препятствия, то алгоритм завершает свою работу – искомый путь найден (и может быть восстановлен с помощью родительских указателей). Если в процессе работы алгоритм список *OPEN* исчерпан, LIAN возвращает *failure*, т.е. сообщение о том, что пути не существует.

В работах [1, 7, 36] показано, что алгоритм LIAN является корректным, полным и оптимальным (что следует из того, что LIAN использует ту же стратегию поиска в пространства состояний, что и A*), однако возможность отыскания пути на конкретной карте (при фиксированном ограничении на угол отклонения) зависит от входного параметра Δ . Заметим также, что на практике построение кратчайших путей нецелесообразно в силу чрезмерного расходования вычислительных ресурсов. Для отыскания субоптимальных решений используется техника взвешивания эвристики, т.е. использование функции $f(a)$ вида $g(a) + w \cdot h(a)$, где $w \geq 1$ – вес эвристики. Как показывают результаты экспериментов, при весе эвристики равном 2 удается существенно образом (более чем в 2–3 раза) сократить пространство поиска и повысить вычислительную эффективность алгоритма (в частности, скорость работы), при этом качество отыскиваемых решений (длины построенных путей) снижается незначительно (менее 2–3%).

Перед тем как перейти к описанию модификации алгоритма, предназначенной для поиска k различных путей, введем дополнительные определения.

Пусть существует некоторое множество промежуточных вершин графа $Q = \{q_{xy}\}$, чьи координаты x и y могут принимать значения отличные от $[0, \dots, m - 1]$, $[0, \dots, n - 1]$ соответственно. В частности, координаты могут быть отрицательными. При этом будем считать, что вершины из множества Q входят в область определения метрической функции d , т.е. известен способ определения расстояния между вершинами из множества Q и/или вершинами **Gr**. Будем считать, что в множество Q входит L элементов.

Рассмотрим теперь более подробно функцию выбора вершины из списка *OPEN*:

$$(2) \quad f(a) = g(a) + w \cdot h(a).$$

Как было сказано выше, $g(a)$ – это длина кратчайшего пути из s в a , найденного к текущей итерации алгоритма. Этот путь представим в виде последовательности секций, или, что то же самое, в виде последовательности клеток (см. раздел 1). Воспользуемся последним представлением и обозначим вершины, входящие в путь, как p_i : $\pi(s, a) = \{p_1, \dots, p_i, \dots, p_k\}$, $p_1 = s$, $p_k = a$. Заметим, что нижние индексы используются для нумерации вершин в последовательности, а не для обозначения индексных координат клеток, как ранее. Модифицируем теперь эвристическую функцию $h(a) (= h(p_k))$ следующим образом:

$$(3) \quad h'(p_k, Q) = d(p_k, g) + \theta \sum_{l=1}^L \min_{i=1, \dots, k} d(q_l, p_i).$$

Первое слагаемое – расстояние от текущей вершины до целевой, т.е. $h(p_k)$. Второе слагаемое есть сумма минимальных расстояний от каждой из вершин множества Q до вершин входящих в путь, умноженная на коэффициент θ . Геометрическая интерпретация этого слагаемого представлена на рис. 2.

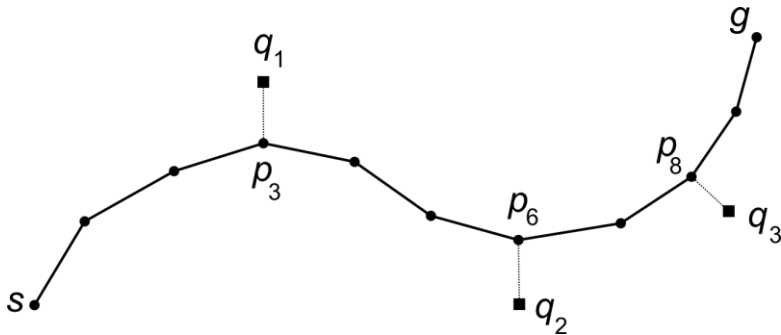


Рис. 2. Путь из s в g и промежуточные вершины (q_1, q_2, q_3)

На приведенном рисунке вершина p_3 является ближайшей вершиной к q_1 , p_6 – к q_2 , p_8 к q_3 . Таким образом, для вершины p_8 имеем:

$$(4) \quad h'(p_8) = \theta[(q_1, p_3) + d(q_2, p_6) + d(q_3, p_8)].$$

Применение данной эвристической функции на практике приводит к такому порядку рассмотрения вершин графа, что, во-первых, минимизируется длина построенного пути (как ранее), во-вторых – построенный путь проходит как можно ближе к промежуточным вершинам. При этом следует заметить, что путь не обязан проходить непосредственно через них. Величина коэффициента θ определяет, насколько высоко будет цениться путь, проходящий близко (через) промежуточные вершины, по сравнению с просто кратчайшим путем. Если множество промежуточных вершин пусто, эвристика имеет в точности такой же вид, как и в оригинальном алгоритме.

Заметим, что использование определенной выше эвристической функции при поиске не нарушает свойств корректности и полноты, т.е. алгоритм завершает работу за конечное число шагов и находит путь, если он существует. Доказательство этих утверждений повторяет доказательство, приведенное в [1, 7, 36], поэтому ограничимся здесь лишь его кратким описанием. Доказательство первого утверждения основывается на том соображении, что на каждом шаге, независимо от эвристики, обрабатывается одна вершина графа, а так как число вершин ограничено, алгоритм завершит работу за конечное число ша-

гов. Доказательство второго утверждения основывается на том, что, если путь существует, то рано или поздно (за определенное число шагов) алгоритм рассмотрит все вершины, образующие его, а следовательно, найдет и сам путь. Следует, однако, заметить, что на практике при решении некоторых задач это число шагов может быть весьма велико.

Теперь приведем описание алгоритма MultiLIAN – алгоритма построения множества путей с использованием промежуточных вершин графа:

1. Построить путь из s в g с помощью алгоритма LIAN.
2. Определить множество промежуточных вершин.
3. Построить путь из s в g с учетом промежуточных вершин (т.е. используя эвристическую функцию описанную выше) – альтернативный путь.
4. Если построено заданное число путей, завершить алгоритм.
5. Вернуться к пункту 2.

Псевдокод алгоритма MultiLIAN приведен на рис. 3. Заметим, что во многом этот код идентичен LIAN (см. [28]). Основные различия: итерация по заданному (в качестве входного параметра) числу путей k (строка 3), наличие процедуры выбора промежуточных вершин (строка 4), способ приоритизации списка *OPEN* с учетом выбранных промежуточных вершин (строка 8–9), наличие переменной *PATHS* – списка, содержащего все построенные пути.

Вид построенных алгоритмом альтернативных путей целиком зависит от того, каким образом выбираются промежуточные вершины. Здесь и далее будем рассматривать задачу построения 2 альтернативных путей, при этом будем считать, что для построения каждого из них используется одна промежуточная вершина (т.е. всего используются две различные промежуточные вершины). Предлагается определять их координаты следующим образом. Строится отрезок, соединяющий начальную и целевую вершины. На данном отрезке отмечается точка, находящаяся на расстоянии a от начальной вершины. От этой точки откладывается два перпендикулярных отрезка длины b , концы которых и определяют координаты промежуточных вершин, см. рис. 4. Значения a и b являются входными параметрами алгоритма MultiLIAN.

| | |
|--|--|
| <pre> 1. MULTILIAN(<i>start</i>, <i>goal</i>, Δ, α_m, <i>k</i>) 2. PATHS := \emptyset 3. for <i>i</i> from 1 to <i>k</i> 4. $Q_i := \text{GetIntermediateAttractors}()$; 5. $bp(\text{start}) := \emptyset$; $g(\text{start}) := 0$; 6. OPEN := \emptyset, CLOSED := \emptyset; 7. OPEN.push(start); 8. while OPEN.size > 0 9. $[a] := \text{argmin}_{[a] \in \text{OPEN}} f([a])$; 10. // $f([a]) = g([a]) + h'([a], Q_i)$ 11. OPEN.remove([a]); 12. if $a = \text{goal}$ 13. $path_i = \text{getPathFromParentPoint}$ 14. $ers([a])$; 15. PATHS.push($path_i$); 16. continue to next path; 17. CLOSED.push([a]); 18. Expand([a], Δ, α_m); 19. if PATHS = \emptyset 20. return "paths not found" 21. return PATHS; 22. end </pre> | <pre> 23. SUCC = 24. calculateCircleSuccessors([a], Δ); 25. if $\text{dist}(a, \text{goal}) < \Delta$ 26. SUCC.push(goal); 27. for each $[a'] \in \text{SUCC}$ 28. if a' is un-traversable 29. continue; 30. if $\alpha(\langle bp(a), a \rangle, \langle a, a' \rangle) > \alpha_m$ 31. continue; 32. if $\text{LineOfSight}(a, a') = \text{false}$ 33. continue; 34. for each $[a''] \in \text{CLOSED}$ 35. if $a' = a''$ and $bp(a') = bp(a'')$ 36. continue; 37. $g(a') = g(bp(a')) + \text{dist}(a', bp(a'))$ 38. OPEN.push([a']); 39. end </pre> |
|--|--|

Рис. 3. Псевдокод алгоритма MultiLIAN

Описанный способ генерации множества промежуточных вершин подходит для построения трех траекторий: основной и двух альтернативных. В случае если требуется построение большего числа альтернативных траекторий, необходимо определить способ генерации промежуточных вершин исходя из условий задачи и свойств карты. Например, можно откладывать с каждой стороны отрезка $[s, g]$ несколько вершин, находящихся на разных расстояниях от отрезка, соединяющего начальную и целевую вершины. В целом, способ генерации промежуточных вершин может являться предметом отдельных исследований.

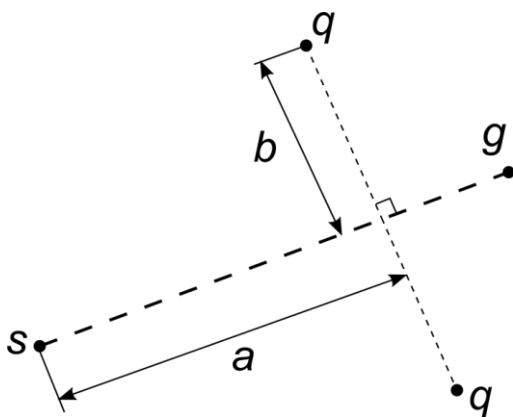


Рис. 4. Параметры a и b , определяющие расположение промежуточных вершин относительно начального и целевого расположения

Заметим, что в некоторых случаях MultiLIAN может строить альтернативные пути, идентичные первому построенному. Поясним на примере, когда множество промежуточных вершин Q для построения очередного альтернативного пути состоит из одной вершины q_1 , и используется стратегия выбора этой вершины, описанная выше. Если препятствие большого размера находится на прямой, соединяющей начальную и целевую вершины, то найденный изначально путь может совпасть с путем, построенной с учетом промежуточных вершины, см. рис 5. Решить эту проблему можно различными способами выбора промежуточных вершин. В рамках данной работы такой подход не рассматривается, но может стать темой последующих исследований.

Представленный подход к построению множества путей может быть охарактеризован как метод грубой силы (англ. – brute force), так как на каждой итерации MultiLIAN альтернативный путь ищется без учета результатов предыдущих итераций. Очевидно, что последние могут быть использованы для повышения вычислительной эффективности алгоритма. Ниже будут описаны две модификации алгоритма MultiLIAN, которые

определенным образом учитывают результаты промежуточных итераций для сокращения пространства поиска.

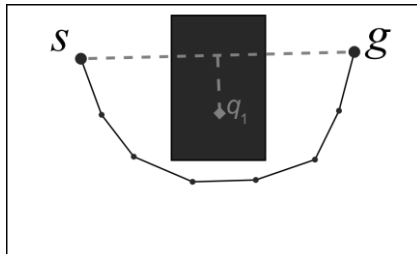


Рис. 5. Случай, когда альтернативный путь совпадает с первым построенным из-за расположения промежуточной вершины q_1

Итак, после завершения первой итерации алгоритма построен путь из s в g , а также образованы списки $OPEN$ и $CLOSED$. Каждый элемент этих списков фактически определяет частичный путь до соответствующей вершины графа из начальной. При этом чем «сложнее» граф, на котором ищется путь, и чем больше на нем непроходимых клеток, тем больше вершин будет в этих списках (см. рис. 6).

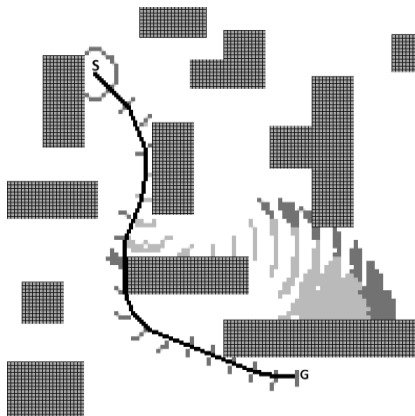


Рис. 6. Визуализация списков $OPEN$ (темно-серые вершины) и $CLOSED$ (светло-серые вершины) после отыскания пути

Логично в качестве промежуточных вершин выбирать вершины из множества $OPEN \cup CLOSE$, так как пути до этих вершин уже известны – в таком случае остается лишь достроить данный частичный путь (а не строить его полностью заново). Выбор вершин из множества $OPEN \cup CLOSE$ осуществляется, как и ранее, с помощью расчета расстояний до определенных промежуточных вершин. При этом при поиске альтернативного пути промежуточные вершины более не используются (т.е. используется оригинальная эвристика алгоритма LIAN).

Описанный подход получил название MultiLIAN-RU (RU – от англ. reuse, повторное использование). Наряду с очевидными достоинствами (повторное использование уже полученных на предшествующих итерациях данных) отметим и недостатки LIAN-RU. Во-первых, при использовании взвешенной эвристики (а на практике чаще всего используется именно она) число вершин в списках $OPEN$ и $CLOSE$ может быть невелико, что естественным образом ограничивает выбор. Во-вторых, путь из этих вершин до целевой может не существовать из-за угловых ограничений (начальная вершина нового поиска теперь характеризуется не только своими координатами, но и родителем, т.е. уже на первом шаге нового LIAN-поиска велика вероятность отсутствия вершин-потомков, удовлетворяющих ограничению на максимальный угол отклонения).

Другая модификация алгоритма MultiLIAN, использующая информацию предыдущих итераций, основывается на следующем предположении. После шага 1 алгоритма список $CLOSE$ содержит в себе, наряду с вершинами, образующими искомый путь, вершины, на обработку которых потрачено время, но при этом не вошедшие в этот самый путь. Можно предположить, что последние являются «лишними» и при очередных итерациях поиска такие вершины должны быть исключены из рассмотрения (образно говоря таким вершинам, соответствуют «тупики» на карте местности, моделью которого служит рассматриваемый граф специальной структуры). Модифицированный соответствующим образом алгоритм будем называть MultiLIAN-CC (CC – от англ. exClusion from Close).

Псевдокод алгоритма MultiLIAN-CC и MultiLIAN-RU приведен на рис. 7 и 8. Заметим, что MultiLIAN-RU использует ту же процедуру Expand(), что и оригинальный MultiLIAN (поэтому на рисунке она не приведена), а MultiLIAN-CC – модифицированную процедуру.

```

1.  MULTILIAN-RU(start, goal, Δ, αm, k)
2.  PATHS := ∅
3.  for i from 1 to k
4.  | [start] = Get_start(i);
5.  | OPEN := ∅, CLOSED := ∅;
6.  | OPEN.push([start]);
7.  | while OPEN.size > 0
8.  | | [a] := argmin[a]∈OPEN f([a]);
9.  | | // f([a]) = g([a]) + h([a])
10. | | OPEN.remove([a]);
11. | | if a = goal
12. | | | pathi = getPathFromParentPoint
13. | | | | | ers([a]);
14. | | | | PATHS.push(pathi);
15. | | | | break;
16. | | | CLOSED.push([a]);
17. | | | Expand([a, Δ, αm]);
18. | | if i=1
19. | | | FormAlternativeStarts(CLOSE, k);
20. | | if PATHS = ∅
21. | | | return “paths not found”;
22. | | return PATHS;
23. end
24. FormAlternativeStarts(CLOSE, k)
25. | for i from 2 to k
26. | | q = GetFirstIntermediateAttractor();
27. | | [start]i := argmin[a]∈CLOSE dist([a], q);
28. | end
29. Get_start(i)
30. | if i = 1
31. | | return [start];
32. | else
33. | | return [start];
34. | end

```

Рис. 7. Псевдокод алгоритма MultiLIAN-RU

```

1. MULTILIAN-CC(start, goal,  $\Delta$ ,  $\alpha_m$ , k)
2. EXCLOSED :=  $\emptyset$ ; PATHS :=  $\emptyset$ ;
3. for i from 1 to k
4.   bp([start]) :=  $\emptyset$ ; g([start]) := 0;
5.   OPEN :=  $\emptyset$ , CLOSED :=  $\emptyset$ ;
6.   OPEN.push([start]);
7.   while OPEN.size > 0
8.     [a] := argmin[a] ∈ OPEN f([a]);
9.     OPEN.remove([a]);
10.    if a = goal
11.      pathi = getPathFromParentPointer
12.        s([a]);
13.      PATHS.push(pathi);
14.      CLOSED.push([a]);
15.      Expand([a,  $\Delta$ ,  $\alpha_m$ ]);
16.    if i=1
17.      EXCLOSED := CLOSED;
18.    end
19. Expand([a],  $\Delta$ ,  $\alpha_m$ )
20. SUCC = calculateCircleSuccessors([a],
21.   $\Delta$ );
22. if dist(a, goal) <  $\Delta$ 
23.   SUCC.push([goal]);
24. for each [a'] ∈ SUCC
25.   if a' is un-traversable
26.     continue;
27.   if | $\alpha$ (⟨bp(a), a⟩, ⟨a, a'⟩)| >  $\alpha_m$ 
28.     continue;
29.   for each [a''] ∈ CLOSED
30.     if a'=a'' and bp(a')=bp(a'')
31.       continue;
32.   for each [a'''] ∈ EXCLOSED
33.     if a'=a'''
34.       continue;
35.   if LineOfSight(a, a') = false
36.     continue;
37.   OPEN.push([a']);
38. end

```

Рис. 8. Псевдокод алгоритма MultiLIAN-CC

4. Экспериментальный анализ

4.1. МЕТОДИКА ПРОВЕДЕНИЯ ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ

Для проведения экспериментальных исследований разработанного алгоритма MultiLIAN (а также его модификаций) рассматривалась модельная задача построения траектории маловысотного полета БЛА мультироторного типа AscTec Hummingbird [11] в городских условиях. Источником исходных данных выступала открытая геоинформационная система

OpenStreetMaps [30]. Графы для проведения экспериментов представляли собой модели фрагментов реальных городских карт (в частности, использовались карты города Москва). Размер отдельного фрагмента составлял $1347 \times 1347 \text{ м}^2$, а соответствующий ГРД/МТ-граф имел размер 501×501 , т.е. каждая вершина соответствовала области размером $\sim 2,7 \times 2,7 \text{ м}^2$. Выбор такого размера единичной области обусловлен физическими размерами мультикоптера AscTec Hummingbird, а также тем фактом, что обеспечить строгое следование по прямолинейной траектории не под силу ни одному известному управляющему регулятору, поэтому необходим некоторый пространственный запас (т.е. каждая единичная область больше по площади, чем площадь проекции объекта управления на рабочую плоскость). Подробнее о принципах выбора шага дискретизации карты сказано в [7]. Вершины, соответствующие зданиям, считались непроходимыми. Исследования проводились на двух коллекциях. В первой коллекции содержалось 80 различных графов, во второй – 100. На каждом графе было выбрано по 5 различных расположений начальной и целевой вершин, таким образом, общее число заданий составило 900. Сами же начальная и целевая вершины выбирались таким образом, что расстояние между ними было примерно одинаковым (образно говоря, начальное и целевое расположение беспилотного летательного аппарата выбирались на противоположных концах карт). Соотношение непроходимых вершин к общему числу вершин в графе (плотность препятствий) в обеих коллекциях составляет примерно 25%, однако в первой коллекции этот показатель варьируется в пределах от 15 до 35%, в то время как во второй коллекции этот показатель варьируется лишь в пределах 20–25% (т.е. вторая коллекция более однородна по плотности). На рис. 9 представлено изображение одного из использованных графов (линии сетки не показаны для наглядности, отмечены также найденные пути).



Рис. 9. Пример задания. Сплошной линией обозначен основной путь, пунктирной – альтернативные пути, найденные с использованием промежуточных вершин

Следуя рекомендациям работ [7, 36] (подтвержденным результатами предварительной серии экспериментов), алгоритм LIAN был инициирован следующим образом: вес эвристики равнялся двум ($w = 2$), параметр Δ равнялся пяти ($\Delta = 5$). Ограничение на максимальный угол отклонения составляло 25° . Такое ограничение (см. [7]) соответствует полету БЛА мультироторного типа AscTec Hummingbird [11] на скорости 4,5 м/с.

На выполнение отдельного задания было установлено временное ограничение – 60 секунд. Если алгоритм не укладывался в отведенное время, то результатом работы считалось *failure* (что учитывалось при подсчете значений интегральных показателей качества работы алгоритма – см. ниже).

Экспериментальное исследование проводилось на персональном компьютере архитектуры x86 с процессором iCore2Quad 2,5 ГГц и 2 Гб RAM, работающим под управлением

ОС Windows 7 64bit SP1. Программная реализация алгоритмов выполнена на языке C++.

В каждом эксперименте отслеживались значения следующих трех выходных параметров алгоритма:

1. M – (от англ. memory) число вершин, рассмотренных (и сохранённых в оперативной памяти) алгоритмом. Это число характеризует емкостную эффективность алгоритма.
2. T – (от англ. time) количество секунд, которое понадобилось алгоритму для нахождения путей. Показатель временной эффективности алгоритма.
3. PL – (от англ. path length) средняя длина найденных путей.

Также в каждой серии экспериментов отслеживались следующие интегральные показатели:

1. SR – (от англ. success rate) процент успеха, равный отношению количества успешно выполненных заданий к их общему числу (900).
2. UPR – (от англ. unique path rate) процент уникальных путей, равный отношению уникальных найденных путей к их общему (максимально возможному) числу ($900 * 3 = 2700$).
3. $PAR-10$ – (от англ. penalized average time) интегральная метрика, отражающая способность алгоритма быстро отыскивать решения задачи планирования в серии экспериментов [20]. $PAR-10 = (\Sigma T_1 + \Sigma T_2)/N$, где N – общее число экспериментов в серии (т.е. 900); T_1 – суммарное время по тем экспериментам, когда алгоритм завершил работу в установленный лимит времени (60 с); T_2 – число экспериментов, которые были прерваны (т.е. не уложились в 60 с), умноженное на 60 (временной лимит), умноженное на 10 (величина штрафа за «неспособность найти решение в отведенное время»).

4.2. ПЕРВАЯ СЕРИЯ ЭКСПЕРИМЕНТОВ

В первой серии экспериментов исследовалось влияние расположения промежуточных вершин (определяемого с помощью параметров a и b – см. раздел 2) на эффективность MultiLIAN. В таблице 1 приведены значения интегральных показателей, при этом цифры в названии алгоритма соответствуют значениям

параметров a и b , использованным при выборе промежуточных вершин.

Таблица 1. Интегральные показатели эффективности алгоритма MultiLIAN при различном расположении промежуточных вершин

| | Moscow Maps 1.0 | | | Moscow Maps 2.0 | | |
|---------------------|-----------------|--------|--------|-----------------|--------|--------|
| | SR | UPR | PAR-10 | SR | UPR | PAR-10 |
| MultiLIAN 0,5-0,5 | 94,5% | 94,5% | 29,55 | 97,6% | 97,6% | 12,91 |
| MultiLIAN 0,25-0,25 | 97% | 96,67% | 14,70 | 98% | 97,53% | 10,27 |
| MultiLIAN 0,5-0,25 | 96% | 96% | 20,45 | 97,8% | 97,8% | 11,64 |
| MultiLIAN 0,75-0,25 | 95,25% | 95,17% | 25,22 | 98% | 97,93% | 10,40 |

Полученные результаты демонстрируют явное преимущество использования параметров «0,25-0,25» и «0,5-0,25». MultiLIAN «0,25-0,25» показал лучший результат по критериям SR и PAR-10 на обеих коллекциях. Однако в некоторых случаях алгоритм с этими настройками находит несколько идентичных путей, в связи с чем показатель UPR ниже, чем SR. Подобная ситуация возникает и при использовании параметров «0,75-0,25». Заметим также, что на второй коллекции все 4 алгоритма продемонстрировали близкие результаты.

Проанализируем более детально критерий SR, а именно, рассмотрим график зависимости успешно выполненных заданий (по обеим коллекциям) от количества времени, которое потребовалось, чтобы завершить каждое задание (см. рис. 10).

График на рис. 10, как и предыдущие результаты, демонстрирует преимущество MultiLIAN с параметрами «0,25-0,25». 63,3% заданий алгоритм MultiLIAN «0,25-0,25» успешно завершил менее чем за 0,1 с. Стоит заметить, что аналогичный результат продемонстрировал MultiLIAN «0,5-0,25». Также интерес представляет количество успешно выполненных заданий,

завершившихся менее чем за 1 мин. У алгоритма MultiLIAN с параметрами «0,25-0,25» этот показатель равен 97,56%, MultiLIAN «0,5-0,25» – 97%, MultiLIAN «0,75-0,25» – 96,8%, MultiLIAN «0,5-0,5» – 96,2%.

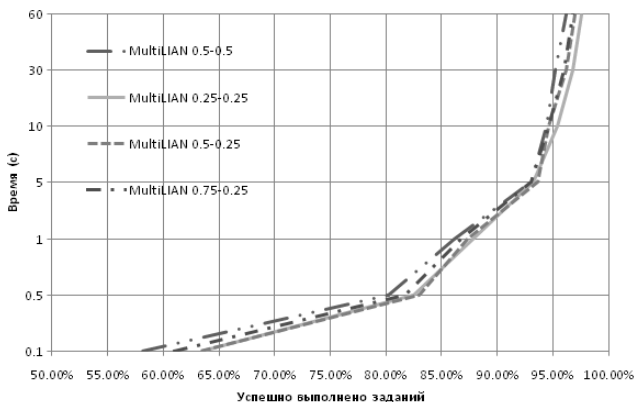


Рис. 10. График зависимости количества выполненных заданий от времени работы алгоритма

В таблице 2 представлены значения индикаторов M , T и PL . Данные получены путем усреднения результатов успешно выполненных заданий, т.е. для каждого алгоритма усреднялись только результаты успешно выполненных этим алгоритмом заданий.

Таблица 2. Индивидуальные показатели эффективности алгоритма MultiLIAN при различном расположении промежуточных вершин.

| | Moscow Maps 1.0 | | | Moscow Maps 2.0 | | |
|---------------------|-----------------|-------|-------|-----------------|-------|-------|
| | M | PL | T | M | PL | T |
| MultiLIAN 0,5-0,5 | 12023 | 715,4 | 1,102 | 8727 | 593,8 | 0,927 |
| MultiLIAN 0,25-0,25 | 11039 | 648,3 | 1,222 | 7350 | 542,0 | 0,684 |
| MultiLIAN 0,5-0,25 | 10772 | 652,4 | 0,983 | 7718 | 545,2 | 0,853 |

| | Moscow Maps 1.0 | | | Moscow Maps 2.0 | | |
|------------------------|-----------------|-----------|----------|-----------------|-----------|----------|
| | <i>M</i> | <i>PL</i> | <i>T</i> | <i>M</i> | <i>PL</i> | <i>T</i> |
| MultiLIAN 0,75-0,25 | 12363 | 664,8 | 1,277 | 8272 | 555,5 | 0,815 |

Результаты, приведенные в таблице 2, соответствуют сделанным ранее выводам о том, что почти по всем показателям на обеих коллекциях наиболее эффективно использование параметров «0,25-0,25». Также хорошие результаты демонстрирует алгоритм MultiLIAN с параметрами «0,5-0,25». Самые низкие показатели демонстрируются алгоритмом при использовании параметров «0,5-0,5». Объяснить это можно тем, что в этом случае промежуточные вершины находятся дальше всего от прямой между начальной и целевой вершинами, вследствие чего алгоритму необходимо рассмотреть больше вершин и потратить на это больше времени (по этой же причине средняя длина пути существенно больше по сравнению с аналогами).

Усредненные по обеим коллекциям и нормированные результаты представлены в виде гистограммы на рис. 11.

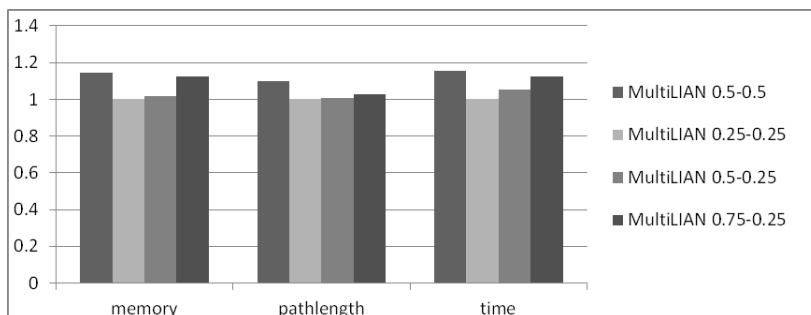


Рис. 11. Гистограммы нормированных усредненных значений индивидуальных индикаторов эффективности алгоритма MultiLIAN при различном расположении промежуточных вершин

Из рис. 6 видно, что алгоритм MultiLIAN с параметрами «0,5-0,5» тратит в среднем на 14% больше памяти и на 15% больше времени, чем MultiLIAN «0,25-0,25». Средняя длина

найденных путей у MultiLIAN «0,5-0,5» больше на 10%, а значения длины пути у остальных алгоритмов практически идентичны. Алгоритм MultiLIAN с параметрами «0,75-0,25» хуже, чем MultiLIAN «0,25-0,25» по показателям временной и ёмкостной эффективности на 12,5%.

Таким образом, в результате проведения первой серии экспериментов удалось определить наилучший способ выбора промежуточных вершин для построения альтернативных путей, а именно – наилучший способ инициализации входных параметров алгоритма MultiLIAN, отвечающих за расположение этих вершин: «0,5-0,25» и «0,25-0,25». При проведении дальнейших экспериментальных исследований использовались лишь эти варианты инициализации.

4.3. ВТОРАЯ СЕРИЯ ЭКСПЕРИМЕНТОВ

Во второй серии экспериментов сравнивались алгоритмы MultiLIAN и MultiLIAN-RU. Задания и настройки алгоритмов аналогичны тем, что были использованы ранее. Параметры выбора промежуточных вершин были установлены в соответствие с результатами первой серии экспериментов («0,5-0,25» и «0,25-0,25»). Ниже представлена таблица индивидуальных показателей эффективности.

Таблица 3. Индивидуальные показатели эффективности алгоритмов MultiLIAN и MultiLIAN-RU

| | Moscow Maps 1.0 | | | Moscow Maps 2.0 | | |
|------------------------|-----------------|-----------|----------|-----------------|-----------|----------|
| | <i>M</i> | <i>PL</i> | <i>T</i> | <i>M</i> | <i>PL</i> | <i>T</i> |
| MultiLIAN 0,25-0,25 | 11039 | 648,3 | 1,215 | 7350 | 542,0 | 0,703 |
| MLIAN-RU 0,25-0,25 | 6079 | 603,6 | 0,639 | 3992 | 503,6 | 0,404 |
| MultiLIAN 0,5-0,25 | 10772 | 652,4 | 0,983 | 7718 | 545,2 | 0,853 |
| MLIAN-RU 0,5-0,25 | 6194 | 604,5 | 0,608 | 4254 | 504,7 | 0,401 |

Основываясь на полученных данных, можно сделать вывод о том, что модификация MultiLIAN-RU существенно улучшает эффективность работы алгоритма. Затраты по памяти уменьшаются на ~80%, затраты по времени уменьшаются на ~60–110% в зависимости от коллекции и выбранных параметров. Средняя длина пути сокращается на ~8% (что весьма существенно). Однако при анализе интегральных результатов – см. таблицу 4 – становится ясно, что вывод сделан поспешно.

Таблица 4. Интегральные показатели эффективности алгоритмов MultiLIAN и MultiLIAN-RU

| | Moscow Maps 1.0 | | | Moscow Maps 2.0 | | |
|---------------------|-----------------|--------|--------|-----------------|--------|--------|
| | SR | UPR | PAR-10 | SR | UPR | PAR-10 |
| MultiLIAN 0,25-0,25 | 97% | 96,67% | 14,68 | 98% | 97,53% | 10,29 |
| MLIAN-RU 0,25-0,25 | 97,75% | 37,25% | 9,63 | 97,8% | 35,07% | 11,19 |
| MultiLIAN 0,5-0,25 | 96% | 96% | 20,45 | 97,8% | 97,8% | 11,66 |
| MLIAN-RU 0,5-0,25 | 97,5% | 37,5% | 11,09 | 98% | 36,27% | 9,99 |

Как видно из таблицы в большинстве случаев алгоритм MultiLIAN-RU не находит альтернативные траектории (см. показатель UPR). Другими словами, алгоритм (в подавляющем большинстве случаев) корректно завершается за отведенное время, однако все три найденных пути совпадают. При более детальном анализе была выяснена причина такого поведения алгоритма. Дело в том, что при использовании взвешенной эвристики размер списков *OPEN* и *CLOSE* (из которых алгоритмом MultiLIAN-RU выбираются вершины для построения альтернативных путей) крайне невелик и ближайшие вершины к промежуточным целям входят в *уже* построенный путь. Таким образом, алгоритм MultiLIAN-RU зачастую строит несколько абсолютно идентичных путей. Во многих случаях алгоритму в принципе не удастся найти альтернативные пути, что связано с тем, что вершина, выбираемая из множества $OPEN \cup CLOSE$,

уже имеет родителя и из-за ограничения на угол, заданного родителем, алгоритм не может обойти ближайшие препятствия. В целом результаты экспериментальных исследований MultiLIAN-RU подтверждают справедливость высказанных ранее (см. раздел 2) опасений, по поводу возможной неэффективности его работы.

4.4. ТРЕТЬЯ СЕРИЯ ЭКСПЕРИМЕНТОВ

В третьей серии экспериментов сравнивались алгоритмы MultiLIAN и MultiLIAN-CC. Задания и настройки алгоритмов аналогичны тем, что были использованы ранее. Параметры выбора промежуточных вершин были установлены в соответствии с результатами первой серии экспериментов («0,5-0,25» и «0,25-0,25»). Ниже представлена таблица интегральных показателей эффективности.

Таблица 5. Интегральные показатели эффективности алгоритмов MultiLIAN и MultiLIAN-CC

| | Moscow Maps 1.0 | | | Moscow Maps 2.0 | | |
|---------------------|-----------------|--------|--------|-----------------|--------|--------|
| | SR | UPR | PAR-10 | SR | UPR | PAR-10 |
| MultiLIAN 0,25-0,25 | 97% | 96,67% | 14,68 | 98% | 97,53% | 10,29 |
| MLIAN-CC 0,25-0,25 | 97,75% | 96,5% | 10,12 | 98,2% | 97,2% | 8,93 |
| MultiLIAN 0,5-0,25 | 96% | 96% | 20,45 | 97,8% | 97,8% | 11,66 |
| MLIAN-CC 0,5-0,25 | 96,5% | 96,5% | 17,49 | 98% | 98% | 10,24 |

Как видно из представленных данных, MultiLIAN-CC (как и MultiLIAN) при значениях входных параметров «0,25-0,25» не всегда находит различные пути, в связи с чем значение показателя UPR несколько ниже значения SR. Впрочем, это не столь критично, как в случае с алгоритмом MultiLIAN-RU, так как при использовании параметров «0,5-0,25» уже все найденные альтернативные пути различны. В целом, если сравнивать MultiLIAN с MultiLIAN-CC при одинаковой параметризации, то

можно сделать вывод о том, что MultiLIAN-CC более эффективен, чем базовый алгоритм. Этот вывод подтверждается и при детальном анализе зависимости количества успешно выполненных заданий от времени работы (см. рис. 12).

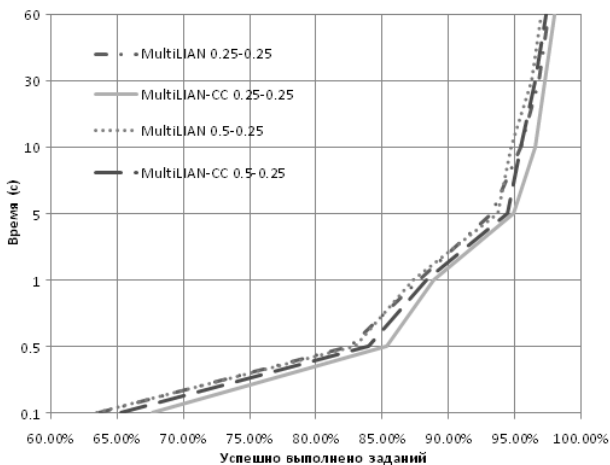


Рис. 12. График зависимости количества выполненных заданий от затраченного времени для алгоритмов MultiLIAN и MultiLIAN-CC

На временном интервале алгоритма MultiLIAN-CC демонстрирует своё преимущество. MultiLIAN-CC с параметрами «0,25-0,25» успешно завершает 67,7% заданий менее чем за 0,1 с; 95% всех заданий MultiLIAN-CC «0,25-0,25» обрабатывает менее чем за 5 с. Базовому алгоритму требуется как минимум 10 с, чтобы преодолеть порог в 95% успешно обработанных заданий.

В таблице 6 представлены значения индикаторов M , T и PL , полученных в результате проведения третьей серии экспериментов. Как и раньше, при подсчете значений индикаторов усреднялись только результаты успешно выполненных заданий.

Как видно из таблицы, значения средней длины пути у MultiLIAN и MultiLIAN-CC при одинаковых настройках практически идентичны, затраты по памяти также различаются не

сильно (в пределах 5,5% на первой коллекции и 3,4% – на второй), но при этом выигрыш MultiLIAN-CC по времени более существенен и составляет в среднем 15%. Более наглядно результаты эксперимента представлены на рис. 13 в виде гистограмм нормированных усредненных значений.

Таблица 6. Индивидуальные показатели эффективности алгоритмов MultiLIAN и MultiLIAN-CC

| | Moscow Maps 1.0 | | | Moscow Maps 2.0 | | |
|---------------------|-----------------|-----------|----------|-----------------|-----------|----------|
| | <i>M</i> | <i>PL</i> | <i>T</i> | <i>M</i> | <i>PL</i> | <i>T</i> |
| MultiLIAN 0,25-0,25 | 11039 | 648,3 | 1,215 | 7350 | 542,0 | 0,703 |
| MLIAN-CC 0,25-0,25 | 11528 | 648,4 | 1,136 | 7107 | 541,8 | 0,543 |
| MultiLIAN 0,5-0,25 | 10772 | 652,4 | 0,983 | 7718 | 545,2 | 0,853 |
| MLIAN-CC 0,5-0,25 | 11359 | 651,7 | 1,026 | 7797 | 544,52 | 0,653 |

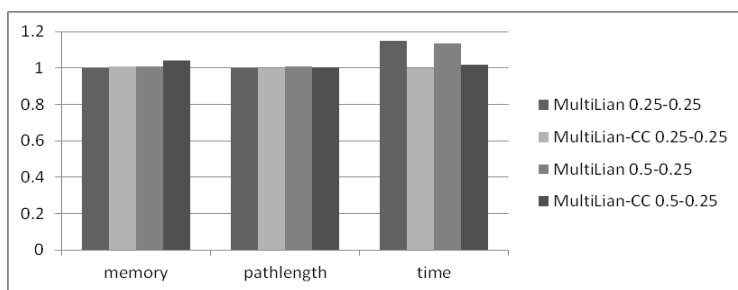


Рис. 13. Гистограммы нормированных усредненных значений индивидуальных индикаторов эффективности алгоритмов MultiLIAN и MultiLIAN-CC

В результате анализа результатов третьей серии экспериментов можно сделать вывод о том, что модификация MultiLIAN-CC превосходит базовый алгоритм по всем рассмотренным показателям эффективности.

4.5. ВЫВОДЫ

В результате проведенных экспериментальных исследований можно сделать следующие выводы:

1. Эффективность алгоритма MultiLIAN (по рассмотренным критериям) существенным образом зависит от выбора промежуточных вершин, которые используются для построения альтернативных путей. При этом на практике целесообразно использование следующих входных параметров MultiLIAN, отвечающих за расположение этих вершин: «0,5-0,25» и «0,25-0,25».

2. Использование модификации MultiLIAN-RU на практике нецелесообразно: при существенно более эффективном расходовании вычислительных ресурсов процент различных путей (т.е. путей отличающихся хотя бы одной вершиной), отыскиваемых алгоритмом, весьма низок.

3. Наиболее эффективным алгоритмом поиска различных альтернативных путей является MultiLIAN-CC, что свидетельствует о перспективности идеи итерационного использования результатов промежуточных вычислений для сокращения пространства поиска.

5. Заключение

В работе рассмотрена одна из ключевых подзадач, возникающих при планировании совокупности неконфликтных пространственных траекторий для коалиции беспилотных летательных аппаратов, а именно, задача автоматического планирования совокупности различных траекторий для отдельного аппарата. Предложен новый метод решения этой задачи, который, в отличие от известных аналогов, позволяет отыскивать траектории особого вида, косвенно учитывающие ограничения на динамику полета. Разработаны также модификации этого метода. Все описанные в работе методы программно реализованы, и проведено экспериментальное исследование алгоритмов в задаче автоматического планирования альтернативных траекторий маловысотного полета мультироторного беспилотного летательного аппарата в городских условиях. Результаты экспериментов показывают целесообразность применения предложен-

ного метода в таких задачах. К дальнейшим направлениям исследований можно отнести повышение вычислительной эффективности алгоритма при решении практических задач, а так же разработку на основе имеющегося задела метода планирования траекторий для коалиции летательных аппаратов, обладающих сложной динамикой движения.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 15-37-20893.

Литература

1. БАСКИН Е.С. *Алгоритм планирования пути lian-search, доказательство его сходимости, полноты и оптимальности* // Задачи современной информатики. – М.: ИПИ РАН, 2014. – С. 59–67.
2. ЖУЛЕВ В.И., ЛЕВУШКИН В.С., НГУЕН Т.Н. *Планирование локальной траектории автомобиля-робота в реальном времени* // Вестник РГРТУ. – 2013. – №4(46). Часть 3. – С. 18–23.
3. МАКАРОВ Д.А., ПАНОВ А.И., ЯКОВЛЕВ К.С. *Архитектура многоуровневой интеллектуальной системы управления беспилотными летательными аппаратами* // Искусственный интеллект и принятие решений. – 2015. – №3. – С. 18–33.
4. ОСИПОВ Г.С., ТИХОМИРОВ И.А., ХАЧУМОВ В.М. и др. *Интеллектуальное управление транспортными средствами: стандарты, проекты, реализации* // Авиакосмическое приборостроение. – 2009. – №6. – С. 34–43.
5. ЯКОВЛЕВ К.С., БАСКИН Е.С. *Графовые модели в задаче планирования траектории на плоскости* // Искусственный интеллект и принятие решений. – 2013. – №1. – С. 5–12.
6. ЯКОВЛЕВ К.С., БАСКИН Е.С. *Эвристическая оценка качества построения пути методом LIAN-Search* // Шестая Международная конференция «Системный анализ и информационные технологии» САИТ-2015 (15-20 июня)

- 2015 г., г. Светлогорск, Россия): Труды конференции. В 2-х т. – Т. 1. – С. 132–138.
7. ЯКОВЛЕВ К.С., МАКАРОВ Д.А., БАСКИН Е.С. *Метод автоматического планирования траектории беспилотного летательного аппарата в условиях ограничений на динамику полета* // Искусственный интеллект и принятие решений. – 2014. – №4. – С. 3–17.
 8. ЯКОВЛЕВ К.С., МАКАРОВ Д.А., ПАНОВ А.И. и др. *Принципы построения многоуровневых архитектур систем управления беспилотными летательными аппаратами* // Авиакосмическое приборостроение. – 2013. – №4. – С. 10–28.
 9. ALBUS J. et al. *4D/Real-time Control System (4D/RCS): A Reference Model Architecture for Unmanned Vehicle Systems v.2.0*, NIST, NISTIR 6910, 2002.
 10. ALJAZZAR H., LEUE S. *K*: A heuristic search algorithm for finding the k shortest paths* // Journal of Artificial Intelligence. – 2011. – Vol. 175. – P. 2129–2154.
 11. *AscTec Hummingbird* – [Электронный ресурс]. – URL: <http://www.asctec.de/en/uav-uas-drone-products/asctec-hummingbird/> (дата обращения: 13.11.2015).
 12. BERTRAM D. et al. *An integrated approach to inverse kinematics and path planning for redundant manipulators* // Proc. IEEE International Conference on Robotics and Automation (ICRA 2006). – 2006. – P. 1874–1879.
 13. BRESENHAM J.E. *Algorithm for computer control of a digital plotter* // IBM Systems Journal. – 1965. – No. 4(1). – P. 25–30.
 14. BRISTEAU P.-J., CALLOU F., VISSIERE D. et al. *The Navigation and Control Technology Inside the AR.Drone Micro UAV* // Preprints of the 18th IFAC World Congress (Milano, Italy). Milano: IFAC, 2011. – Vol. 18, Part 1. – P. 1477–1484.
 15. DIJKSTRA E.W. *A note on two problems in connexion with graphs* // Numerische mathematik. – 1959. – Vol. 1. №1. – P. 269–271.
 16. EPPSTEIN D. *Finding the k shortest paths* // SIAM Journal on Computing. – 1998. – No. 28(2). – P. 652–673.

17. FREDERICKSON G.N. *An optimal algorithm for selection in a min-heap* // Information and Computation. – 1993. – No. 104(2). – P. 197–214.
18. HARABOR D., GRASTIEN A.. *Online graph pruning for pathfinding on grid maps*. 2011 In AAAI-11 – [Электронный ресурс] – URL: <http://grastien.net/ban/articles/hg-aaai11.pdf> (дата обращения: 13.11.2015).
19. HART P.E., NILSSON N.J., RAPHAEL B. *A formal basis for the heuristic determination of minimum cost paths* // IEEE Trans. on Systems Science and Cybernetics. – 1968. – No. 4(2). – P. 100–107.
20. HUTTER F., HOOS H.H., LEYTON-BROWN K. et al. *ParamILS: an automatic algorithm configuration framework* // Journal of Artificial Intelligence Research. – 2009. – No. 36(1). – P. 267–306.
21. JIM'ENEZ V.M., MARZAL A. *A lazy version of Eppstein's K shortest paths algorithm* // Experimental and Efficient Algorithms. – 2003. – Vol. 2647. – P. 179–191.
22. JIM'ENEZ V.M., MARZAL A. *Computing the k shortest paths: A new algorithm and an experimental comparison* // Algorithm Engineering – 1999. – Vol. 1668. – P. 15–29.
23. KIM H., KIM D., SHIN J.U. et al. *Angular rate-constrained path planning algorithm for unmanned surface vehicles* // Ocean Engineering. – 2014. – No. 84. – P. 37–44.
24. KIM H., PARK B., MYUNG H. *Curvature Path Planning with High Resolution Graph for Unmanned Surface Vehicle* // Robot Intelligence Technology and Applications. – 2012. – Vol. 208. – P. 147–154.
25. LAVALLE S.M., HUTCHINSON S.A. *Optimal motion planning for multiple robots having independent goals* // IEEE Transactions on Robotics and Automation. – 1998. – Vol. 14, Issue 6. – P. 912–925.
26. LAVALLE S.M., KUFFNER JR J.J. *Rapidly-exploring random trees: Progress and prospects* // Proc. Workshop on the Algorithmic and Computational Robotics: New Directions. – 2000. – [Электронный ресурс] – URL: <http://msl.cs.uiuc.edu/~lavalle/papers/LavKuf01.pdf> (дата обращения: 13.11.2015).

27. LIKHACHEV M., STENTZ A. *R* Search* // Proc. Twenty-Third AAAI Conference on Artificial Intelligence. Menlo Park, Calif.: AAAI press. – 2008. – [Электронный ресурс] – URL: http://repository.upenn.edu/cgi/viewcontent.cgi?article=1030&context=grasp_papers (дата обращения: 13.11.2015).
28. *Mikrokopter* – [Электронный ресурс] – URL: <http://mikrokopter.de> (дата обращения 13.11.2015).
29. NASH A., DANIEL K., KOENIG S. *Theta*: Any-Angle Path Planning on Grids* // Proc. National Conference on Artificial Intelligence – 2007. – Vol. 22, No. 2. – P. 1177–1183.
30. *OpenStreetMaps* – [Электронный ресурс] – URL: <http://wiki.openstreetmap.org/wiki/Database> (дата обращения: 13.11.2015).
31. *Parrot Ar.Drone* – [Электронный ресурс] – URL: <http://ardrone2.parrot.com/> (дата обращения: 13.11.2015).
32. SILVER D. *Cooperative pathfinding* // AI Programming Wisdom. – 2006. – P. 99–111.
33. STANDLEY T. *Finding optimal solutions to cooperative pathfinding problems* // In AAAI. – 2010. – P. 173–178.
34. WANG K.-H.C., BOTEVA A. *Fast and Memory-Efficient Multi-Agent Pathfinding* // ICAPS. – 2008. – P. 380–387.
35. WANG K.-H.C., BOTEVA A. *Tractable Multi-Agent Path Planning on Grid Maps* // Proc. International Joint Conference on Artificial Intelligence (IJCAI). – 2009. – P. 1870–1875.
36. YAKOVLEV K., BASKIN E., HRAMOIN I. *Grid-based angle-constrained path planning* // Proc. 38th Annual German Conference on AI, Dresden, Germany, September 21-25, 2015. – Springer International Publishing. doi 10.1007/978-3-319-24489-1_16. – P. 208–221.

DYNAMICS CONSTRAINT-AWARE PLANNING OF MULTIPLE PATHS FOR UNMANNED VEHICLE

Konstantin Yakovlev, Institute for Systems Analysis of Federal Research Centre “Computer Science and Control” of Russian Academy of Sciences (ISA RAS), Cand.Sc. (yakovlev@isa.ru).

Egor Baskin, Institute for Systems Analysis of Federal Research Centre “Computer Science and Control” of Russian Academy of Sciences (ISA RAS), PhD student (baskin@isa.ru).

Anton Andreychuk, Peoples Friendship University of Russia, Moscow, student (andreychuk@mail.com).

Abstract: We study alternative paths' planning problem for an unmanned vehicle. It constitutes a key part of a bigger problem – that of multi-agent path planning, i.e., finding a non-conflicting path set for a coalition of vehicles. We propose a new path-planning technique, which indirectly takes into account vehicle movement dynamics and guarantees feasibility of resulting paths. We also elaborate a number of extensions of the method proposed. We conduct an empirical study of all introduced algorithms by running the large number of experiments simulating nap-of-the-earth flight of a compact multi-rotor unmanned aerial vehicle in urban environment.¹

Keywords: unmanned vehicle, multirotor, quadrotor, navigation, intelligent control system, planning, path planning, path finding, alternative paths, k shortest paths, angle-constrained path planning, heuristic search.

Статья представлена к публикации членом редакционной коллегии Н.И. Базенковым

Поступила в редакцию 28.10.2015.

Опубликована 30.11.2015.

¹ *The reported study was partially supported by RFBR, research project No. 15-37-20893.*