

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

УПРАВЛЕНИЕ БОЛЬШИМИ СИСТЕМАМИ

***Выпуск 75
Сентябрь 2018***

**СБОРНИК
ТРУДОВ**

ISSN 1819-2467

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

Москва – 2018

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

**УПРАВЛЕНИЕ
БОЛЬШИМИ
СИСТЕМАМИ**

СБОРНИК ТРУДОВ

Выпуск 75

Москва – 2018

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: 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 Калимулина Э.Ю. **Редактор:** канд. техн. наук Квинто Я.И.

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-р 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 интернете: ubs.mtas.ru.

СОДЕРЖАНИЕ

Системный анализ

Кузнецов А.В.

Динамическая модель системы связи группы агентов..... 6

Математическая теория управления

Семенов Д.М.

Управление синхронизацией двух связанных неидентичных систем Хиндмари–Роуз 30

Анализ и синтез систем управления

Надеждин И.С., Горюнов А.Г., Маненти Ф.

Системы управления нестационарным объектом на основе МРС-регулятора и ПИД-регулятора с нечеткой логикой 50

Сетевые модели в управлении

Горбатенко Д.Е., Семенов А.А.

Противодействие сговору в дискретных динамических моделях компьютерных сетей..... 76

Управление в социально-экономических системах

Давыдов В.А.

Комбинирование метода встречных планов и метода обратных приоритетов для распределения фонда мотивации научных лабораторий..... 103

**Управление в медико-биологических
и экологических системах**

Засухин С.В.

*Нахождение испарения как решение задачи
оптимального управления 128*

**Управление техническими системами
и технологическими процессами**

Колодкина А.С., Марьясин О.Ю., Огарков А.А.

*Оптимальное управление энергопотреблением
и микроклиматом больших многозонных зданий..... 146*

**Программы и системы моделирования объектов,
средств и систем управления**

Скворцов Е.Б., Шелехова А.С.

*Начала теории концептуального проектирования
с приложениями в области авиационной науки
и технологий..... 170*

ДИНАМИЧЕСКАЯ МОДЕЛЬ СИСТЕМЫ СВЯЗИ ГРУППЫ АГЕНТОВ¹

Кузнецов А. В.²

(Воронежский государственный университет, Воронеж)

Построена модель задержек сообщений в сети мобильной связи агентов, в которой связность и виды каналов между агентами изменяются во времени. Агенты перемещаются по местности, на которой расположены препятствия разной проходимости, в различной степени затрудняющие связь между агентами. Движение агентов по местности моделируется с помощью клеточного автомата. Агенты являются вершинами графа связи. Каждый агент владеет набором средств связи. Ребрами графа связи являются каналы связи, организованные с помощью средств связи агентов, располагающихся в вершинах графа. Каналы разных видов имеют различную предельную дальность и допустимую высоту препятствий между агентами. Если между агентами находится препятствие, исключающее связь по существующему каналу связи, канал связи изменяется на более подходящий с учетом возможностей агентов. Каждому агенту соответствует расписание рассылки сообщений, а каждому каналу связи – максимальная пропускная способность. При рассылке сообщений одновременно несколькими агентами с использованием одного канала связи учитывается уменьшение пропускной способности этого канала. Рассматривается ретрансляция сообщения цепочкой агентов. Моделируется изменение трафика в каналах связи системы при движении агентов и возможных обрывах каналов. Рассмотрены возможные способы применения модели.

Ключевые слова: модель движения, модель системы связи, клеточный автомат, теория расписаний.

1. Введение

В предыдущих работах [4, 5] автором были предложены модель движения иерархически организованных агентов и модель конфликта таких агентов. При описании модели конфликта агентов была сконструирована упрощенная система связи, позволяющая агентам обмениваться сигналами о наступлении или отступлении. Данная статья посвящена моделированию системы связи, подобной тем, что используются в тактическом звене управле-

¹ Работа выполнена при финансовой поддержке РФФИ, грант №18-07-01240 А.

² Александр Владимирович Кузнецов, к.ф.-м.н., доцент (avkuz@bk.ru).

ния воинским формированием у нас в стране и за рубежом. Одной отличительной чертой моделируемой системы связи является то, что канал связи между двумя агентами может состоять из нескольких каналов, организованных совершенно различно как на физическом уровне, так и на уровне протоколов. Другой чертой является неустойчивость системы военной связи (которая относится в целом к классу MANET [8], хотя и может иметь некую стационарную компоненту). В один момент времени между агентами, которые, как правило, оснащены как высокоскоростными средствами с протоколами типа WiMAX или Wi-Fi, так и низкоскоростными узкополосными радиостанциями диапазона МВ или ДКМВ с нестандартизированными протоколами передачи данных, может быть организован высокоскоростной канал, а в другой – только низкоскоростной, а в некоторые промежутки времени невозможно организовать никакого канала или возможно организовать только составной канал. По очевидным причинам, в открытой печати не существует подробного описания систем военной связи, однако функции и архитектуру таких систем можно понять из многочисленных популярных статей, например [2].

Вышесказанное делает довольно затруднительным моделирование канала связи на сетевом или канальном уровне. Более того, в военных условиях интерес представляет не доставка пакета, а именно доставка сообщения в установленный срок, так как ценность доставленной со значительной задержкой информации может оказаться нулевой. Поэтому моделирование будет проводиться на уровне сообщений, которыми обмениваются агенты, с упрощением очень многих деталей.

Предлагаемая модель сети связи состоит из двух частей: из меняющегося во времени в зависимости от местоположения связывающихся агентов графа системы связи, имеющего ребра, соответствующие каналам связи разного вида, и присвоенных вершинам графа системы связи расписаний рассылки сообщений по другим вершинам, причем скорость доставки сообщения зависит от вида канала между узлом-отправителем и узлом-адресатом. Таким образом, задачи, возникающие при исследовании модели,

в некоторой степени напоминают динамические задачи теории расписаний, типа задач управления транспортными системами, однако, в отличие от многих задач этого типа, допустимо одновременное обслуживание произвольного количества требований. Классические модели из теории телеграфика не вполне подходят – поступление сообщений в канал не случайно, а обусловлено расписанием и потеря сообщения исключена. С другой стороны, в разных каналах применяются различные протоколы, не всегда документированные, с разными максимальными скоростями передачи и дисциплинами обслуживания пакетов, что затрудняет имитационное моделирование в известных средах типа ns-3 или OMNeT++. Все вышеперечисленное и подтолкнуло автора к созданию предлагаемой феноменологической модели сети военной связи. Отметим, что аналогичная задача, но без учета движения агентов по местности, была рассмотрена автором в [6]. Также стоит указать, что автором было реализовано программное обеспечение «Психодод» (<https://bitbucket.org/bokohodteam/bokohod>), реализующее на настоящий момент алгоритмы, приведенные в работе.

2. Ячейки и состояния.

Будем исследовать клеточный автомат, моделирующий движение и конфликт агентов, ранее описанного вида [5]. Предположим, что множество ячеек нашего автомата – это \mathbb{Z}^2 , а в качестве множества состояний будем рассматривать множество векторов типа

$$\mathcal{E} = F \times W \times Q \times U = \{(f_{ij}, w_{ij}, q_{ij}; u_{ij}) | (i, j) \in \mathbb{Z}^2\},$$

где для всех $i, j \in \mathbb{Z}^2$ $f_{ij} \in F = \{0, 1, 2\}$, $u_{ij} \in U \subset \mathbb{R}$ – ограниченное множество, $w_{ij} \in W \subseteq \mathbb{Z}^g$ – набор чисел, необходимый для реализации внутренней динамики клеточного автомата (подробнее данный набор описан в [5]),

$$\sum_{i,j \in \mathbb{Z}} f_{ij} > \infty.$$

Значение $f_{ij} = 0$ соответствует отсутствию, $f_{ij} = 1$ или $f_{ij} = 2$ – наличию в клетке (i, j) агента, принадлежащему к первой или второй противоборствующей стороне соответственно. Величина $q_{ij} \in Q \subset \mathbb{Z}^p$ соответствуют некоторому уникальному идентификатору агента, находящегося в клетке с координатами (i, j) . Этот идентификатор является вектором целых чисел, так как в предыдущих работах автора рассматривались модели, учитывающие внутреннюю p -уровневую иерархию агентов, компонента с номером $k = \overline{1, p-1}$ вектора q_{ij} соответствует номеру агента (группы агентов) k -го уровня иерархии внутри группы $(k+1)$ -го уровня иерархии, p -я компонента вектора q_{ij} соответствует номеру группы агентов (агента) самого верхнего уровня.

Значения u_{ij} соответствуют некоторой характеристике местности, влияющей на распространение радиоволн – например, высоте середины квадрата с координатами (i, j) над уровнем моря, лесистости, наличию зданий и т.п.

Для целей практической реализации можно считать, что в множестве клеток \mathbb{Z}^2 выделен квадрат K из клеток с положительными координатами, не превосходящими длины стороны квадрата L_K , причем для любой клетки (i, j) вне этого квадрата всегда $f_{ij} = 0$, $w_{ij} = \mathbf{0}$, $q_{ij} = \mathbf{0}$, вне квадрата $u_{ij} = 0$, т.е. клетки вне квадрата K находятся всегда в состоянии покоя.

Поскольку сама модель движения и конфликта была уже рассмотрена в предыдущих статьях [4, 5], то можно считать, что наборы f_{ij}, w_{ij}, q_{ij} являются исходными данными и уже заданы во всех $i, j \in \mathbb{Z}^2$ в любой момент функционирования автомата.

3. Граф системы связи

Должен быть задан граф потребностей в связи, указывающий на то, какие вообще агенты должны быть связаны друг с другом. Пусть $Q_f \subseteq Q$ – множество вершин-идентификаторов агентов, принадлежащих к стороне f , $E_f \subseteq Q_f \times Q_f$ – множество ребер, определяющее связность агентов. Определим граф потребностей в связи стороны f как $\Pi_f = (Q_f, E_f)$.

Система связи может изменяться в каждый момент $t \in \mathbb{Z}$,

$t = \overline{0, N}$, функционирования клеточного автомата в зависимости от дальности агентов, особенностей местности, применения средств радиоэлектронной борьбы и т.п. Зададим граф системы связи соответствующим образом, чтобы это отразить:

$$\Gamma_f(t) = (V_f(t), C, \varphi_f(t, \cdot)),$$

где $f = \overline{1, 2}$, $V_f(t) = \{(i, j) \in \mathbb{Z}^2 \mid f_{ij}(t) = f\}$ – множество клеток, в которых в данный момент t находятся агенты стороны f , $C = \{c_1, \dots, c_l\}$ – множество каналов, $\varphi_f(t, \cdot) : C \rightarrow V_f(t) \times V_f(t)$ – функция инцидентности. Каждый элемент множества C может означать, например, канал (Hotspot) некоторой модификации Wi-Fi 802.11b ($\sim 2,4$ ГГц) в диапазоне ДМВ (300 МГц – 3 ГГц), канал определенной модуляции, например, ЧМ в диапазоне МВ (30–300 МГц), канал определенной модуляции, например *USB*, в диапазоне ДКМВ (3 – 30 МГц). Таким образом, $c \in C$ является числовым идентификатором определенного канала связи.

Будем считать, что все каналы имеют некоторый m -мерный «вес» – абстракцию свойств канала – предельной дальности, необходимости прямой видимости и т.п., определенный функцией $\psi_1 : C \rightarrow \mathbb{R}^m$, $\psi_1 = (\psi_1^1, \dots, \psi_1^m)$. Длина вектора веса m зависит количества выделяемых у канала связи признаков, т.е. от степени желаемой детализации модели системы связи.

Также для каждого канала $c \in C$ задана предельная скорость передачи данных (пропускная способность) $v_{max}(c)$.

Обозначим через $L(i_1, j_1; i_2, j_2)$ множество всех клеток, пересекаемых в более чем одной точке отрезком прямой между точками $(i_1 + 0,5, j_1 + 0,5)$ и $(i_2 + 0,5, j_2 + 0,5)$. Будем считать, что задана функция видимости $\psi_2 : 2^{\mathbb{Z}^2} \rightarrow \mathbb{R}^m$, определяющая, каковы условия для организации канала, $\psi_2 = (\psi_2^1, \dots, \psi_2^m)$.

Приведем пример такой функции видимости для $m = 2$. Если величина u_{ij} из вектора состояния ячейки соответствует высоте местности, то уравнение прямой, проходящей через середины клеток (i_1, j_1) , (i_2, j_2) на высоте $u_{i_1 j_1}$ и $u_{i_2 j_2}$ над серединами первой и второй клетки соответственно будет (в координатах ι ,

$\kappa, v(l, \kappa)$.

$$\frac{l - i_1 - 0,5}{i_2 - i_1} = \frac{\kappa - j_1 - 0,5}{j_2 - j_1} = \frac{v - u_{i_1, j_1}}{u_{i_2, j_2} - u_{i_1, j_1}}.$$

Соответственно, можно рассчитать, насколько высота клеток по середине отрезка, соединяющего середины клеток (i_1, j_1) , (i_2, j_2) , превышает линию, соединяющую клетки на его концах, вычисляя величину типа

$$(1) \quad \psi_2^1(L(i_1, j_1; i_2, j_2)) = \max_{(l, \kappa) \in L(i_1, j_1; i_2, j_2)} u_{l, \kappa} - v(l, \kappa).$$

Очевидно, что если значение ψ_2^1 будет много больше нуля, то, например, каналы ДМВ между соответствующими клетками организовать будет нельзя, хотя возможно будет – до какого-то значения ψ_2^1 – организовать каналы МВ и т.п. Соответствующим образом должны быть назначены и веса для каналов данных типов.

Также можно ограничить каналы некоторой предельной длиной между серединами квадратов (i_1, j_1) и (i_2, j_2) и положить

$$(2) \quad \psi_2^2(L(i_1, j_1; i_2, j_2)) = \sqrt{(i_2 - i_1)^2 + (j_2 - j_1)^2}.$$

Описание так сконструированной функции видимости можно пояснить с помощью рис. 1. На нем высоты незаштрихованных клеток равны 0, высоты клеток, штрихованных одинарной штриховкой, равны 1. «Вес» канала c определим как $\psi_1(c) = (1, 200)$, т.е. канал может быть организован через клетки, «возвышающиеся» над связываемыми не более, чем на 1, и максимальная дальность канала – приблизительно 200 клеток (точная оценка зависит от того, как именно расположены клетки). Как показано на рис. 1, $\psi_2(L(2, 3; 2, 5)) = (\psi_2^1(L(2, 3; 2, 5)), \psi_2^2(L(2, 3; 2, 5))) = (1, 2)$.

Определим функцию совместимости каналов $com : C \times C \rightarrow \{0, 1\}$. Каналы c_i, c_j будут совместимыми ($com(c_i, c_j) = 1$), если возможна одновременная работа обоих каналов без помех и несовместимыми ($com(c_i, c_j) = 0$) – в противоположном случае. Если организовано n несовместимых каналов, по которым одновременно осуществляется отправка сообщений, то скорость передачи данных по каждому из них падает в n раз.

Функция инцидентности отвечает следующим условиям. Пусть

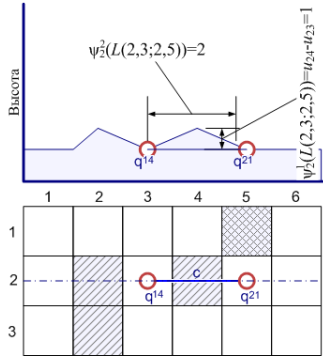


Рис. 1. Пример функции видимости

- 1) $\varphi_f(t, c_k) = ((i_1, j_1), (i_2, j_2))$;
- 2) клетка (i_1, j_1) имеет состояние $(f, w_{i_1 j_1}, q_{i_1 j_1}; u_{i_1 j_1})$;
- 3) клетка (i_2, j_2) имеет состояние $(f, w_{i_2 j_2}, q_{i_2 j_2}; u_{i_2 j_2})$.

Тогда

1. $(q_{i_1 j_1}, q_{i_2 j_2}) \in E_f$. Это значит, что в любой момент времени связаны не произвольные агенты, а лишь те, которые изначально должны быть связаны в соответствии с графом Π_f , что соответствует логике планирования связи.

2. Для всех $i = \overline{1, m}$, где m – длина вектора веса, $\psi_1^i(c_k) \geq \psi_2^i(L(i_1, j_1; i_2, j_2))$. Это означает, что вес канала не может быть меньше, чем допускают условия видимости.

3. Максимальная пропускная способность канала c_k – наибольшая из возможных. Это означает, что сначала пытаются назначить каналы с максимальной v_{max} , а если это не получается (каналы данной пропускной способности кончились или назначение канала запрещает функция видимости), то с меньшей пропускной способностью.

Агенты с идентификаторами из множества Q_f могут являться адресатами или отправителями сообщений друг для друга. Сообщения рассылаются каждым агентом с идентификатором $q \in Q_f$ по определенному расписанию

$$(3) \quad R(q) = \{(t_i, q_i, b_i, \tilde{t}_i, \gamma_i) | i \in \mathbb{Z}\},$$

где $(t_i, q_i, b_i, \tilde{t}_i, \gamma_i) \in R(q)$ имеет следующий смысл: t_i – момент времени, в который рассылается сообщение длиной b_i , которое должно быть доставлено за время \tilde{t}_i адресату с идентификатором q_i , $\gamma_i \in \mathbb{N}$ – важность сообщения, определяющая, какая часть канала достанется на передачу сообщения, если через один и тот же канал одновременно пересылается несколько сообщений. При некоторых состояниях клеточного автомата (соответствующих обнаружению противника, попаданию в окружение и т.п.) сообщения рассылаются внепланово и имеют зависящую от состояния клеток в окрестности агента длину и срок доставки.

Идентификатор агента, находящегося в момент времени в клетке с координатами z в момент времени t , будем обозначать $q(t, z)$. Клетку, в которой в момент времени t находится агент с идентификатором q , будем обозначать $z(t, q)$.

Канал, соединяющий z и z' в момент времени t , обозначим как $c(t, z, z') = c(t, q, q')$, $q = q(t, z)$, $q' = q(t, z')$. Если вершины z, z' не смежные в графе $\Gamma_f(t)$ и существует некоторый путь, их соединяющий (кратчайший или иной обусловленный выбранной моделью передачи информации), то $c(t, z, z')$ – множество всех каналов-ребер данного пути.

Количество несовместимых с $c = c(t, q, q')$ каналов, по которым происходит передача информации в момент времени t , обозначим как $n(t, c)$.

Заметим, что в [7] была вкратце описана подобная конструкция коммуникационного графа, не связанная, однако, с клеточным автоматом.

4. Модель передачи информации

Длину сообщения вместе с накладными расходами на пересылку, отправленного агентом с идентификатором q агенту с идентификатором q' в момент времени t , будем обозначать как $b(t, q, q')$. Если $t = t_i$ из расписания $R(q)$ и q' входит в число адресатов в момент t_i , то объем пересылаемого сообщения в «чистом виде» будет, согласно (3), b_i , а вместе с накладными

расходами станет

$$(4) \quad b(t, q, q') = K_1(t, c(t, q, q'))b_i + K_2,$$

иначе $b(t, q, q') = 0$. Здесь $K_1(t, c(t, q, q')) > 0$, $K_2 > 0$ – коэффициенты накладных расходов на пересылку. Далее будем полагать, что обозначения типа b , b' обозначают уже длины сообщений с учетом накладных расходов.

Будем обозначать скорость в канале c в момент времени t как $v(t, c)$. Скорость зависит от пропускной способности канала $v_{max}(c)$ и от некоторых нижеперечисленных случаев. Если по каналу c передается в некоторый интервал времени только одно сообщение, то в этот интервал $v(t, c) = v_{max}(c)$, иначе скорость падает прямо пропорционально количеству n одновременно передаваемых в один интервал сообщений как $v(t, c) = v_{max}(c)/n$ (см. далее).

Если $c(t_0, q_0, q_r)$ содержит в себе некий путь из более чем одного ребра (который обозначим $q_0 \dots q_i \dots q_r$) и передача сообщения длиной b_0 в момент времени t_0 осуществляется именно по этому пути, то в узле $z_i = z(t, q_i)$ может устанавливаться задержка $d_i(t)$ перед отправкой сообщения из транзитного узла далее по маршруту. Будем считать, что сообщение начнет исходить из i -й вершины маршрута, $i = \overline{1, r-1}$, в момент времени

$$t_i = t_{i-1} + d_i(t_{i-1}) + \frac{b\alpha}{v(t_{i-1}, c(t_{i-1}, q_{i-1}, q_i))},$$

где t_{i-1} – время, в которое сообщение начало передаваться из предыдущего узла маршрута, α – параметр системы, отвечающий за то, после приема какой части сообщения транзитный узел начинает ретрансляцию, $0 < \alpha \leq 1$, $b = K_1(t_{i-1}, c(t_{i-1}, q_{i-1}, q_i))b_0 + K_2$.

Скорость передачи сообщения по i -му каналу – ребру маршрута в момент времени t_i , $i = \overline{1, r}$, будем считать равной наименьшей из скоростей передачи сообщения по всем предыдущим ребрам маршрута и самого i -го ребра:

$$v(t_i, c(t_i, q_{i-1}, q_i)) = \min_{j=\overline{1, i}} v(t_j, c(t_j, q_{j-1}, q_j)).$$

Скоростью всего маршрута будем считать скорость в его последнем ребре.

Пусть в канале $c = c(t, q, q')$ начиная с момента t передается сообщение β длиной в b с приоритетом γ . В процессе передачи в момент времени t' возможны три «элементарных случая», влияющих на время доставки сообщения (пока предполагается, что в достаточно большой интервал времени происходит только один случай из нижеперечисленных):

1. В момент времени t' сменяется канал: $c' = c(t', q, q')$, тогда время доставки сообщения становится

$$(5) \quad \hat{T}_1(\beta) = t' + \frac{b - (t' - t)v(t, c)}{v(t', c')}.$$

2. В момент времени $t' \geq t$ в канал поступает сообщение β' длиной b' и приоритетом γ' , таким что $\gamma \neq 0$ или $\gamma' \neq 0$. Тогда время доставки сообщения β составит

$$(6) \quad \hat{T}_2(\beta) = t' + \frac{(\gamma + \gamma')(b - (t - t')v(t, c))}{\gamma v(t, c)},$$

а продолжительность доставки сообщения β' составит

$$(7) \quad \hat{T}_2(\beta') = \frac{(\gamma + \gamma')b'}{\gamma'v(t, c)}.$$

При этом, если сообщение β будет доставлено до окончания доставки сообщения β' , т.е. если $\hat{T}_2(\beta) < t' + \hat{T}_2(\beta')$, то выражение (7) изменится как

$$(8) \quad \begin{aligned} \hat{T}_2(\beta') &= \hat{T}_2(\beta) - t' + \frac{b' - \frac{(\hat{T}_2(\beta) - t')v(t, c)\gamma'}{\gamma + \gamma'}}{v(t, c)} = \hat{T}_2(\beta) - t' + \\ &+ \frac{b'(\gamma + \gamma') - (\hat{T}_2(\beta) - t')v(t, c)\gamma'}{(\gamma + \gamma')v(t, c)} = \\ &= \frac{b + b' - v(t, c)(t - t')}{v(t, c)} = \frac{b + b'}{v(t, c)} + t' - t, \end{aligned}$$

а если сообщение β' будет доставлено до окончания доставки сообщения β , т.е. $\hat{T}_2(\beta) > t' + \hat{T}_2(\beta')$, то выражение (6) изменится как

$$(9) \quad \hat{T}_2(\beta) = t' + \hat{T}_2(\beta') + \frac{b - t'v(t, c) - \frac{\hat{T}_2(\beta')v(t, c)\gamma}{(\gamma + \gamma')}}{v(t, c)} = \frac{b + b'}{v(t, c)}.$$

3. В несовместимый с c канал начинают в момент времени t' передавать сообщение β' длиной b' . Тогда время доставки сообщения β становится

$$(10) \quad \hat{T}_3(\beta) = t' + \frac{b - (t' - t)v(t, c)}{v(t', c)},$$

где $v(t', c) = v(t, c)/2$ (несложно видеть, что если сообщение начинает передаваться по $n = n(t, c)$ несовместимым каналам, то $v(t', c) = v(t, c)/n$). Если сообщение β' будет доставлено до окончания доставки сообщения β (тогда продолжительность доставки сообщения β' будет $t'' = b'/v(t', c)$), то время доставки сообщения β станет

$$(11) \quad \hat{T}_3(\beta) = t' + t'' + \frac{b - (t' - t)v(t, c) - (t'' - t')v(t', c)}{v(t, c)}.$$

Алгоритм исследования задержек в системе связи вкратце следующий (перед началом работы алгоритма полагаем множество $TC_f^{(i)} = \{\}, i = 0$):

Шаг 1. Разбить граф $\Gamma_f(t)$ на подграфы (если это возможно) $\Gamma_f^{(i)}(t)$, такие что для любой вершины $z = z(t, q)$ графа $\Gamma_f^{(i)}(t)$ справедливо, что, если q' является адресатом в расписании $R(q(t, z))$, то $z' = z(t, q')$ является вершиной графа $\Gamma_f^{(i)}(t)$ и что подграфы $\Gamma_f^{(i)}(t)$ и $\Gamma_f^{(j)}(t)$ имеют максимально одну общую вершину. Очевидно, что каналы в таких подграфах освобождаются и занимаются независимо и вместо изучения всего графа целиком можно изучать передачу информации по меньшим по размеру подграфам.

Шаг 2. В каждом графе $\Gamma_f^{(i)}(t)$ для каждого t построить упорядоченный список $TC1_f^{(i)} = \{\tau_1, \dots, \tau_m\}$ моментов, в которых происходит «элементарный случай» 1 (смена канала).

Шаг 3. Для каждой вершины q графа $\Gamma_f^{(i)}(t)$ просмотреть расписание $R(q)$ и построить упорядоченный по возрастанию t список

$$TT_r_f^{(i)} = \{(t_j; q_j, q'_j, b_j, \tilde{t}_j, \gamma_j) | j = \overline{1, l}\}$$

моментов, в которых происходит рассылка сообщений с адресатом q' в $\Gamma_f^{(i)}(t)$.

Шаг 4. Для каждого элемента $(t_j; q_j, q'_j, b, \tilde{t}_j, \gamma_j)$ из $TTr_f^{(i)}$ последовательно вычислить время прохождения сообщения T_j , учитывая по формуле (4) накладные расходы и увеличивая или уменьшая время прохождения в соответствии с формулами (5)–(11), если на время передачи информации по расписанию придется момент из $TC_f^{(i)}$ или же возникнут «элементарные случаи» 1 или 2, моменты возникновения которых последовательно рассчитываются.

С этой целью полагаем $j = 1$, множество моментов случая 2 $TC2_f^{(i)} = \{\}$, множество моментов случая 3 $TC3_f^{(i)} = \{\}$, множество времен прохода сообщений через транзитные узлы $TM_f^{(i)} = \{\}$ и совершаем:

Шаг 4.1. Упорядочиваем $TTr_f^{(i)} \cup TM_f^{(i)}$ по возрастанию t . Выбираем $t_j \in TTr_f^{(i)} \cup TM_f^{(i)}$. Вычисляем $T_j = t_j + (b(t_j, q_j, q'_j)/v(t_j, c(t_j, q_j, q'_j)))$.

Шаг 4.2. Полагаем $TC_f^{(i)} = TC1_f^{(i)} \cup TC2_f^{(i)} \cup TC3_f^{(i)}$, упорядочиваем $TC^{(i)}$ по возрастанию. Если существуют $t_k^c, t_{k+1}^c \in TC^{(i)}$, такие что

$$t_k^c < t_j < t_{k+1}^c \text{ и } T_j > t_{k+1}^c \text{ или}$$

$$0 < t_j < t_1^c = \min TC_f^{(i)} \text{ и } T_j > t_1^c,$$

то изменяем T_j в соответствии с формулами (6)–(11), причем нужную формулу выберем в зависимости от того, моментом какого «элементарного случая» является t_{k+1}^c или t_1^c .

Шаг 4.3. Если $T_j > t_{j+1}$ и q_{j+1}, q'_{j+1} таковы, что при пересылке между ними информации в момент времени t_{j+1}

- задействуется канал между q_j, q'_j , то добавляем t_{j+1} в $TC2_f^{(i)}$,
- задействуется зависимый от $c(T_j, q_j, q'_j)$ канал, то добавляем t_{j+1} в $TC3_f^{(i)}$,

- $c(t_j, q_j, q'_j)$ – путь из более, чем одного ребра и q_s, q'_s – соседние узлы в этом пути, между которыми в момент t_{j+1} передается сообщение размером b (включая накладные расходы), то добавляем $t_s + b\alpha/v(t_s, c(t_s, q_s, q'_s)) + d_s(t_s)$, где t_s – время попадания сообщения в узел $q_s, t_j < t_s < t_{j+1}$, $d_s(t_s)$ – задержка сообщения в узле q_s , в $TM_f^{(i)}$,
- переходим к шагу 4.2.

Иначе, если $j \leq l$ – увеличиваем j на единицу, переходим к шагу 4.1, если $j > l$ – к шагу 5.

Шаг 5. Конец.

Можно заметить, что основные элементы модели – сообщения, каналы и скорость передачи аналогичны соответственно требованиям, машинам и ресурсам задачи построения расписания выполнения работ проекта с учетом отношений предшествования и ограничения на ресурсы (RCPSР [1]).

В разделе 5 покажем работу алгоритма на конкретном примере.

5. Демонстрация работы алгоритма

Для проверки модели выберем типичный фрагмент военной сети связи $\Gamma_1(t)$, в котором агенты q^{11} , q^{12} , q^{13} – командиры отделений – связаны с командиром взвода q^{14} , который, в свою очередь, соединен с командиром роты q^{21} , а последний – с командиром батальона q^{31} . Нижестоящие агенты периодически докладывают вышестоящим об изменениях в обстановке и об исполнении приказов, а вышестоящие направляют нижестоящим различные указания. Агенты обычно пользуются для обмена сообщениями высокоскоростными каналами одного типа, а в случае невозможности их применения из-за помех или особенностей рельефа – используют низкоскоростные каналы, также одного типа (рис. 2).

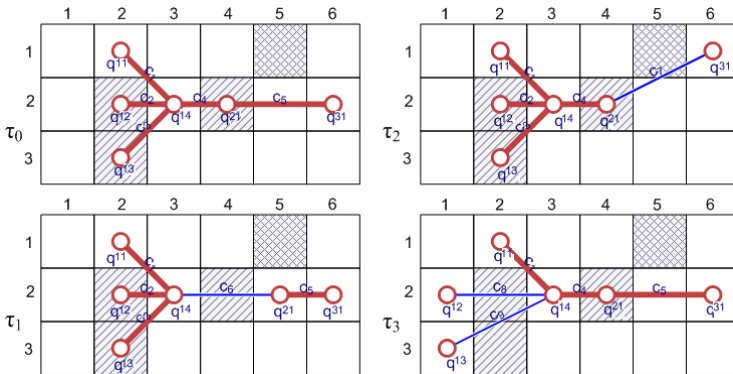


Рис. 2. Поведение модельной системы связи. Штриховкой обозначена степень проницаемости клетки для радиоволн: чем гуще штриховка, тем хуже проницаемость

Для того чтобы избежать излишних вопросов, длины сообщений, время рассылки и максимальную пропускную способность каналов будем исчислять в условных безразмерных единицах.

Здесь $C = \{c_1, \dots, c_9\}$, $Q = Q_1 = \{q^{11}, q^{12}, q^{13}, q^{14}, q^{21}, q^{31}\}$,

$$v_{max}(c_i) = \begin{cases} 400 \text{ единиц в такт,} & i = \overline{1, 5}, \\ 4 \text{ единицы в такт,} & i = \overline{6, 9}. \end{cases}$$

Функция совместимости каналов com задается таблицей 1.

Таблица 1. Значения функции com

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9
c_1	1	0	0	1	1	1	1	1	1
c_2	0	1	0	1	1	1	1	1	1
c_3	0	0	1	1	1	1	1	1	1
c_4	1	1	1	1	1	1	1	1	1
c_5	1	1	1	1	1	1	1	1	1
c_6	1	1	1	1	1	1	1	1	1
c_7	1	1	1	1	1	1	1	1	1
c_8	1	1	1	1	1	1	1	1	0
c_9	1	1	1	1	1	1	1	0	1

Моменты $\tau_i \in \mathbb{Z}$, $i = \overline{0, 3}$, изменения графа таковы: $\tau_0 = 0$, $\tau_1 = 350$, $\tau_2 = 801$, $\tau_3 = 821$.

Функция весов каналов задается как

$$(12) \quad \psi_1(c_i) = \begin{cases} (0, 100), & i = \overline{1, 5}, \\ (1, 5, 200), & i = \overline{6, 9}. \end{cases}$$

Первая компонента $\psi_1(c_i)$ соответствует способности канала пропускать клетки, вторая – максимальной дальности (в клетках), на которой может работать канал. Клетки (i_1, j_1) и (i_2, j_2) могут быть соединены каналом c только если

$$\psi_1^k(c) \geq \psi_2^k(L(i_1, j_1; i_2, j_2)), \quad k = \overline{1, 2}.$$

где ψ_2 – функция, определяемая соотношениями (1), (2).

Предполагается, что накладные расходы несущественны. Расписание агентов пусть имеет вид:

- для q^{ij} , $i = 1$, $j = \overline{1, 3}$, $R(q^{ij}) = \{(600 + 100p, q^{14}, 300, 30, 1) | p = \overline{0, 4}\}$;
- $R(q^{14}) = \{(400 + 100p, q^{1k}, 300, 30, 1) | p = \overline{0, 6}, k = \overline{1, 3}\} \cup \{(800 + 150p, q^{21}, 900, 30, 1) | p = \overline{0, 2}\}$;
- $R(q^{21}) = \{(200 + 150p, q^{14}, 900, 30, 1) | p = \overline{0, 6}\} \cup \{(1000 + 200p, q^{31}, 850, 30, 1) | p = \overline{0, 1}\}$;
- $R(q^{31}) = \{(200p, q^{21}, 900, 30, 1) | p = \overline{0, 6}\} \cup \{(800, q^{14}, 950, 30, 1)\}$.

Функция $\mathcal{Q} : Q \times \overline{1, N} \rightarrow \mathbb{Z}^2$ задается таблицей 2. Эта функция позволяет получать местоположение агента q^{ij} в некоторый момент времени $t = \overline{1, N}$.

Таблица 2. Значения функции \mathcal{Q}

	$\tau_0 \leq t < \tau_1$	$\tau_1 \leq t < \tau_2$	$\tau_2 \leq t < \tau_3$	$t \geq \tau_3$
q^{11}	(1, 2)	(1, 2)	(1, 2)	(1, 2)
q^{12}	(2, 2)	(2, 2)	(2, 2)	(2, 1)
q^{13}	(3, 2)	(3, 2)	(3, 2)	(3, 1)
q^{14}	(2, 3)	(2, 3)	(2, 3)	(2, 3)
q^{21}	(2, 4)	(2, 5)	(2, 4)	(2, 4)
q^{31}	(2, 6)	(2, 6)	(1, 6)	(2, 6)

Свойства местности, изображенной на рис. 2, будут описываться следующей матрицей:

$$(13) \quad \mathcal{U} = (u_{ij})_{j=\overline{1,6}}^{i=\overline{1,3}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Множество ребер графа потребностей в связи $\Pi_1 = (Q, E_1)$ будет иметь вид

$$E_1 = ((q^{11}, q^{14}), (q^{12}, q^{14}), (q^{13}, q^{14}), (q^{14}, q^{21}), (q^{21}, q^{31})).$$

На основе графа Π_1 , матрицы \mathcal{U} (13) и функции весов каналов ψ_1 (12) конструируется функция инцидентности φ_1 графа

связи Γ_1 , вид которой показан в таблице 3. Например, в момент времени τ_0 клетки (1, 2) и (2, 3) соединены каналом c_1 , так как в этих клетках, согласно таблице 2, находятся агенты q^{11} и q^{14} , $(q^{11}, q^{14}) \in E_1$, $\psi_1^1(c_1) = 0 \geq \psi_2^1(L(1, 2; 2, 3)) = 0$ и $\psi_1^2(c_1) = 100 \geq \psi_2^2(L(1, 2; 2, 3)) = \sqrt{2}$ и

$$c_1 \in \arg \max_{c \in C} v_{max}(c).$$

Таблица 3. Значения функции φ_1

	τ_0	τ_1	τ_2	τ_3
c_1	((1, 2),(2, 3))	((1, 2),(2, 3))	((1, 2),(2, 3))	((1, 2),(2, 3))
c_2	((2, 2),(2, 3))	((2, 2),(2, 3))	((2, 2),(2, 3))	–
c_3	((3, 2),(2, 3))	((3, 2),(2, 3))	((3, 2),(2, 3))	–
c_4	((2, 3),(2, 4))	((2, 3),(2, 4))	–	((2, 3),(2, 4))
c_5	((2, 4),(2, 6))	–	–	((2, 4),(2, 6))
c_6	–	((2, 3),(2, 5))	–	–
c_7	–	–	((2, 4),(1, 6))	–
c_8	–	–	–	((2, 1),(2, 3))
c_9	–	–	–	((3, 1),(2, 3))

В соответствии с шагом 1 алгоритма, граф $\Gamma_1(t)$ может быть в любой такт разбит на два подграфа

$$\Gamma_1^{(0)}(t) = (\{z(t, q^{11}), z(t, q^{12}), z(t, q^{13}), z(t, q^{14}), z(t, q^{21})\}, C, \varphi_1),$$

$$\Gamma_1^{(1)}(t) = (\{z(t, q^{21}), z(t, q^{31}), z(t, q^{14})\}, C, \varphi_1).$$

Легко видеть, что обмен начнется в каналах между (q^{31}, q^{21}) и сообщения, отправленные от q^{31} к q^{21} в моменты времени $t_{2p+1} = 200p$, $p = \bar{0}, \bar{3}$, будут приходить в срок

$$T_{2p+1} = 200p + \frac{900}{400} = 202,25.$$

В момент $t_8 = 800$ произойдет «элементарный случай» 2, скорость передачи упадет в 2 раза и сообщение от q^{31} к q^{21} должно

было бы прийти за

$$T_8 = 800 + 2 \cdot \frac{900}{400} = 804,5.$$

Так как в момент $\tau_2 = 801 > T_8$ происходит «элементарный случай» 1 – смена канала между (q^{31}, q^{21}) , то сообщение от q^{31} к q^{21} придет за время

$$T_8 = 801 + \frac{900 - (801 - 800)200}{4} = 976.$$

В момент $\tau_3 = 821$ опять произойдет «элементарный случай» 1, так как $T_8 > \tau_3$ и поэтому необходимо изменить значение T_8 так, что

$$T_8 = 821 + \frac{900 - (801 - 800)200 - (821 - 801)2}{200} = 824,3.$$

Аналогично, время доставки сообщения от q^{31} к q^{14} составит (при сделанном для упрощения предположении, что сообщение было задержано в узле q^{21} до момента $t_{11} = 811$ в виду перегрузки канала)

$$T_{11} = T_8 + \frac{950 - 900}{400} = 824,425.$$

В последующие моменты времени $t_{14}, t_{15} = 1000, t_{18}, t_{19} = 1200$ будет происходить только «элементарный случай» 2. Поэтому время доставки

$$T_i = t_i + 2 \cdot \frac{850}{400} + \frac{900 - 850}{400} = t_i + 4,375, \quad i = \overline{14, 15},$$

$$T_i = t_i + 2 \cdot \frac{850}{400} = t_i + 4,25, \quad i = \overline{18, 19}.$$

Обмен по каналам, связывающим пару (q^{14}, q^{21}) в соответствующие моменты времени будет (стрелочка означает вычисление нового времени доставки сообщения после тех или иных событий в соответствии с алгоритмом):

- $t_2 = 200, T_2 = 202,25.$

- $t_4 = 350, T_4 = 350 + \frac{900}{4} > 500 \rightarrow T_4 = 500 + \frac{900 - (500 - 350)4}{2} = 650.$
- $t_6 = 500, T_6 = 800 + \frac{900 - (800 - 650)2 - (650 - 650)4 - (650 - 500)2}{4/5} = 1175 > 801 \rightarrow \frac{T_i}{4/5} = 801 + \frac{900 - (801 - 800)4/5 - (800 - 650)2 - (650 - 500)2}{400/5} = 804,74.$
- $t_7 = 650, T_7 = 804,74 + \frac{900 - (804,74 - 801)400/5 - (801 - 800)4/5 - (800 - 650)2}{400/4} = 807,74$
- $t_9 = 800, T_9 = 807,74 + \frac{900 - (807,74 - 804,74)400/4 - (804,74 - 801)400/5 - (801 - 800)4/5}{400/3} = 809,99.$
- $t_{10} = 800,$

$$T_{10} = 809,99 + \frac{950 - (809,99 - 807,74)400/3}{400} - \frac{(807,74 - 804,74)400/4 + (804,74 - 801)400/5}{400} + \frac{(801 - 800)4/5}{400} = 810,115.$$

- $t_{12} = t_{13} = 950, T_{12} = T_{13} = 950 + 2 \cdot 900/400 = 954,5.$
- $t_{16} = t_{17} = 1100, T_{16} = T_{17} = 1104,5.$

В графе $\Gamma_1^{(0)}(t)$ в канале (q^{1j}, q^{14}) обмен будет таков:

- $t_i \in \{400, 500\}$, «элементарный случай» 3, $T_i = t_i + 3/4.$
- $t_i \in \{600, 600, 700, 700, 800, 800\}$, «элементарный случай» 3, $T_i = t_i + \frac{300}{400/4} = t_i + 3.$

Далее, в канале (q^{11}, q^{14}) :

$$t_i \geq 900, T_i = t_i + \frac{300}{400/2} = t_i + 1,5.$$

В каналах $(q^{12}, q^{14}), (q^{13}, q^{14})$, после $\tau_3 = 821$:

- $t_{10} = 900$, «элементарные случаи» 2 и 3, $T_{10} = 900 + \frac{300}{4/3} = 1125$.
- $t_{11} = 1000$, «элементарные случаи» 2 и 3, $T_{11} = 1000 + \frac{300 - (1000 - 900)4/3}{4/6} = 1250$.

Видно, что, например, в моменты t_{10} , t_{11} сообщения не доставляются в установленный срок в точности таким же образом и по той же причине, что и в многократно повторенных автором наблюдениях реальной сети военной связи.

Результаты, полученные в настоящем разделе, хорошо согласуются с экспериментами на стенде. Самый простой способ смоделировать фрагмент системы связи с помощью стенда, не прибегая к крайне дорогостоящей и неоднозначно функционирующей военной технике, – соединить 6 компьютеров с двумя сетевыми платами, одна из которых обеспечивает передачу данных со скоростью 100 Мбит/с, а вторая – 10 Мбит/с, причем нужно пересылать данные между компьютерами и включать и отключать сетевые платы по определенному расписанию. Разумеется, необходимо скорректировать объем пересылаемых данных пропорционально столь высоким скоростям передачи. При этом следует иметь, однако, в виду, что в реальности стандартный протокол TCP в глобальной сети (wide area network) военной связи может не употребляться, а вместо него могут использовать различные специальные протоколы разной степени продуманности и документированности на основе UDP.

6. Задачи, для решения которых проводится моделирование

Модель, предложенная в данной статье, позволяет дать ответы на следующие вопросы, представляющие большой практический интерес:

1. В любой ли момент времени t заданный граф системы связи $\Gamma_f(t)$ обеспечивает плановую передачу сообщений от всех источников за заданное время?

2. Какое количество информации передается по заданному ребру $\Gamma_f(t)$ в любой заданный момент времени? Не выбраны ли каналы так, что это количество всегда много меньше или больше того, что может пропустить ребро, т.е. не будет ли время доставки сообщения по данному ребру всегда много больше или много меньше максимально допустимого? Иначе говоря, адекватен ли подбор средств связи у агента его потребностям?

3. Возможно ли подобрать некую общую характеристику местности (т.е. значений компоненты состояния u_{ij}) и графа потребностей в связи Π_f так, чтобы от них зависела вероятность отказа системы связи (т.е. невозможности обеспечить передачу данных за заданное время) в одном или нескольких ребрах?

4. Сколько внеплановых сообщений можно одновременно разослать без отказа системы связи? Возможно ли подобрать множество каналов (т.е. оснастить соответствующим образом агентов средствами связи) так, чтоб плановая и внеплановая рассылка сообщений осуществлялась бы без отказов системы связи или с минимальным количеством отказов?

5. Как оптимально установить задержки в транзитных узлах и скорректировать расписание, так, чтобы при выдерживании установленного времени доставки минимизировать возникновение задержек из-за «элементарных случаев»?

Для ответов на эти вопросы необходимо знать, во-первых, существует ли путь в графе системы связи между клетками, занимаемыми теми или иными агентами, а во-вторых – сколько в действительности займет передача информации по данному пути. Отметим, что в используемых в Вооруженных силах РФ телекоммуникационных сетях, являвшихся основным предметом исследований автора (ОКР «Созвездие М2» [3]), применяются протоколы, исключющие отказ в передаче данных из-за сильного превышения объема этих данных пропускной способности канала, однако остро стоят проблемы устаревания информации и наличия в сложном канале узких мест, через которые в разумный срок можно передать только очень ограниченный объем данных.

При этом предлагаемая модель не требует знания используемых протоколов передачи данных, что может оказаться немаловажным как в случае исследования системы связи вероятного противника, так и в случае моделирования своей собственной сети связи, когда число используемых протоколов может исчисляться десятками и все они сильно нестандартны.

Дальнейшим направлением развития будет являться исследование случаев, когда можно полагать, что время доставки и размер сообщений будут являться случайными величинами с заданным распределением. Также планируется изучать системы, в которых время передачи в расписании одного агента будет увязано с временем получения этим агентом сообщения от другого агента.

Литература

1. АНИЧКИН А.С., СЕМЕНОВ В.А. *Современные модели и методы теории расписаний* // Труды ИСП РАН. – 2014. – Т. 26, №3. – С. 5–50. – URL: http://www.ispras.ru/proceedings/docs/2014/26/3/isp_26_2014_3_5.pdf.
2. ВАСИЛЬЕВ АНДРЕЙ *Концепция автоматизированной системы управления боевыми действиями Сухопутных войск АСУВ 2.0* // Армейский Вестник. – 2013. – Октябрь. – URL: <http://army-news.ru/2013/10/koncepciya-avtomatizirovannoj-sistemy-upravleniya-boevymi-dejstviyami-suxoputnyx-vojsk-asuv-2-0/> (дата обращения: 11.08.2015).
3. КАНДАУРОВ ДМИТРИЙ *АСУВ в Алабино: желаемое и действительное* // Независимое военное обозрение. – 2010. – Ноябрь. – URL: http://nvo.ng.ru/armament/2010-11-19/1_asuv.html (дата обращения: 21.08.2015).
4. КУЗНЕЦОВ А.В. *Модель совместного движения агентов с трехуровневой иерархией на основе клеточного автомата* // Журнал вычислительной математики и математической физики. – 2017. – № 2. – С. 339–349.

5. КУЗНЕЦОВ А.В. Упрощенная модель боевых действий на основе клеточного автомата // Известия РАН: Теория и системы управления. – 2017. – №2. – С. 397–409.
6. КУЗНЕЦОВ А.В. Распределение ограниченных ресурсов в системе с устойчивой иерархией (на примере перспективной системы военной связи) // Управление большими системами: сборник трудов. – 2017. – Вып. 66. – С. 68–93.
7. ТИМОФЕЕВ А.В., ЮСУПОВ Р.М.. Принципы построения интегрированных систем мультиагентной навигации и интеллектуального управления мехатронными роботами // Int. Journal «Information Technologies & Knowledge». – 2011. – Т. 5, № 3. – С. 237–244. – URL: <http://www.foibg.com/ijitk/ijitk-vol05/ijitk05-3-p04.pdf>.
8. CORSON S., MACKER J. *Request for Comments: 2501. Mobile Ad hoc Networking (MANET): Routing Protocol Performance Issues and Evaluation Considerations* // IETF Tools. – 1999. – January. – URL: <https://tools.ietf.org/html/rfc2501> (дата обращения: 21.08.2015).

DYNAMIC MODEL OF THE COMMUNICATION SYSTEM OF A GROUP OF AGENTS

Alexander Kuznetsov, Voronezh State University, Voronezh,
Cand.Sc., associate professor (avkuz@bk.ru).

Abstract: We construct a model of message delays in the mobile communication network of agents. In the model, the connectivity and types of channels between agents change over time. Agents move around the terrain, on which there are obstacles of various impassability that impede communication between agents in various degree. We use a cellular automaton model for the movement of agents along the terrain. Agents are the vertices of the communication graph. Each agent owns a set of communication devices. The edges of the communication graph are the communication channels organized by communication devices of agents located at the vertices of the graph. Channels of different types have a different limiting range and permissible height of obstacles between agents. If there is an obstacle between agents that excludes communication on the existing communication channel, the communication channel is changed to the more suitable one, taking into account the capabilities of the agents. Each agent has a message schedule, and each communication channel has the maximum bandwidth. When sending messages by several agents simultaneously using one communication channel, the reduction in the capacity of this channel is taken into account. We also consider the relaying of a message by a chain of agents. We simulate a traffic dynamic in telecommunication channels of the system with agents motion and possible channel breaks and changes. Further, we discuss different ways of the model's application.

Keywords: model of movement, model of communication system, cellular automaton, scheduling theory.

УДК 621.391

ББК 32.81

*Статья представлена к публикации
членом редакционной коллегии В.М. Вишневым.*

Поступила в редакцию 29.04.2018.

Дата опубликования 30.09.2018.

УПРАВЛЕНИЕ СИНХРОНИЗАЦИЕЙ ДВУХ СВЯЗАННЫХ НЕИДЕНТИЧНЫХ СИСТЕМ ХИНДМАРШ–РОУЗ¹

Семенов Д. М.²

(Санкт-Петербургский государственный университет,
Институт проблем машиноведения РАН, Санкт-Петербург)

Рассмотрена задача управления синхронизацией между двумя связанными неидентичными системами Хиндмарш–Роуз, каждая из которых описывает поведение отдельного биологического нейрона. Важность решения данной задачи обусловлена множеством медико-биологических исследований, которые напрямую связывают появление некоторых заболеваний нервной системы (таких как эпилепсия) с наличием патологической синхронизации между нейронами в отдельных участках головного мозга. Таким образом, возможность управления синхронизацией между нейронами является перспективным методом лечения эпилепсии и активно применяется в медицинской практике. Очевидно, что разработка такого метода лечения требует использования качественных математических инструментов. В качестве таких инструментов в данной работе используются инструменты теории управления. Кроме того, необходимо учитывать наличие неточности современных моделей биологических нейронов. Для этого в настоящей работе предлагается рассматривать эту неточность как непрерывные, равномерно ограниченные функции, описывающие некоторые возмущения. Таким образом, используя предлагаемый в работе закон управления и соблюдая выполнение сформулированных теорем, можно гарантированно достичь синхронизированного поведения между исследуемыми системами в условиях отсутствия и наличия модельных возмущений в системах. Полученные результаты доказаны и подтверждены численными экспериментами.

Ключевые слова: синхронизация, связанные осцилляторы, система Хиндмарш–Роуз.

¹ Работа выполнена в ИПМаш РАН при финансовой поддержке Российского научного фонда (грант 14-29-00142).

Автор выражает благодарность своему научному руководителю, доктору технических наук, профессору Фрадкову Александру Львовичу за ценные указания в ходе работы над статьёй и всяческую поддержку.

² Данила Михайлович Семенов, студент магистратуры, стажёр-исследователь (semenovdm90@gmail.com).

Введение

Благодаря огромному количеству исследований, посвящённых синхронизации в биологических системах, можно с уверенностью сказать, что явление синхронизации играет ключевую роль в функционировании биологических систем и их взаимодействии между собой [1, 2, 13, 14]. В качестве примеров можно привести популяции светлячков, синхронно вспыхивающих в ночном небе, или синхронную активность сердечных клеток, позволяющую сердцу сокращаться [20]. Кроме того, к таким системам относится мозг человека или животного. На настоящий момент известно, что синхронизация больших популяций нейронов является основным механизмом образования макроколебаний в мозге [14, 21, 24]. Также было установлено, что ряд патологических состояний и заболеваний центральной нервной системы, например таких, как эссенциальный тремор, эпилепсия, болезни Паркинсона и Альцгеймера, напрямую связаны с аномальной синхронизацией отдельных нейронных популяций [5, 12, 22]. Сейчас активно ведётся разработка методов лечения данных заболеваний путём подавления патологической синхронизации в мозге. Причём особый интерес представляют собой неинвазивные методы, основанные на биологической обратной связи [3, 4, 9]. Очевидно, что разработка таких методов требует качественного математического описания как моделей нейронов, так и условий их синхронизации в сетях.

На сегодняшний день существует множество исследований, посвящённых поиску условий синхронизации сетей, узлами которых являются математические модели биологических нейронов. При этом до недавнего времени практически все исследования ограничивались лишь поиском условий синхронизации для случая идентичных узлов в сети и отсутствия модельных возмущений в каждом отдельном узле [6, 10, 11, 18, 23]. Но поскольку популяции нейронов в мозге представляют собой гетерогенные сети (т.е. сети, состоящие из неидентичных узлов), возник интерес перенесения накопленного опыта на случай синхронизации в дан-

ных сетях. Например, продемонстрирована возможность синхронизации гетерогенной сети, узлы которой описываются динамической системой ФитцХью–Нагумо [15, 16, 17]. Настоящая работа продолжает исследования гетерогенных сетей и рассматривает синхронизацию простейшей сети, состоящей из двух связанных неидентичных динамических систем Хиндмарш–Роуз, причём как в условиях отсутствия, так и в условиях наличия возмущений в системах.

Статья организована следующим образом: в разделе 1 рассматривается динамическая система Хиндмарш–Роуз как модель биологического нейрона; в разделе 2.1 формулируется понятие синхронизации и приводятся её основные виды; в разделах 2.2 и 2.3 представлены доказательства синхронизации двух взаимосвязанных неидентичных систем Хиндмарш–Роуз в условиях отсутствия и наличия возмущений; результаты численного моделирования, подтверждающие теоретические результаты, приведены в разделе 3.

1. Система Хиндмарш–Роуз

Система Хиндмарш–Роуз представляет собой модель биологического нейрона и описывается нелинейной системой дифференциальных уравнений 3-го порядка:

$$\begin{aligned} \dot{x}(t) &= y(t) - ax^3(t) + bx^2(t) - z(t) + u(t), \\ (1) \quad \dot{y}(t) &= c - dx^2(t) - y(t), \\ \dot{z}(t) &= \varepsilon[s(x(t) - r) - z(t)]. \end{aligned}$$

Здесь переменная $x(t)$ описывает динамику мембранного потенциала нейрона, а переменные $y(t)$ и $z(t)$ описывают калий-натриевый насос. Поскольку скорость изменения переменной $z(t)$ определяется малым параметром $0 < \varepsilon \ll 1$, то $z(t)$ описывает динамику медленного калиевого тока, а $y(t)$ – быстрого натриевого тока. Внешнее воздействие на нейрон определяется переменной $u(t)$.

Изначально система Хиндмарш–Роуз описывалась нелинейной системой дифференциальных уравнений 2-го порядка и явля-

лась упрощённым вариантом модели Ходжкина–Хаксли [7]. Однако после ряда экспериментов, проведённых над клетками мозга пресноводной улитки *Lymnaea stagnalis*, были обнаружены новые режимы функционирования нейронов, которые не учитывались в модели [8]. Потому авторы модели, Джеймс Хиндмарш и Мальком Роуз, приняли решение о включении в модель третьего уравнения, что позволило учесть большинство новых режимов поведения биологического нейрона. В частности, модель (1) способна демонстрировать пачечную активность и посттормозную отдачу, а также адаптироваться к воздействию внешней среды [8, 19].

2. Синхронизация систем Хиндмарш–Роуз

2.1. ПОНЯТИЕ СИНХРОНИЗАЦИИ

Настоящая работа в значительной степени посвящена явлению синхронизации. Потому важно определить, что в дальнейшем понимается под термином «синхронизация». *Синхронизацией* будем называть согласованное во времени функционирование двух или нескольких объектов или процессов [2]. Также определение синхронизации может быть сформулировано математически.

Определение 1 [2, с. 78]. Будем говорить, что имеет место синхронизация процессов или объектов $x_i(t) \in \mathcal{X}$, $i = 1, 2, \dots, k$, относительно характеристики $C_t(\cdot)$ и вектор-функции $F_i(\cdot)$, если существуют числа $\tau_i \in \mathbb{R}$, $i = 1, 2, \dots, k$, такие, что $\forall t \geq 0$ выполняются соотношения

$$(2) \quad F_1(C_{t+\tau_1}(x_1)) = F_2(C_{t+\tau_2}(x_2)) = \dots = F_k(C_{t+\tau_k}(x_k)).$$

Здесь характеристика $C_t(\cdot)$ называется *показателем (индексом) синхронизации* и определяется как отображение $C_t : \mathcal{X} \rightarrow \mathcal{C}$, где \mathcal{C} есть множество возможных значений $C_t(\cdot)$. Необходимо отметить, что характеристика $C_t(\cdot)$ предполагается одинаковой для всех объектов или процессов. Принимаемое характеристикой $C_t(\cdot)$ значение может быть скаляром, вектором, матрицей и даже функцией. Сравнить характеристики различных процессов

между собой позволяют вектор-функции $F_i : \mathcal{C} \rightarrow \mathbb{R}^m$, называемые *функциями сравнения*.

Под *приближённой синхронизацией* (ε -синхронизацией) понимается случай, когда соотношения (2) выполняются приближённо:

$$(3) \quad \|F_i(C_{t+\tau_i}(x_i)) - F_j(C_{t+\tau_j}(x_j))\|_2 \leq \varepsilon \quad \forall i, j, \quad t \geq 0,$$

а под *асимптотической синхронизацией* – случай, когда погрешность соотношений (2) со временем исчезает:

$$(4) \quad \lim_{t \rightarrow \infty} \|F_i(C_{t+\tau_i}(x_i)) - F_j(C_{t+\tau_j}(x_j))\|_2 = 0 \quad \forall i, j, \quad t \geq 0.$$

Представленное выше определение является достаточно общим и охватывает большинство из основных видов синхронизации. В данной работе же рассматривается вид синхронизации, называемый *координатной синхронизацией*. Под координатной синхронизацией взаимосвязанных подсистем понимается совпадения координат векторов состояний этих подсистем. Координатная синхронизация согласуется с приведённым выше определением, если определить показатель синхронизации как $C_{t+\tau_i}(x_i) = x_i(t + \tau_i)$, $i = 1, 2, \dots, k$, а функции сравнения взять тождественными своему аргументу: $F_i(x) = x$, $i = 1, 2, \dots, k$.

2.2. СИНХРОНИЗАЦИЯ СИСТЕМ БЕЗ ВОЗМУЩЕНИЙ

Рассмотрим гетерогенную сеть нейронов, представленную двумя связанными системами Хиндмарш-Роуз:

$$(5) \quad \begin{cases} \dot{\mathbf{x}}_1(t) = F_1(\mathbf{x}_1(t)) + B[\sigma(x_2(t) - x_1(t)) + u(t)], \\ \dot{\mathbf{x}}_2(t) = F_2(\mathbf{x}_2(t)) + B\sigma(x_1(t) - x_2(t)), \end{cases}$$

где

$$F_i(\mathbf{x}(t)) = \begin{pmatrix} y(t) - ax^3(t) + bx^2(t) - z(t) \\ c - dx^2(t) - y(t) \\ \varepsilon[s(x(t) - r_i) - z(t)] \end{pmatrix}, \quad i = 1, 2.$$

Здесь $\mathbf{x}_i(t) = \text{col}(x_i(t), y_i(t), z_i(t))$, $i = 1, 2$, – вектора состояний связанных систем (5); $B = \text{col}(1, 0, 0)$; $u(t)$ – управляющее воздействие; σ – сила связи между нейронами; $a, b, c, d, \varepsilon, s, r_1$ и

r_2 – известные параметры нейронов. Будем проводить синхронизацию систем, предполагая известность параметров. Последовательно вычтя уравнения второй системы из первой и положив $\psi(t) = x_1^2(t) + x_1(t)x_2(t) + x_2^2(t)$, $\varphi(t) = x_1(t) + x_2(t)$, получим систему

$$(6) \quad \dot{\delta}(t) = \widehat{F}(\delta(t)) + Bu(t),$$

где

$$\widehat{F}(\delta(t)) = \begin{pmatrix} -(a\psi(t) - b\varphi(t) + 2\sigma)\delta_x(t) + \delta_y(t) - \delta_z(t) \\ -d\varphi(t)\delta_x(t) - \delta_y(t) \\ \varepsilon(s\delta_x(t) - \delta_z(t) - sr) \end{pmatrix}.$$

Здесь $\delta(t) = \text{col}(\delta_x(t), \delta_y(t), \delta_z(t))$ – вектор состояния системы (6); $\delta_x(t) = x_1(t) - x_2(t)$, $\delta_y(t) = y_1(t) - y_2(t)$, $\delta_z(t) = z_1(t) - z_2(t)$ и $r = r_1 - r_2$. Теперь, чтобы гарантировать синхронизацию систем (5), необходимо обеспечить устойчивость системы (6). Это достигается посредством управления, определяемого функцией $u(t)$. Очевидно, что выбор конкретного вида синхронизации определяет конкретный вид цели управления, а значит и функции $u(t)$. В настоящей работе цель управления сформулирована следующим образом:

$$(7) \quad |\delta_x(t)| \rightarrow 0, \quad |\delta_y(t)| \rightarrow 0, \quad |\delta_z(t)| \rightarrow \text{const}, \quad \text{при } t \rightarrow \infty.$$

Далее перейдем к новым координатам: $e_x = \delta_x$, $e_y = \delta_y$, $e_z = \delta_z + sr$. Система (6) примет следующий вид:

$$(8) \quad \dot{e}(t) = \widetilde{F}(e(t)) + Bu(t),$$

где

$$\widetilde{F}(e(t)) = \begin{pmatrix} -(a\psi(t) - b\varphi(t) + 2\sigma)e_x(t) + e_y(t) - e_z(t) + sr \\ -d\varphi(t)e_x(t) - e_y(t) \\ \varepsilon(se_x(t) - e_z(t)) \end{pmatrix}.$$

Здесь $e(t) = \text{col}(e_x(t), e_y(t), e_z(t))$ – вектор состояния системы (8).

Введем функцию Ляпунова:

$$(9) \quad V(e(t)) = \frac{1}{2} \left(e_x^2(t) + e_y^2(t) + \frac{1}{\varepsilon s} e_z^2(t) \right).$$

Найдем производную $\dot{V}(e(t))$ в силу системы (8). Получим следующее:

$$(10) \quad \begin{aligned} \dot{V}(e(t)) = & -(a\psi(t) - b\varphi(t) + 2\sigma)e_x^2(t) - e_y^2(t) - \frac{1}{s}e_z^2(t) + \\ & + (1 - d\varphi)e_x(t)e_y(t) + sre_x(t) + u(t)e_x(t). \end{aligned}$$

В виду того, что функция $\psi(t)$ неотрицательна $\forall t \geq 0$, производную (10) можно оценить сверху:

$$(11) \quad \begin{aligned} \dot{V}(e(t)) \leq & -(2\sigma - b\varphi(t))e_x^2(t) - e_y^2(t) - \frac{1}{s}e_z^2(t) + \\ & + (1 - d\varphi)e_x(t)e_y(t) + sre_x(t) + u(t)e_x(t). \end{aligned}$$

Зададим следующий закон управления:

$$(12) \quad u(t) = -(\gamma + b\varphi(t))\delta_x(t) + d\varphi(t)\delta_y(t) - sr, \quad \gamma \geq 0.$$

Переходя к новым координатам e_x , e_y и e_z , управление (12) примет вид

$$(13) \quad u(t) = -(\gamma + b\varphi(t))e_x(t) + d\varphi(t)e_y(t) - sr, \quad \gamma \geq 0.$$

Подставляя закон управления (13) в оценку (11), получаем

$$(14) \quad \dot{V}(e(t)) \leq -(\gamma + 2\sigma)e_x^2(t) + e_x(t)e_y(t) - e_y^2(t) - \frac{1}{s}e_z^2(t).$$

Перепишем правую часть неравенства (14) в следующем виде:

$$(15) \quad \begin{aligned} \dot{V}(e(t)) \leq & -\left(\gamma + 2\sigma - \frac{1}{2}\right)e_x^2(t) + e_x(t)e_y(t) - \frac{1}{2}e_y^2(t) - \\ & - \frac{1}{2s}e_z^2(t) - \frac{1}{2}\left(e_x^2(t) + e_y^2(t) + \frac{1}{s}e_z^2(t)\right). \end{aligned}$$

Для правой части неравенства (15) имеет место оценка сверху:

$$(16) \quad \begin{aligned} \dot{V}(e(t)) \leq & -\left(\gamma + 2\sigma - \frac{1}{2}\right)e_x^2(t) + e_x(t)e_y(t) - \frac{1}{2}e_y^2(t) - \\ & - \frac{1}{2s}e_z^2(t) - \varepsilon V(e(t)). \end{aligned}$$

Далее в неравенстве (16) выделим квадратичную форму и перепишем её в матричном виде.

$$(17) \quad \dot{V}(e(t)) \leq e^*(t)W e(t) - \varepsilon V(e(t)),$$

где

$$W = \begin{bmatrix} -(\gamma + 2\sigma - 1/2) & 1/2 & 0 \\ & 1/2 & -1/2 & 0 \\ & & 0 & 0 & -1/2s \end{bmatrix}.$$

Теперь найдем такие γ и σ , при которых квадратичная форма в неравенстве (17) является отрицательно определённой. Используя критерий Сильвестра, получим следующее условие:

$$(18) \quad \gamma + 2\sigma - 1 > 0.$$

Если условие (18) выполнено, тогда квадратичная форма в неравенстве (16) является отрицательно определённой, а значит, её можно оценить сверху нулём. Получим следующее неравенство:

$$(19) \quad \dot{V}(e(t)) \leq -\varepsilon V(e(t)).$$

В силу неравенства (19), траектории системы (8) стремятся к нулевому решению. А это означает, что нулевое решение системы (8) является асимптотически устойчивым. Переходя к исходным координатам, получаем:

$$(20) \quad \overline{\lim}_{t \rightarrow \infty} |\delta_x(t)| = 0, \overline{\lim}_{t \rightarrow \infty} |\delta_y(t)| = 0, \overline{\lim}_{t \rightarrow \infty} |\delta_z(t)| = s|r|.$$

Полученный результат можно сформулировать в виде теоремы.

Теорема 1. Пусть σ – сила связи между системами (5), а γ – коэффициент усиления в законе управления $u(t)$. Тогда $\forall \mathbf{x}_i(0)$, $i = 1, 2$, систем, закон управления $u(t)$ в форме (13) обеспечивает достижение цели управления (7), если выполнено условие (18).

Таким образом, выполнение условий теоремы 1 является достаточным условием для синхронизации систем (5). Причем относительно переменных $x_i(t)$ и $y_i(t)$, $i = 1, 2$, синхронизация является асимптотической. А из неравенства (19) следует, что скорость синхронизации подчиняется экспоненциальному закону.

2.3. СИНХРОНИЗАЦИЯ СИСТЕМ С ВОЗМУЩЕНИЯМИ

Несмотря на всю схожесть поведения системы Хиндмарш-Роуз с поведением биологических нейронов, её универсальность не позволяет учесть все физиологические особенности отдельного нейрона. Учесть данные особенности возможно, если известна некоторая априорная информация о них. Например известно, что они описываются возмущениями, которые представляют собой равномерно ограниченные функции от времени. Таким образом,

сеть (5), состоящая из двух неидентичных систем Хиндмарш-Роза, может быть переписана в следующем виде:

$$(21) \quad \begin{cases} \dot{\mathbf{x}}_1(t) = F_1(\mathbf{x}_1(t)) + \xi_1(t) + B[\sigma(x_2(t) - x_1(t)) + u(t)], \\ \dot{\mathbf{x}}_2(t) = F_2(\mathbf{x}_2(t)) + \xi_2(t) + B\sigma(x_1(t) - x_2(t)), \end{cases}$$

где

$$F_i(\mathbf{x}(t)) = \begin{pmatrix} y(t) - ax^3(t) + bx^2(t) - z(t) \\ c - dx^2(t) - y(t) \\ \varepsilon[s(x(t) - r_i) - z(t)] \end{pmatrix}, \quad i = 1, 2.$$

Здесь $\xi_i(t) = \text{col}(\xi_{x,i}(t), \xi_{y,i}(t), \xi_{z,i}(t))$, $i = 1, 2$, – вектор-функции возмущений связанных систем (21), причём

$$|\xi_{x,i}(t)| \leq \frac{\Delta_x}{2}, \quad |\xi_{y,i}(t)| \leq \frac{\Delta_y}{2}, \quad |\xi_{z,i}(t)| \leq \frac{\Delta_z}{2}, \quad i = 1, 2, \quad \forall t \geq 0.$$

Наличие возмущений в системах (21) требует такого управления, которое бы смогло обеспечить синхронизацию систем даже при их наличии. Далее покажем, что полученный закон управления (12) обладает данным свойством и обеспечивает синхронизацию систем (21). Для этого вычтем уравнения первой системы из второй и используем ранее введенные обозначения $\psi(t)$ и $\varphi(t)$:

$$(22) \quad \dot{\delta}(t) = \widehat{F}(\delta(t)) + \eta(t) + Bu(t),$$

где

$$\widehat{F}(\delta(t)) = \begin{pmatrix} -(a\psi(t) - b\varphi(t) + 2\sigma)\delta_x(t) + \delta_y(t) - \delta_z(t) \\ -d\varphi(t)\delta_x(t) - \delta_y(t) \\ \varepsilon(s\delta_x(t) - \delta_z(t) - sr) \end{pmatrix}.$$

Здесь $\eta(t) = \text{col}(\eta_x(t), \eta_y(t), \eta_z(t))$ – вектор-функция возмущений системы (22), такая, что

$$|\eta_x(t)| \leq \Delta_x, \quad |\eta_y(t)| \leq \Delta_y, \quad |\eta_z(t)| \leq \Delta_z \quad \forall t \geq 0.$$

Переходя к координатам $e_x = \delta_x$, $e_y = \delta_y$, $e_z = \delta_z + sr$, получим следующую систему:

$$(23) \quad \dot{e}(t) = \widetilde{F}(e(t)) + \eta(t) + Bu(t),$$

где

$$\tilde{F}(e(t)) = \begin{pmatrix} -(\alpha\psi(t) - b\varphi(t) + 2\sigma)e_x(t) + e_y(t) - e_z(t) + sr \\ -d\varphi(t)e_x(t) - e_y(t) \\ \varepsilon(se_x(t) - e_z(t)) \end{pmatrix}.$$

Здесь $e(t) = \text{col}(e_x(t), e_y(t), e_z(t))$ – вектор состояния системы (23).

Теперь покажем, что управление (12) обеспечивает синхронизацию системы (21). Для доказательства этого факта найдем производную функции Ляпунова (9) в силу системы (23). Получим следующее выражение:

$$\begin{aligned} \dot{V}(e(t)) &= -(\alpha\psi(t) - b\varphi(t) + 2\sigma)e_x^2(t) - e_y^2(t) - \frac{1}{s}e_z^2(t) + \\ (24) \quad &+ (1 - d\varphi)e_x(t)e_y(t) + sre_x(t) + u(t)e_x(t) + \\ &+ \eta_x(t)e_x(t) + \eta_y(t)e_y(t) + \frac{1}{\varepsilon s}\eta_z(t)e(t). \end{aligned}$$

Подставим закон управления (13) в выражение (24):

$$\begin{aligned} \dot{V}(e(t)) &= -(\alpha\psi(t) + \gamma + 2\sigma)e_x^2(t) + e_x(t)e_y(t) - e_y^2(t) - \\ (25) \quad &- \frac{1}{s}e_z^2(t) + \eta_x(t)e_x(t) + \eta_y(t)e_y(t) + \frac{1}{\varepsilon s}\eta_z(t)e_z(t). \end{aligned}$$

Далее выражение (25) будем оценивать сверху. Исключим слагаемое, содержащее функцию $\psi(t)$:

$$\begin{aligned} \dot{V}(e(t)) &\leq -(\gamma + 2\sigma)e_x^2(t) + e_x(t)e_y(t) - e_y^2(t) - \frac{1}{s}e_z^2(t) + \\ (26) \quad &+ \eta_x(t)e_x(t) + \eta_y(t)e_y(t) + \frac{1}{\varepsilon s}\eta_z(t)e_z(t). \end{aligned}$$

Значения слагаемых, содержащих функции $\eta_x(t)$, $\eta_y(t)$ и $\eta_z(t)$, можно оценить сверху значениями модулей этих слагаемых. Тогда получим оценку:

$$\begin{aligned} \dot{V}(e(t)) &\leq -(\gamma + 2\sigma)e_x^2(t) + e_x(t)e_y(t) - e_y^2(t) - \frac{1}{s}e_z^2(t) + \\ (27) \quad &+ |\eta_x(t)e_x(t)| + |\eta_y(t)e_y(t)| + \frac{1}{\varepsilon s}|\eta_z(t)e_z(t)|. \end{aligned}$$

Рассмотрим вспомогательное неравенство $\left(\sqrt{\nu}\alpha + \frac{1}{\sqrt{\nu}}\beta\right)^2 \geq 0$,

оно равносильно неравенству $|\alpha\beta| \leq \frac{\nu}{2}\alpha^2 + \frac{1}{2\nu}\beta^2$. А значит, име-

от место следующее выражение:

$$(28) \quad \begin{aligned} \dot{V}(e(t)) \leq & -(\gamma + 2\sigma) e_x^2(t) + e_x(t)e_y(t) - e_y^2(t) - \frac{1}{s} e_z^2(t) + \\ & + \frac{\nu_x}{2} e_x^2(t) + \frac{1}{2\nu_x} \eta_x^2(t) + \frac{\nu_y}{2} e_y^2(t) + \\ & + \frac{1}{2\nu_y} \eta_y^2(t) + \frac{\nu_z}{2\varepsilon s} e_z^2(t) + \frac{1}{2\varepsilon s\nu_z} \eta_z^2(t). \end{aligned}$$

Представим правую часть неравенства (28) в следующем виде:

$$(29) \quad \begin{aligned} \dot{V}(e(t)) \leq & -\left(\gamma + 2\sigma - \frac{1}{2}\right) e_x^2(t) + e_x(t)e_y(t) - \frac{1}{2} e_y^2(t) - \\ & - \frac{1}{2s} e_z^2(t) - \frac{1}{2} \left[\left(e_x^2(t) + e_y^2(t) + \frac{1}{s} e_z^2(t) \right) + \nu_x e_x^2(t) + \right. \\ & \left. + \frac{1}{\nu_x} \eta_x^2(t) + \nu_y e_y^2(t) + \frac{1}{\nu_y} \eta_y^2(t) + \frac{\nu_z}{\varepsilon s} e_z^2(t) + \frac{1}{\varepsilon s\nu_z} \eta_z^2(t) \right]. \end{aligned}$$

Как и ранее, выделим квадратичную форму в неравенстве (29), а ν_x , ν_y и ν_z положим равными $1/2$, $1/2$ и $\varepsilon/2$ соответственно. Отсюда получаем следующую оценку:

$$(30) \quad \begin{aligned} \dot{V}(e(t)) \leq & e^*(t)W e(t) - \frac{\varepsilon}{2} V(e(t)) + \\ & + \left(\eta_x^2(t) + \eta_y^2(t) + \frac{1}{\varepsilon^2 s} \eta_z^2(t) \right), \end{aligned}$$

где

$$W = \begin{bmatrix} -(\gamma + 2\sigma - 1/2) & 1/2 & 0 \\ & 1/2 & -1/2 \\ & 0 & 0 & -1/2s \end{bmatrix}.$$

Ранее показано, что при выполнении условий (18), квадратичная форма $e^*(t)W e(t)$ является отрицательно определённой. Кроме того, функции $\eta_x(t)$, $\eta_y(t)$ и $\eta_z(t)$ являются равномерно ограниченными. Поэтому правую часть неравенства (30) можно вновь оценить сверху:

$$(31) \quad \dot{V}(e(t)) \leq -\frac{\varepsilon}{2} V(e(t)) + h,$$

где

$$h = \left(\Delta_x^2 + \Delta_y^2 + \frac{1}{\varepsilon^2 s} \Delta_z^2 \right).$$

Разрешая неравенство (31) и устремляя t к бесконечности, получаем верхний предел:

$$(32) \quad \overline{\lim}_{t \rightarrow \infty} V(e(t)) = \frac{2h}{\varepsilon}.$$

Следовательно, траектории системы (22) ограничены эллипсоидом:

$$(33) \quad \frac{\delta_x^2(t)}{\left(\sqrt{\frac{4h}{\varepsilon}}\right)^2} + \frac{\delta_y^2(t)}{\left(\sqrt{\frac{4h}{\varepsilon}}\right)^2} + \frac{(\delta_z(t) + sr)^2}{\left(\sqrt{4sh}\right)^2} = 1.$$

Ограниченность траекторий в системе (22) свидетельствует о наличии приближённой синхронизации между связанными системами (21). Следует отметить, что точность такой синхронизации прямо пропорциональна величинам Δ_x , Δ_y и Δ_z . Как и ранее, сформулируем полученный результат в виде теоремы.

Теорема 2. Пусть $\xi_{x,i}(t)$, $\xi_{y,i}(t)$ и $\xi_{z,i}(t)$, $i = 1, 2$, – равномерно ограниченные функции возмущений в системе (21). Тогда $\forall \mathbf{x}_i(0)$, $i = 1, 2$, системы, закон управления $u(t)$ в форме (13) обеспечивает достижение приближённой синхронизации, если выполнены условия теоремы 1.

Таким образом, результат, полученный в пункте 3.2, удастся расширить на случай наличия модельных возмущений, описываемых детерминированными, равномерно ограниченными функциями.

3. Численные эксперименты

В рамках данной работы было проведено два численных эксперимента, целью которых являлось подтверждение адекватности полученных теорем, а следовательно и подтверждение существования такого закона управления $u(t)$, который обеспечит синхронизацию между двумя связанными системами Хиндмарш-Роуз как при отсутствии возмущений в этих системах (5), так и при их наличии (21). Далее рассмотрим каждый из экспериментов в отдельности, но прежде определим априорные данные эксперимента: начальные значения векторов состояния систем

$\mathbf{x}_1(0) = \text{col}(-1,81; -1; 1,6008)$ и $\mathbf{x}_2(0) = \text{col}(-1,8; -1,005; 0,8)$, а также их параметры $a = 1$, $b = 3$, $c = 1$, $d = 5$, $\varepsilon = 3 \cdot 10^{-3}$, $s = 4$, $r_1 = -1$, $r_2 = -0,8$, $\gamma = 4$, $\sigma = 0,4$.

Теперь перейдем к рассмотрению результатов первого эксперимента. Поскольку значения γ и σ заранее известны и удовлетворяют условиям теоремы 1, то для выбранных начальных значений переменных состояния систем (5) закон управления (12) гарантирует достижение цели управления (7). Другими словами, системы (5) под воздействием управления (7) переходят в синхронный режим функционирования за конечное время, что проиллюстрировано на рис. 1. Причём относительно переменных $x_i(t)$ и $y_i(t)$, $i = 1, 2$, синхронизация является асимптотической (рис. 1б и 1г). Отсюда можно заключить верность теоремы 1.

Как отмечалось ранее, модели Хиндмарш-Роуз свойственна некоторая неопределённость. В данном эксперименте будем предполагать, что эта неопределённость представляет собой возмущения, описываемые непрерывными и равномерно ограниченными функциями от времени. Основываясь на данных предположениях, для проведения второго эксперимента были выбраны следующие функции, моделирующие возмущения:

$$\begin{aligned}\xi_{x,1}(t) &= 0,02\varepsilon^2 \sin(100t), & \xi_{x,2}(t) &= 0,01\varepsilon^2 \sin(100t), \\ \xi_{y,1}(t) &= 0,02\varepsilon^2 \sin(t), & \xi_{y,2}(t) &= 0,01\varepsilon^2 \sin(t), \\ \xi_{z,1}(t) &= 0,2\varepsilon^2 \sin(0,01t), & \xi_{z,2}(t) &= 0,1\varepsilon^2 \sin(0,01t).\end{aligned}$$

Из равномерной ограниченности функций-возмущений и выполнения условий теоремы 1 следует выполнение теоремы 2. А значит, закон управления (12) обеспечивает достижение приближённой синхронизации систем в сети (21), что показано на рис. 2. Как можно видеть, область допустимой разницы между траекториями систем ограничена эллипсоидом (рис. 2б, 2г, 2е и 2ж), который, в свою очередь, описывается уравнением (33). А следовательно, теорема 2 верна.

4. Заключение

В статье рассмотрена задача управления синхронизацией двух неидентичных взаимосвязанных систем Хиндмарш-Роуз в условиях отсутствия и наличия модельных возмущений, описываемых непрерывными, равномерно ограниченными функциями. Математически доказана достижимость такой синхронизации посредством управления, предложенного в данной работе. Причём в случае отсутствия модельных возмущений наблюдается асимптотическая синхронизация систем относительно переменных $x_1(t)$, $x_2(t)$ и $y_1(t)$, $y_2(t)$. Кроме того, для случая присутствия возмущений в системах была получена множественная оценка области допустимой разницы между траекториями синхронизируемых систем, представляющая собой эллипсоид.

Проведённые численные эксперименты подтвердили адекватность полученных теоретических результатов для обоих исследуемых случаев.

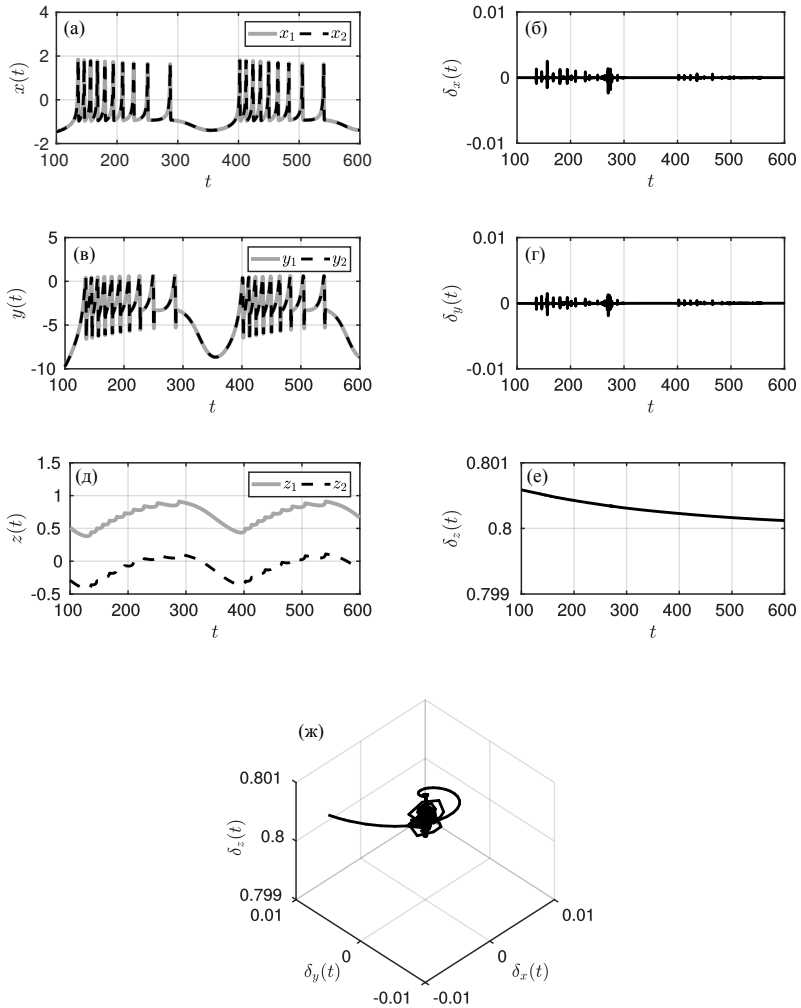


Рис. 1: Синхронизация двух связанных систем Хиндмарш-Роуз без возмущений: (а) поведение переменных $x_1(t)$ и $x_2(t)$; (б) разница между $x_1(t)$ и $x_2(t)$; (в) поведение переменных y_1 и y_2 ; (г) разница между y_1 и y_2 ; (д) поведение переменных z_1 и z_2 ; (е) разница между z_1 и z_2 ; (ж) разница между траекториями

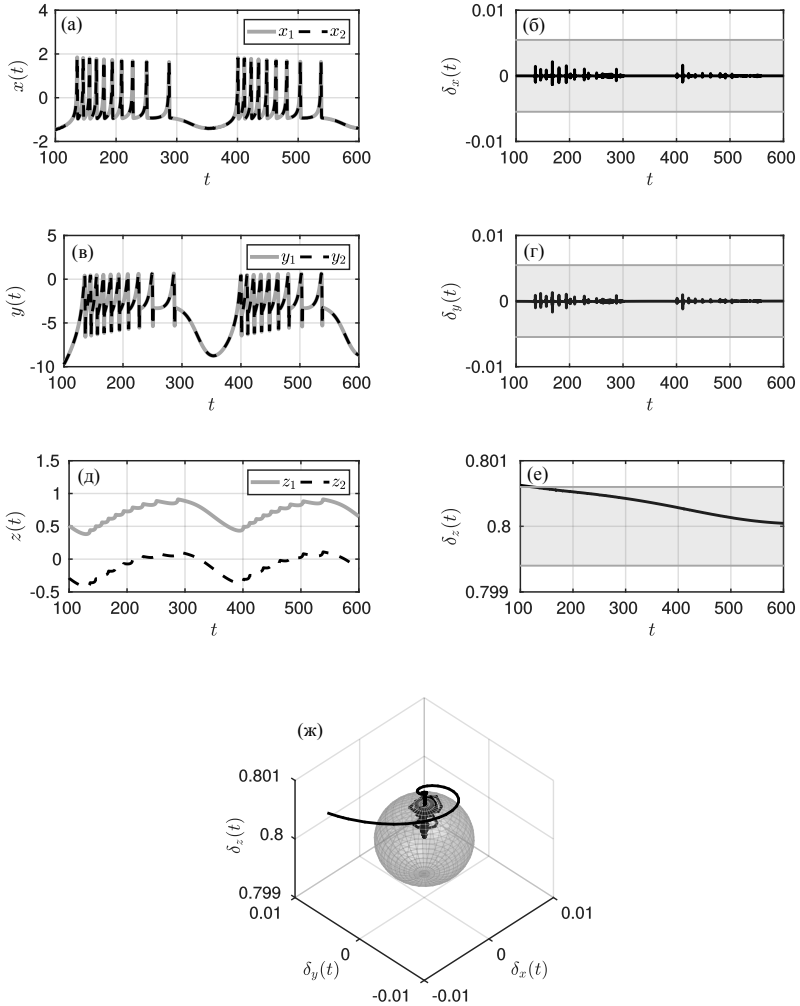


Рис. 2: Синхронизация двух связанных систем Хиндмарш-Роуз с возмущениями: (а) поведение переменных $x_1(t)$ и $x_2(t)$; (б) разница между $x_1(t)$ и $x_2(t)$; (в) поведение переменных y_1 и y_2 ; (г) разница между y_1 и y_2 ; (д) поведение переменных z_1 и z_2 ; (е) разница между z_1 и z_2 ; (ж) разница между траекториями

Литература

1. БЛЕХМАН И.И. *Синхронизация в природе и технике*. – М.: Наука, 1981. – 352 с.
2. ФРАДКОВ А.Л. *Кибернетическая физика: принципы и примеры*. – СПб.: Наука, 2003. – 208 с.
3. CORTESE S., FERRIN M., BRANDEIS D. *Neurofeedback for attention-deficit/hyperactivity Disorder: meta-analysis of clinical and neuropsychological outcomes from randomized controlled trials* // Journal of The American Academy of Child and Adolescent Psychiatry. – 2016. – Vol. 55, Issue 6. – P. 444–455.
4. FERNÁNDEZ T., BOSCH-BAYARD J., HARMONY T. *Neurofeedback in learning disabled children: visual versus auditory reinforcement* // Applied Psychophysiology And Biofeedback. – 2016. – Vol. 41, Issue 1. – P. 27–37
5. GOLOMB D., HANSEL D., MATO G. *Mechanisms of synchrony of neural activity in large networks* // Neuroinformatics and neural modeling, Handbook of biological physics. – Vol. 4. / Eds.: F. Moss, S. Gielen. – Amsterdam: Elsevier, 2001. – P. 887–968.
6. GUZENKO P.Y., LEHNERT J., SCHÖLL E. *Application of adaptive methods to chaos control of networks of Rössler systems* // Cybernetics and Physics. – 2013. – Vol. 2, Issue 1. – P. 15–24.
7. HINDMARSH J.L., ROSE R.M. *A model of the nerve impulse using two first-order differential equations* // Nature. – 1982. – Vol. 296, Issue 5853. – P. 162–164.
8. HINDMARSH J.L., ROSE R.M. *A model of neuronal bursting using three coupled first order differential equations* // Royal Society. – 1984. – Vol. 221, Issue 1222. – P. 87–102.

9. LEEMAN-MARKOWSKI B.A., SCHACHTER S.C. *Cognitive and behavioral interventions in epilepsy* // Current Neurology And Neuroscience Reports. – 2017. – Vol. 17, Issue 5. – P. 42.
10. LEHNERT J., HÖVEL A., SELIVANOV A.A., FRADKOV A.L., SCHÖLL E. *Controlling cluster synchronization by adapting the topology* // Physical Review E. – 2014. – Vol. 90, Issue 4. – 042914.
11. LU J., KURTHS J., CAO J. *Synchronization control for nonlinear stochastic dynamical networks: pinning impulsive strategy* // IEEE Trans. on Neural Networks and Learning Systems – 2012. – Vol. 23, Issue 2. P. 285–292.
12. MILTON J., JUNG P. *Epilepsy as a dynamic disease*. – Berlin: Springer, 2003. – 417 p.
13. OSIPOV G.V., KURTHS J., ZHOU C. *Synchronization in oscillatory networks*. – Berlin: Springer Verlag, 2007. – 370 p.
14. PESKIN C.S. *Mathematical aspects of heart physiology*. – New York: Courant Institute of Mathematical Sciences, 1975. – 278 p.
15. PLOTNIKOV S. *Controlled synchronization in two Fitzhugh–Nagumo systems with slowly-varying delays* // Cybernetics and Physics. – 2015. – Vol. 4, Issue 1. – P. 21–25.
16. PLOTNIKOV S.A., LEHNERT J., FRADKOV A.L., SCHÖLL E. *Adaptive control of synchronization in delay-coupled heterogeneous networks of FitzHugh–Nagumo nodes* // Int. Journal Bifurcation and Chaos. – 2016. – Vol. 26, Issue 4. – 1650058.
17. PLOTNIKOV S.A., LEHNERT J., FRADKOV A.L., SCHÖLL E. *Synchronization in heterogeneous FitzHugh–Nagumo networks with hierarchical architecture* // Physical Review E. – 2016. – Vol. 94 – 012203.

18. SELIVANOV A.A., LEHNERT J., DAHMS T., HÖVEL P., FRADKOV A.L., SCHÖLL E. *Adaptive synchronization in delay-coupled networks of Stuart-Landau oscillators* // Physical Review E. – 2012. – Vol. 85, Issue 1. – 016201.
19. SHILNIKOV A., KOLOMIETS M. *Methods of the qualitative theory for the Hindmarsh–Rose model: A case study. A tutorial* // Int. Journal Bifurcation and Chaos. – 2008. – Vol. 18, Issue 8. – P. 2141–2168.
20. STEUR E., TYUKIN I., NIJMEIJER H. *Semi-passivity and synchronization of diffusively coupled neuronal oscillators* // Physica D. – 2009. – Vol. 238, Issue 21. – P. 2119–2128.
21. STROGATZ S., STEWART I. *Coupled oscillators and biological synchronization* // Scientific American. – 1993. – Vol. 269, Issue 6. – P. 102–109.
22. TASS P.A. *Phase resetting in medicine and biology: stochastic modelling and data analysis*. – Berlin: Springer, 1999. – 329 p.
23. WANG Z., SHI X. *Lag synchronization of two identical Hindmarsh-Rose neuron systems with mismatched parameters and external disturbance via a single sliding mode controller* // Applied Mathematics and Computation. – 2012. – Vol. 218, Issue 22. – P. 10914–10921.
24. WINFREE A.T. *The geometry of biological time* – 2nd edition. – Berlin: Springer, 2001. – 803 p.

SYNCHRONIZATION CONTROL OF TWO COUPLED NON-IDENTICAL HINDMARSH–ROSE SYSTEMS

Danila Semenov, Saint Petersburg State University, student;
Institute for Problems of Mechanical Engineering of RAS,
Saint Petersburg, trainee researcher (semenovdm90@gmail.com).

Abstract: The problem of controlled synchronization between two coupled non-identical Hindmarsh-Rose systems, each of which describes the behavior of a biological neuron, was considered. The importance of solving this problem is caused by a variety of medical and biological research that determines the correlation between the certain diseases of the nervous system (such as epilepsy) and pathological synchronization between neurons in the brain areas. Thus, the ability to control synchronization between neurons is a promising method for the therapy of epilepsy and has been actively used in medical practice. Obviously, the development of this method of therapy requires the usage of qualitative mathematical tools. Our approach is based on the applying of the tools of control theory. In addition, it is necessary to take into account an inaccuracy of the modern neural models. In order to do this, we propose to consider this inaccuracy in the form of the continuous functions which describe the disturbances. Thus, using the control law, which was suggested in this article, and abiding by the theorems, which were formulated here, it is possible to achieve synchronized behavior of the systems in the conditions of absence and presence of the disturbances. The results were proved and confirmed by simulations.

Keywords: synchronization, coupled oscillators, Hindmarsh–Rose system.

УДК 62.50

ББК Ж 30

*Статья представлена к публикации
членом редакционной коллегии В.Г. Лебедевым.*

Поступила в редакцию 20.10.2017.

Дата опубликования 30.09.2018.

СИСТЕМЫ УПРАВЛЕНИЯ НЕСТАЦИОНАРНЫМ ОБЪЕКТОМ НА ОСНОВЕ МРС-РЕГУЛЯТОРА И ПИД-РЕГУЛЯТОРА С НЕЧЕТКОЙ ЛОГИКОЙ

Надеждин И. С.¹, Горюнов А. Г.²

*(Национальный исследовательский
Томский политехнический университет, Томск)*

Маненти Ф.³

(Миланский политехнический университет, Милан)

Как правило, в современной промышленности на любом этапе производственного процесса существует высокий уровень автоматизации. При этом технологические процессы и установки, встречающиеся в промышленности, являются нестационарными объектами управления, параметры которых могут изменяться в ходе технологического процесса. С целью повысить эффективность систем автоматического управления нестационарными объектами управления исследователи по всему миру проводят исследования, пытаясь объединить стандартный ПИД-регулятор с нечетким адаптивным контроллером. Наиболее часто настройка нечеткого регулятора производится на базе контроллера Мамдани. Кроме того, в последнее время для управления нестационарными объектами управления нашли широкое применение регуляторы на базе прогнозирующих моделей (Model Predictive Control). Данная статья посвящена сравнению систем автоматического управления нестационарным динамическим объектом на базе МРС-регулятора и ПИД-регулятора с нечеткой логикой. В ходе исследований производилось изменение параметров объекта управления и подавались возмущающие воздействия определенной амплитуды и длительности. В результате были получены переходные процессы, для которых были определены показатели качества. Сопоставление и анализ полученных результатов позволили сделать заключение о преимуществах одного регулятора перед другим.

Ключевые слова: системы автоматического управления, ПИД-регулятор, регулятор с нечеткой логикой, управление с прогнозирующей моделью.

¹ Игорь Сергеевич Надеждин, аспирант (kun9@list.ru).

² Алексей Германович Горюнов, д.т.н. (alex1479@tpu.ru).

³ Флавио Маненти, к.х.н., проф. (flavio.manenti@polimi.it).

1. Введение

В настоящее время в промышленности на любом этапе обработки материала/вещества существует высокий уровень автоматизации. Использование систем автоматического управления производственными процессами обуславливается необходимостью повышения уровня безопасности протекания технологического процесса, а также повышением экономических показателей установки и производства в целом. Как правило, большинство технологических процессов (установок), встречающихся в промышленности, представляют из себя нестационарные объекты управления [4]. Нестационарности могут быть причиной изменения свойств устойчивости и даже порядка (сингулярности) динамической модели объекта [3, 5]. Как показала практика, управление такими объектами с помощью традиционных ПИД-регуляторов не позволяет добиться требуемого качества управления.

С целью повысить эффективность систем автоматического управления нестационарными объектами исследователи по всему миру занимались исследованиями, пытаясь объединить стандартный ПИД-регулятор с нечетким адаптивным контроллером [9, 15]. В настоящее время существуют различные виды нечетких регуляторов, но нечеткие регуляторы на базе ПИД-регуляторов являются наиболее распространенными [6, 8]. Как правило, настройка нечеткого регулятора производится на базе контроллера Мамдани. При этом контроллер Мамдани находится непосредственно в канале управления, как это представлено в работах [7, 14]. Для данного исследования используется нечеткий ПИД-регулятор, представленный в работе [12], но без процедуры идентификации параметров объекта управления. Отличительная особенность данного нечеткого ПИД-регулятора в том, что в нем на основе экспертных оценок определяются коэффициенты ПИД-регулятора.

Также в последнее время для управления нестационарными объектами нашли широкое применение регуляторы на основе прогнозирующей модели – MPC-регуляторы (Model Predictive Control) [11, 13]. Для синтеза такого регулятора необходимо

составить математическую модель объекта управления, на основе которой регулятор будет прогнозировать изменение контролируемой переменной на определенный промежуток времени вперед и вычислять оптимальное управляющее воздействие для обеспечения наилучшей траектории контролируемой переменной.

Цель данной работы заключается в сравнительном анализе систем автоматического управления с ПИД-регулятором на базе нечеткой логики и с регулятором на базе прогнозирующей модели (МРС-регулятор). При этом будут изменяться параметры объекта управления, а также в систему будут вноситься ступенчатые возмущающие воздействия не известной величины и длительности.

2. Теоретическая часть

2.1. ОБЪЕКТ УПРАВЛЕНИЯ

В работе рассматриваются системы автоматического управления нестационарным динамическим объектом, поэтому его параметры могут изменяться в ходе режимов технологического процесса.

Динамика объекта управления описывается как линейная система с передаточной функцией $W_{p0}^u(s)$, которая характеризует канал по управлению и с передаточной функцией $W_{p0}^f(s)$, которая определяет канал по возмущению:

$$(1) \quad W_{p0}^u(s) = \frac{k_p}{T_p \cdot s + 1} \cdot e^{-\tau_p s},$$
$$W_{p0}^f(s) = \frac{k_p^f}{T_p^f \cdot s + 1} \cdot e^{-\tau_p^f s}.$$

В течение технологического процесса при переключении установки с одного режима на другой, при изменении характеристик материала (сырья) может изменяться порядок передаточной функции, описывающей объект управления по каналу управления $W_p^u(s)$. В рассматриваемом случае динамика объекта

управления будет описываться следующими передаточными функциями:

$$(2) \quad W_p^u(s) = \frac{k_p}{(T_{p1} \cdot s + 1) \cdot (T_{p2} \cdot s + 1) \cdot (T_{p3} \cdot s + 1)} \cdot e^{-\tau_p s},$$

$$W_p^f(s) = \frac{k_p^f}{T_p^f \cdot s + 1} \cdot e^{-\tau_p^f s}.$$

Практическим примером рассматриваемого случая является химико-технологический процесс радиохимического производства. Имеется три технологических аппарата A , B и C . В аппаратах A и B протекает одинаковый химический процесс, но при этом концентрации химических компонентов отличаются. Соответственно на выходе аппаратов получается продукт с разными химическими свойствами. Полученный в аппаратах A и B продукт выгружают в аппарат C . Для обеспечения непрерывности химико-технологического процесса выгрузка продукта происходит поочередно: сначала из аппарата A , а затем из аппарата B . Переключение с одного аппарата на другой и нестабильность характеристик перегружаемого продукта приводит к тому, что изменяется порядок передаточной функции, описывающей аппарат C как объект управления.

Кроме изменения параметров объекта управления (изменение порядка и значений коэффициентов передаточной функции), в течение технологического процесса с систему вносятся ступенчатые возмущающие воздействия неизвестной величины и длительности.

2.2. СИСТЕМА АВТОМАТИЧЕСКОГО УПРАВЛЕНИЯ С MPC-РЕГУЛЯТОРОМ

Схема предлагаемой системы автоматического управления с MPC-регулятором представлена на рис. 1.

Предложенная система управления (рис. 1) была реализована в пакете MATLAB/Simulink. С помощью программного инструмента MPC Toolbox был синтезирован MPC-регулятор.

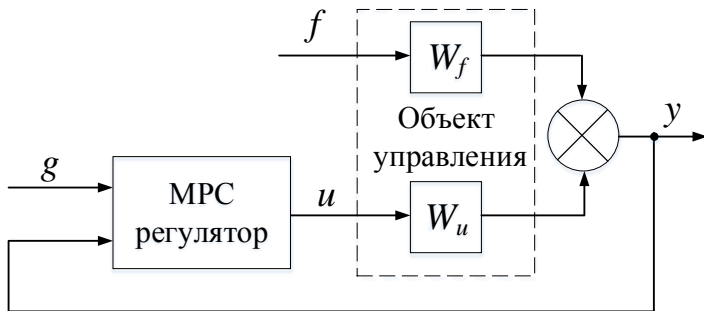


Рис. 1. Система автоматического управления с MPC-регулятором

Структурная схема реализованного MPC-регулятора представлена на рис. 2. В состав регулятора входит: прогнозирующая модель и блок оптимизации.

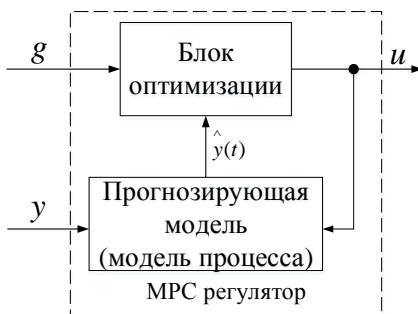


Рис. 2. Структурная схема MPC регулятора

Основную идею управления с прогнозирующей моделью можно представить следующим образом: имеются вход объекта по управлению $u(t)$ и выходная контролируемая переменная $y(t)$, $g(t)$ является желаемой величиной (зависимостью) изменения контролируемой переменной. Рассмотрим систему в дискретном времени, т.е. только в моменты времени $t = k \cdot \Delta T$, при этом ΔT – некоторый период квантования, а k – некоторое целое число. Для удобства графического представления будем считать $\Delta T = 1$.

Основной особенностью управления с помощью MPC-регулятора является наличие математической модели объекта управления (процесса, аппарата), которая достаточно точно описывает его поведение. Наличие адекватной математической модели объекта управления позволяет прогнозировать значения контролируемой переменной на определенное число шагов вперед (рис. 3).

В рамках данной работы математическая модель объекта управления была представлена передаточной функцией первого порядка (1) с определенными коэффициентом передачи и постоянной времени.

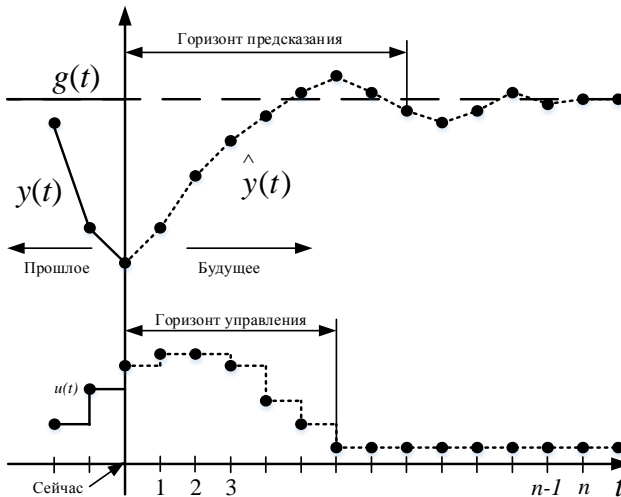


Рис. 3. Графическое представление концепции управления с помощью MPC-регулятора

Значения контролируемой переменной $y(t)$, предсказанные в некоторый момент времени t , на рис. 3 обозначены как $\hat{y}(t)$. Горизонт предсказания строится на определенное число шагов. Прогнозируемая траектория контролируемой переменной будет зависеть от значений управляющего воздействия $u(t)$.

Суть метода заключается в нахождении некоторой последовательности значений управляемой переменной $u(t)$, позволяющей обеспечить наилучшую траекторию для контролируемой переменной $y(t)$. Длина последовательности рассчитываемых управляющих воздействий $u(t)$ является фиксированной величиной и называется горизонтом управления. Искомая последовательность значений управляющего воздействия определяется в результате решения некоторой задачи оптимизации.

Для оптимизации используется целевой функционал, предложенный академиком А.А. Красовским [2]. Функционал содержит квадрат рассогласования между прогнозируемой выходной переменной объекта управления $y(t)$ и желаемой траекторией $g(t)$. При выборе оптимальных значений управляемой переменной $u(t)$ регулятор стремится минимизировать функционал, представленный выражением вида

$$(3) \quad J = \sum_{i=k}^{k+he} (g(i) - y(i))^2 + \sum_{i=k}^{k+hu} (u(i) - u(k))^2,$$

где $k = 1, 2, \dots, \infty$; he – число шагов, на которые строится прогноз поведения контролируемой переменной $y(t)$ (горизонт предсказания); hu – длина последовательности будущих значений управляющего воздействия $u(t)$ (горизонт управления).

После подачи на объект управления вычисленного оптимального управляющего воздействия $u(t)$ на следующем такте вся процедура повторяется заново с учетом вновь поступившей информации. Алгоритм функционирования MPC-регулятора представлен на рис. 4.



Рис. 4. Алгоритм функционирования MPC-регулятора

2.3. СИСТЕМА АВТОМАТИЧЕСКОГО УПРАВЛЕНИЯ С НЕЧЕТКИМ ПИД-РЕГУЛЯТОРОМ

Система автоматического управления с нечетким ПИД-регулятором аналогична схеме с MPC-регулятором (рис. 5).

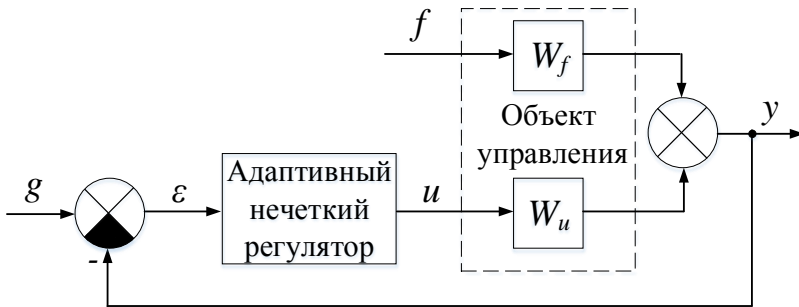


Рис. 5. Адаптивная нечеткая система автоматического управления

Показанные на рис. 5 переменные имеют следующее значение: g – уставка; f – измеряемое возмущение; u – управляющее воздействие; W_u – канал объекта по управлению; W_f – канал

объекта по возмущению; y – контролируемая переменная; ε – ошибка управления, определяемая как $\varepsilon = g - y$.

Следует более подробно рассмотреть адаптивный нечеткий регулятор, входящий в состав системы управления, представленной на рис. 5. Схема адаптивного нечеткого регулятора представлена на рис. 6.

Адаптивный нечеткий регулятор состоит из следующих блоков: генератор базы нечетких правил, нечеткий контроллер Мамдани и система для расчета J_n , J_e и J_u .

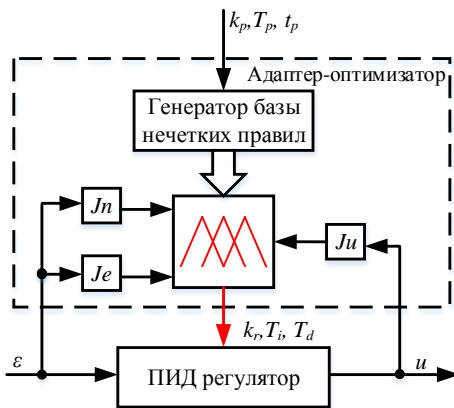


Рис. 6. Адаптивный нечеткий регулятор

Задача оптимизации состоит в минимизации функционала, который играет ключевую роль при настройке адаптивных и оптимальных систем управления. Оптимизируемый функционал имеет следующий вид:

$$(4) \quad \min(Je_k + Ju_k + Jn_k),$$

где

$$(5) \quad Je_k = \sqrt{\frac{\sum_{j=k}^{k+he} (\varepsilon_j)^2}{he - 1}},$$

$$(6) \quad Ju_k = \sqrt{\frac{\sum_{j=k}^{k+hu} (u_j - u_k)^2}{hu - 1}},$$

Jn_k – число ошибок управления, колеблющееся в интервале he , $k = 1, 2, \dots \infty$; ε_j – ошибка управления; u_j – управляющая переменная; he – интервал ошибок управления; hu – интервал контроля; j – индекс времени дискретизации.

Принцип функционирования генератора базы нечетких правил представлен на рис. 7. Параметры передаточной функции первого порядка задаются в качестве исходных данных генератору нечетких правил.

Для расчета параметров ПИД-регулятора используется контроллер Мамдани с нечеткими правилами, полученными путем минимизации функционала (4). Члены функции генерируются на основании двух групп параметров ПИД-регулятора. Первая группа параметров ПИД-регулятора (Kr , Ti , Td) рассчитываются с помощью метода оптимального модуля. А вторая группа параметров ПИД-регулятора вычисляется с помощью амплитудно-фазовой частотной характеристики (АФЧХ) разомкнутой системы $W_{\text{раз}}(j\omega)$. На рис. 8 представлена АФЧХ системы с разомкнутым контуром. ПИД-регулятор настраивается методом оптимального модуля (линия «Оптимальный модуль») и частотным методом (линия «Максимальный») для объекта управления, поведение которого описывается передаточными функциями первого порядка (выражения (1)). Далее происходит итерационный поиск параметров ПИД-регулятора с помощью адаптер-оптимизатора. При этом график АФЧХ разомкнутой системы должен находиться между кривыми «Оптимальный модуль» и «Максимальный» (рис. 8).

Используя обе группы рассчитанных параметров ПИД-регулятора генерируется база нечетких правил Мамдани. Этот набор правил и необходимых алгоритмов для их оценки был создан в пакете MATLAB. Конфигурация контроллера Мамдани для расчета параметров ПИД-регулятора на примере Kr представлена на рис. 9. Аналогичная конфигурация используется и для расчета Ti и Td .

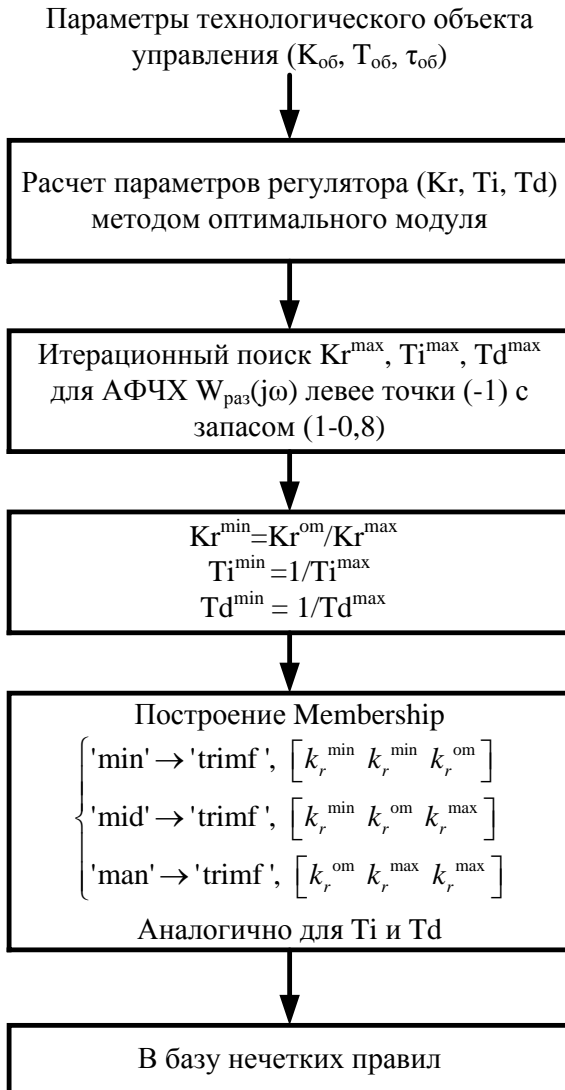


Рис. 7. Принцип работы генератора нечетких правил

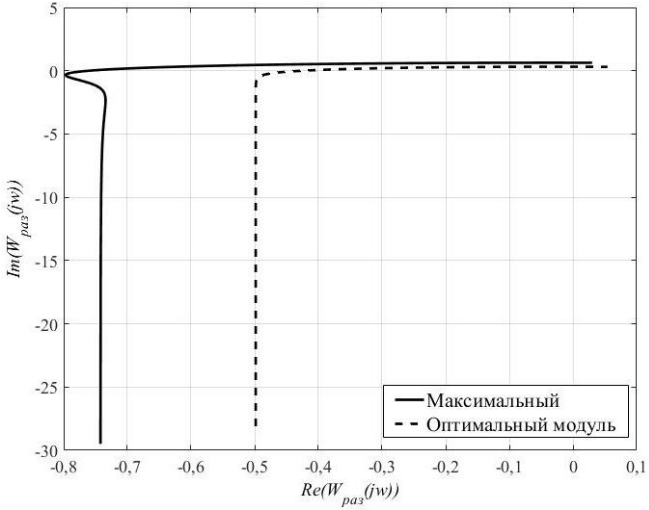


Рис. 8. Амплитудно-фазовая частотная характеристика системы с разомкнутым контуром

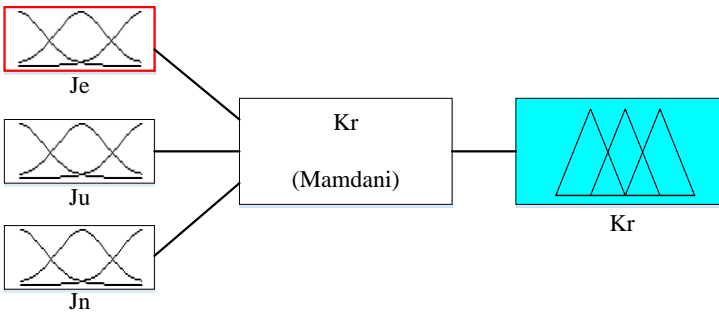


Рис. 9. Конфигурация контроллера Мамдани для вычисления параметров ПИД-регулятора в MATLAB (на примере Kr)

В общем виде ПИД-регулятор описывается функцией $W_r(s)$, представленной в виде

$$(7) \quad W_r(s) = k_r + \frac{k_r}{T_i \cdot s} + k_r \cdot T_d \cdot s,$$

где k_r – коэффициент передачи; T_i – постоянная интегрирования; T_d – постоянная дифференцирования. Параметры k_r , T_i и T_d рассчитывающиеся с помощью контроллера Мамдани (рис. 6).

Цифровая реализация ПИД-регулятора (7) описывается с помощью конечно-разностной формы

$$(8) \quad \begin{cases} u_j = k_r \cdot \varepsilon_j + C_j + \frac{k_r \cdot T_d}{\Delta t} \cdot (\varepsilon_j - \varepsilon_{j-1}), \\ C_j = C_{j-1} + \frac{k_r \cdot \Delta t}{T_i} \cdot (\varepsilon_j + \varepsilon_{j-1}); \end{cases}$$

где Δt – время дискретизации.

3. Результаты моделирования

Описанные ранее системы автоматического управления с разными регуляторами были реализованы в пакете MATLAB/Simulink. Для настройки нечеткого ПИД-регулятора в системе автоматического управления (рис. 5) были заданы параметры объекта управления, описываемого передаточной функцией первого порядка (1). Для синтеза регулятора с прогнозирующей моделью также использовалась передаточная функция первого порядка (1) с теми же параметрами, что и при настройке нечеткого ПИД-регулятора. В обеих системах автоматического управления накладывалось ограничение на управляющее воздействие.

В момент времени 0,0028 ч (10 с) было задано желаемое значение для контролируемой переменной (уставка $Y = 50\%$). Полученные в результате задающего воздействия переходные процессы представлены на рис. 10.

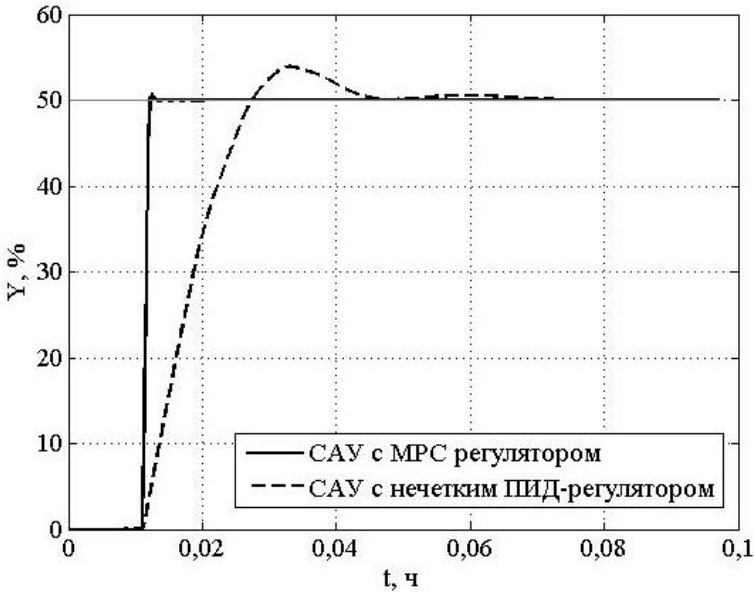


Рис. 10. Переходные процессы по задающему воздействию

Как видно из переходных процессов (рис. 10), обе системы управления выводят контролируемую переменную на заданный уровень. При этом время регулирования системы автоматического управления с нечетким ПИД-регулятором составило 0,036 ч (128,9 с), а с MPC-регулятором – 0,0094 ч (34 с). Перерегулирование системы автоматического управления с нечетким ПИД-регулятором составила 7,8%, а система управления с MPC-регулятором вышла на заданный уровень без перерегулирования. Также для переходных процессов, представленных на рис. 10, был рассчитан интегральный квадратичный критерий (*QIC*) качества переходных процессов по следующему выражению:

$$(9) \quad QIC = \int_{t_1}^{t_2} (g(t) - y(t))^2 dt.$$

Результаты расчетов представлены в таблице 1.

Таблица 1. Интегральные показатели качества

Регулятор	Интегральный квадратичный критерий (QIC)
Нечеткий ПИД-регулятор	$1,14 \cdot 10^5$
МРС-регулятор	$7,78 \cdot 10^4$

Проанализировав полученные показатели качества, отметим, что наилучший переходный процесс по задающему воздействию обеспечивает система автоматического управления с МРС-регулятором.

Затем в момент времени 0,14 ч в систему было внесено 30%-е ступенчатое возмущающее воздействие. Полученные переходные процессы представлены на рис. 11.

Для полученных переходных процессов (рис. 11) были определены показатели качества переходных процессов системы автоматического управления по возмущающему воздействию.

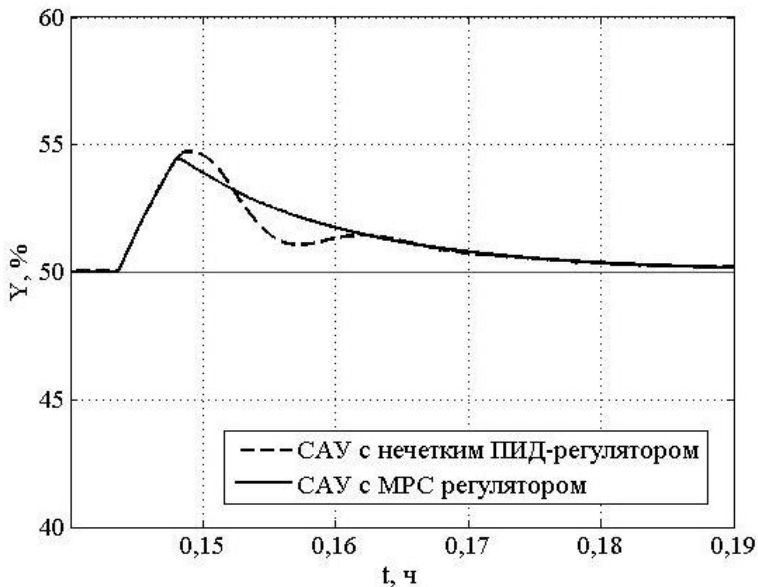


Рис. 11. Переходные процессы при 30 % ступенчатом возмущающем воздействии

Время регулирования системы автоматического управления с нечетким ПИД-регулятором составило 0,0275 ч (99 с), а с МРС-регулятором – 0,0318 ч (114,48 с). Вместо перерегулирования для оценки максимального отклонения контролируемой переменной от установившегося значения рассчитали относительное максимальное отклонение по следующей формуле:

$$(10) \sigma = \frac{y_{\max}}{g} \cdot 100[\%],$$

где y_{\max} – максимальное отклонение контролируемой переменной, g – заданное значение контролируемой переменной.

Результаты расчетов представлены в таблице 2. Также для переходных процессов, представленных на рис. 10, был рассчитан интегральный квадратичный критерий качества переходных процессов по выражению (9).

Таблица 2. Показатели качества переходных процессов

Регулятор	Относительное максимальное отклонение, σ (%)	Интегральный квадратичный критерий (QIC)
Нечеткий ПИД-регулятор	9,48	$1,09 \cdot 10^3$
МРС-регулятор	8,92	$1,16 \cdot 10^3$

Анализируя полученные значения показателей качества переходных процессов можно сделать вывод, что нечеткий ПИД-регулятор немного лучше обрабатывает возмущающее воздействие.

Далее в момент времени 0,28 ч было подано 50%-е ступенчатое возмущающее воздействие и изменили порядок передаточной функции, описывающей объект по каналу управления, с первого порядка на третий (2). В момент времени 0,38 ч было подано еще одно 50%-е возмущающее воздействие, при этом контролируемая переменная еще не стабилизировалась. Полученные в результате этого экстремального возмущающего воздействия переходные процессы представлены на рис. 12.

Для полученных переходных процессов (рис. 12) также были определены показатели качества переходных процессов. Время регулирования системы автоматического управления с

нечетким ПИД-регулятором составило 0,2384 ч (858 с), а с МРС-регулятором – 0,2836 ч (1021 с).

Рассчитали относительное максимальное отклонение контролируемой переменной от заданного значения согласно выражению (10), а также определили интегральный квадратичный критерий (QIC) качества переходных процессов по выражению (9). Результаты расчетов представлены в таблице 3.

Из полученных показателей качества переходных процессов очевидно, что с внесенными в систему возмущающими воздействиями лучше справляется система автоматического управления с нечетким ПИД-регулятором.

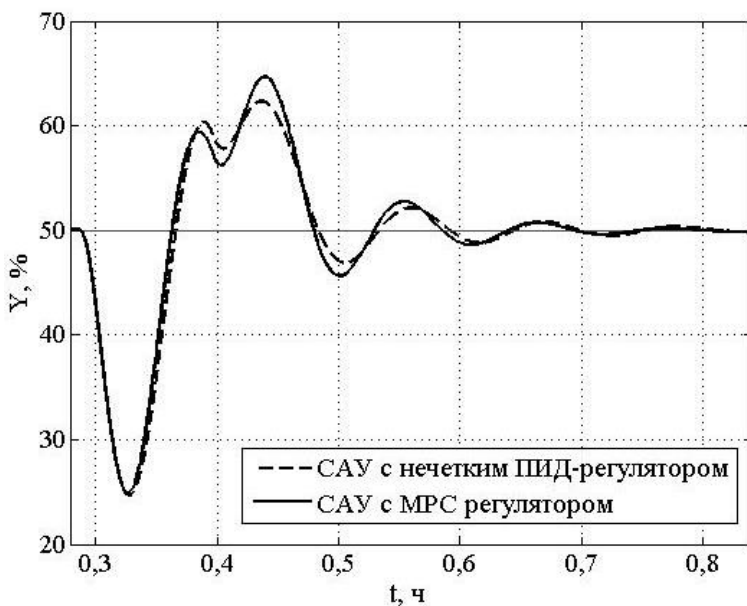


Рис. 12. Переходные процессы при 50 % ступенчатом возмущающем воздействии и изменении параметров объекта управления

Таблица 3. Показатели качества переходных процессов

Регулятор	Относительное максимальное отклонение, σ (%)	Интегральный квадратичный критерий (QIC)
Нечеткий ПИД-регулятор	50,68	$1,1543 \cdot 10^5$
МРС-регулятор	50,22	$1,1571 \cdot 10^5$

В момент времени 0,83 ч в систему было внесено 30%-е возмущающее воздействие, при этом объект по каналу управления описывался передаточной функцией третьего порядка. Так как оба регулятора были настроены на объект управления первого порядка, то качество управления несколько ухудшилось по сравнению с переходными процессами, представленным на рис. 11. Полученные переходные процессы представлены на рис. 13.

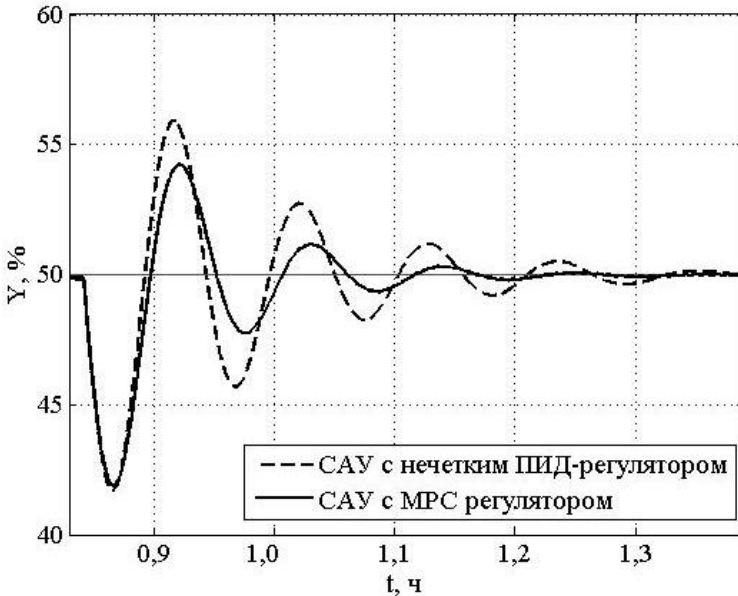


Рис. 13. Переходные процессы при 30 % ступенчатом возмущающем воздействии и новых параметрах объекта управления

Как видно на рис. 13, полученные переходные процессы отличаются наличием колебаний контролируемой переменной. Для представленных переходных процессов определили прямые показатели качества.

Время регулирования системы автоматического управления с нечетким ПИД-регулятором составило 0,1957 ч (704 с), а с МРС-регулятором – 0,1053 ч (379 с).

Аналогично предыдущим опытам, рассчитали относительное максимальное отклонение контролируемой переменной от заданного значения по выражению (10), а также определили интегральный квадратичный критерий качества переходных процессов по выражению (9).

Кроме того, для оценки качества колебательных переходных процессов по возмущающему воздействию вычислили степень затухания колебаний согласно следующему выражению:

$$(11) \psi = \frac{A_1 - A_2}{A_1},$$

где A_1 и A_2 – амплитуды двух соседних колебаний, направленных в одну сторону.

Результаты расчетов представлены в таблице 4.

Таблица 4. Показатели качества переходных процессов

Регулятор	Относительное максимальное отклонение, σ (%)	Интегральный квадратичный критерий (QIC)	Степень затухания, ψ
Нечеткий ПИД-регулятор	16,62	$1,2569 \cdot 10^4$	0,48
МРС регулятор	16,26	$9,4949 \cdot 10^3$	0,72

Рассчитанные показатели качества переходных процессов (рис. 13) численно подтверждают, что система автоматического управления с МРС-регулятором обладает значительно лучшим качеством управления.

Затем в момент времени 1,5 ч при новых параметрах передаточной функции объекта управления было подано управляющее воздействие. Желаемое значение для контролируемой

переменной было задано $Y = 60\%$. Полученные в результате задающего воздействия переходные процессы представлены на рис. 14.

Как видно из переходных процессов (рис. 14), обе системы управления выводят контролируемую переменную на заданный уровень. При этом время регулирования системы автоматического управления с нечетким ПИД-регулятором составило 0,175 ч (630 с), а с МРС-регулятором – 0,120 ч (432 с).

Для полученных переходных процессов были рассчитаны прямые показатели качества такие как, максимальное отклонение контролируемой переменной от заданного значения (10), интегральный квадратичный критерий качества переходных процессов (9) и степень затухания колебаний (11). Результаты расчетов представлены в таблице 5.

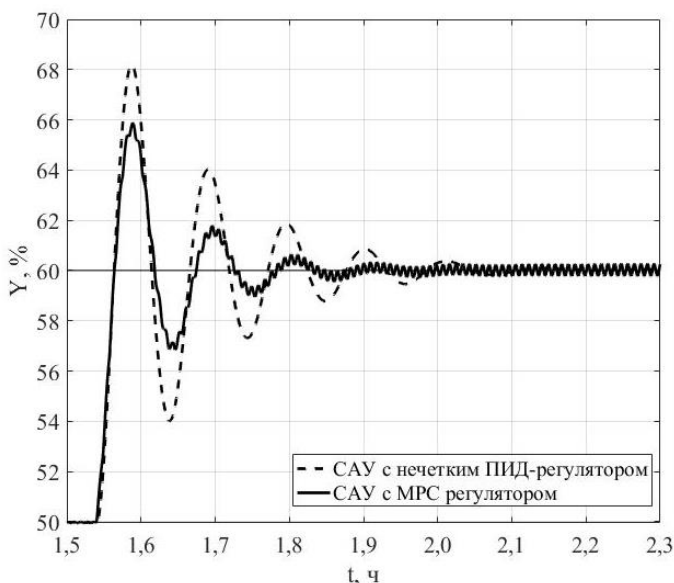


Рис. 14. Переходные процессы по задающему воздействию при новых параметрах объекта управления

Таблица 5. Показатели качества переходных процессов

Регулятор	Относительное максимальное отклонение, σ (%)	Интегральный квадратичный критерий (QIC)	Степень затухания, ψ
Нечеткий ПИД-регулятор	13,65	$1,4621 \cdot 10^4$	0,51
МРС регулятор	9,8	$1,0523 \cdot 10^4$	0,70

Полученные показатели качества переходных процессов (рис. 14) численно подтверждают, что система автоматического управления с МРС-регулятором обладает лучшим качеством управления по задающему воздействию.

4. Обсуждение полученных результатов

Сравнивая переходные процессы, представленные на рис. 10–14, и анализируя рассчитанные показатели качества для этих переходных процессов, определим регулятор, который обеспечивает наилучшее качество управления.

При выводе контролируемой переменной на заданный уровень наилучшие показатели качества переходных процессов обеспечивает система управления с МРС-регулятором. При использовании МРС-регулятора отсутствует перерегулирование контролируемой переменной (рис. 10), а также время регулирования меньше в 3,8 раза, чем при использовании нечеткого ПИД-регулятора.

Однако при внесении в систему 30%-го возмущающего воздействия были получены переходные процессы (рис. 11), демонстрирующие, что время регулирования нечеткого ПИД-регулятора в 1,15 раза меньше, чем при использовании МРС-регулятора, а также интегральный квадратичный критерий в 1,1 раза меньше. Но при этом относительное максимальное отклонение при использовании нечеткого ПИД-регулятора больше в 1,1 раза. Как видно, система управления с нечетким ПИД-регулятором обеспечивает переходные процессы с лучшими показателями.

Затем в систему было внесено 50%-е возмущающее воздействие, а также изменились параметры объекта управления и порядок передаточной функции, описывающей объект по каналу управления, с первого порядка на третий (2). По полученным переходным процессам (рис. 12) определили, что система управления с нечетким ПИД-регулятором обеспечивает в 1,2 раза меньшее время регулирования, но обладает на 0,5% большим относительным максимальным отклонением. Рассчитанные для переходных процессов (рис. 12) интегральные квадратичные критерии, представленные в таблице 3, отличаются незначительно. В данном случае система управления с нечетким ПИД-регулятором немного лучше обрабатывает возмущающие воздействия.

Далее при новых параметрах объекта управления было внесено 30%-е возмущающее воздействие. Из представленных на рис. 13 переходных процессов определили, что система управления с МРС-регулятором обеспечивает лучшие показатели качества (таблица 4). Время регулирования системы управления с МРС-регулятором в 1,86 раза меньше, чем у системы управления с нечетким ПИД-регулятором. А также меньше относительное максимальное отклонение в 1,02 раза и меньше интегральный квадратичный критерий в 1,37 раза. Кроме того, МРС-регулятор обеспечивает в 1,5 раза большую степень затухания колебаний контролируемой переменной, чем нечеткий ПИД-регулятор.

Также при новых параметрах объекта управления было подано управляющее воздействие. Из полученных переходных процессов (рис. 14) видно, что система автоматического управления с МРС-регулятором обладает лучшим качеством управления. Кроме того, это подтверждают рассчитанные показатели качества переходных процессов (таблица 5).

Таким образом, сравнивая и анализируя полученные результаты при моделировании системы автоматического управления с двумя разными регуляторами (нечеткий ПИД-регулятор и МРС-регулятор), невозможно выбрать какой-то один.

Отдельно отметим достоинства и недостатки рассматриваемых регуляторов.

К достоинствам рассматриваемого в статье нечеткого ПИД-регулятора относится возможность его реализации с помощью типовых промышленных элементов. Для этого необходим ПИД-регулятор и промышленный контроллер, в котором в виде DLL-библиотеки будет реализован адаптер-оптимизатор. Программное обеспечение контура адаптации не требует жесткого режима реального времени и может быть реализовано, например, на системе с платформой MS Windows. В этом случае также нет жесткой привязки к программному обеспечению нижнего уровня. Система управления на базе аналогичного нечеткого регулятора для управления каскадом центробежных экстракторов была реализована на радиохимическом заводе ОАО «Сибирский химический комбинат». Результаты этой работы нашли отражение в [1].

Необходимость в составлении точных и адекватных математических моделей объекта управления можно отнести к недостаткам МРС-регулятора. Во-первых, не всегда возможно составить адекватную математическую модель объекта управления, во-вторых, чем сложнее математическая модель объекта управления, тем больше вычислительных мощностей необходимо для реализации МРС-регулятора. Кроме того, МРС-подход требует существенных вычислительных ресурсов при управлении объектами, подобными рассматриваемому. Оптимизатор и прогнозирующая модель должны работать в темпе с управляемым процессом, и их затруднительно разделить на две независимые подзадачи.

5. Заключение

Данная работа посвящена синтезу и сравнению систем автоматического управления с нечетким ПИД-регулятором и с МРС-регулятором. При сравнении показателей качества, полученных переходных процессов, не удалось сделать однозначный выбор в пользу одного из сравниваемых регуляторов. Отмечены достоинства и недостатки рассмотренных регуляторов. С точки зрения практического применения и реализации рассмотренных регуляторов наиболее простым является нечеткий ПИД-

регулятор, так как для реализации этого регулятора могут быть использованы типовые промышленные элементы.

Литература

1. ГОРЮНОВ А.Г. Система автоматического управления каскадом центробежных экстракторов // Управление большими системами. – 2012. – №40. – С. 257–290.
2. КРАСОВСКИЙ А.А., БУКОВ В.Н., ШЕНДРИК В.С. Универсальные алгоритмы управления непрерывными процессами. – М.: Наука, 1977. – 272 с.
3. ПОЛЯК Б.Т., ЩЕРБАКОВ П.С. Робастная устойчивость и управление. – М.: Наука, 2002. – 303 с.
4. ПУПКОВ К.А., ЕГУПОВ Н.Д. Нестационарные системы автоматического управления: анализ, синтез и оптимизация. – М.: Издательство МГТУ им. Н.Э. Баумана, 2007. – 632 с.
5. ЦЫКУНОВ А.М. Адаптивное и робастное управление динамическими объектами по выходу. – М.: ФИЗМАТЛИТ, 2009. – 268 с.
6. ANG L.Y., JAFAR F.A. Simulation analysis of non-linear fuzzy PID temperature controller // Applied Mechanics and Materials. – 2014. – Vol. 465–466. – P. 677–681.
7. EKER I., TORUN Y. Fuzzy logic control to be conventional method // Energy Conversion and Management. – 2006. – Vol. 47. – P. 377–394.
8. FERREIDOUNI A., MASOUM M.A.S., MOGHBEL M. A new adaptive configuration of PID type fuzzy logic controller // ISA Transactions. – 2015. – Vol. 56. – P. 222–240.
9. GUZELKAYA M., EKSIN I., YESIL E. Self-tuning of PID-type fuzzy logic controller coefficients via relative rate observer // Engineering Applications of Artificial Intelligence. – 2003. – Vol. 16. – P. 227–236.
10. JOLEVSKI D., BEGO O. Model predictive control of gantry/bridge crane with anti-sway algorithm // Journal of Mechanical Science and Technology. – 2015. – Vol. 29. – P. 827–834.

11. MANENTI F. *Considerations on nonlinear model predictive control techniques* // Computers and Chemical Engineering. – 2011. – Vol. 35. – P. 2491–2509.
12. MANENTI F., ROSSI F., GORYUNOV A.G., DYADIK V.F., KOZIN K.A., NADEZHDIR I.S., MIKHALEVICH S.S. *Fuzzy adaptive control system of a non-stationary plant with closed-loop passive identifier* // Resource-Efficient Technologies. – 2015. – Vol. 1. – P. 10–18.
13. NAYHOUSE M., TRAN A., KWON J.S.I., CROSE M., ORKOULAS G., CHRISTOFIDES P.D. *Modeling and control of ibuprofen crystal growth and size distribution* // Chemical Engineering Science. – 2015. – Vol. 134. – P. 414–422.
14. PANA I., DASA S., GUPTA A. *Tuning of an optimal fuzzy PID controller with stochastic algorithms for networked control systems with random time delay* // ISA Transactions. – 2011. – Vol. 50. – P. 28–36.
15. SIDDIQUE N. *Fuzzy control* // Studies in Computational Intelligence. – 2014. – Vol. 517. – P. 95–135.

CONTROL SYSTEMS OF A NON-STATIONARY PLANT BASED ON MPC AND PID TYPE FUZZY LOGIC CONTROLLER

Igor Nadezhdin, National Research Tomsk Polytechnic University, Tomsk, Ph.D. student (kun9@list.ru).

Aleksey Goryunov, National Research Tomsk Polytechnic University, Tomsk, Ph.D., Head of Department (alex1479@tpu.ru).

Flavio Manenti, Polytechnic University of Milan, Milan, professor (flavio.manenti@polimi.it).

Abstract: Currently in the industry, at any stage of production process there is a high level of automation. As a rule, the majority of technological processes, occurring today in the industry, are non-stationary control objects. In order to improve the efficiency of automatic control system of non-stationary control object, researchers around the world, doing research, trying to combine the standard PID regulator with fuzzy adaptive controller. As a rule, setting a fuzzy controller is made

based on the Mamdani controller. Also, recently, for the control of non-stationary control object are widely used controllers on the basis of a predictive model. The purpose of this paper is the comparative analysis of automatic control systems with PID regulator based on fuzzy logic and regulator on the basis of a predictive model (MPC-controller). In the course of research, stepwise disturbance of a certain amplitude and duration were applied to the control object. As a result, transient processes were obtained and quality indicators were determined. By comparing and analyzing the results obtained, a conclusion was made about the advantages of one controller over another.

Keywords: automatic control system, PID controller, fuzzy logic controller, model predictive control.

УДК 681.5

ББК 32.965, 32.966

*Статья представлена к публикации
членом редакционной коллегии А.П. Курдюковым.*

Поступила в редакцию 02.06.2016.

Опубликована 30.09.2018.

ПРОТИВОДЕЙСТВИЕ СГОВОРУ В ДИСКРЕТНЫХ ДИНАМИЧЕСКИХ МОДЕЛЯХ КОМПЬЮТЕРНЫХ СЕТЕЙ¹

Горбатенко Д. Е.²

(Институт математики, экономики и информатики, Иркутск)

Семенов А. А.³

(Институт динамики систем и теории управления
им. В.М. Матросова, Иркутск)

Исследуется проблема противодействия сговору субъектов в компьютерных сетях. Для моделирования соответствующих ситуаций используются дискретные модели автоматного типа, близкие по своей сути к синхронным булевым сетям. В реальных ситуациях целью сговора является возникновение некоторых прав между двумя выделенными субъектами сети. В рамках введенной в статье модели сговор рассматривается как динамический процесс, протекающий в дискретном времени. Исходной сети сопоставляется дискретная динамическая система (ДДС), в которой сговору соответствует некоторая последовательность переходов между состояниями. Сговор считается успешным, если в результате некоторой последовательности переходов происходит передача прав от одного выделенного субъекта другому. Противодействие сговору осуществляется за счет деактивации ряда узлов. В реальной сети деактивации соответствует понижение уровня доступа между некоторыми субъектами. Конкретному примеру деактивации нескольких узлов будет соответствовать новая ДДС. Заключительным состоянием данной ДДС (как и исходной) обязательно будет неподвижная точка. Если в неподвижной точке новой ДДС нет передачи прав между двумя выделенными субъектами, то это означает, что деактивация блокировала сговор. Рассматривается комбинаторная задача выбора наименьшего деактивирующего множества. Данная задача сводится к задаче о булевой выполнимости (SAT) и решается с использованием современных SAT-решателей. В качестве модели распространения прав доступа рассматривается модель «Take-Grant» как одна из самых простых и хорошо изученных моделей компьютерной безопасности.

Ключевые слова: дискретные динамические системы, модель «Take-Grant», граф сговора, SAT.

¹ Работа выполнена при финансовой поддержке Российского научного фонда, проект №16-11-10046.

² Дмитрий Евгеньевич Горбатенко, аспирант (gorbadima@yandex.ru).

³ Александр Анатольевич Семенов, к.т.н., доцент (biclop.rambler@yandex.ru).

1. Введение

В связи с интенсивным развитием информационных сетей вопросы обеспечения различных аспектов безопасности в таких сетях приобретают особую актуальность. В данном контексте весьма ценными представляются любые формальные методы описания, моделирования и анализа действий злоумышленников, атакующих рассматриваемую сеть. При работе с соответствующими моделями могут возникать задачи комбинаторной природы, для эффективного решения которых «вычислительного потенциала» самой рассматриваемой модели может оказаться недостаточно. Более конкретно, мы подразумеваем ситуации, когда наблюдение за поведением сети при помощи, скажем, компьютерной симуляции не позволяет получить в достаточной степени полную картину о некоторых принципиально возможных в рамках данной сети процессах.

Простой пример такого типа – возникновение несанкционированных прав доступа вследствие неумышленных или умышленных действий некоторых участников сети. К неумышленным действиям можно отнести программно-аппаратные сбои, тогда как к умышленным – сговор между недобросовестными пользователями. Мэтт Бишоп в [13] построил эффективные алгоритмы выявления ситуаций сговора в рамках известной модели распространения прав доступа «Take-Grant». Однако нам ничего не известно о работах, в которых рассматривались бы задачи блокирования сговора. При детализации постановок такого рода задач становится ясна их комбинаторная природа. И если для выявления множеств объектов, которые в принципе могут вступить в сговор, может быть использовано имитационное моделирование (о чем М. Бишоп, кстати, не упоминает), то для блокирования ситуаций сговора обычного «проигрывания модели во времени», вообще говоря, недостаточно. Как мы увидим далее, ключевое свойство объектов в сети, способствующее возникновению сговора, – это свойство быть активными, т.е. выступать в роли проводников информации. Соответственно, противодействовать сго-

вору можно деактиваций некоторых объектов. Деактивация без ограничений, однако, может привести к потере важных функций сети. В контексте сказанного возникают задачи оптимизационного характера с дополнительными условиями: для противодействия сговору требуется деактивировать как можно меньше объектов, притом что деактивировать некоторые объекты в принципе недопустимо. Такого рода задачи имеют комбинаторную природу (требуется перебирать большое число различных кандидатов в решения). Конечно же, в подобных ситуациях применение имитационного моделирования, вообще говоря, не гарантирует приемлемый результат.

С нашей точки зрения наиболее естественный подход к описанным проблемам состоит в их сведении к комбинаторным задачам, для решения которых существуют алгоритмы с обоснованной аргументацией их практической эффективности.

В качестве основы для формального описания понятий и процессов, напрямую относящихся к компьютерной безопасности, была взята известная модель «Take–Grant» (TG) [1, 2, 13, 18, 21]. В роли комбинаторной задачи с хорошо развитой алгоритмикой использовалась задача о булевой выполнимости (Boolean Satisfiability Problem), сокращенно обозначаемая SAT [12]. Наконец, для исследования механизмов противодействия ситуациям, приводящим к сговору в сети, применялся подход, близкий к подходу, использованному ранее для анализа феномена конформного коллективного поведения [6, 20].

Следует отметить, что описательные возможности TG уже давно не отвечают современным реалиям. С другой стороны, TG, в силу своей простоты, является хорошим объектом для отработки новых подходов к анализу безопасности компьютерных сетей. В отношении результатов, полученных в настоящей работе для модели TG, мы надеемся, что смогли продемонстрировать принципиальную возможность применения современных комбинаторных алгоритмов к анализу ряда феноменов, относящихся к безопасности в компьютерных сетях. Именно такова была изначальная цель написания настоящей статьи. При этом мы, ко-

нечно же, осознаем тот факт, что для использования в реальных сетях описанные идеи и методы нуждаются в серьезной детализации. Для этого, как минимум, необходима тесная кооперация со специалистами-практиками в области компьютерной безопасности.

Приведем краткое описание структуры статьи. Второй раздел содержит необходимые в дальнейшем понятия, касающиеся дискретных динамических систем (ДДС), которые порождаются сетями. Здесь же приведены основные определения, относящиеся к задаче о булевой выполнимости (SAT), а также схематично описана техника сведения к SAT задач поиска циклов в ДДС. В третьем разделе приведено краткое описание известной модели «Take–Grant» (TG) передачи прав доступа в компьютерных сетях, а также результатов М. Бишоп по анализу ситуаций сговора в рамках расширенной TG-модели. Четвертый раздел является основным. В нем вводится понятие графа сговора и описывается алгоритм выделения графа сговора из графа исходной модели, сложность которого линейна от размера исходного графа. Данный алгоритм является модификацией алгоритма Бишоп. Затем ставится задача устранения сговора за счет деактивации как можно меньшего числа активных вершин в графе сговора (задача построения наименьшего по мощности множества, устраняющего сговор). Рассматриваются дискретные динамические системы, порождаемые графами сговора и их модификациями, в которых некоторые изначально активные вершины деактивированы. Установлена прямая связь между тем, что множество деактивированных вершин устраняет сговор, и видом неподвижной точки соответствующей ДДС. Затем задача поиска наименьшего по мощности множества, устраняющего сговор, сводится к SAT с использованием разработанных ранее техник пропозиционального кодирования ДДС. Пятый раздел содержит результаты вычислительных экспериментов.

2. Теоретические основы

2.1. СЕТИ И ПОРОЖДАЕМЫЕ ИМИ ДИСКРЕТНЫЕ ДИНАМИЧЕСКИЕ СИСТЕМЫ

Везде далее сеть – это произвольный граф (как правило, ориентированный и размеченный), вершины которого интерпретируются как участники некоторого коллектива, часто называемые агентами, а ребра или дуги интерпретируются как бинарные отношения на множестве агентов (например, «друзья», «влияние», «доступ» и т.п.). Наша ближайшая цель – дать содержательное определение динамических процессов, при помощи которых в дальнейшем можно будет представлять распространение прав по сети. С этой целью мы свяжем с рассматриваемой сетью дискретную динамическую систему (ДДС) автоматного типа [7]. Такие системы могут определяться различными способами. Мы начнем с описания класса ДДС, нашедших широкое применение при изучении коллективного поведения.

Итак, пусть $G = (V, A)$ – простой ориентированный граф, в котором V – множество вершин, A – множество дуг. Везде далее граф G называем сетью. Произвольная дуга $(w, v) \in A$ будет интерпретироваться как отношение «агент w влияет на агента v ». Для произвольной вершины $v \in V$ определим ее окрестность следующим образом:

$$V_v = \{w \in V, w \neq v | (w, v) \in A\}.$$

Можно сказать, что окрестность v образована теми вершинами сети, которые напрямую влияют на данную вершину. Под этим влиянием вершина v может менять свое поведение, например, переходить в активное или неактивное состояние. Для отображения такого рода процессов будем связывать с произвольной вершиной $v \in V$ специальную функцию, значениями которой являются элементы некоторого конечного множества Ω_v . Будем рассматривать нашу сеть в дискретные моменты времени $t \in \{0, 1, 2, \dots\}$, полагая, что в каждый конкретный момент t вершине v сопоставлен некоторый $\omega_v(t) \in \Omega_v$ (вообще говоря, некоторый символ – например, 0 или 1).

Пусть $v \in V$ – произвольная вершина сети и $V_v = \{v'_1, \dots, v'_l\}$ – окрестность v . Определим правило f_v , в соответствии с которым в момент времени $t + 1$ элементам $\omega_{v'_1}(t), \dots, \omega_{v'_l}(t)$ сопоставляется элемент $\omega_v(t + 1)$:

$$(1) \quad \omega_v(t + 1) = f_v(\omega_{v'_1}(t), \dots, \omega_{v'_l}(t)).$$

Определение 1. Функцию $f_v : \Omega_{v'_1} \times \dots \times \Omega_{v'_l} \rightarrow \Omega_v$ вида (1) будем называть *весовой функцией вершины v* . Явное задание всех весовых функций вершин сети задает функцию

$$(2) \quad F_G : \Omega \rightarrow \Omega,$$

где $\Omega = \Omega_{v_1} \times \dots \times \Omega_{v_n}$. Значения F_G , рассматриваемые в моменты времени $t \in \{0, 1, 2, \dots\}$, будем называть *состояниями сети*. Произвольное состояние сети в момент t будем обозначать через $W_G(t)$. Состояние $W_G(0)$ называется *начальным*.

Если для каждой $v \in V$ имеет место $\Omega_v = \{0, 1\}$, то рассматриваемая сеть называется *синхронной булевой сетью* (Synchronous Boolean Network, SBN) или сетью Кауффмана. Таким образом, в SBN весовые функции – это булевы функции, для задания которых можно использовать таблицы истинности или формулы.

В оригинальной статье Стюарта Кауффмана [19] рассматривались SBN, которые использовались в роли простых моделей динамических процессов в генных сетях. Однако природа коллективов, к которым могут применяться ключевые понятия и идеи этой работы, существенно разнообразней. Следуя [19], мы будем представлять функцию (2) в виде специального графа Γ , называемого графом переходов между состояниями, или кратко графом состояний (State Transition Graph, STG). Вершины Γ интерпретируются как состояния сети, т.е. n -кортежи из Ω . Дуги Γ интерпретируются как переходы между состояниями сети из момента t в момент $t + 1$. В некоторых работах Γ называют пространством фазовых переходов, подразумевая аналогию с непрерывными динамическими системами. Часто говорят, что граф Γ представляет поведение дискретной динамической системы (ДДС), заданной сетью G .

В случае ДДС граф Γ всегда содержит конечное число вершин. Так, если G – SBN, то граф Γ содержит 2^n вершин, каждой из которых соответствует двоичное слово длины n , представляющее собой набор значений весовых функций всех вершин сети.

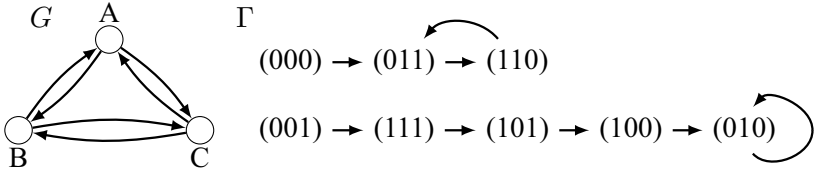
Определение 2. Ввиду конечности числа состояний ДДС, для любого начального состояния $W_G(0)$ обязательно найдутся такие натуральные числа L и M , $0 \leq L < M$, что в последовательности состояний

$$W_G(0), W_G(1), \dots, W_G(M)$$

имеет место: $W_G(L) = W_G(M)$. Обозначим через L_0, M_0 наименьшие L и M , обладающие таким свойством. Последовательность $W_G(L_0), \dots, W_G(M_0)$ будем называть циклом (в некоторых работах используется термин «аттрактор») длины $M_0 - L_0$. Цикл длины 1 называется стационарным состоянием или неподвижной точкой функции F_G .

На рис. 1 приведен пример SBN, иллюстрирующий введенные выше понятия. Приведено два способа задания весовых функций вершин рассматриваемой сети – таблицами и формулами.

Поскольку в общем случае имеет место экспоненциальная зависимость числа вершин в графе Γ от числа вершин в сети G , то задачи поиска стационарных состояний и циклов отображений вида (2) относятся к комбинаторным. При исследовании конкретных сетей для решения данных задач можно использовать различные эффективные на практике комбинаторные алгоритмы. В ряде работ для этих целей использовались алгоритмы решения проблемы булевой выполнимости (SAT) [3, 6, 16, 17, 20]. В следующем разделе мы кратко опишем техники сведения задач поиска циклов отображений вида (2) к SAT.



$f_B(t)$	$f_C(t)$	$f_A(t+1)$
0	0	0
0	1	1
1	0	0
1	1	1

$f_A(t)$	$f_C(t)$	$f_B(t+1)$
0	0	1
0	1	1
1	0	1
1	1	0

$f_A(t)$	$f_B(t)$	$f_C(t+1)$
0	0	1
0	1	0
1	0	0
1	1	1

$$f_A(t+1) = (f_B(t) \vee f_C(t)) \wedge (\neg f_B(t) \vee f_C(t))$$

$$f_B(t+1) = \neg(f_A(t) \wedge f_C(t))$$

$$f_C(t+1) = (f_A(t) \equiv f_B(t))$$

Рис. 1. Пример синхронной булевой сети, весовые функции вершин в которой заданы таблицами и булевыми формулами

2.2. ЭФФЕКТИВНЫЕ СВОДИМОСТИ К SAT ЗАДАЧ ПОИСКА ЦИКЛОВ ДИСКРЕТНЫХ ФУНКЦИЙ, ЗАДАВАЕМЫХ СЕТЯМИ

Везде далее функции, преобразующие двоичные слова в двоичные слова, будем называть дискретными функциями. Далее нас будут интересовать дискретные функции типа (2). Как мы увидим, задачи поиска циклов и неподвижных точек функции F_G удобно связывать с задачей обращения данной функции, которая заключается в нахождении по известному образу F_G произвольного его прообраза.

Пусть $\{0, 1\}^n$ – множество всех двоичных последовательностей длины n ($n \in \mathbb{N}$), а $\{0, 1\}^*$ – множество всех двоичных последовательностей произвольной конечной длины.

Рассмотрим произвольную всюду определенную (тотальную) дискретную функцию $f : \{0, 1\}^* \rightarrow \{0, 1\}^*$, заданную некоторым алгоритмом $A(f)$. Алгоритм $A(f)$ задает счетное семейство тотальных функций вида $f_n : \{0, 1\}^n \rightarrow \{0, 1\}^*$. Везде далее

мы будем рассматривать только такие ситуации, когда сложность $A(f)$ ограничена полиномом от n .

Задача обращения произвольной функции

$$(3) \quad f_n : \{0, 1\}^n \rightarrow \{0, 1\}^m$$

состоит в том, чтобы по известному $\gamma \in \text{Range} f_n$ найти такой $\alpha \in \{0, 1\}^n$, что $f_n(\alpha) = \gamma$.

Хорошо известно, что для рассматриваемого класса функций описанная задача обращения эффективно (за полиномиальное от n время) сводится к поиску выполняющего набора выполнимой конъюнктивной нормальной формы (КНФ). По сути, данный факт является прямым следствием известной теоремы Кука–Левина [4, 15].

Задача SAT (Boolean Satisfiability Problem) заключается в ответе на вопрос, выполнима ли произвольная КНФ C . SAT в распознавательной формулировке – классическая NP-полная задача. Тем не менее в последние 20 лет наблюдается выдающийся прогресс в разработке вычислительных алгоритмов решения SAT, применимых к различным тестам индустриального характера. В настоящее время лучшие алгоритмы решения SAT справляются с КНФ, содержащими десятки тысяч переменных и сотни тысяч дизъюнктов. Такие алгоритмы используются в символьной верификации, биоинформатике, комбинаторике и криптоанализе. Опишем кратко суть процедур перехода от задачи обращения дискретной функции к SAT.

Итак, рассматриваем задачу обращения произвольной функции вида (3). Заметим, что произвольную такую функцию мы можем рассматривать как набор из m булевых функций n -арности. Соответственно, функцию (3) мы можем задать при помощи схемы из функциональных элементов S_{f_n} над любым полным базисом [5]. У схемы S_{f_n} будет n входов и m выходов. Существует специальная техника, называемая преобразованием Цейтина [8], которая произвольной схеме S_{f_n} ставит в соответствие CNF C_{f_n} над множеством булевых переменных U , $X \subset U$. Здесь $X = \{x_1, \dots, x_n\}$ – переменные, приписанные входам схемы S_{f_n} , а $U \setminus X$ – переменные, приписанные всем гей-

там S_{f_n} . Сложность (time complexity) процедуры построения C_{f_n} по S_{f_n} линейна от общего числа узлов схемы S_{f_n} . В U выделим подмножество Y переменных, приписанных выходам S_{f_n} . Пусть $\gamma \in \{0, 1\}^m$ – произвольный набор значений переменных из Y . Означим в C_{f_n} все переменные из Y их значениями из γ и проделаем все возможные элементарные преобразования над полученной формулой. Обозначим результирующую CNF через $C_{f_n}(\gamma)$. Из свойств преобразований Цейтина следует, что если $\gamma \in Range f_n$, то $C_{f_n}(\gamma)$ выполнима, и из любого выполняющего ее набора можно выделить такой набор α значений переменных из X , что $f_n(\alpha) = \gamma$. Тем самым задача обращения произвольной функции вида (3) сведена к SAT.

В соответствии с описанной процедурой можно эффективно преобразовать задачу поиска цикла произвольного отображения вида (2) к SAT. Проиллюстрируем это на примере SBN. Пусть $G = (V, A)$ – SBN на n вершинах, весовые функции которых заданы булевыми формулами. Легко видеть, что функция (2) в данном случае – это дискретная функция вида

$$(4) \quad F_G : \{0, 1\}^n \rightarrow \{0, 1\}^n,$$

которую можно задать простым алгоритмом. Используя описанную выше технику, построим CNF C_{F_G} . Пусть U – множество булевых переменных, фигурирующих в C_{F_G} , и пусть $X = \{x_1, \dots, x_n\}$ и $Y = \{y_1, \dots, y_n\}$ – переменные, приписанные входам/выходам схемы S_{F_G} . Рассмотрим следующую CNF:

$$(5) \quad C_{F_G} \wedge (x_1 \equiv y_1) \wedge \dots \wedge (x_n \equiv y_n),$$

где через \equiv обозначена логическая эквивалентность. В свете сказанного выше несложно видеть, что CNF (5) выполнима тогда и только тогда, когда отображение F_G имеет неподвижные точки. Если (5) выполнима и μ – выполняющий её набор, то из него можно выделить набор α переменных из X , который определяет стационарное состояние в STG $\Gamma(F_G)$. Подобным же образом к SAT может быть сведена и задача поиска циклов произвольной длины в графах состояний отображений типа (2). В случае, когда G – это SBN на n вершинах, для этой цели можно рассмотреть функцию F_G^t , являющуюся t -кратной суперпозицией функции вида (4), где t – длина искомого цикла. Соответствующая

схема $S_{F_G^t}$ – это, по сути, цепочка из t схем вида S_{F_G} , в которой входы каждой последующей схемы соединены с выходами предшествующей. К КНФ $C_{F_G^t}$ по аналогии с (5) потребует приписать условие на совпадение состояний на входе и выходе F_G^t .

Описанный здесь алгоритм перехода от задачи поиска циклов в STG отображений типа (2) к SAT относится к классу процедур, называемых пропозициональными кодированиями. В работах [6, 20] более детально описаны процедуры пропозиционального кодирования ряда задач активационной динамики на сетях Кауффмана.

3. Модели компьютерной безопасности семейства «Take–Grant» и их основные свойства

Базовая модель «Take–Grant» (TG) [18, 21] – это формальная модель, интерпретирующая процессы передачи прав доступа в компьютерной сети. Сеть представляется в рамках TG-модели в виде простого ориентированного размеченного графа $G = (O, A)$.

O – множество всех вершин графа, интерпретирующих объекты. В множестве O выделяют специальное множество S «активных объектов», называемых также субъектами: они могут совершать некоторые действия в сети, например, брать или передавать другим объектам/субъектам какие-либо права доступа. Субъекты обозначаются в графе как ●. Объекты, не являющиеся субъектами (т.е. неактивные), обозначаются как ○.

$R = \{r_1, r_2 \dots r_n\} \cup \{t, g\}$ – множество видов прав доступа, в котором выделены следующие права:

- t – право брать права доступа;
- g – право давать (делегировать) права доступа;

A – множество дуг графа: $A \subseteq O \times O \times R$. Таким образом, дуги означают в TG права доступа между объектами.

Изменениям состояний системы соответствуют изменения графа. В классической TG-модели выделено два типа базовых

правил, определяющих передачу прав доступа между объектами: $take(P, x, y, z)$ и $grant(P, x, y, z)$. Через x, y, z здесь обозначены объекты, причем x – обязательно субъект. Через P обозначено множество прав доступа, которые в случае правила «take» x берет у y и передает z ; в случае правила «grant» x передает свои права доступа P к объекту z некоторому объекту y (в соответствии с отношением $x grant y$). Базовые правила $take$ и $grant$ называются де-юре правилами.

На рис. 2 показано, что в некоторый момент времени субъект x , имеющий право брать у объекта y некоторые права доступа на объект z , получает часть прав ($A \subseteq B$) на объект z . Как результат, в графе доступов возникает новая дуга (x, z) помеченная A . Так работает правило $take$. Работа правила $grant$ продемонстрирована на втором рисунке в аналогичном стиле.

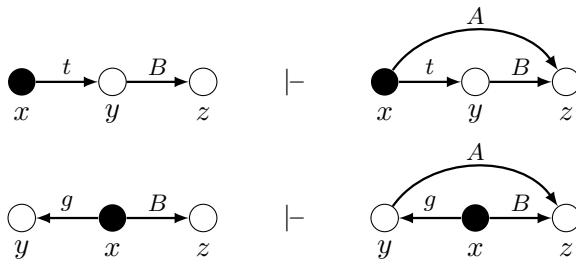


Рис. 2. Иллюстрация работы правил «take» и «grant» в базовой TG-модели

В статье [14] была описана расширенная модель «Take–Grant», в которой присутствовали де-факто правила, а также регламентировалась работа с информационными потоками. Далее мы будем использовать терминологию из [1], в рамках которой при наличии информационных потоков вида $p read q$ и $p write q$ будем говорить о соответствующих правах $read(r)$ и $write(w)$. Дуги, которые соответствуют этим правам в графе расширенной TG-модели, называются мнимыми.

Весьма детально базовая модель TG и ее расширения рассмотрены в [1, 2]. Для варианта «Take–Grant» без возможности

создания и удаления объектов в [18, 21] были предложены эффективные алгоритмы проверки ряда свойств, определяющих безопасность системы.

Основные интересующие нас вопросы будут касаться ситуаций получения в рамках модели расширенной TG прав вида $p\ read\ q$ в результате вступления в сговор некоторых объектов сети. Соответствующие проблемы были поставлены и изучены М. Бишопом в работе [13]. Сговор между участниками сети в [13] явно не определяется, однако по смыслу это ситуация, в которой сотрудничество ряда объектов/субъектов может привести к появлению информационных потоков на чтение или запись и соответствующих им прав на некоторый критический объект в сети. Важный факт заключается в том, что сговор может быть успешным лишь тогда, когда среди объектов, вступающих в сговор, имеется необходимое число активных объектов, в отношении которых М. Бишоп использует термин «*actors*». Именно такие объекты выполняют роль «проводников» информации, обеспечивая передачу прав r и w по сети. Основные достижения работы [13] – это эффективные (полиномиальные от числа вершин в графе) процедуры проверки возможности получения доступа к данным тех или иных вершин в результате сговора. Также в [13] устанавливается явная связь между числом активных объектов (*actors*) в специальном графе действий (*acting graph*) и успешностью сговора. Проиллюстрируем сказанное на следующем примере из [13].

В верхней части рис. 3 изображен фрагмент графа доступа компьютерной сети. Вершины графа помечены латинскими буквами. В данном примере по сети распространяются только два права: $read(r)$ и $write(w)$. Активные объекты (*actors*) обозначены сплошными точками, неактивные – пустыми. В нижней части рис. 3 изображен соответствующий граф действий (*acting graph*). Произвольная дуга (v_1, v_2) в графе действий означает получение права $read$ объекта v_1 на данные объекта v_2 . Можно заметить, что вершина p получает право r к данным вершины b , поскольку b пишет в a (которая сама по себе неактивна),

а p читает из a . Аналогично, c читает из d и затем пишет в b , соответственно, у p возникает право читать данные d . И так далее. Как итог, p может читать данные объекта q . Описанная ситуация дает пример сговора по похищению права на чтение данных вершины q : если вершины из множества $\{p, b, c, d, e\}$ действуют так, как только что было сказано, то p получает возможность читать данные q . С другой стороны, вершина s не может получить доступ к данным q . И, что очень важно, этому препятствуют две неактивные вершины i, j . Несмотря на право j писать в i , вершина j этого не делает, поскольку неактивна (здесь принципиальное отличие от предыдущего случая, где от вершины a не требовалось никаких действий). Соответственно, информация от q не проходит дальше вершины j .

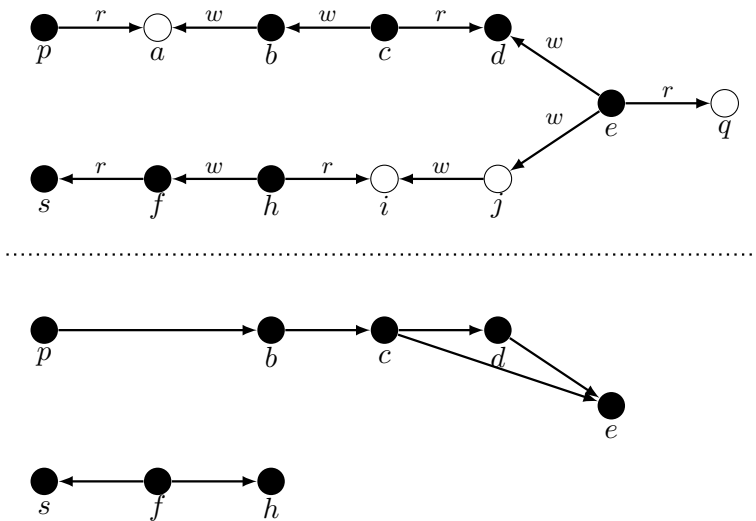


Рис. 3. Фрагмент сети «Take-Grant» и пример графа действий из работы [13]

4. Противодействие сговору в ДДС на базе модели «Take–Grant» (качественный анализ)

Еще раз отметим, что некоторые задачи анализа свойств безопасности сетей в рамках TG-модели, рассмотренные в [13, 18, 21], решаются за полиномиальное время. Тем не менее несложно описать задачи, естественным образом связанные с рассмотренными, но при этом являющиеся комбинаторными по своей природе. Далее мы концентрируемся на изучении задач такого типа. Следуя [13], будем рассматривать распространение по сети только прав *read* и *write*. Дополнительно отметим, что (как и в работах [13, 18, 21]) во всех рассматриваемых далее сетях запрещено создавать новые и удалять имеющиеся объекты. Это позволяет обеспечить так называемое «свойство монотонности» [9], наличие которого, собственно, и позволяет рассматривать распространение информации по сети как конечный процесс.

Сначала обратимся к задаче выявления множеств объектов, сговор между которыми приводит к появлению права *read* у объекта p на объект q . Предварительно уточним ряд понятий.

Определение 3. Пусть $G(O, A)$ – граф расширенной модели «Take–Grant» до начала каких-либо действий по распространению прав. Все дуги из A будем называть базовыми. Появление отношения *read* или *write* между двумя объектами будем интерпретировать новой дугой, не входящей в A . Будем называть такие дуги приобретенными.

Теперь предположим, что в результате распространения прав в рамках рассматриваемой модели возникло право p *read* q . В [13] был предложен эффективный (на самом деле линейный от размера исходного графа) алгоритм построения графа действий (acting graph) G' . Все вершины, входящие в G' , – это активные объекты, принимающие участие в передаче прав от q к p . Для дальнейших целей нам более удобен граф, который содержит вообще все объекты (как активные, так и неактивные), при участии которых возникает отношение p *read* q .

Определение 4. Граф $G_{pq} = (O_{pq}, A_{pq})$, который содержит

все вершины и базовые дуги G , участвующие в передаче прав от q к p , будем называть графом сговора. Произвольное множество $O', O' \subseteq O_{pq}$, объекты которого, взаимодействуя в рамках правил расширенной модели TG , обеспечивают передачу прав от q к p , будем называть множеством сговора. Множество O_{pq} будем называть полным множеством сговора.

Еще раз подчеркнем, что множество O_{pq} может содержать неактивные объекты (в рассмотренном выше примере такой объект представлен вершиной a). Анализируя алгоритм, приведенный в [12], несложно его модифицировать с целью построения графа G_{pq} . Соответствующий алгоритм неформально описан ниже.

Алгоритм 1 (Выделение графа сговора G_{pq}).

1. При всех перечисленных выше условиях (в том числе с учетом свойства монотонности) проследить распространение прав r и w в сети, запоминая для каждой приобретенной дуги те дуги (как базовые, так и приобретенные), благодаря которым она возникла (такие дуги называем предками).

2. В конечном состоянии сети (после которого не возникает новых дуг) убедиться, что между вершинами p, q существует отношение $p \text{ read } q$, которое отсутствовало в начальном состоянии.

3. Рекурсивно обойти предков приобретенной дуги (p, q).

4. Те вершины графа G и его базовые дуги, которые были пройдены в процессе выполнения шага 3, образуют граф сговора $G_{pq} = (O_{pq}, A_{pq})$.

Нетрудно убедиться, что для конкретных заданных $p, q \in O$ сложность описанного алгоритма построения G_{pq} зависит линейным образом от числа базовых дуг в графе G .

Определение 5. Пусть $G_{pq} = (O_{pq}, A_{pq})$ – граф сговора по передаче прав от q к p . Пусть $O', O' \subseteq O_{pq}$, произвольное множество сговора. В O' исключим из рассмотрения произвольную вершину v , обозначив полученное множество через O'' . Если для любой вершины $v \in O'$ соответствующее множество O'' не является множеством сговора по передаче прав от q к p , то множество O' назовем каналом сговора в графе G_{pq} .

Грубо говоря, канал – это один конкретный пример реализации сговора. Множество различных каналов, обеспечивающих отношение $p \text{ read } q$ в графе G_{pq} , может быть очень большим. Наша ближайшая задача – научиться блокировать такие каналы. Как уже было сказано, в рамках рассматриваемой модели передача прав в сети возможна только благодаря активным объектам. Соответственно, естественный способ блокирования каналов сговора – это придание некоторым активным объектам неактивного статуса (деактивация). Заметим, однако, что на практике далеко не каждый объект может быть деактивирован.

Определение 6. Пусть $G_{pq} = (O_{pq}, A_{pq})$ – граф сговора по передаче прав от q к p . Деактивируем один или несколько объектов из O_{pq} . Любое подмножество в O_{pq} , которое до деактивации было каналом сговора, а после деактивации перестало им быть, назовем фиктивным каналом. Множество вершин $D \subseteq O_{pq}$, деактивация которых делает фиктивными все каналы сговора в G_{pq} , будем называть множеством, устраняющим сговор по передаче прав от q к p .

Далее мы будем рассматривать задачу построения множества D наименьшей возможной мощности, а также задачу построения множеств типа D при дополнительных запретах на деактивацию некоторых вершин. Заметим, что сформулированные задачи интуитивно являются комбинаторными – так например, задача поиска деактивирующего множества наименьшей мощности похожа по постановке на известную NP-трудную задачу поиска наименьшего независимого множества. В настоящей статье мы, однако, не ставим вопрос выяснения структурной сложности изучаемых задач, а предлагаем вычислительные алгоритмы, которые могут быть использованы для их решения.

В данном контексте важно отметить следующий нюанс: деактивация некоторого активного объекта в канале, обеспечивающем право $p \text{ read } q$, может не дать фиктивный канал. Соответствующий пример приведен на рис. 3: в канале (p, a, b, c, d, e, q) вершина a неактивна, но данный канал не является фиктивным. Таким образом, на первый взгляд непонятно, как можно было

бы свести рассматриваемую задачу к «традиционным» комбинаторным задачам на графах (типа поиска максимальных клик или независимых множеств).

Далее для решения задач блокирования каналов сговора в сети обратимся к дискретным динамическим системам и описанным выше техникам их пропозиционального кодирования. Мы начнем с того, что свяжем с рассматриваемым вариантом TG-модели дискретную динамическую систему. Более точно, нас будет интересовать ДДС, порождаемая графом сговора. Ниже приведено описание этой ДДС.

Итак, пусть $G = (O, A)$ – исходная сеть, рассматриваемая в рамках описанной выше модели распространения прав *read* и *write*. Пусть $p, q \in O$ – фиксированные вершины сети и $G_{pq} = (O_{pq}, A_{pq})$ – эффективно выделенный из исходной сети граф сговора по передаче прав от q к p . Будем рассматривать распространение информации по графу G_{pq} как дискретный динамический процесс, шаги которого осуществляются в моменты времени $t \in \{0, 1, \dots\}$.

Выделим в множестве $O_{pq} \times O_{pq}$ подмножество A_{pq}^0 , образованное всеми теми парами вида (u, v) , в которых u не имеет права *read* на v в графе G_{pq} . Каждому моменту $t \in \{0, 1, \dots\}$ поставим в соответствие граф G_{pq}^t . Граф G_{pq}^0 совпадает с графом G_{pq} . При переходе от $t = 0$ к $t = 1$ для любой пары вершин $(u, v) \in A_{pq}^0$ проверяется возможность возникновения права u *read* v . Если такое право возникает, оно интерпретируется новой дугой (u, v) , помеченной правом r . Множество пар из A_{pq}^0 , для которых отношение « u *read* v » не возникло, обозначается через A_{pq}^1 . При переходе от $t = 1$ к $t = 2$ аналогичным образом проверяются все пары из A_{pq}^1 . И так далее. Ввиду свойства монотонности через конечное число M моментов времени данный процесс перестанет порождать новые дуги. Рассмотрим граф G_{pq}^M . Все дуги в G_{pq}^M , отличные от дуг из G_{pq} , порожденные описанным выше процессом, будем называть пунктирными. Очевидно, что каждой пунктирной дуге соответствует некоторая приобретенная дуга в смысле определения 3. Удалим из G_{pq}^M все дуги графа G_{pq} . Обо-

значим получившийся граф через $G_{pq}^* = (O_{pq}, A^*)$. Множество A^* , таким образом, включает только пунктирные дуги.

Пусть $A^* = \{a_1, \dots, a_K\}$, причем $a_K = (p, q)$ (очевидно, что пунктирная дуга a_K заведомо существует, поскольку G_{pq} – граф сговора по передаче прав от q к p). В каждый момент времени $t \in \{0, 1, \dots, M\}$ произвольной дуге $a \in A^*$ поставим в соответствие вес $w_a(t) \in \{0, 1\}$. Факт $w_a(t) = 0$ интерпретирует отсутствие в момент t права *read* между соответствующими объектами, а $w_a(t) = 1$ интерпретирует наличие данного права. Вектор $W(t) = (w_{a_1}(t), \dots, w_{a_K}(t))$ – это состояние нашей ДДС в момент t . Начальное состояние соответствует состоянию рассматриваемой модели до начала каких-либо действий по распространению прав. В этот момент всем пунктирным дугам приписан вес 0. В моменты времени $t \in \{1, 2, \dots, M\}$ веса пунктирных дуг пересчитываются фактически в соответствии с правилами выделения графа сговора приведенным выше алгоритмом (алгоритм 1). Иными словами, для каждого момента в отношении конкретной пунктирной дуги проверяется, возникла ли в рамках TG-модели соответствующая приобретенная дуга; в случае ответа «да» рассматриваемая пунктирная дуга получает вес 1, в противном случае она получает вес 0. Как итог имеем ДДС, которую будем обозначать через $\Delta(G_{pq}^*)$. Ввиду всего сказанного выше, несложно заметить, что данная ДДС не имеет циклов длины > 1 и имеет единственную неподвижную точку – состояние $W(M)$.

Рассмотрим граф $G_{pq}^* = (O_{pq}, A^*)$ и рассмотрим произвольное множество $D, D \subseteq O_{pq}$. В множестве O_{pq} промаркируем все вершины, которые входят в D , как неактивные. Полученное множество обозначим через \tilde{O}_{pq} . Рассмотрим граф $\tilde{G}_{pq} = (\tilde{O}_{pq}, A^*)$. Построим на основе графа \tilde{G}_{pq} ДДС, действующую в точности по тем же правилам, по которым действует ДДС $\Delta(G_{pq}^*)$. Обозначим данную ДДС через $\Delta(\tilde{G}_{pq})$. Установим справедливость следующего факта.

Теорема 1. *Множество вершин $D \subseteq O_{pq}$ устраняет сговор по передаче прав от q к p тогда и только тогда, когда в неподвижной точке системы $\Delta(\tilde{G}_{pq})$ вес дуги a_K равен 0.*

Доказательство. Докажем необходимость. Пусть D – множество, устраняющее сговор по передаче прав от q к p . Это, согласно определению 4, означает, что деактивация вершин из D делает фиктивными все каналы в G_{pq} , обеспечивающие передачу прав от q к p . Если предположить, что в неподвижной точке системы $\Delta(\tilde{G}_{pq})$ дуга a_K имеет вес 1, то это, согласно правилам построения $\Delta(\tilde{G}_{pq})$, будет означать, что в процессе перехода данной системы от начального состояния к неподвижной точке у объекта p возникло право *read* на объект q . Следовательно, некоторый канал в графе сговора после деактивации вершин из D не является фиктивным. Полученное противоречие доказывает необходимость. Наоборот, пусть в неподвижной точке системы $\Delta(\tilde{G}_{pq})$ дуга a_K имеет вес 0. Предположим, что при этом существует канал сговора в G_{pq} , не являющийся фиктивным. Но тогда в процессе входа системы $\Delta(\tilde{G}_{pq})$ в неподвижную точку обязательно должно возникнуть право *read* q , соответственно, дуга a_K в этом случае получила бы вес 1. Данное противоречие доказывает достаточность. Теорема 1 доказана.

Рисунки 4 и 5 иллюстрируют понятия, использованные в доказательстве теоремы 1.

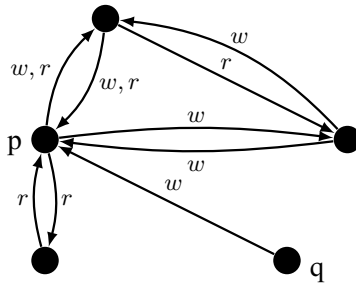


Рис. 4. Пример графа сговора

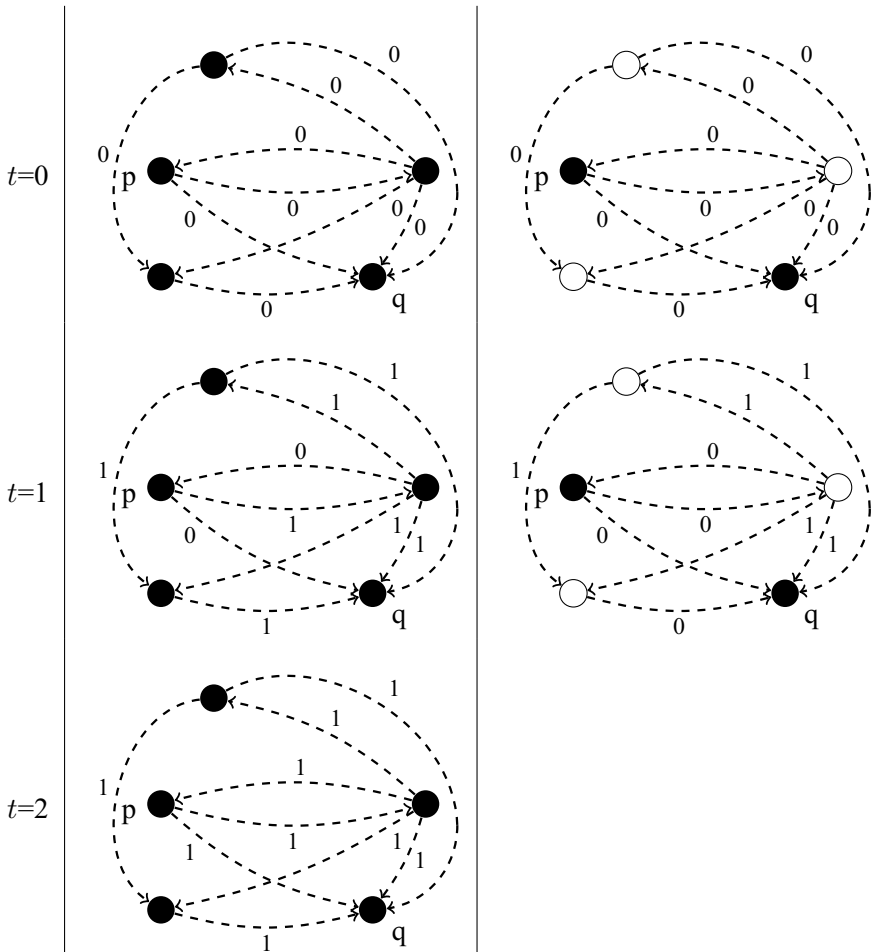


Рис. 5. Слева изображена ДДС $\Delta(G_{pq}^{r*})$, построенная на основе графа сговора, приведенного на рис. 4. Справа ДДС $\Delta(\tilde{G}_{pq})$. В первом случае ДДС входит в неподвижную точку в момент $t = 2$, во втором случае – в момент $t = 1$

5. Результаты вычислительных экспериментов

Опишем кратко вычислительные алгоритмы, которые позволяют строить устраняющие сговор множества наименьшей мощности. В основе разработанных алгоритмов лежат те же идеи, которые были использованы в статье [20]. Более точно, рассматривается конкретный граф сговора G_{pq} . Относительно каждой вершины из O_{pq} за исключением специально выделенных (вершина q , а также вершины, находящиеся слишком близко к q в смысле расстояния в графе) делается предположение о ее потенциально возможном попадании в D . Данный факт кодируется при помощи специально вводимых булевых переменных (значение 1 такой переменной соответствует тому факту, что соответствующая вершина попадает в D , значение 0 означает, что рассматриваемая вершина не попадает в D). Затем мы кодируем процесс входа системы $\Delta(\tilde{G}_{pq})$ в неподвижную точку. Из сказанного выше следует, что для этой цели достаточно закодировать последовательность переходов $\Delta(\tilde{G}_{pq})$ между состояниями $W(0), \dots, W(M)$. Затем мы добавляем условие $w_{a_K}(M) = 0$, а также условие, что число точек из O_{pq} , оказавшихся в D , не превосходит некоторого числа d . Также мы используем дополнительные ограничения на возможность вершин из O_{pq} попадать в D . В частности, мы запрещаем попадать в D вершинам, находящимся от q (целевой вершины) на расстоянии, меньшем заданного значения (в смысле стандартного расстояния в графе). Для кодирования последнего условия так же, как и в [20], используются известные техники работы с мощностными ограничениями (cardinality constraints) [10].

В тестовых экспериментах в роли исходных сетей использовались случайные графы на 200 вершинах, сгенерированные в соответствии с моделью Барабаши–Альберт [11]. Дополнительно предполагалось, что число активных вершин (субъектов) в сети не должно превышать 40. В каждом таком графе строился граф сговора G_{pq} (вершины p и q выбирались случайным образом). Всего в соответствии с описанной схемой было сгенерировано 700 графов, представляющих сети, для которых число вершин в

графе G_{pq} варьировалось от 30 до 50. Для каждого такого G_{pq} ставилась задача построения множества D (устраняющего сговор по передаче прав от q к p) наименьшей мощности. При этом мы накладывали запрет на деактивацию любой вершины в G_{pq} , из которой существует путь в q по дугам G_{pq} длины ≤ 6 . Поставленная так задача сводилась к SAT с использованием описанной выше техники. Средний размер (по 700 задачам) построенных КНФ составил 126 мегабайт. Для решения SAT-задач использовался SAT-решатель *minisat2.2* [22]. Все SAT-задачи были решены. Среднее время решения составило 0,11 секунды. Использовалась вычислительная платформа следующей конфигурации: процессор Core i7-2670QM, 8Gb RAM. Примерно в 50 случаях наименьшее множество, устраняющее сговор, имело размер от 3 до 5 вершин. В большинстве случаев находилось D мощности менее чем из 3 вершин. Примерно в 100 тестах SAT-решатель выдал UNSAT. Это означает, что в соответствующих случаях при имеющихся ограничениях (на расстояние от q до вершин, которые допустимо деактивировать) устранить сговор невозможно.

6. Заключение

В статье представлен подход к анализу и устранению ситуаций сговора, возникающих в рамках известной модели «Take-Grant» распространения прав в компьютерных сетях. В основе подхода лежит эффективный алгоритм выделения в исходном графе модели т.н. «графа сговора». Данный алгоритм является модификацией известного алгоритма построения графа действий, предложенного М. Бишопом в 1996 году. Деактивируя в графе сговора некоторые вершины, можно полностью устранить ситуацию сговора по передаче прав между двумя выделенными вершинами. Соответствующее множество деактивируемых вершин называется множеством, устраняющим сговор. Мы рассматриваем задачу построения наименьшего по мощности множества, устраняющего сговор. Данная задача является комбинаторной по своей природе, и для ее вычислительного решения мы сводим ее к задаче о булевой выполнимости (SAT). Для сведения мы стро-

им на основе графа сговора дискретную динамическую систему и используем известные техники пропозиционального кодирования таких систем. Описанный в статье подход к устранению ситуаций сговора был программно реализован и протестирован на случайных графах TG-модели размерностью 200 вершин. Задачи построения устраняющих сговор множеств наименьшей мощности удавалось успешно решать на обычном ПК.

Дополнительно мы хотим еще раз подчеркнуть, что основная цель статьи заключалась в демонстрации принципиальной возможности применения современных алгоритмов вычислительной логики (конкретно алгоритмов решения SAT) к комбинаторным задачам, возникающим при изучении ряда аспектов безопасности компьютерных сетей. На наш взгляд поставленная цель была достигнута, поскольку комбинаторная размерность задач, которые удастся успешно решать на обычном ПК с использованием предложенных алгоритмов, существенно превосходит возможности простого перебора. Конечно, применение идей и методов, представленных в статье, к реальным сетям потребует большой исследовательской работы и тесной кооперации со специалистами по практической компьютерной безопасности. Однако, с нашей точки зрения, этот шаг представляется осуществимым, и мы надеемся в дальнейшем работать над его реализацией.

Мы выражаем благодарность неизвестному рецензенту за целый ряд конструктивных замечаний, учет которых способствовал повышению качества статьи.

Литература

1. ДЕВЯНИН П.Н. *Модели безопасности компьютерных систем*: учеб. пособие. – 2005.
2. ДЕВЯНИН П.Н. *Модели безопасности компьютерных систем. Управление доступом и информационными потоками*: учеб. пособие. – 2011.

3. ЕВДОКИМОВ А.А., КОЧЕМАЗОВ С.Е., СЕМЕНОВ А.А. *Применение символьных вычислений к исследованию дискретных моделей некоторых классов генных сетей* // Вычислительные технологии. – 2011. – Т. 16, №1. – С. 30–47.
4. ЛЕВИН Л.А. *Универсальные задачи перебора* // Проблемы передачи информации. –1973. – Т. 9, №3. – Р. 265–266.
5. НИГМАТУЛЛИН Р.Г. *Сложность булевых функций*. – М.: Наука, Гл. ред. физ.-мат. лит-ры, 1991. – 240 с.
6. СЕМЁНОВ А.А., КОЧЕМАЗОВ С.Е. *О дискретно-автоматных моделях конформного поведения* // Управление большими системами. – 2013. – Вып. 46. – С. 266–292.
7. *Системная компьютерная биология* / Под ред. Н.А. Колчанова. – Новосибирск: Изд-во Сибирского отделения Российской академии наук, 2008. – 769 с.
8. ЦЕЙТИН Г.С. *О сложности вывода в исчислении высказываний* // Записки научных семинаров ЛОМИ АН СССР. – 1968. – Т. 8. – С. 234–259.
9. AMMANN P., WIJESEKARA D., KAUSHIK S. *Scalable, Graph-Based Network Vulnerability Analysis* // Proc. of the 9th ACM Conference on Computer and Communication Security (CCS'02), Washington, DC, USA, November 17–21, 2002. – Р. 217–224.
10. ASIN R., NIEUWENHUIS R., OLIVERAS A., RODRIGUEZ-CARBONELL E. *Cardinality networks: a theoretical and empirical study* // Constraints. – 2011. – Vol. 16, No. 2. – Р. 195–221.
11. BARABASI A.L., ALBERT R. *Emergence of scaling in random networks* // Science. – 1999. – Vol. 286. – P. 509–512.
12. BIERE A., HEULE M., VAN MAAREN H., WALSH T. *Handbook of Satisfiability* // Frontiers in Artificial Intelligence and Applications. – 2009. – Vol. 185. – P. 1–980.
13. BISHOP M. *Conspiracy and Information Flow in the Take-Grant Protection Model* // Journal of Computer Security. – 1996. – Vol. 4, No. 4. – P. 331–359.

14. BISHOP M., SNYDER L. *The Transfer of Information and Authority in a Protection System* // Proc. of the 7th Symp. on Operating Systems Principles, Pacific Grove, California, USA, December 10–12, 1979. – P. 45–54.
15. COOK S. *The complexity of theorem-proving procedures* // Proc. of the third annual ACM symposium on Theory of computing (STOC'71), Shaker Heights, Ohio, USA, May 03–05, 1971. – P. 151–158.
16. DUBROVA E., TESLENKO M. *A SAT-based algorithm for finding attractors in synchronous boolean networks* // IEEE/ACM Trans. on Computational Biology and Bioinformatics. – 2011. – Vol. 8. – P. 1393–1399.
17. EVDOKIMOV A.A., KOICHEMAZOV S.E., OTPUSSHCHENNIKOV I.V., SEMENOV A.A. *Study of discrete automaton models of gene networks of non-regular structure using symbolic calculations* // Journal of Applied and Industrial Mathematics. – 2014. – Vol. 8, No. 3. – P. 307–316.
18. JONES A., LIPTON R., SNYDER L. *A Linear Time Algorithm for Deciding Security* // Proc. of the 17th Annual Symp. on the Foundations of Computer Science, Houston, TX, USA, October 25–27, 1976. – P. 33–41.
19. KAUFFMAN S. *Metabolic stability and epigenesis in randomly constructed genetic nets* // Journal of Theoretical Biology. – 1969. – Vol. 22. – P. 437–467.
20. KOICHEMAZOV S., SEMENOV A. *Using Synchronous Boolean Networks to Model Several Phenomena of Collective Behavior.* // PLOS ONE. – 2014. – Vol. 9, No. 12, e115156. – P. 1–28.
21. LIPTON R., SNYDER L. *A Linear Time Algorithm for Deciding Subject Security* // Journal of the ACM. – 1977. – Vol. 24, No. 3. – P. 455–464.
22. *The MiniSat Page* [Электронный ресурс]. – <http://minisat.se>.

COUNTERACTION TO CONSPIRACY IN DISCRETE DYNAMICAL MODELS OF COMPUTERS NETWORK

Dmitriy Gorbatenko, Institute of Mathematics, Economics and Informatics, Irkutsk, postgraduate student (gorbadima@yandex.ru).

Alexander Semenov, Matrosov Institute for System Dynamics and Control Theory, Irkutsk, Cand.Sc., assistant professor (biclop.rambler@yandex.ru).

Abstract: The paper studies the problem of counteracting the collusion of subjects in computer networks. In real world situations the goal of the collusions is to obtain some access rights between two network subjects. Within the proposed model, collusion is considered as a dynamic process taking place at discrete time. An original network is associated with a discrete dynamic system (DDS), in which collusion corresponds to a sequence of transitions between states. Collusion is successful if as a result of a sequence of steps the access rights are transferred from one particular subject to another. A counteraction to such collusion is achieved by deactivating some nodes. In a real network a deactivation corresponds to reducing access level between subjects. Deactivation of several nodes induces a new DDS. A final state of such DDS is a stationary point. If in this point the access rights are not transferred between selected subjects, then deactivation successfully blocked collusion. In the paper we consider a problem of finding minimal deactivating set. It is reduced to Boolean satisfiability problem and solved using state-of-the-art SAT solvers. We use the Take-Grant model of computer security because it is one of the most simple and well-studied models.

Keywords: discrete dynamical systems, "Grant model", conspiracy graph, SAT.

УДК 519.7

ББК 22.17

*Статья представлена к публикации
членом редакционной коллегии М.В. Губко.*

Поступила в редакцию 06.03.2018.

Дата опубликования 30.09.2018.

КОМБИНИРОВАНИЕ МЕТОДА ВСТРЕЧНЫХ ПЛАНОВ И МЕТОДА ОБРАТНЫХ ПРИОРИТЕТОВ ДЛЯ РАСПРЕДЕЛЕНИЯ ФОНДА МОТИВАЦИИ НАУЧНЫХ ЛАБОРАТОРИЙ

Давыдов В. А.¹

(Московский институт экономики и математики
НИУ ВШЭ, Москва)

Рассматривается модель стимулирования агента организационной системы (ОС), использующая комбинацию метода обратных приоритетов и модифицированного метода встречных планов. Полученная модель мотивирует агента на сообщение планового значения КРІ, которое совпадает с прогнозом агента и одновременно стимулирует агента на подачу адекватной заявки на ресурс. Модель позволяет центру ОС распределить выделенный ресурс для выполнения каждого КРІ между агентами так, чтобы кроме сообщения адекватных планов агенты стремились к их выполнению. Вводятся ограничения на параметры модели, при которых выбор агента, равный его собственному прогнозу, является доминантной стратегией (ДС), при условии подачи агентом оптимальной по методу обратных приоритетов заявки на ресурс. Доказывается, что такой выбор для всех агентов ОС является равновесием в ДС. Описывается процедура определения плановых значений КРІ для агентов ОС и распределения ресурса между агентами. Параметры модели, определяемые центром ОС с учетом введенных ограничений, позволяют выбирать требуемые приоритеты для мотивации агентов и распределять в соответствии с определенными центром приоритетами мотивационный фонд. Приводится пример применения модели для стимулирования научных лабораторий по выполнению КРІ «число подготовленных студентов».

Ключевые слова: метод встречных планов, метод обратных приоритетов, стимулирование, КРІ.

1. Введение

Для мотивации агентов организационной системы (ОС), а также для процедуры планирования показателей КРІ агентов в статье [6] предложено использование метода встречных планов. Распределение ограниченного ресурса центра ОС в зависимости от заявленных агентами плановых показателей

¹ Вячеслав Анатольевич Давыдов, Советник НИУ ВШЭ
(novdav2017@yandex.ru, v.davydov@hse.ru).

КРІ может быть осуществлено методом обратных приоритетов [3]. В настоящей статье рассматривается комбинация данных двух методов, позволяющая использовать их преимущества для решения задач планирования показателей КРІ и мотивации агентов (в виде распределения между ними ограниченного бюджета).

2. Обзор литературы

В работе [9] отмечается, что задача распределения ограниченных ресурсов является актуальной задачей теории управления организационными системами (ТУОС) [4], математической экономики [7], микроэкономической теории [16], теории игр и теории выбора [11]. Кратко эта проблема может быть сформулирована следующим образом. Для каждого агента существует наилучшее с его точки зрения количество ресурсов (точка пика), которое он хотел бы получить, и сумма точек пика превышает количество ресурсов, имеющегося в системе. Управляющему органу – центру – необходимо распределить ресурсы между агентами, обеспечив при этом эффективность их использования в соответствии с теми или иными критериями. Процедура принятия решений центром, ставящая в соответствие вектору заявок агентов количество ресурсов, выделяемое тому или иному агенту, называется механизмом распределения ресурсов. Если агенты сообщают непосредственно требуемое им количество ресурсов, то механизм называется прямым. В классификации ТУОС механизмы распределения ресурсов принадлежат классу механизмов планирования, для которых помимо эффективности, важным свойством является их манипулируемость/неманипулируемость. При фиксированном механизме распределения ресурсов агенты являются вовлеченными в игру. Количество ресурсов, получаемое каждым из агентов, зависит в общем случае от заявок всех агентов. При этом в равновесии этой игры не всем агентам может быть выгодно честно сообщать информацию о своих точках пика. Механизм называется неманипулируемым, если при его использовании в равновесии всем агентам выгодно сообщать достоверную информацию.

Важным свойством механизмов распределения ресурсов является анонимность. Механизм является анонимным, если он симметричен относительно перестановок агентов, – итоговое распределение ресурсов зависит только от заявок агентов. В [2] было доказано, что при заданном количестве ограниченных ресурсов все анонимные монотонные по заявкам агентов механизмы эквивалентны, и, как следствие, обладают одинаковой эффективностью

Упомянутые выше результаты исследования механизмов распределения ресурсов считаются «классическими» в ТУОС [4, 12]. Однако эти результаты, как правило, ограничиваются только классом анонимных механизмов. В зарубежной литературе, посвященной данной проблематике, основной акцент делался также на анонимные механизмы. Наиболее полный обзор полученных результатов можно найти в [15]. Особо следует выделить работу [17], в которой был получен общий вид аналитической записи анонимных неманипулируемых механизмов распределения ресурсов. Данный результат был распространен на неанонимные механизмы в [14], где был получен общий вид записи любого неманипулируемого и неанонимного механизма распределения ресурсов и было доказано, что все такие механизмы являются механизмами последовательного распределения ресурсов.

Традиционно в ТУОС рассматриваются механизмы планирования (распределения ресурсов), удовлетворяющие следующим требованиям [3, 4]:

Р1. Процедура планирования непрерывна и монотонна по заявкам агентов.

Р2. Если агент получил некоторое количество ресурсов, то, изменяя свою заявку, он может получить любое меньшее количество ресурсов.

Р3. Если количество ресурсов, распределяемое между группой агентов, увеличилось, то каждый из агентов этой группы получит не меньшее количество ресурсов, чем раньше.

Достаточно широким и популярным классом механизмов распределения ресурсов, удовлетворяющим требованиям Р1–Р3, является класс приоритетных механизмов, в которых решение о том, как должен быть распределен ресурс между агентами,

определяется на основании их функций приоритета, аргументом которых являются заявки агентов на ресурс. Выделяют три класса приоритетных механизмов: прямых приоритетов, в которых функция приоритета каждого агента является возрастающей функцией его заявки на ресурс; обратных приоритетов, в которых функция приоритета убывает с ростом заявки агента на ресурс; и абсолютных приоритетов, в которых функция приоритета каждого агента не зависит от его заявки. Приоритетный механизм является анонимным, если все агенты имеют одинаковые функции приоритета.

В [2] было доказано, что все анонимные механизмы распределения ресурсов эквивалентны. Это означает, что все анонимные механизмы обладают одинаковой эффективностью. Отказ от анонимности делает актуальной задачу поиска эффективного по заданному критерию (например, максимума суммарной полезности всех агентов) механизма из класса механизмов последовательного распределения ресурсов (эквивалентных неанонимным механизмам).

Механизм обратных приоритетов, как отмечается в [4], обладает рядом преимуществ по сравнению с механизмом прямых приоритетов. Данный механизм достаточно часто используется для решения задачи оптимального распределения ресурса [5, 8, 1, 10]. Однако для эффективного управления ресурсом ОС задачу распределения ресурса необходимо увязывать с мотивацией агентов системы на выбор наиболее напряженного плана при планировании и последующее достижение заявленного плана. Модификация метода обратных приоритетов, которая решает данную задачу для неанонимного случая, рассматривается в разделе 5 настоящей работы. В разделах 3 и 4 описываются соответственно задача формирования плановых значений КРІ методом встречных планов и классическая задача распределения ресурса методом обратных приоритетов. В разделе 6 приводится пример использования предлагаемого модифицированного метода распределения ресурса для бюджетирования научно-исследовательских лабораторий.

3. Описание задачи формирования плановых значений KPI методом встречных планов

Опишем модель определения плановых значений одного KPI для нескольких подразделений образующих ОС, а также модель распределения ресурса для подразделений, входящих в ОС. Для этого будем использовать систему обозначений из [13]. В данной работе указанная задача называется «Задача стимулирования»

Пусть есть n подразделений – агентов с номерами $i \in N = \{1, 2, \dots, n\}$, объединение которых с центром образуют двухуровневую ОС. Первый уровень ОС образуют указанные агенты. Второй уровень – центр, который делает в игре первый ход, определяя для агентов их мотивационные факторы.

Каждое подразделение-агент i знает собственный реальный прогноз KPI, который равен R_i и неизвестен другим подразделениям, а также неизвестен центру ОС. При этом у каждого агента i есть множество различных потенциальных вариантов прогноза A_i . В нашей постановке задачи, когда все агенты устанавливают прогноз по одному и тому же KPI, можно считать, что множества вариантов прогноза совпадают для различных агентов. Другими словами $A_i = A_j, 1 \leq i, j \leq N$.

Рассмотрим ситуацию, когда у каждого агента есть единственный вариант прогноза. Будем полагать, что фактическое значение KPI, которое будет достигнуто агентом i , совпадает со значением прогноза R_i . Другими словами это означает, что каждый агент адекватно оценивает то значение KPI, которое может быть им достигнуто и может осуществить все действия для его достижения.

Обозначим $y_i \in A$ действие агента i по выбору прогноза KPI, или проще: y_i – значение прогноза KPI, которое заявляет центру агент i . Тогда

$$y = (y_1, y_2, \dots, y_n) \in \prod_{j \in N} A = A^*$$

– вектор действий агентов. Будем называть

$$\mathbf{y}_{-i} = (y_1, y_2, \dots, y_{i-1}, y_{i+1}, \dots, y_n) \in \prod_{\substack{j \in N \\ j \neq i}} A_j = A_{-i}$$

обстановкой игры для агента i .

Сделаем два предположения относительно рассматриваемой ОС.

Во-первых, будем считать, что затраты агентов не зависят от того, какой прогноз они заявят и какой будет получен фактический результат, т.е. какое значение КРІ получит агент в конце отчетного периода. Данное допущение в случае планирования КРІ обосновано, в том случае, когда все агенты-подразделения, имеют повременную оплату труда, фиксированный график работы и даже в том случае, когда возникает необходимость в привлечении какого-либо внешнего ресурса для более активной работы по выполнению КРІ, затраты на такое привлечение компенсируются центром ОС в полном объеме. Тогда, по сравнению с общим случаем, описанным в работе [13], можно считать, что целевая функция $f_i(\sigma_i, \mathbf{y})$ агента i , зависящая от механизма стимулирования агента σ_i и вектора действий агентов \mathbf{y} , фактически состоит только из стимулирования $\sigma_i(\mathbf{y})$. Или $f_i(\sigma_i, \mathbf{y}) = \sigma_i(\mathbf{y})$, $i \in N$.

Во-вторых, будем считать в рассматриваемой задаче, что единственная роль центра заключается в осуществлении управления. То есть у центра отсутствует собственный (не опосредованный агентом) результат деятельности. Такое предположение обосновано в том случае, когда результат центра равен суммарному результату всех агентов, и весь получаемый результат распределяется центром на стимулирование агента.

При стандартном механизме планирования «сверху-вниз» агенту выгодно получить от центра как можно менее напряженный план. Заведомо перевыполняя такой план, агент гарантирует себе дополнительное вознаграждение. Для получения менее напряженного плана агенту не выгодно раскрывать центру всю имеющуюся информацию относительно перспектив выполнения КРІ. Более того, агент может заведомо исказить (в свою пользу) имеющуюся у него информацию

относительно перспектив значения КРІ, которые могут быть им достигнуты.

Задача центра ОС состоит в том, чтобы предложить такую конструкцию установки неманипулируемого механизма стимулирования целевых значений КРІ для всех агентов $\sigma(\mathbf{y}) = (\sigma_1(\mathbf{y}), \dots, \sigma_n(\mathbf{y}))$, чтобы в результате каждый агент при проведении планирования с подходом «снизу–вверх» установил себе значение $y_i^* = R_i$. В работе [6] для решения данной задачи предложена конструкция модификации традиционного метода встречных планов. Соответствующая лемма в [6] доказывает, что предложенная конструкция имеет равновесие в доминантных стратегиях агентов.

Лемма 1. Пусть для одноуровневой ОС с n агентами заданы параметры $\alpha \geq 1$, $\beta > \alpha > \gamma > 0$ и вектор значений реальных прогнозов (R_1, R_2, \dots, R_n) . Тогда для целевой функции $\sigma(R_i, y_i, y_{-i}^0) = O_i K_i$ агента $1 \leq i \leq n$ где

$$O_i = \begin{cases} \gamma \left(\frac{R_i}{y_i} - 1 \right) + 1, & \frac{R_i}{y_i} \geq 1 \\ \beta \left(\frac{R_i}{y_i} - 1 \right) + 1, & \frac{R_i}{y_i} \leq 1 \end{cases}, K_i = \left(\frac{y_i}{y_{-i}^0} \right)^\alpha,$$

$$y_{-i}^0 = \frac{\sum_{j=1}^n y_j - y_i}{n-1},$$

существует равновесие в доминантных стратегиях (РДС), равное $y_i^* = R_i, 1 \leq i \leq n$.

Центр может модифицировать метод установления значения y_{-i}^0 . Это необходимо в том случае, например, если агенты организационной системы могут договориться между собой и в результате совместно дать заниженные прогнозы. В таких случаях значение y_{-i}^0 определяется по формуле

$$y_{-i}^{\circ} = \text{Max} \left\{ \frac{\sum_{j=1}^n y_j - y_i}{n-1}, y_{-i}^{\min} \right\},$$

Где y_{-i}^{\min} – некоторое минимальное значение КРІ, устанавливаемое центром для агента i .

4. Описание задачи распределения ресурса методом обратных приоритетов

Помимо получения от агентов плановых значений $y_i^* = R_i$, центру необходимо мотивировать агентов на достижение данных заявленных значений. Будем считать, что у центра имеется некоторый ограниченный объем ресурса S , который центр может использовать для их мотивации.

В качестве примера рассмотрим S как совокупный объем премии за определенный отчетный период, например, за год. Данный объем премии центр распределяет между агентами в зависимости от того вклада в выполнение некоторого КРІ, который внес каждый агент.

В [3] приводится механизм обратных приоритетов для распределения ресурса S . Согласно данному механизму, перед началом квартала каждый агент сообщает центру вместе с заявляемым плановым значением y_i также свою заявку на распределение ресурса s_i . Получаемый агентом i ресурс x_i вычисляется по формуле

$$x_i = \begin{cases} s_i, & \text{если } \sum_{j=1}^n s_j \leq S, \\ \min(s_i; \varphi \mu_i(s_i)), & \text{если } \sum_{j=1}^n s_j > S; \end{cases}$$

$$\mu_i(s_i) = \frac{y_i^2}{s_i}.$$

Здесь $\mu_i(s_i)$ – функция приоритета агента i в зависимости от его заявки на ресурс s_i и от заявленного планового значения y_i . Операция минимума отражает простое содержательное условие – агент получает ресурс в объеме не более запрашиваемой величины. Параметр φ выбирается из условия

$$\sum_{j=1}^n \min [s_j; \varphi \mu_j(s_j)] = S.$$

При достаточно большом числе n влияние отдельного агента на величину φ мало. Гипотезой слабого влияния в [1] называется ситуация выбора оценки ожидаемого эффекта, когда агенты не учитывают собственного влияния на параметр φ и считают его просто константой. В [1] доказывается, что механизм обратных приоритетов с функцией приоритета $\mu_i(s_i)$ при гипотезе слабого влияния обеспечивает оптимальное распределение ресурса. Также в [3] доказывается, что для агента i гарантирующей является стратегия

$$s_i^* = \sqrt{\varphi^* y_i}, \text{ где } \varphi^* = \left(\frac{S}{\sqrt{\sum_{i=1}^n y_i}} \right)^2.$$

5. Комбинирование метода встречных планов и метода обратных приоритетов

Рассмотрим функцию произведения $\eta(R_i, y_i, y_i^0, s_i) = O_i K_i x_i$. Данная функция объединяет свойства целевой функции агента i $\sigma(R_i, y_i, y_i^0) = O_i K_i$, и свойства функции приоритета ресурса, получаемого агентом i . Назовем данную функцию модифицированным ресурсом, получаемым агентом i .

Рассмотрим два варианта выбора

$$x_i = \begin{cases} s_i, \text{ если } \sum_{j=1}^n s_j \leq S, \\ \min(s_i; \varphi \mu_i(s_i)), \text{ если } \sum_{j=1}^n s_j > S. \end{cases}$$

Вариант 1. $x_i = s_i$.

Для данного варианта функция модифицированного ресурса принимает вид $\eta(R_i, y_i, y_{-i}^0, s_i) = O_i K_i s_i$. Отметим, что величина заявки на распределение ресурса s_i не зависит от величин R_i, y_i, y_{-i}^0 и может рассматриваться для данных параметров, как константа. Таким образом, РДС функции $\eta(R_i, y_i, y_{-i}^0, s_i) = O_i K_i s_i$ для параметров R_i, y_i, y_{-i}^0 не зависит от s_i и определяется в соответствии с условиями леммы 1.

Данный факт означает, что агенту с номером i , которому известен реальный прогноз КРП R_i , выгодно заявлять значение плана по КРП $y_i^* = R_i$ при проведении планирования методом встречных планов по методу, описанному в лемме 1. При этом заявляемая агентом i заявка на ресурс $x_i = s_i$ не зависит от заявляемого значения y_i^* .

Вариант 2. $x_i = \min(s_i; \varphi \mu_i(s_i))$.

Данный вариант состоит из двух случаев.

Для случая 1 $s_i < \varphi \mu_i(s_i)$ получаем $x_i = s_i$ и приходим к варианту 1.

Для случая 2 $s_i \geq \varphi \mu_i(s_i)$ получаем $x_i = \frac{y_i^2}{s_i}$,

и функция модифицированного ресурса принимает вид

$$\eta(R_i, y_i, y_{-i}^0, s_i) = O_i K_i \frac{y_i^2}{s_i}.$$

Определим условия для случая 2, при которых данная функция принимает максимум. Сначала докажем

Свойство 1. Если выполняются условия $\frac{R_i}{y_i} \geq 1$ и $\gamma \leq \alpha + 2$,

то функция $\eta(R_i, y_i, y_{-i}^0, s_i)$ является монотонно возрастающей.

Доказательство. Пусть

$$O_i = \gamma \left(\frac{R_i}{y_i} - 1 \right) + 1, \quad \frac{R_i}{y_i} \geq 1.$$

Тогда функция модифицированного ресурса принимает вид

$$\eta(R_i, y_i, y_{-i}^{\circ}, s_i) = \left(\gamma \left(\frac{R_i}{y_i} - 1 \right) + 1 \right) \left(\frac{y_i^{\circ}}{y_{-i}} \right)^{\alpha} \frac{y_i^2}{s_i},$$

$$y_{-i}^{\circ} = \text{Max} \left\{ \frac{\sum_{j=1}^n y_j - y_i}{n-1}, y_{-i}^{\text{min}} \right\},$$

Вычислим производную для данной функции

$$\frac{\partial \eta(R_i, y_i, y_{-i}^{\circ}, s_i)}{\partial y_i} =$$

$$= \frac{(1-\gamma) y_i \left(\frac{y_i^{\circ}}{y_{-i}} \right)^{\alpha}}{s_i} + \frac{(\gamma R_i - \gamma y_i + y_i) \left(\frac{y_i^{\circ}}{y_{-i}} \right)^{\alpha}}{s_i} +$$

$$+ \frac{(\gamma R_i - \gamma y_i + y_i) \left(\frac{y_i^{\circ}}{y_{-i}} \right)^{\alpha}}{s_i} \alpha.$$

Данная производная обращается в ноль при условии

$$y_i = \frac{0}{\frac{\gamma R_i (1+\alpha)}{2\gamma - 2 + \gamma\alpha - \alpha}}.$$

Корнем уравнения

$$\frac{\gamma R_i (1+\alpha)}{2\gamma - 2 + \gamma\alpha - \alpha} = R_i$$

Является $\gamma = \alpha + 2$. •

Свойство 1 отражает стимулирование агента при перевыполнении заявляемого агентом плана y_i .

Теперь докажем

Свойство 2. Если выполняются условия $\frac{R_i}{y_i} \leq 1$ и $\beta \geq \alpha + 2$,

то функция $\eta(R_i, y_i, y_{-i}^{\circ}, s_i)$ является монотонно убывающей.

Доказательство. Пусть

$$O_i = \beta \left(\frac{R_i}{y_i} - 1 \right) + 1, \quad \frac{R_i}{y_i} \leq 1.$$

Тогда функция модифицированного ресурса принимает вид

$$\eta(R_i, y_i, y_{-i}^\circ, s_i) = \left(\beta \left(\frac{R_i}{y_i} - 1 \right) + 1 \right) \left(\frac{y_i}{y_{-i}^\circ} \right)^\alpha \frac{y_i^2}{s_i}.$$

Вычислим производную для данной функции.

$$\begin{aligned} \frac{\partial \eta(R_i, y_i, y_{-i}^\circ, s_i)}{\partial y_i} &= \\ &= \frac{(1-\beta)y_i \left(\frac{y_i}{y_{-i}^\circ} \right)^\alpha}{s_i} + \frac{(\beta R_i - \beta y_i + y_i) \left(\frac{y_i}{y_{-i}^\circ} \right)^\alpha}{s_i} \\ &+ \frac{(\beta R_i - \beta y_i + y_i) \left(\frac{y_i}{y_{-i}^\circ} \right)^\alpha}{s_i} \alpha. \end{aligned}$$

Данная производная обращается в ноль при условии

$$y_i = \left[\begin{array}{c} 0 \\ \frac{\beta R_i (1+\alpha)}{2\gamma - 2 + \gamma\alpha - \alpha} \end{array} \right].$$

Корнем уравнения

$$\frac{\beta R_i (1+\alpha)}{2\gamma - 2 + \gamma\alpha - \alpha} = R_i$$

является $\beta = \alpha + 2$. •

Свойство 2 отражает стимулирование агента при невыполнении заявляемого агентом плана y_i . В [13] доказывается

Лемма 2. Если в игре n лиц $y_i \in [a_i, b_i]$, функции выигрыша непрерывны по совокупности стратегий для каждого игрока, частная производная $\frac{\partial \sigma(R_i, y_i, y_{-i}^\circ)}{\partial y_i}$ существует и знакопостоянна, то существует равновесие в доминантных

стратегиях (РДС). При этом доминантной стратегией y_i^* игрока i будет стратегия

$$y_i^* = \begin{cases} a_i, & \frac{\partial \sigma(R_i, y_i, y_{-i}^\circ)}{\partial y_i} < 0, \\ b_i, & \frac{\partial \sigma(R_i, y_i, y_{-i}^\circ)}{\partial y_i} > 0; \end{cases} \quad 1 \leq i \leq n.$$

Сформулируем основное утверждение данного раздела.

Теорема. Пусть для одноуровневой ОС с n агентами заданы параметры $\gamma > 0$; $\beta \geq \alpha + 2$; $\alpha > \gamma$; $\alpha \geq 1$, у центра имеется некоторый ограниченный объем ресурса S . Обозначим вектор значений реальных прогнозов агентов (R_1, R_2, \dots, R_n) и вектор заявок агентов на ресурс (s_1, s_2, \dots, s_n) .

Тогда для целевой функции модифицированного ресурса агента $1 \leq i \leq n$, $\eta(R_i, y_i, y_{-i}^\circ, s_i) = O_i K_i x_i$ где

$$O_i = \begin{cases} \gamma \left(\frac{R_i}{y_i} - 1 \right) + 1, & \frac{R_i}{y_i} \geq 1 \\ \beta \left(\frac{R_i}{y_i} - 1 \right) + 1, & \frac{R_i}{y_i} \leq 1 \end{cases}, \quad K_i = \left(\frac{y_i}{y_{-i}^\circ} \right)^\alpha,$$

$$y_{-i}^\circ = \text{Max} \left\{ \frac{\sum_{j=1}^n y_j - y_i}{n-1}, y_{-i}^{\min} \right\},$$

где y_{-i}^{\min} – некоторое минимальное значение КРІ, устанавливаемое центром для агента $1 \leq i \leq n$, и

$$x_i = \begin{cases} s_i, & \text{если } \sum_{j=1}^n s_j \leq S, \\ \min \left(s_i; \varphi \frac{y_i^2}{s_i} \right), & \text{если } \sum_{j=1}^n s_j > S, \end{cases}$$

существует равновесие в доминантных стратегиях (РДС), равное

$$y_i^* = R_i, s_i^* = \sqrt{\varphi^* y_i}, \text{ где } \varphi^* = \left(\frac{S}{\sqrt{\sum_{i=1}^n y_i}} \right)^2.$$

Доказательство. С учетом леммы 1, леммы 2 и доказанных свойства 1 и свойства 2, при которых функция $\eta(R_i, y_i, y_{-i}^0, s_i)$ по переменной y_i является монотонно убывающей на интервале $[R_i; +\infty[$ и монотонно возрастающей на интервале $]0; R_i]$, получаем утверждение теоремы. •

Отметим, что сумма

$$\sum_{i=1}^n \eta(R_i, y_i, y_{-i}^0, s_i) = P$$

может быть в общем случае как больше, так и меньше распределяемого ресурса S . Будем обозначать нормированным модифицированным ресурсом величину, рассчитываемую по формуле

$$\eta^*(R_i, y_i, y_{-i}^0, s_i) = \frac{O_i K_i x_i}{P}.$$

Из определения нормированного модифицированного ресурса следует, что

$$\sum_{i=1}^n \eta^*(R_i, y_i, y_{-i}^0, s_i) = S,$$

а также что РДС для $\eta^*(R_i, y_i, y_{-i}^0, s_i)$ совпадает с РДС для $\eta(R_i, y_i, y_{-i}^0, s_i)$.

Опишем процедуру использования функции нормированного модифицированного ресурса для процедуры планирования и распределения ресурса.

1. До начала процедуры планирования результата по каждому из КРП Центр устанавливает значения базовых параметров $\gamma > 0$; $\beta \geq \alpha + 2$; $\alpha > \gamma$; $\alpha \geq 1$ для данного КРП, а также объем ресурса S , который будет являться премией участников за достижение результатов данного КРП, и сообщает их агентам.

2. При планировании результата по КРП (в начале отчетного периода) каждый из агентов с номером $1 \leq i \leq n$ сообщает Центру свой план y_i , а также размер вознаграждения s_i , который

планирует получить в случае достижения запланированного результата.

3. Получив данные от агентов, Центр вычисляет коэффициенты напряженности K_i и значения распределения ресурса x_i для каждого агента и сообщает их агентам.

4. В конце отчетного периода каждый агент получает фактическое значение KPI R_i , которое сравнивается Центром с заявленным агентом планом y_i . В результате Центром вычисляется значение $\eta^*(R_i, y_i, y_i^0, s_i)$, которое выплачивается агенту.

В случае, когда все агенты заявили одинаково напряженные планы и выполнили их на 100% получаем, что $K_i = O_i = 1$ и распределение ресурса совпадает с распределением методом обратных приоритетов. Если агенты заявляют планы разной напряженности и выполняют их по-разному, то это вносит соответствующие корректировки в распределение ресурса S между агентами.

6. Пример распределения фонда мотивации между научными лабораториями

При подведении на Совете итогов работы научных лабораторий, как правило, в качестве основных показателей работы обсуждаются публикационная активность, число защит, а также число обучаемых студентов. Со стороны руководителей лабораторий приводятся аргументы, что заявляемые ими планируемые значения – это «нижняя оценка», что «невозможно планировать научную деятельность», а также ссылки на сокращаемый по сравнению с предыдущим периодом бюджет лаборатории. Со стороны членов Совета высказываются предложения о взятии лабораторией обязательств по увеличению данных показателей хотя бы до уровня, достигнутого в отчетном периоде в обмен на увеличение финансирования.

Таким образом, имеет место задача теории активных систем по мотивации агентов (научных лабораторий) центром (Советом) на достижение максимально возможных для каждого агента результатов по трем показателям (публикации, защита

аспирантов, работа со студентами), а также задача распределения центром ограниченного ресурса (бюджета). Решение данной задачи может быть осуществлено методом, описанным в разделе 5 настоящей статьи

Для решения данной задачи вначале необходимо установить показатели КРІ для научных лабораторий, а также разделить общий бюджет центра на бюджеты по каждому КРІ. Все показатели для лаборатории рассчитываются за один год. Такими КРІ могут быть:

- 1) суммарный индекс публикаций сотрудников лаборатории;
- 2) число защитившихся аспирантов у руководителей – сотрудников лаборатории;
- 3) число студентов, работающих в лаборатории.

Нормативы по числу защитившихся аспирантов и числу работающих студентов на одного научного работника лаборатории должны зависеть от:

- ученой степени (кандидат, доктор)
- ученого звания (доцент, профессор, академик)

Норматив по публикационной активности на одного научного сотрудника лаборатории дополнительно должен зависеть от предметной области, в которой работает сотрудник (экономика, математика, физика, техника и пр.).

Итоговое нормативное значение КРІ Лаборатории для каждого из трех показателей зависит от числа и состава сотрудников лаборатории и определяется как сумма нормативных значений для всех участников лаборатории. Нормативные значения на одного научного сотрудника по всем трем КРІ устанавливаются центром и доводятся до всех лабораторий перед формированием планов лабораторий на следующий отчетный период.

Для определения нормативного значения по публикационной активности на одного научного сотрудника центр анализирует статистику всех публикаций отечественных авторов в разрезе предметных областей, а также ученой степени и звания автора публикации. Выборка определяется по публикациям в журналах Q1, Q2, Q3, Q4 за последние несколько лет. В качестве итогового нормативного значения центр может устанавливать, как полученное среднее значение, так и

полученное значение, умноженное на поправочный коэффициент.

Для определения нормативного значения по числу защитившихся аспирантов на одного научного сотрудника центр анализирует статистику защит за последние несколько лет в разрезе ученой степени и ученого звания руководителя защитившегося аспиранта. Нормативное значение устанавливается центром как полученное среднее, умноженное на поправочный коэффициент.

Для определения нормативного значения по числу работающих студентов центр анализирует статистику лабораторий по данному показателю за последние несколько лет. Анализ определяет значение коэффициентов участия в привлечении работающих студентов для разных категорий научных сотрудников в линейном уравнении, показывающее влияние числа сотрудников каждой категории на число студентов, работающих в лаборатории. Значения коэффициентов выбираются путем нахождения решения задачи оптимизации, которое минимизирует среднеквадратичное отклонение числа работающих студентов, полученных по уравнению для каждой лаборатории, от фактических значений числа работающих студентов в этой лаборатории. Нормативное значение коэффициента участия для каждой категории научных сотрудников устанавливается центром как полученный в результате оптимизации результат, умноженный на поправочный коэффициент.

Дальнейшие действия центра и агентов осуществляются в соответствии с алгоритмом, определяемом принципом планирования методом встречных планов. Последовательность таких действий состоит из следующих четырех этапов.

Этап 1. Перед началом нового отчетного периода, определив нормативные значения величин по нормативному значению КРІ для каждой лаборатории (исходя из численности, контингента научных сотрудников и тематики лаборатории), центр (Совет) доводит информацию о нормативных значениях до всех агентов (лабораторий). Также центр доводит до лабораторий размеры бюджета, выделяемые центром на каждое из трех направлений (публикации, аспирантов, студентов). Кроме этого центр

сообщает лабораториям общие для всех лабораторий параметры функций мотивации $\gamma > 0$; $\beta \geq \alpha + 2$; $\alpha > \gamma$; $\alpha \geq 1$ по методу встречных планов.

Этап 2. Лаборатории, получив информацию от Совета, сообщают свои плановые цифры u_i по каждому из трех КРІ, а также размер финансирования от центра по каждому из трех направлений s_i , который лаборатория хотела бы получить, если заявленные значения КРІ будут достигнуты.

Этап 3. Совет обрабатывает полученные от лабораторий данные, производит расчет параметров коэффициента напряженности планов K_i и значения распределения ресурса x_i для каждого КРІ каждого агента (лаборатории) и сообщает их агентам.

Этап 4. В конце отчетного периода Совет подводит итоги работы по каждой лаборатории. Минимальное вознаграждение (постоянная часть) выплачивается лабораториям ежемесячно в течение года. Выплата основной доли вознаграждения (переменная часть) лабораториям путем распределения бюджета по каждому из направлений за минусом уже выплаченной суммы, производится в соответствии с комбинированным методом обратных приоритетов и встречных планов.

Пример распределения ресурса для КРІ «Число работающих студентов», а также графическое представление функций мотивации агентов, для набора из пяти условных лабораторий, приведены в Приложении. Содержательная постановка примера заключена в задаче распределения 100% ресурса между лабораториями, исходя из числа работающих студентов в лаборатории и числа сотрудников, имеющих звание профессора и доцента. Для решения этой задачи Центр устанавливает норматив по числу работающих студентов для одного профессора и доцента и доводит его значения до лабораторий.

В рассмотренном примере методом обратных приоритетов Лаборатория 1 могла максимально получить 23,77% от общего объема ресурса. Пример показывает, как при заявке Лаборатории 1 равной 20% от общего объема ресурса фактический объем получаемого Лабораторией 1 ресурса составляет 30,19%, что объясняется более напряженным планом

Лаборатории 1 (140% по отношению к нормативу, установленному для Лаборатории 1 центром).

7. Заключение

Предложенная система работы с КРІ и общим ресурсом (бюджетом) научных лабораторий позволяет достичь следующих результатов:

1. Мотивирует лаборатории на сообщение в качестве плановых значений КРІ тех значений, которые реально могут быть достигнуты, по мнению самих лабораторий.

2. Мотивирует лаборатории на достижение и перевыполнение заявленных плановых значений КРІ.

3. Дает центру гибкий инструмент распределения приоритетов между тремя направлениями работы лабораторий и устанавливает индикативы для достижения по каждому направлению для каждой лаборатории.

4. Делает процесс планирования и распределения средств бюджета между лабораториями прозрачным, а принципы такого распределения - заранее известными всем участникам.

При построении системы использовались два предположения:

– у центра отсутствует собственный (не опосредованный лабораториями) результат деятельности;

– затраты лабораторий не зависят от того, какой прогноз они заявят и какой будет получен фактический результат.

В основе предложенной системы лежит доказанная автором Теорема, задающая ограничения на параметры целевых функции агента, позволяющие Центру получить решение, являющееся равновесием в доминантной стратегии (РДС).

Литература

1. БОНДАРИК В.Н., КОРГИН Н.А. *Механизмы распределения ресурсов на основе неманипулируемых симметричных анонимных процедур голосования с делегированием* // Проблемы управления. – 2012. – Вып. 5. – С. 26–32.
2. БУРКОВ В.Н., ГОРГИДЗЕ И.И., НОВИКОВ Д.А., ЮСУПОВ Б.С. *Модели и механизмы распределения затрат и доходов в рыночной экономике*. – М.: ИПУ РАН, 1997. – 61 с.
3. БУРКОВ В.Н., ДАНЕВ Б., ЕНАЛЕЕВ А.К. и др., *Большие системы: моделирование организационных механизмов*. – М.: Наука, 1989. – 248 с.
4. БУРКОВ В.Н., КОРГИН Н.А., НОВИКОВ Д.А. *Введение в теорию управления организационными системами* / Под ред. чл.-корр. РАН Д.А. Новикова. – М.: Либроком, 2009. – 264 с.
5. БУРКОВА И.В., КРЮКОВ С.В., ЗУБАРЕВ В.В., ШУМАРИН В.В. *Механизм распределения корпоративных заказов* // Вестник Воронежского государственного технического университета, 2010.
6. ДАВЫДОВ В.А. *Методы встречных планов для определения целевого значения KPI в двухуровневой организационной системе* // Управление большими системами. – 2018. – Вып. 73. – С. 27–54.
7. ИНТРИЛЛИГАТОР М. *Математические методы оптимизации и экономическая теория*. – М.: Прогресс, 1975. – 606 с.
8. КИМ Е.Р., ШУКАЕВ Д.Н., ЛАМАШЕВА Ж.Б. *Моделирование распределения и размещения ресурсов оборудования в производственных системах* // Фундаментальные исследования. – 2016. – №10-1. – С. 48–52
9. КОРГИН Н.А. *Эквивалентность и неманипулируемость неанонимных приоритетных механизмов распределения ресурсов* // МТИП. – 2009. – №1:3. – С. 46–70.

10. ЛЕВЧЕНКО С.П., ЗИМИН В.В., КУЛАКОВ С.М.
Применение принципа обратных приоритетов для распределения ресурсов при управлении жизненным циклом ИТ-сервисов. – Новокузнецк: ФГБОУ ВПО «Сибирский государственный индустриальный университет», 2007.
11. МУЛИН Э. *Кооперативное принятие решений: Аксиомы и модели.* – М.: Мир, 1991. – 464 с.
12. НОВИКОВ Д.А. *Теория управления организационными системами. 2-е изд.* – М.: Физматлит, 2007. – 584 с.
13. *Механизмы управления: Учебное пособие* / Под ред. Д.А. Новикова. – М.: УРСС, 2011.
14. BARBERA S., JACKSON M., NEME A. *Strategy-Proof Allotment Rules* // *Games and Economic Behavior.* – 1997. – Vol. 18, No.1. – P. 1–21.
15. BOSSERT W., WEYMARK J.A. *Social choice (new developments)* // In: *The New Palgrave Dictionary of Economics, Second Edition* / Eds.: S.N. Durlauf, L.E. Blume. – Palgrave Macmillan, 2008.
16. MAS-COLLEL A., WHINSTON M.D., GREEN J.R. *Microeconomic theory.* – N.Y.: Oxford Univ. Press, 1995. – 981 p.
17. SPRUMONT Y. *The division problem with single-peaked preferences: A characterization of the uniform rule* // *Econometrica.* – 1991. – No. 59. – P. 509–519.

COMBINING THE COUNTER-PLAN METHOD AND THE INVERTED PRIORITIES METHOD FOR DISTRIBUTION OF THE MOTIVATION FUND FOR SCIENTIFIC LABORATORIES

Viacheslav Davydov, Moscow Institute of Electronics and Mathematics HSE, Moscow (v.davydov@hse.ru).

Abstract: An incentive model for an organizational system (OS) agent is considered, using a combination of the method of inverted priorities and a modified counter-plan method. The resulting model motivates the agent to report the planned KPI value, which coincides with the agent's forecast and at the same time stimulates the agent to submit an adequate application for the resource. The model allows the OS center to allocate a dedicated resource to run each KPI between agents so that in addition to reporting adequate plans, agents are committed to their implementation. Restrictions are imposed on the model parameters, in which the choice of an agent equal to its own forecast is a dominant strategy (DS), provided that the agent submits the optimal (according to the method of inverted priorities) requests for the resource. It is proved that such a choice for all OS agents is an equilibrium in DS. The procedure for determining the planned KPI values for OS agents and resource allocation between agents is described. The model parameters determined by the OS center, taking into account the imposed restrictions, allow selecting the required priorities for agent motivation and distributing a motivational fund in accordance with the priorities identified by the center. An example of the application of the model for stimulating scientific laboratories for the preparation of students is given.

Keywords: method of counter plans, method of inverted priorities, incentive, KPI.

УДК 338.984

ББК 65.291

*Статья представлена к публикации
членом редакционной коллегии Н.А. Коргиным.*

Поступила в редакцию 18.12.2017.

Опубликована 30.09.2018.

Приложение.

Таблица 1.

Альфа (степень коэффициента напряженности прогноза плана)						1,00
Гамма (коэффициент премии за перевыполнение плана)						0,20
Бета (коэффициент штрафа за невыполнение плана)						3,00
Плановое задание центра по числу работающих студентов на 1 профессора						2,00
Плановое задание центра по числу работающих студентов на 1 доцента						1,00
	Лаб. 1	Лаб. 2	Лаб. 3	Лаб. 4	Лаб. 5	ИТОГО
Численность профессоров	2,00	5,00	1,00	10,00	4,00	22,00
Численность доцентов	11,00	5,00	3,00	20,00	2,00	41,00
План Центра для Лаборатории по числу работающих студентов	15,00	15,00	5,00	40,00	10,00	85,00
Доля вознаграждения, заявляемая Лабораторией	20,00%	18,70%	12,71%	28,41%	16,40%	96,23%
Прогноз Лабораторией числа работающих студентов	21	13	6	30	10	80
Плана Лаборатории / План Центра	140%	87%	120%	75%	100%	94%
Оптимальное распределение 100% ресурса методом обратных приоритетов	23,77%	18,70%	12,71%	28,41%	16,40%	100,00%
Фактическое распределение ресурса методом обратных приоритетов	20,00%	18,70%	12,71%	28,41%	16,40%	96,23%
Распределение 100% ресурса с учетом К нап. и нормирования	30,19%	16,07%	16,37%	20,58%	16,78%	100,00%

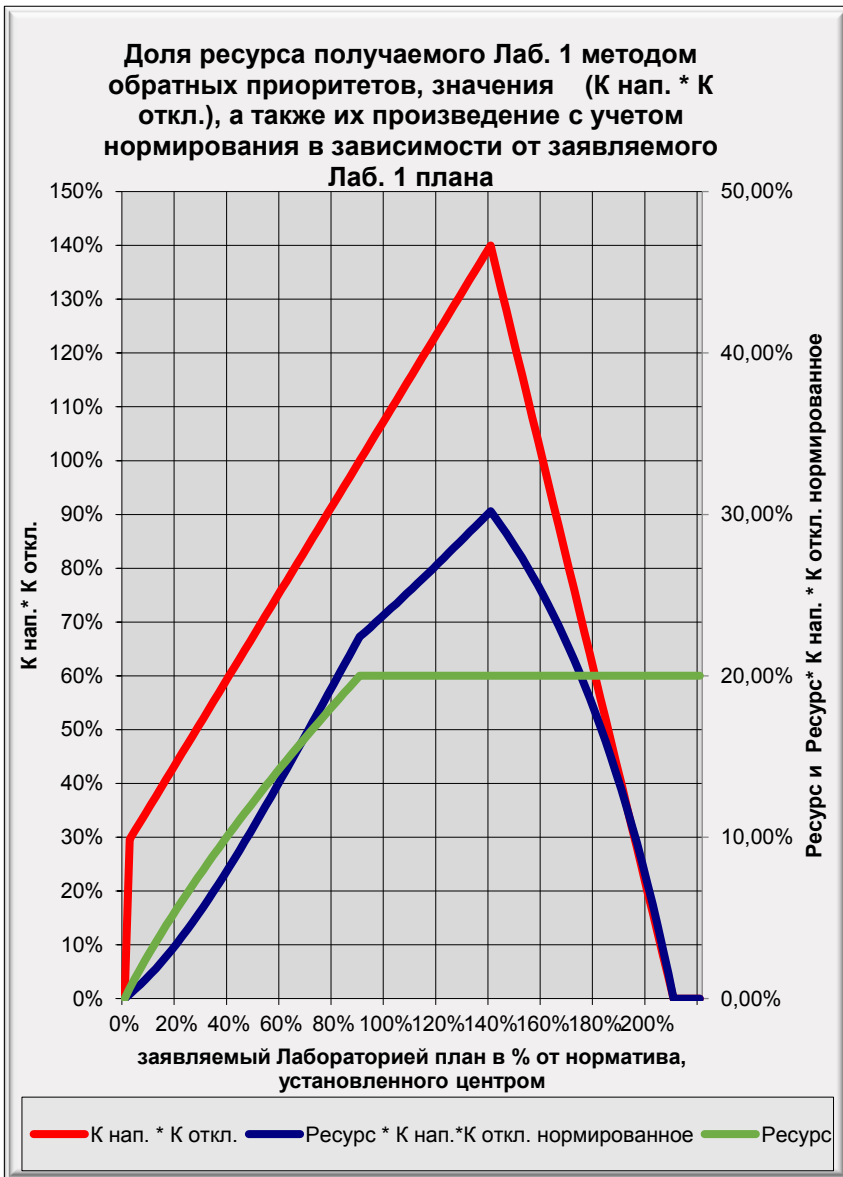


Рис. 1.

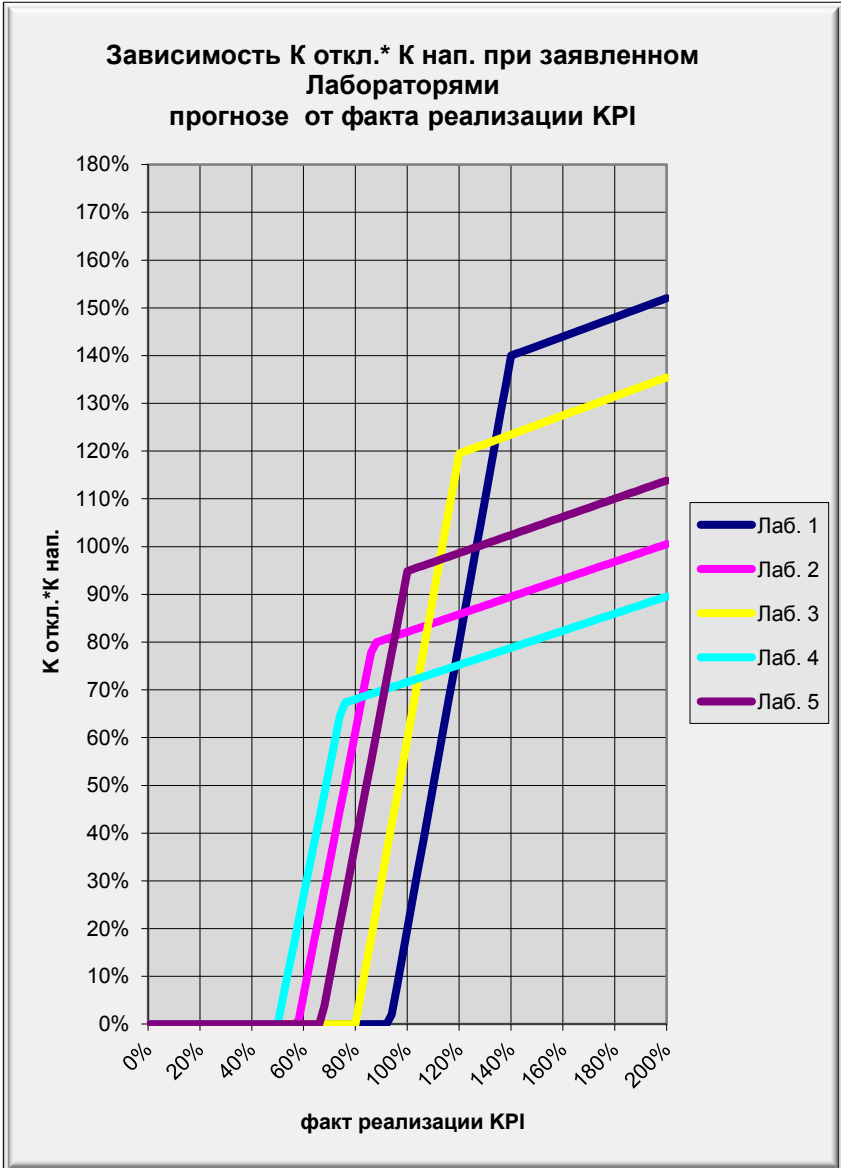


Рис. 2.

НАХОЖДЕНИЕ ИСПАРЕНИЯ КАК РЕШЕНИЕ ЗАДАЧИ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ

Засухин С. В.¹

(Вычислительный центр им. А.А. Дородницына РАН,
Московский физико-технический институт, Москва)

Задача нахождения испарения формулируется как задача оптимального управления, в которой управляемый процесс вертикального передвижения влаги в почве описывается одномерным нелинейным уравнением с частными производными второго порядка параболического типа. Управлением является суточное испарение, а целевой функционал есть среднеквадратическое отклонение вычисленных значений влажности почвы от некоторых предписанных значений. В результате проведения конечно-разностной аппроксимации задача редуцируется к задаче нелинейного программирования. Искать решение полученной задачи предлагается с помощью метода наискорейшего спуска. При этом градиент целевой функции вычисляется с применением точных формул быстрого автоматического дифференцирования. Предполагается, что предписанные значения совпадают с некоторым решением прямой задачи в области сравнения вычисленных значений влажности с предписанными значениями. Исследовался вопрос о том, как вид этого множества и, значит, вид данных, по которым определяется искомое испарение, влияет на точность получаемого решения и ход оптимизационного процесса. Было рассмотрено несколько вариантов такого множества. Каждое множество представляло собой совокупность узлов сетки равномерной по времени и по пространству, покрывающей всю рассматриваемую область. Анализ результатов численного решения соответствующих задач позволил из всех рассматриваемых вариантов выбрать наилучший в смысле близости полученного оптимального и истинного управлений и в смысле количества измерений влажности почвы, проводимых с целью получения данных для задачи определения испарения.

Ключевые слова: целевая функция, метод наискорейшего спуска, нелинейное параболическое уравнение, быстрое автоматическое дифференцирование.

1. Введение

Испарение с поверхности почвы входит в состав многих гидрологических и метеорологических моделей. Из всех составляющих гидрологического цикла испарение, пожалуй, является

¹ Сергей Владимирович Засухин, аспирант (s.zasukhin@yandex.ru).

наиболее трудно определяемой вследствие сложного взаимодействия между компонентами системы «почва–растительность–атмосфера». Непосредственное измерение его в естественных условиях представляет собой трудную задачу. Поэтому для определения испарения часто прибегают к расчетным методам.

Точных и универсальных методов расчета испарения в настоящее время не существует. Обычно удается оценить испарение за большие промежутки времени (недели, месяцы) и на больших территориях. Существующие расчетные методы испарения могут быть условно выделены в следующие группы.

Эмпирические методы используют формулы, получаемые на основании связи между измеренным испарением и метеорологическими факторами, которая обычно устанавливается с помощью регрессионного анализа [4, 6]. Эти формулы содержат, как правило, небольшое количество входных переменных и просты в применении. Недостаток их состоит в том, что они применимы лишь в условиях, близких к тем, в которых они были получены.

Методы водного баланса основаны на водобалансовых соотношениях поверхности земли с учетом характеристик водообмена между почвой и растительностью. С помощью этих методов вычисляется осредненная величина суммарного испарения. Среди работ, посвященных методам водного баланса, следует указать на [2, 5, 7, 15]. К недостаткам этих методов относятся их трудоемкость и невозможность определения суммарного испарения за краткосрочный период времени.

Методы теплового баланса используют уравнения теплового баланса поверхности земли с учетом тепло- и водообмена в приземном слое воздуха [5, 8, 12, 13, 14]. За рубежом широко распространен метод Пенмана–Монтейта [12, 13], в котором используются уравнения энергетического баланса на поверхности суши и уравнения переноса водяного пара и тепла между поверхностью суши и атмосферой. Модифицированная модель Пенмана–Монтейта по результатам полевой проверки в США оказалась лучшей по сравнению с 20 другими методами, основанными на регрессионном анализе.

С развитием методов и систем дистанционного зондирования Земли получают распространение **методы расчета испарения, учитывающие спутниковую информацию**. Вопросам разработки таких методов и соответствующей обработки спутниковых данных посвящены, например, работы [10, 11].

В представляемой работе продолжено исследование вопроса о нахождении испарения с поверхности почвы с применением модели вертикального переноса влаги в почве, начатое в [3]. Задача определения испарения формулируется как задача оптимального управления, в которой управляемым процессом являются моделируемые значения влажности почвы. В отличие от работы [3], где в качестве управления выступает интенсивность испарения, здесь управлением является суммарное суточное испарение. Это изменение вызвано тем, что в практической жизни гидрологи имеют дело с суточными испарением и осадками, т.е., как правило, они наблюдают и оперируют суммарными суточными значениями испарения и осадков. Целевая функция есть среднеквадратическое отклонение моделируемых значений влажности почвы от некоторых предписанных значений. Сравнение моделируемых и предписанных значений влажности происходит в некоторой области, выбор которой определяет поведение целевой функции. Ранее в [3] был исследован вопрос о том, в каких точках рассматриваемой области влажность почвы наиболее чувствительна к изменению испарения. Целью настоящей работы является выяснение вопроса о том, как вид множества, на котором происходит сравнение моделируемых и предписанных значений влажности, и, значит, вид начальных данных о влажности почвы, влияет на точность решения и на ход процесса численной оптимизации. Одной из целей такого изучения является понимание, какой вариант начальных данных обеспечит приемлемую точность решения, и при этом потребует сравнительно небольших усилий при проведении измерительных экспериментов для получения этих начальных данных. Причем эти эксперименты соответствовали бы имеющимся техническим возможностям. В качестве множеств, на которых происходит сравнение моделируемых значений влажности почвы с предписанными значениями, рассматривались множества, со-

стоящие из узлов сетки, покрывающей всю рассматриваемую область. Шаги сетки по пространству и по времени менялись от одного множества к другому. То есть эти множества представляли всю рассматриваемую область с разной степенью подробности. Этот вопрос детально освещается в разделе 5.

2. Постановка задачи

Предположим, что почва представляет собой изотермическую недеформируемую однородную пористую среду. При соблюдении этих предположений вертикальное передвижение влаги в почве хорошо описывается одномерным нелинейным уравнением параболического типа, которое приводится ниже. Это уравнение выводится на основе закона Дарси, утверждающего, что поток влаги через единичную поверхность пропорционален градиенту потенциала всех действующих в почве сил, и уравнения неразрывности для несжимаемой жидкости, являющегося выражением закона сохранения массы. Коэффициент между величиной потока влаги в ненасыщенной почве и градиентом потенциала называется гидравлической проводимостью, а в условиях насыщения почвы влагой – коэффициентом фильтрации. Полный потенциал складывается из капиллярно-сорбционного или влажностного потенциала почвы и гравитационного потенциала. Физический смысл уравнения состоит в том, что движение влаги в почве складывается из диффузии почвенной влаги и движения под действием гравитационных сил.

Далее нас будет интересовать процесс движения влаги в ненасыщенной почве. Рассмотрим следующую начальнокреповую задачу:

$$\begin{aligned}
 \frac{\partial \theta(z, t)}{\partial t} &= \frac{\partial}{\partial z} \left(D(\theta(z, t)) \frac{\partial \theta(z, t)}{\partial z} - K(\theta(z, t)) \right), & (z, t) \in Q, \\
 \theta(z, 0) &= \varphi(z), & z \in (0, L), \\
 \theta(L, t) &= \psi(t), & t \in (0, T), \\
 - \left(D(\theta(z, t)) \frac{\partial \theta(z, t)}{\partial z} - K(\theta(z, t)) \right) \Big|_{z=0} &= R(t) - E(t), & t \in (0, T), \\
 \theta_{\min} \leq \theta(0, t) \leq \theta_{\max}, & & t \in (0, T),
 \end{aligned}
 \tag{1}$$

где z – пространственная переменная, причем ось z направлена сверху-вниз; t – время; $\theta(z, t)$ – искомая влажность в точке (z, t) , так называемая «объемная» влажность почвы, выражаемая в единицах объема воды в единичном объеме почвы (безразмерная величина); $Q = (0, L) \times (0, T)$; $\varphi(z)$ и $\psi(t)$ – заданные функции; $D(\theta)$ и $K(\theta)$ – коэффициент диффузии и гидравлическая проводимость – гидрофизические характеристики почвы; $\theta_{\min} = \theta_r + \varepsilon$, $\theta_{\max} = \theta_s - \varepsilon$, где θ_r и θ_s – остаточная влажность и влажность насыщения соответственно, зависящие от типа почвы, ε – некоторая постоянная, такая, что $0 < \varepsilon \ll \theta_r$; $R(t)$ и $E(t)$ – интенсивности осадков и испарения – линейные потоки влаги; $0 \leq E(t) \leq M$, $t \in (0, T)$, M – некоторая константа, $M > 0$.

Функция $\varphi(z)$ задает влажность почвы во всех точках рассматриваемого интервала $(0, L)$ в начальный момент времени $t = 0$. Функция $\psi(t)$, $t \in (0, T)$, определяет нижнее граничное условие и задает значения влажности почвы на глубине $z = L$. Приведенное в (1) верхнее (при $z = 0$) краевое условие означает, что поверхностный сток отсутствует, и поток влаги в почве при $z = 0$ равен суммарному входящему потоку влаги на границе атмосфера-почва.

Коэффициент диффузии и гидравлическая проводимость вычисляются по широко применяемым формулам ван Генухтена [16]:

$$\begin{aligned}
 K(\theta) &= K_0 S^{0.5} \left[1 - \left(1 - S^{1/m} \right)^m \right]^2, \\
 D(\theta) &= K_0 \frac{1 - m}{\alpha m (\theta_s - \theta_r)} S^{0.5 - 1/m} \times \left[\left(1 - S^{1/m} \right)^{-m} + \left(1 - S^{1/m} \right)^m - 2 \right],
 \end{aligned}
 \tag{2}$$

где $S = (\theta - \theta_r)/(\theta_s - \theta_r)$; $K_0, \alpha, m, \theta_r, \theta_s$ – некоторые параметры.

Разделим интервал $(0, T)$ на q равных подынтервалов, $q > 1$, с концевыми точками $t_n, 0 \leq n \leq q$. Введем множество $U = \{u : u \in R^q, 0 \leq u^n \leq M, n = 1, \dots, q\}$. Каждому $u \in U$ поставим в соответствие $E(u, t)$ таким образом, что $E(u, t) = u^n, t \in (t_{n-1}, t_n), n = 1, \dots, q$. Будем рассматривать U в качестве множества допустимых управлений, а под функциями $E(t)$ из (1) будем понимать функции $E(u, t), u \in U$.

Назовем описанную задачу (1)–(2) прямой задачей.

Зададимся целью найти испарение $E(u, t), u \in U$, из описанного выше класса кусочно-постоянных функций по некоторой наблюдаемой в области $Q_0 \subseteq Q$ функции влажности, назовем ее $\Theta(z, t)$, найти в том смысле, чтобы решение $\theta(z, t)$ прямой задачи (1)–(2) при определяемом испарении было бы как можно ближе к наблюдаемой функции влажности $\Theta(z, t)$ в области Q_0 . В качестве оценки такой близости выберем интегральную норму отклонения решения $\theta(z, t)$ прямой задачи (1)–(2) от наблюдаемой (или предписанной) функции влажности $\Theta(z, t)$ в области Q_0 .

Сформулируем задачу нахождения испарения $E(u, t), u \in U$, как следующую задачу оптимального управления.

Задача 1. Пусть на некотором множестве $Q_0 \subseteq Q$ задана функция $\Theta(z, t)$. Найти управление $u^{opt} \in U$, при котором соответствующее решение $\theta^{opt}(t)$ прямой задачи (1)–(2) при $E(u^{opt}, t)$ приводит к минимуму функционала:

$$J = \frac{1}{2} \int_{Q_0} (\theta - \Theta)^2 dz dt .$$

3. Дискретный аналог задачи оптимального управления

Разобьем интервалы $(0, T)$ и $(0, L)$ на N и I равных подынтервалов с концевыми точками $t^n = \tau n, 0 \leq n \leq N$ и $z_i = hi, 0 \leq i \leq I$ соответственно, где $\tau = T/N$ и $h = L/I$. Пусть $N/q = \gamma$ – целое число. Аппроксимируем прямую задачу (1)–(2) с помощью следующей конечно-разностной схемы:

$$\frac{\theta_i^n - \theta_i^{n-1}}{\tau} = \frac{1}{h} \left(D_{i+1/2}^n \frac{\theta_{i+1}^n - \theta_i^n}{h} - K_{i+1/2}^n - D_{i-1/2}^n \frac{\theta_i^n - \theta_{i-1}^n}{h} + K_{i-1/2}^n \right),$$

$$1 \leq i < I, \quad 0 < n \leq N,$$

$$\theta_i^0 = \varphi_i, \quad 0 \leq i \leq I, \quad \theta_i^n = \psi^n, \quad 1 \leq n \leq N,$$

где

$$\theta_i^n = \theta(z_i, t^n), \quad D_{i+1/2}^n = D(\theta((i+1/2)h, t^n)), \quad K_{i-1/2}^n = K(\theta((i-1/2)h, t^n)),$$

а φ_i и ψ^n – значения функций $\varphi(z)$ и $\psi(t)$ в точках z_i и t^n соответственно. Конечно-разностная аппроксимация левого краевого условия выглядит следующим образом:

$$\frac{\theta_0^{n+1} - \theta_0^n}{\tau} = \frac{2}{h} \left(D_{1/2}^{n+1} \frac{\theta_1^{n+1} - \theta_0^{n+1}}{h} - K_{1/2}^{n+1} + R^{n+1} - E^{n+1} \right), \quad 0 \leq n < N,$$

где R^{n+1} , E^{n+1} – значения функций $R(t)$ и $E(u, t)$ в точке $t^{n+1} = \tau(n+1)$. В результате приходим к следующему конечно-разностному аналогу прямой задачи (1)–(2):

$$\Phi_0^n = - \left(\frac{1}{\tau} + \frac{2}{h^2} D_{1/2}^n \right) \theta_0^n + \frac{2}{h^2} D_{1/2}^n \theta_1^n + \frac{1}{\tau} \theta_0^{n-1} + \frac{2}{h} (-K_{1/2}^n + R^n - E^n) = 0,$$

$$\theta_{\min} \leq \theta_0^n \leq \theta_{\max}, \quad 1 \leq n \leq N,$$

$$(3) \quad \Phi_i^n = \frac{1}{h^2} D_{i-1/2}^n \theta_{i-1}^n - \left[\frac{1}{\tau} + \frac{1}{h^2} (D_{i+1/2}^n + D_{i-1/2}^n) \right] \theta_i^n + \frac{1}{h^2} D_{i+1/2}^n \theta_{i+1}^n + \left[\frac{\theta_i^{n-1}}{\tau} + \frac{1}{h} (K_{i-1/2}^n - K_{i+1/2}^n) \right] = 0, \quad 1 \leq i < I, \quad 1 \leq n \leq N,$$

$$\Phi_i^n = \psi^n - \theta_i^n = 0, \quad 1 \leq n \leq N,$$

$$\theta_i^0 = \varphi_i, \quad 0 \leq i \leq I,$$

где $E = E(u) = [E^1, \dots, E^N]$, $E^n = E(u, t^n)$, $n = 1, \dots, N$, $u \in U$. При этом коэффициент диффузии D и гидравлическую проводимость K в промежуточных точках будем вычислять по следующим формулам:

$$(4) \quad D_{i+1/2}^n = 2 \frac{D_i^n D_{i+1}^n}{D_i^n + D_{i+1}^n}, \quad K_{i+1/2}^n = 2 \frac{K_i^n K_{i+1}^n}{K_i^n + K_{i+1}^n}, \quad 1 \leq n \leq N, \quad 0 \leq i < I.$$

Введем в рассмотрение множества $A_0 = \{0, 1, \dots, I\}$ и $B_0 = \{1, \dots, N\}$. Пусть $Q_0 = \{(z, t) : z = ih, t = lt, (i, l) \in C\}$, где C –

некоторое множество, такое что $C \subseteq A_0 \times B_0$. Определим целевую функцию в виде:

$$(5) \quad W(\theta(u), u) = \frac{1}{2} \sum_{(j, n) \in C} (\theta_j^n - \Theta_j^n)^2 h \tau,$$

где $\theta(u)$ – решение системы (3)–(4) при $E(u)$, $u \in U$, Θ_j^n – значение функции $\Theta(z, t)$ в точке (z_j, t^n) .

Сформулируем дискретную задачу оптимального управления.

Задача 2. Найти оптимальное управление $u^{opt} \in U$ и соответствующее этому управлению решение θ^{opt} системы (3)–(4), при которых функционал $W(\theta(u), u)$ достигал бы минимального значения.

4. Решение дискретной задачи оптимального управления

В результате дискретизации задача оптимального управления, как видим, сводится к задаче нелинейного программирования. Решать численно эту конечномерную задачу оптимизации предлагается методом наискорейшего спуска. Градиент целевой функции (5) при этом вычисляется с применением метода быстрого автоматического дифференцирования (БАД) [1, 9]. Согласно БАД, градиент функции $W(\theta(u), u)$ (5) вычисляется по формуле

$$(6) \quad dW(\theta(u), u)/du = W_u(\theta(u), u) + \Phi_u^T(\theta(u), u)p.$$

Входящий в эту формулу вектор $p \in R^n$ — множитель Лагранжа, который определяется в результате решения следующей линейной относительно p системы уравнений:

$$(7) \quad W_\theta(\theta(u), u) + \Phi_\theta^T(\theta(u), u)p = 0_q,$$

где

$$\Phi^T = [\Phi_0^1, \Phi_1^1, \dots, \Phi_l^1, \dots, \Phi_0^N, \Phi_1^N, \dots, \Phi_l^N], \quad q = (I+1)N,$$

$$\theta^T = [\theta_0^1, \theta_1^1, \dots, \theta_l^1, \dots, \theta_0^N, \theta_1^N, \dots, \theta_l^N], \quad u \in U.$$

Система уравнений (7) является сопряженной к системе (3). В [3] было показано, что процедура вычисления градиента по формулам (6)–(7) является устойчивой.

5. Численные результаты

Описанная конечномерная задача решалась при следующих значениях входных параметров:

$$L = 90 \text{ см}, \quad T = 61 \text{ сут}, \quad \varepsilon = 1 \cdot 10^{-10}, \quad \varphi(z) = 0.3, \quad z \in (0, L),$$

$$K_0 = 103.68 \text{ см/сут}, \quad \alpha = 0.075 \text{ см}^{-1}, \quad m = 0.471,$$

$$\theta_r = 0.065 \text{ см}^3/\text{см}^3, \quad \theta_s = 0.51 \text{ см}^3/\text{см}^3, \quad M = 2 \text{ см/сут}.$$

Так как на практике наблюдаются, как правило, суммарные за сутки значения осадков и испарения, было принято допущение о том, что интенсивности испарения и осадков в течение суток сохраняют постоянные значения. Управление представляет собой вектор из 61 компоненты, каждая из которых есть суточный слой испарения, и численно равна интенсивности испарения в течение соответствующих суток. Шаг по времени составил 1/100 сут, а шаг по пространству – 1 см. Таким образом, $N = 6100$, $q = 61$ и $I = 90$ в нашем случае.

График зависимости влажности почвы при $z = 90$ от времени (функции $\psi(t)$) представлен на рис. 1.

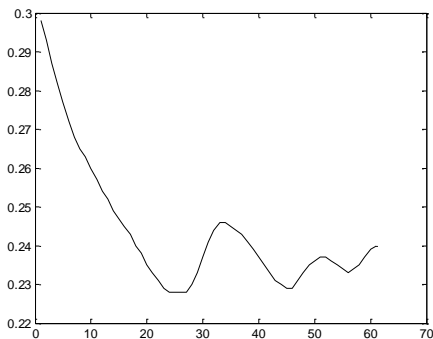


Рис. 1. Зависимость влажности почвы при $z = 90$ от времени

График зависимости слоя выпавших за сутки осадков от времени показан на рис. 2.

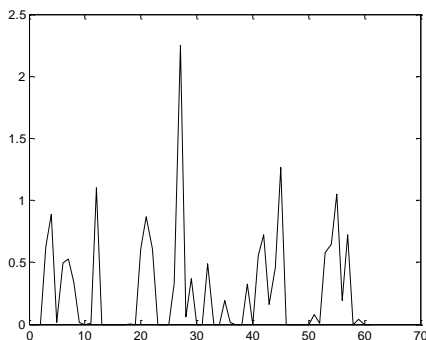


Рис. 2. Зависимость суточного слоя выпавших осадков в сантиметрах от времени

Численные расчеты проводились в несколько этапов.

5.1. ПЕРВЫЙ ЭТАП

На этом этапе было выбрано некое управление, назовем его $u^{true} = [u_1^{true}, \dots, u_{61}^{true}]^T$, $u^{true} \in U$. Далее решалась прямая конечно-разностная задача (3) с выбранным $E(u^{true})$. Понятно, что система (3) допускает расщепление на N подсистем, каждая из которых относится к n -му, $n = 1, \dots, N$, временному слою и содержит в качестве неизвестных значения влажности только на этом временном слое. Двигаясь в направлении от первого временного слоя к N -му временному слою, мы последовательно для каждого временного слоя решаем соответствующую систему уравнений отдельно от других. Решение каждой такой системы находится в результате выполнения следующего итерационного процесса. На каждой итерации решается соответствующая система, в которой значения коэффициента диффузии и гидравлической проводимости вычисляются по формулам (2), (4) с использованием значений влажности, полученных на предыдущей итерации. При известных коэффициентах система является линейной, при этом основная матрица системы – трехдиагональная. Система

решается методом прогонки. На первой итерации значения коэффициента диффузии и гидравлической проводимости вычисляются с применением значений влажности на предыдущем временном слое. Этот итерационный процесс продолжался до тех пор, пока евклидово расстояние между решениями, полученным на соседних итерациях, не станет менее 10^{-4} . При выбранных параметрах задачи для нахождения решения требовалось, как правило, не более пяти итераций. Полученное решение задачи (3) было названо «экспериментальными данными» и обозначено $\Theta(z, t)$, $(z, t) \in Q$.

5.2. ВТОРОЙ ЭТАП

На втором этапе решалась задача 2 с различными целевыми функциями. Целевые функции вычислялись по формуле (5), в которой множество C имело вид:

$$(8) \quad \begin{aligned} C &= A \times B, \quad A = \{i : i = i_0 k, k = 0, 1, \dots, [I/i_0]\}, \\ B &= \{n : n = n_0 l, l = 1, \dots, [N/n_0]\}. \end{aligned}$$

Рассматривались все сочетания n_0 и i_0 , когда $i_0 = 1, 5, 10$, $n_0 = 1, 5, 10, 20, 50, 100$. Численная оптимизация проводилась методом наискорейшего спуска, градиент целевой функции вычислялся по формулам БАД (6)–(7). Величина шага вдоль выбранного направления определялась в результате проведения одномерной оптимизации функции, интерполирующей целевую функцию вдоль указанного направления с помощью сплайнов, построенных по 40 точкам. Итерационный процесс продолжался до тех пор, пока чебышевская норма градиента целевой функции не становилась меньше 10^{-14} . В качестве начального управления было выбрано u^{init} , $u_n^{init} = 0,3$, $n = 1, \dots, 61$. Задача численной оптимизации на всем промежутке времени допускает декомпозицию на совокупность отдельных задач оптимизации в соответствии с выбранным разбиением всего временного промежутка. С учетом вида управления рассматривалось множество из 61 задачи, каждая из которых соответствует временному интервалу в одни сутки.

Результаты численных расчетов приведены в таблице 1.

Как показали численные расчеты, для каждой задачи требовалось приблизительно от 250 до 330 итераций для нахождения решения. Как правило, это число было равно 300 или очень незначительно отличалось от 300. В задаче, соответствующей 13-м суткам, количество итераций, приводящих к решению, было равно 23-26. Это объясняется тем, что начальное приближение $u_{13}^{init} = 0,3$ и истинное управление $u_{13}^{true} = 0,2777$ оказались близки друг другу. В некоторых задачах итерационный процесс не приводил к решению и за 500 итераций. Эта ситуация возникла при $n_0 = 50, 100$. Тогда приходилось искать приемлемое начальное приближение. Количество таких осложненных дней (в случае их возникновения) изменялось от 2 до 15. В частности, таких дней оказалось 15 в случае $n_0 = 50, i_0 = 1$, а именно: это были сутки с порядковыми номерами 11, 14, 15, 16, 17, 18, 19, 25, 37, 38, 40, 49, 50, 52 и 61, и 2 дня в случае $n_0 = 50, i_0 = 10$ (сутки с порядковыми номерами 18, 19).

Таблица 1. Результаты второго этапа

n_0	i_0	Отклонение	Итерации	Осложненные дни
1	1	0	290-300	0
	5	$1,00 \cdot 10^{-12}$	290-300	0
	10	$1,00 \cdot 10^{-12}$	290-300	0
5	1	$1,00 \cdot 10^{-12}$	290-300	0
	5	$2,00 \cdot 10^{-12}$	290-301	0
	10	$2,00 \cdot 10^{-12}$	290-302	0
10	1	$1,00 \cdot 10^{-12}$	290-300	0
	5	$1,30 \cdot 10^{-11}$	290-302	0
	10	$1,00 \cdot 10^{-12}$	290-304	0
20	1	$3,00 \cdot 10^{-12}$	290-303	0
	5	$1,81 \cdot 10^{-11}$	300-306	0
	10	$2,23 \cdot 10^{-11}$	300-309	0
50	1	$1,84 \cdot 10^{-11}$	290-307	15
	5	$3,10 \cdot 10^{-11}$	300-309	2
	10	$7,00 \cdot 10^{-12}$	300-319	2
100	1	$4,00 \cdot 10^{-12}$	290-302	15
	5	$6,90 \cdot 10^{-11}$	300-314	14
	10	$1,80 \cdot 10^{-11}$	300-328	8

Проведенные расчеты показали, что во всех рассматриваемых случаях максимальное отклонение найденного оптимального управления от истинного управления, как правило, возрастало с увеличением n_0 . При этом во всех случаях оно оказалось незначительным, а именно:

$$\left| u_n^{true} - u_n^{opt} \right| \leq 6,09 \cdot 10^{-11}, \quad n = 1, \dots, 61. .$$

Графики начального, истинного и найденного в результате численных расчетов оптимального управлений для случаев $n_0 = 100$, $i_0 = 1$ и $n_0 = 10$, $i_0 = 10$ представлены на рис. 3 и рис. 4 соответственно. На этих рисунках начальное управление обозначено толстой сплошной линией, истинное управление — сплошной линией, а оптимальное управление отмечено окружностями. Видно, что графики истинного и оптимального управлений в обоих случаях практически совпадают.

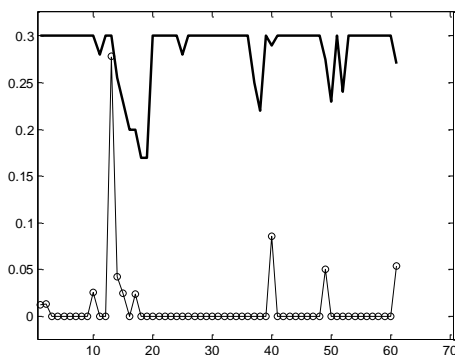


Рис. 3. Графики истинного, начального и оптимального управлений в случае $n_0 = 100$, $i_0 = 1$

Описанная схема проведения численных расчетов может применяться для оценивания того, как использование того или иного набора начальных данных, определяющего множество Q_0 , повлияет на близость получаемого оптимального управления к его истинному значению и на процесс численной оптимизации. И, значит, еще на стадии планирования экспериментов по изме-

рению влажности почвы эти численные результаты могут быть использованы для выбора оптимального плана экспериментов по измерению влажности почвы.

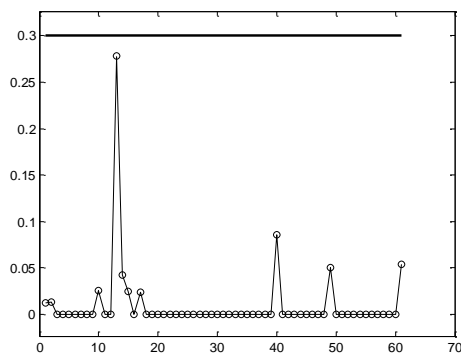


Рис. 4. Графики истинного, начального и оптимального управлений в случае $n_0 = 10$, $i_0 = 10$

Поясним сказанное на следующем примере. Предположим, нам известно, что искомое испарение близко к рассмотренному ранее испарению $E^{true}(t)$. Рассмотрим несколько вариантов проведения измерительных экспериментов, когда измерения влажности почвы производятся 100, 20, 10, 5, 2 и 1 раз в сутки через равные промежутки времени и с шагом по глубине в 1, 5 и 10 см. Полученные при проведении таких экспериментов данные могут быть использованы при решении задачи оптимального управления с целевой функцией, вычисляемой по формулам (5), (8), где $n_0 = 1, 5, 10, 20, 50, 100$ и $i_0 = 1, 5, 10$. Зададимся вопросом, какой вариант следует выбрать, руководствуясь естественным стремлением минимизировать затраты на проведение такого эксперимента, но при этом обеспечить достаточную близость к искомому испарению численного решения соответствующей дискретной задачи оптимального управления. Кроме того, желательно было бы также понять, будет ли процесс оптимизации при поиске численного решения задачи оптимального управления проходить без осложнений. Проведя численные расчеты по описанной выше схеме и проанализировав их ре-

зультаты, мы можем выбрать подходящий вариант проведения измерений.

Так, анализ полученных выше численных результатов позволяет сказать, что таким вариантом является измерение влажности почвы через равные интервалы времени 5 раз в сутки с шагом по глубине в 10 см, так как численное решение будет очень незначительно отличаться от истинного испарения, оптимизация будет проходить без осложнений и, кроме того, при этом количество измерений будет минимальным.

6. Заключение

Анализ полученных результатов показывает, что в случае, когда начальные данные совпадают с решением задачи (3)–(4) в соответствующих точках, можно сделать следующие выводы.

- Применение предлагаемого подхода позволяет восстанавливать испарение с поверхности почвы с хорошей точностью.
- Результаты численных экспериментов, проведенных по подобному описанному выше сценарию, позволяют оценить, как вид множества, в котором сравниваются моделируемые и предписанные значения влажности (а значит, и вид используемых в задаче экспериментальных данных), повлияет на близость соответствующего численного решения к истинному значению испарения, а также понять, насколько неосложненным окажется процесс численной оптимизации.

Литература

1. АЙДА-ЗАДЕ К. Р., ЕВТУШЕНКО Ю.Г. *Быстрое автоматическое дифференцирование на ЭВМ // Математическое моделирование.* – 1989. – Т. 1. – С. 121–139.
2. АЛПАТЬЕВ А.М. *К Влагообороты в природе и их преобразования.* – Л.: Гидрометеиздат, 1969. – 322 с.
3. ДИКУСАР В.В., ЗАСУХИН С.В. *Применение быстрого автоматического дифференцирования при нахождении испарения с поверхности почвы // Вестник МГТУ им. Н. Э. Баумана. Серия: Естественные науки.* – 2016. – Vol. 69, №6. – С. 42–55.
4. ИВАНОВ Н.Н. *Об определении величин испаряемости // Изв. ВГО.* – 1954. – Т. 86, №2. – С. 189–196.
5. КОНСТАНТИНОВ А.Р. *Нормирование орошения: методы, их оценка, пути уточнения // Гидротехника и мелиорация.* – 1986. – №1. – С. 19–28.
6. ЛЬГОВ Г.К. *Орошаемое земледелие.* – М.: Колос, 1979. – 191 с.
7. ЧЕРЕМИСИНОВ А.Ю., ЖЕРДЕВ В.Н., ЧЕРЕМИСИНОВ А.А. *Динамика климата, водных балансов и ресурсов Центрального Черноземья.* – Воронеж: Воронежский ГАУ, 2013. – 326 с.
8. ALLEN R.G., PRUITT W.O., BUSINGER J.A., FRITSCHEN L. J., JENSEN M.E., QUINN F.H. *Evaporation and Transpiration : ASCE Handbook of Hydrology.* – N.Y.: ASCE, 1996. – P. 125–252.
9. GRIEWANK A. *Evaluating derivatives.* – Philadelphia: SIAM, 2000.
10. MARTENS B., MIRALLES D., LIEVENS H., FERNANDEZ-PRIETO D., VERHOEST N. *Improving terrestrial evaporation estimates over continental Australia through assimilation of SMOS soil moisture // Intern. J. Applied Earth Observation and Geoinformation – Special Issue: “Advances in the Validation and Application of Remotely Sensed Soil Moisture”. Part 2.* – 2016. – Vol. 48. – P. 146–162.

11. MIRALLES D.G., HOLMES T.R.H., DE JEU R.A.M., GASH J.H., MEESTERS A.G.C.A., DOLMAN A.J. *Global land-surface evaporation estimated from satellite-based observations* // Hydrol. Earth Syst. Sci. – 2011. – Vol. 15. – P. 453–469.
12. MONTEITH J. *Evaporation and the environment* // 19th Symposium of the Society of Experimental Biology roc. Royal Society. – Cambridge, UK: Cambridge University Press, 1965. – P. 205–234.
13. PENMAN H.L. *Natural evaporation from open water, bare soil and grass* // Proc. Royal Society. – 1948. – Vol. 193. – P. 120–145.
14. PRIESTLEY C.M.B., TAYLOR R.J. *On the assessment of surface heat flux and evaporation using large-scale parameters* // Monthly Weather Review. – 1972. – Vol. 100, №2. – P. 81–92.
15. SAVENIJE H.H.G. *Determination of evaporation from a catchment water balance at a monthly time scale* // Hydrology and Earth System Sciences. – 1997. – Vol. 1, №1. – P. 93–100.
16. VAN GENUCHTEN M.TH. *A closed form equation for predicting the hydraulic conductivity of unsaturated soils* // Soil. Sci. Soc. Am. J. – 1980. – Vol. 44. – P. 892–898.

DETERMINING EVAPORATION AS A SOLUTION OF THE OPTIMAL CONTROL PROBLEM

Sergey Zasukhin, Dorodnicyn Computing Centre of RAS, Moscow Institute of Physics and Technology, Moscow, post-graduate student (s.zasukhin@yandex.ru).

Abstract: The problem of determining evaporation is stated as an optimal control problem. The controlled process of vertical water transfer in soil is described by one-dimensional nonlinear parabolic equation. The daily evaporation is a control and, the objective function is the mean-square deviation of calculated values of the soil moisture from some prescribed values. As a result of finite difference approximation, the optimal control problem is reduced to a nonlinear programming problem. The problem obtained is proposed to be solved by the steepest descent method. The objective function gradient is computed using exact formulas of fast automatic differentiation. It is assumed that the prescribed values coincide with some solution of the direct problem on the set of comparisons of the calculated values and prescribed values of soil moisture. How does the type of this set affect the accuracy of the solution? Several variants of such a set are considered. The analysis of numerical solutions of the corresponding problems allows to choose the optimal variant in the sense of the accuracy of the solution and the number of measurements of the required soil moisture data.

Keywords: objective function, steepest descent method, nonlinear parabolic equation, fast automatic differentiation.

УДК 519.6 + 519.85
ББК 22.19 + 22.18

Статья представлена к публикации членом редакционной коллегии Э.Ю. Калимулиной.

*Поступила в редакцию 26.09.2017.
Опубликована 30.09.2018.*

ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ ЭНЕРГОПОТРЕБЛЕНИЕМ И МИКРОКЛИМАТОМ БОЛЬШИХ МНОГОЗОННЫХ ЗДАНИЙ

Колодкина А. С.¹

(Ярославский промышленно-экономический колледж, Ярославль)

Марьясин О. Ю.², Огарков А. А.³

(Ярославский государственный технический университет,
Ярославль)

Рассмотрена задача энергоэффективного управления микроклиматом больших многозонных зданий. Для решения указанной задачи предлагается использовать метод иерархического распределенного прогнозирующего оптимального управления (МРС-подход). Данный метод позволяет достичь минимума глобального критерия качества и выполнение ограничений для всей системы с учетом взаимосвязей между подсистемами. При реализации иерархического распределенного МРС-алгоритма возникает проблема скоординированного решения задачи математического программирования для каждой из подсистем. Для решения глобальной задачи математического программирования авторы предлагают метод, основанный на методе декомпозиции путем разделения ресурсов. Авторами доказано, что если локальные задачи оптимизации имеют решение при определенных допущениях на множества допустимых решений локальных задач, то и задача координации будет иметь допустимое оптимальное решение. Результаты численных экспериментов показали преимущества использования предложенного авторами подхода для управления микроклиматом больших многозонных зданий. Сравнение различных вариантов реализации МРС-алгоритма показало, что иерархический распределенный МРС-алгоритм обеспечивает требуемое качество поддержания микроклимата с учетом выполнения глобальных ограничений при наименьшем энергопотреблении. Наличие различных видов энергоресурсов позволяет, например, при резком увеличении потребления тепловой энергии в часы пик, связанное с ее расходом на бытовые нужды, увеличить, для поддержания требуемого микроклимата, расход электроэнергии.

Ключевые слова: микроклимат, энергоресурсоэффективность, Model Predictive Control, иерархическое оптимальное управление, MATLAB.

¹ Анна Сергеевна Колодкина, преподаватель (anuta_25@mail.ru).

² Олег Юрьевич Марьясин, к.т.н., доцент (maryasin2003@yandex.ru).

³ Андрей Алексеевич Огарков, аспирант (drivemox@mail.ru).

Введение

В области жилищно-коммунального хозяйства (ЖКХ) постоянно происходят большие энергетические потери. Только на энергоснабжение жилых, промышленных и общественных зданий расходуется около 35% потребляемых в России энергетических ресурсов. При этом потенциал энергосбережения в области ЖКХ составляет более 30%, из которых около 30% приходится на коммунальные инженерные системы и до 70% – на здания и сооружения [2]. От эффективного использования энергоресурсов в ЖКХ зависит не только сохранение окружающей природной среды, но и уменьшение стоимости предоставляемых потребителям жилищно-коммунальных услуг.

Одной из основных причин больших энергетических потерь в области ЖКХ является низкая автоматизация существующих систем тепло- и энергоснабжения зданий, не позволяющая производить погодозависимое, оптимальное управление энергопотреблением и микроклиматом с учетом индивидуальных характеристик зданий. Особенно это касается больших многоэтажных, многосекционных зданий, зданий, соединенных между собой системой переходов, больших спортивных, культурных и торгово-развлекательных центров.

В настоящее время, с одной стороны, наблюдаются тенденции снижения стоимости потребительской электроники и микропроцессорной техники, увеличения уровня автоматизации и интеллектуализации инженерных систем зданий, широкое распространение распределенных систем управления, систем типа «Интеллектуальное здание», «Зеленое здание» и «Умный дом». С другой, происходит переход от централизованных систем энергообеспечения к децентрализованным системам, в том числе с использованием нетрадиционных источников энергии. Все это приводит к тому, что практически в каждом помещении здания может работать своя локальная система управления микроклиматом.

Новые тенденции дают возможность совершенствования си-

стем управления инженерным оборудованием зданий и снижения больших энергетических потерь в области ЖКХ. В частности, вместо традиционных двухпозиционных/трехпозиционных или ПИД-регуляторов для управления микроклиматом помещений могут использоваться более сложные алгоритмы автоматического и оптимального управления.

В зарубежных публикациях широко представлены работы, в которых для управления микроклиматом зданий предлагаются системы адаптивного и оптимального регулирования, системы на базе нечеткой логики и с использованием нейросетевых и генетических алгоритмов. Обзор таких работ можно найти, например, в [9]. В настоящее время в зарубежной научной прессе наиболее активно обсуждается подход, использующий прогнозирующие модели для управления энергопотреблением и микроклиматом зданий. Этот подход получил название Model Predictive Control (MPC) [8] и уже давно хорошо зарекомендовал себя при применении в других областях, например в нефтехимии [12]. К достоинствам MPC-подхода можно отнести то, что оптимальный регулятор, синтезированный в соответствии с данным подходом, обеспечивает соблюдение ограничений на управляющие и выходные переменные и позволяет достичь компромисс между робастностью и качеством регулирования. Однако, как и большинство оптимальных систем, он требует знания модели объекта управления. И здесь достоинством MPC-подхода является то, что поскольку он реализует закон управления с обратной связью по состоянию и управление рассматривается и реализуется на очень коротком промежутке времени, то для синтеза управления достаточно использовать лишь приближенные, линейные модели. Кроме того, прогнозирующие свойства MPC-регулятора позволяют эффективно подавлять контролируемые возмущения, учитываемые в модели.

1. Математическая модель микроклимата многозонного здания

Рассмотрим задачу управления микроклиматом здания, состоящего из множества зон (секций, помещений). Условием выделения зоны будем считать наличие у зоны собственной системы управления микроклиматом. Это может быть свой индивидуальный тепловой пункт (ИТП), электрический обогреватель или конвектор, газовый обогреватель, это может быть любая смешанная система отопления, вентиляции и кондиционирования (ОВК или HVAC, как их называют на западе). Предполагается, что ИТП или другое климатическое оборудование имеет средства автоматического управления по заданной программе.

Так как зоны здания имеют общие ограждающие конструкции, то на микроклимат в каждой зоне оказывает влияние микроклимат в соседней зоне. Это влияние тем сильнее, чем выше разность параметров микроклимата в соседних зонах, больше площадь общих ограждающих конструкций и выше их коэффициенты тепло и массопередачи.

Составим математическую модель микроклимата многозонного здания. Для простоты будем полагать, что зоны имеют одинаковый материал внешних стен и перекрытий между зонами. В основу математической модели положены уравнения теплового и материального баланса здания [3].

Динамика микроклимата здания без учета термической массы ограждающих конструкций (стен, окон, пола, потолка), инфильтрации тепла, влаги и газов через ограждающие конструкции, солнечной радиации и бытовых тепло-, влаго- и газовыделений может быть описана системой дифференциальных уравнений

$$(1) \quad \left\{ \begin{array}{l} c\rho V_i \frac{dT_i(t)}{dt} = \sum_{s=1, s \neq i}^{S_i} k_{in} F_s^{in} (T_s(t) - T_i(t)) + \\ + \sum_{r=1}^{R_i} k_e F_r^{ext} (T_E(t) - T_i(t)) + cG_{ni} (T_{ni}(t) - T_i(t)) + \\ + q_i, i = 1, \dots, N, \\ \rho V_i \frac{dW_i(t)}{dt} = G_{ni} (W_{ni}(t) - W_i(t)), i = 1, \dots, N, \\ \rho V_i \frac{dC_i(t)}{dt} = G_{ni} (C_{ni}(t) - C_i(t)), i = 1, \dots, N, \end{array} \right.$$

где c – теплоемкость воздуха; ρ – плотность воздуха; V_i – объем i -й зоны, $i = 1, 2, \dots, N$; T_i – температура i -й зоны, °C; t – время; S_i – число внутренних ограждающих конструкций i -й зоны; k_{in} – коэффициент теплопередачи ограждающих конструкций между зонами; F_s^{in} – площадь ограждающих конструкций между зонами, $s = 1, 2, \dots, S_i$; R_i – число внешних ограждающих конструкций i -й зоны; k_e – коэффициент теплопередачи внешних ограждающих конструкций i -й зоны; F_r^{ext} – площадь внешних ограждающих конструкций i -й зоны, $r = 1, 2, \dots, R_i$; T_E – температура наружной среды; G_{ni} – расход приточного воздуха i -й зоны; T_{ni} – температура приточного воздуха i -й зоны; q_i – тепло, полученное от системы теплоснабжения i -й зоны; W_i – влажность воздуха i -й, г; W_{ni} – влажность приточного воздуха i -й зоны; C_i – концентрация CO₂ в воздухе i -й зоны, ppm; C_{ni} – концентрация CO₂ в приточном воздухе i -й зоны. Параметры T_{ni} , W_{ni} , C_{ni} зависят от температуры T_E , влажности W_E , концентрации CO₂ наружного воздуха C_E и реализации системы ОВК.

Структура модели микроклимата здания, соответствующая системе (1), для двухмерного случая описывается схемой, фрагмент которой показан на рис. 1. Здесь индекс i для удобства восприятия разделен на два индекса: индекс i действует по горизонтали, j – по вертикали. Вектор $X_{ij} = \{T_{ij}, W_{ij}, C_{ij}\}$ – вектор параметров микроклимата здания. Вектор $X_E = \{T_E, W_E, C_E\}$ – вектор параметров наружной среды. Такая структура соответствует, например, многоэтажному зданию с последовательным расположением зон. В общем случае структура модели будет описываться трехмерной схемой.

Более точную модель температурного режима здания можно получить, если учесть термическую массу ограждающих кон-

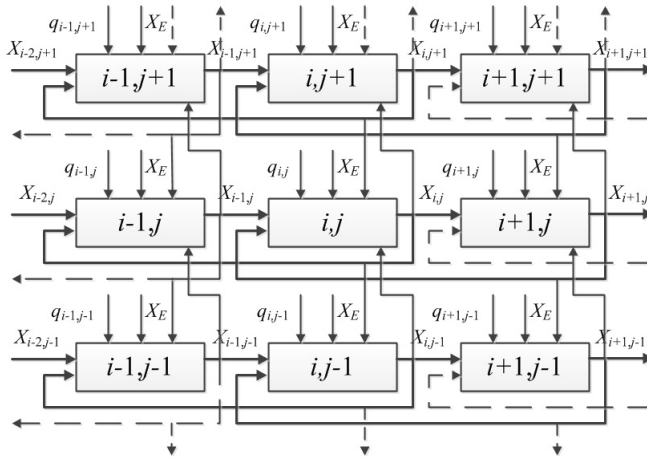


Рис. 1. Структура модели многозонного здания

струкций (стен, окон, пола, потолка). В этом случае к системе уравнений (1) добавляются уравнения

$$(2) \quad \begin{cases} c_{fij} \rho_{fij} V_{fij} \frac{dT_{fij}(t)}{dt} = \alpha_{fij} F_{ij}^{in} (T_i(t) - T_{fij}(t)) - \\ - k_{fij} F_{ij}^{in} (T_{fij}(t) - T_{ei}(t)), i = 1, \dots, N, \\ j = 1, \dots, n, \end{cases}$$

где n – число ограждающих конструкций i -й зоны; c_{fij} – теплоемкость j -й ограждающей конструкции i -й зоны; ρ_{fij} – плотность j -й ограждающей конструкции i -й зоны; V_{fij} – объем j -й ограждающей конструкции i -й зоны; T_{fij} – температура внутренней стенки j -й ограждающей конструкции i -й зоны, °C; α_{fij} – коэффициент теплоотдачи с внутренней поверхности стены j -й ограждающей конструкции i -й зоны; k_{fij} – коэффициент теплопередачи j -й ограждающей конструкции i -й зоны от внутренней поверхности ограждающей конструкции к наружному воздуху; T_{ei} – температура внешней среды: $T_{ei} = T_E$, если j -я ограждающая конструкция является внешней, и $T_{ei} = T_l$, $l \neq i, l = 1, 2, \dots, N$, если j -я ограждающая конструкция является внутренней. Коэффициенты теплопередачи k_{in} и k_e в системе уравне-

ний (1) при этом заменяются на коэффициенты теплоотдачи α_{fij} с внутренней поверхности стены j -й ограждающей конструкции i -й зоны. Массопередачей через ограждающие конструкции мы пренебрегаем в силу ее малости.

Учет термической массы ограждающих конструкций приводит, с одной стороны, к повышению достоверности модели [5], а с другой – к значительному увеличению размерности модели, и, следовательно, к увеличению ее сложности.

В уравнениях (1) не учитывается влияние таких факторов, как инфильтрация, солнечная радиация и бытовые тепловыделения, что естественно приводит к погрешностям определения параметров микроклимата в помещении i -й зоны. Оценка таких факторов осложняется их случайным характером и сложностью измерения. Поэтому учесть их влияние можно только путем мониторинга параметров микроклимата в каждой зоне. Это легко позволяют сделать современные системы типа «Интеллектуальное здание» или «Умный дом».

2. Оптимальное управление энергопотреблением и микроклиматом здания

2.1. ПОСТАНОВКА ЗАДАЧИ

Поставим задачу оптимального управления микроклиматом здания. В качестве критерия оптимальности будем использовать критерий типа

$$(3) \quad J = \sum_{i=1}^N J_i,$$

где $J_i = J_i(q_i)$ – критерий оптимальности i -й зоны, где $q_i = \{q_{i1}, q_{i2}, \dots, q_{im}, \dots, q_{iM_i}\}$ – расход энергоресурсов для i -й зоны; q_{im} – расход m -го вида энергоресурсов i -й зоны; M_i – число видов энергоресурсов, потребляемых i -й зоной. Критерий (3) должен соответствовать цели, связанной с поддержанием требуемых значений параметров микроклимата здания и одновременной минимизацией расхода энергоресурсов с учетом их стоимости.

Тогда задача оптимального управления энергопотреблением и микроклиматом здания может быть записана в виде

$$(4) \quad J = \sum_{i=1}^N J_i \rightarrow \min_{q_i},$$

при ограничениях задаваемых уравнениями (1), (2), локальными ограничениями

$$(5) \quad \begin{cases} T_{imin}(t) \leq T_i \leq T_{imax}(t), & i = 1, \dots, N, \\ \varphi_{imin} \leq \varphi_i \leq \varphi_{imax}, & i = 1, \dots, N, \\ C_i \leq C_{imax}, i = 1, 2, \dots, N, & i = 1, \dots, N, \\ q_{immin}(t) \leq q_{im}(t) \leq q_{immax}(t), & m = 1, \dots, M_i, \\ i = 1, \dots, N, & \end{cases}$$

и глобальным ограничением

$$(6) \quad \sum_{i=1}^N q_{im}(t) \leq q_{mmax}(t), m = 1, \dots, M,$$

где T_{imin}, T_{imax} – минимальное и максимальное значение температур в помещении i -й зоны; $\varphi_{imin}, \varphi_{imax}$ – минимальное и максимальное значение относительной влажности в помещении i -й зоны; C_{imax} – максимальное содержание CO_2 в воздухе i -й зоны; q_{immin}, q_{immax} – минимальное и максимальное количество m -го вида энергоресурса на энергоснабжение i -й зоны; q_{mmax} – максимальное количество m -го вида энергоресурса на энергоснабжение всего здания. Пределы q_{immax}, q_{mmax} задаются с учетом ограничений мощности соответствующих инженерных систем и климатического оборудования.

Если пренебречь несущественными нестационарными эффектами, заключающимися в зависимости коэффициентов модели (1), (2) от ее параметров, и нелинейностями, то уравнения модели микроклимата (1), (2) могут быть записаны в виде системы линейных уравнений состояния

$$(7) \quad \begin{cases} \dot{x}_i(t) = A_i x_i(t) + \sum_{l=1, l \neq i}^{L_i} A_l x_l(t) + B_i u_i(t) + G_i w_i(t), \\ y_i(t) = C_i x_i(t), i = 1, \dots, N, \end{cases}$$

где $x_i = \{T_i, T_{fij}, \varphi_i, C_i\}$ – вектор переменных состояния i -й зоны; $x_l = \{T_l\}$ – вектор переменных состояния l -й зоны, $l \neq i, l = 1, 2, \dots, L_i$; L_i – число зон, граничащих с i -й зоной; $u_i = \{q_i\}$ – вектор управляющих переменных i -й зоны;

$w_i = \{T_E, T_{ni}, \varphi_{ni}, C_{ni}\}$ – вектор внешних контролируемых возмущений для i -й зоны; $y_i = \{T_i, \varphi_i, C_i\}$ – вектор выходных измеряемых переменных i -й зоны; $A_i, A_l, B_i, G_i, C_i, i = 1, 2, \dots, N$, – соответствующие матрицы уравнений состояния.

Тогда задача оптимального управления энергопотреблением и микроклиматом здания заключается в минимизации критерия (3) при локальных ограничениях (7) и

$$(8) \quad \begin{cases} y_{imin}(t) \leq y_i \leq y_{imax}(t), & i = 1, \dots, N, \\ u_{imin}(t) \leq u_i \leq u_{imax}(t), & i = 1, \dots, N, \end{cases}$$

и глобальном ограничении

$$(9) \quad \sum_{i=1}^N u_{im}(t) \leq U_m(t), m = 1, \dots, M.$$

Перейдем от непрерывного описания системы к дискретному с сохранением обозначения соответствующих векторов и матриц. Уравнения (7), (8), (9) в этом случае заменяются на уравнения

$$(10) \quad \begin{cases} x_{ik+1} = A_i x_{ik} + \sum_{l=1, l \neq i}^{L_i} A_l x_{lk} + B_i u_{ik} + G_i w_{ik}, \\ y_{ik} = C_i x_{ik}, i = 1, \dots, N, k = 0, \dots, N_p, \end{cases}$$

$$(11) \quad \begin{cases} y_{ikmin} \leq y_k \leq y_{ikmax}, i = 1, \dots, N, k = 0, \dots, N_p, \\ u_{ikmin} \leq u_k \leq u_{ikmax}, i = 1, \dots, N, k = 0, \dots, N_p - 1, \end{cases}$$

$$(12) \quad \sum_{i=1}^N u_{ikm} \leq U_{mk}, m = 1, \dots, M, k = 0, \dots, N_p - 1,$$

где N_p – интервал времени на котором осуществляется управление.

2.2. ВАРИАНТЫ РЕШЕНИЯ ЗАДАЧИ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ

Для решения задачи оптимального управления, описываемой системой (10), (11), (12) и критерием (3), будем использовать МРС-подход. Этот подход может быть реализован в различной форме: централизованной, децентрализованной и распределенной [16].

Для централизованного МРС-алгоритма (сМРС) поставленная задача оптимального управления рассматривается в целом,

как задача централизованного управления многомерным объектом. Однако при большом числе зон из-за высокой размерности задачи увеличиваются вычислительные затраты на ее решение, а следовательно, увеличивается время на реализацию оптимального управления. Это может привести к снижению качества управления. Поэтому для решения поставленной задачи мы будем использовать принцип декомпозиции, в соответствии с которым общая задача разбивается на ряд локальных задач оптимизации (подзадач), размерность каждой из которых значительно меньше размерности общей задачи [6]. При этом каждая подзадача реализуется с помощью своей локальной подсистемы управления. Локальные подсистемы могут быть автономными или обмениваться информацией между собой и с верхним уровнем управления. В последнем случае может быть построена иерархическая структура управления. Вопрос о рациональном разбиении общей задачи на подзадачи в данной статье не рассматривается.

Рассмотрим так называемую неперекрывающуюся декомпозицию, когда каждой зоне здания будет соответствовать своя отдельная подсистема. В этом случае каждая i -я локальная подсистема будет решать задачу оптимального управления:

$$(13) \quad J_i \rightarrow \min_{u_i}$$

при ограничениях (10), (11). Если связи между подсистемами и глобальное ограничение (12) игнорируются, то в таком случае получается децентрализованная система оптимального управления. Децентрализованная система функционирует как совокупность автономных подсистем. Игнорирование взаимосвязей приводит к ошибкам, которые тем больше, чем сильнее характер связей между подсистемами. Игнорирование глобального ограничения может нанести урон системе в целом и привести к отключению ее от централизованной системы энергоснабжения. Децентрализованной системе соответствует децентрализованный MPC-алгоритм (dMPC). Данный вариант управления будет эффективным и применим тогда, когда параметры микроклимата в соседних зонах здания постоянно равны или близки по значению, а суммы потребления энергоносителей локальными подсистемами

далеки от предельно допустимых.

Если подсистемы могут обмениваться между собой информацией о значениях параметров микроклимата в каждой зоне по каналам связи, то тогда возможно построение распределенной системы оптимального управления. В распределенной системе учитываются связи между подсистемами, что повышает качество управления микроклиматом, но локальные задачи оптимизации (13) решаются подсистемами нескоординированно. Это может привести к тому, что минимум суммы локальных критериев оптимальности не будет достигать минимума критерия J для централизованной задачи управления. Кроме того, в данном случае необходимость передачи информации между подсистемами требует организации сети связи между ними. Однако в настоящее время ввиду широкого распространения сетевых технологий это обстоятельство не является ограничивающим фактором. Все сказанное относительно глобального ограничения (12) для децентрализованной системы оптимального управления остается справедливым и для распределенной системы. Распределенной системе соответствует распределенный MPC-алгоритм (DMPC). Такой вариант управления может использоваться тогда, когда в целях упрощения система управления не разделяется на уровни и не ставится задача координации локальных подсистем. Пример использования распределенного MPC-алгоритма для задачи оптимального управления тепловым режимом здания можно найти в работе [13].

Если в распределенной системе оптимального управления учитываются связи между подсистемами и необходима координация решения локальных задач оптимизации (13) с целью достижения минимума глобального критерия J , то в этом случае требуется построение иерархической распределенной системы оптимального управления. В такой системе на нижнем уровне будут решаться локальные задачи оптимизации, а на верхнем уровне – задача координации. Данному варианту соответствует иерархический распределенный MPC-алгоритм (HDMPC).

2.3. ИЕРАРХИЧЕСКИЙ РАСПРЕДЕЛЕННЫЙ MPC-АЛГОРИТМ

Известно [17], что решение задачи оптимального управления с ограничениями, с использованием MPC-подхода, сводится к решению на каждом шаге задачи математического программирования. Требование достижения минимума глобального критерия, учета связей между подсистемами и выполнения глобальных ограничений приводит к необходимости совместного решения локальных задач оптимизации в рамках единой глобальной задачи. При высокой размерности глобальной задачи ее решение может быть связано с большими вычислительными затратами. Для решения сложных задач оптимизации, задач с аддитивно-сепарабельным видом критерия оптимальности и с блочно-диагональной структурой части ограничений традиционно используются методы иерархической оптимизации (декомпозиционные методы оптимизации). Эти методы в основном были разработаны в 60-80-х годах и детально описаны в [1, 4, 7, 10].

Применение методов иерархической оптимизации для решения задач оптимального управления энергопотреблением и микроклиматом зданий с использованием MPC-алгоритмов описано в [11, 14, 15]. В работе [14] описывается применение классического метода Данцига–Вульфа для решения задачи оптимального управления тепловым режимом многозонного здания. При этом задача, решаемая MPC-алгоритмом на каждом шаге, приводится к задаче линейного программирования. В работе [15] для решения задачи, подобной задаче в [14], используется метод декомпозиции Бендерса (Benders' decomposition). Здесь так же, как и в [14], решается задача линейного программирования.

В работе [11] для решения задачи оптимального управления микроклиматом здания используется метод декомпозиции путем распределения ресурсов, ранее описанный в [4]. При этом локальные задачи и задача координации представляют собой задачи линейного программирования. Так как целевая функция координирующей задачи может быть не дифференцируемой в отдельных точках, то для ее решения используется Бандл-метод (bundle

method).

Использование линейного критерия в работах [11, 14, 15] изменяет цель задачи оптимального управления и при переходе к задаче линейного программирования приводит к увеличению числа ограничений. Поэтому авторы предлагают использовать естественный для задач оптимального управления линейными динамическими системами квадратичный критерий оптимальности

$$(14) \quad \begin{cases} J_i = \sum_{k=0}^{k=N_p} ([y_{ik} - y_{zik}]^\top Q_i [y_{ik} - y_{zik}] + \\ + u_i^\top R_i u_i), i = 1, \dots, N, \end{cases}$$

где y_{zi} – заданное значение параметров микроклимата i -й зоны; Q_i – неотрицательно-определенная матрица; R_i – положительно-определенная матрица. Квадратичный критерий (14) позволяет не только экономить энергоресурсы, но и обеспечивать требуемое качество параметров микроклимата в динамических режимах, например, при переходе с одного режима на другой, более энергосберегающий, при отсутствии людей в помещении в нерабочее время. При этом решение задачи оптимального управления i -й подсистемы, описываемой уравнениями (10) с ограничениями (11) и критерием (14), сводится к решению на каждом шаге задачи квадратичного программирования

$$(15) \quad \begin{cases} J_i = v_i^\top H_i v_i + f_i^\top v_i \rightarrow \mathbf{min}_{v_i}, \\ L_{yi} v_i \leq b_{yi}, i = 1, \dots, N, \\ L_{ui} v_i \leq b_{ui}, i = 1, \dots, N, \end{cases}$$

где

$$v_i = \begin{pmatrix} u_{i0} \\ u_{i1} \\ \vdots \\ u_{iN_p-1} \end{pmatrix},$$

а матрицы $H_i, f_i, L_{yi}, b_{yi}, L_{ui}, b_{ui}$ зависят от матриц модели A_i, A_l, B_i, G_i, C_i , весовых матриц Q_i, R_i и векторов ограничений $y_{ikmin}, y_{ikmax}, u_{ikmin}, u_{ikmax}$. Здесь необходимо отметить что вектор v_i формируется из значений управления в текущий момент времени u_{i0} и предсказанных значений управления на $N_p - 1$

шагов вперед. При этом состояние системы на k -м шаге $k = 1, \dots, N_p - 1$ будет определяться как

$$(16) \quad \begin{cases} x_{ik} = A_i^k x_{i0} + \sum_{j=0}^{k-1} A_i^{k-1-j} \sum_{l=1, l \neq i}^{L_i} A_l x_{lj} + \\ + \sum_{j=0}^{k-1} A_i^{k-1-j} B_i u_{ij} + \sum_{j=0}^{k-1} A_i^{k-1-j} G_i w_{ij}, \\ i = 1, \dots, N, k = 0, \dots, N_p. \end{cases}$$

Напомним, что МРС-алгоритмом на текущем шаге реализуется только управление u_{i0} , после чего горизонт N_p сдвигается на следующий такт времени, он становится текущим и процесс повторяется.

Задачу (15) мы будем считать локальной задачей оптимизации, решаемой на каждом шаге i -й подсистемой. Тогда глобальная задача для всех N подсистем будет иметь вид

$$(17) \quad \begin{cases} J = \sum_{i=1}^N J_i \rightarrow \mathbf{min}_{v_i} \\ \sum_{i=1}^N v_{im} \leq V_m, m = 1, \dots, M. \end{cases}$$

Для решения задачи (17) авторы предлагают метод, основанный на методе явной декомпозиции [1] (известным также под названием метода предсказания взаимодействий, прямого метода) и методе декомпозиции путем разделения ресурсов. Согласно предлагаемому методу локальные подсистемы используют информацию о текущих измеренных значениях параметров x_{i0} i -й зоны, текущих измеренных и прогнозных значениях параметров x_l , l -й зоны $l \neq i, l = 1, 2, \dots, L_i$, текущих измеренных и прогнозных значениях контролируемых возмущений w_i для решения локальных задач

$$(18) \quad \begin{cases} J_i(b_{uiz}) = v_i^\top H_i v_i + f_i^\top v_i \rightarrow \mathbf{min}_{v_i}, \\ L_{yi} v_i \leq b_{yi}, i = 1, \dots, N, \\ L_{ui} v_i \leq b_{uiz}, i = 1, \dots, N, \end{cases}$$

где b_{uiz} – заданные значения ограничений на энергоресурсы для i -й зоны. Здесь в качестве координирующих переменных выступают b_{uiz} и $y_{ip} = \{T_{ip}, \varphi_{ip}, C_{ip}\}$ – заданные (прогнозные) значения переменных связей между подсистемами.

Выполнение условий $y_i = y_{ip}$ для координирующих переменных y_{ip} равносильно введению дополнительных ограничений $|y_i - y_{ip}| \leq \varepsilon_{1i}$, где ε_{1i} – точность выполнения условий $y_i = y_{ip}$. Эти дополнительные ограничения в виде $y_i - \varepsilon_{1i} \leq y_{ip}$

и $-y_i - \varepsilon_{1i} \leq -y_{ip}$ могут быть введены внутрь условий $L_{ui}v_i \leq b_{uiz}, i = 1, \dots, N$. С помощью данного приема авторы фактически приводят метод явной декомпозиции к методу декомпозиции путем разделения ресурсов. Это позволяет, в рамках единого подхода, учитывать и связи между подсистемами и объем расходуемых ими энергоресурсов. Для улучшения разрешимости локальных задач коэффициенты ε_{1i} могут быть увеличены.

В результате решения локальных задач получаем значения оптимального расхода m -го вида энергоресурсов на энергоснабжение i -й зоны v_{im}^* при ограничениях $b_{uimz}, m = 1, \dots, M$, и значение критерия оптимальности i -й зоны $J_i(b_{uiz})$. Координатор использует информацию о $J_i(b_{uiz})$ от подсистем нижнего уровня для решения задачи координации в виде

$$(19) \quad \begin{cases} J = \sum_{i=1}^N J_i(b_{uiz}) \rightarrow \min_{b_{uiz}} \\ \sum_{i=1}^N b_{uimz} \leq U_m, m = 1, \dots, M. \end{cases}$$

Обозначим множества $D_i = \{v_i \mid L_{xi}v_i \leq b_{xi}\}$ и $V_i = \{b_{uiz} \mid \exists v_i \in D_i, \text{ такое, что } L_{ui}v_i \leq b_{uiz}\}$. Тогда можно сформулировать следующую теорему.

Теорема 1. Пусть выполнены следующие условия:

- а) множества D_i непусты, компактны и выпуклы;
- б) функции $J_i(b_{uiz})$ и $L_{ui}v_i$ непрерывны и выпуклы на D_i ;
- в) задача (18) имеет допустимое решение.

Тогда задача координации (19) имеет допустимое оптимальное решение.

Доказательство. Предположения о выпуклости $J_i(b_{uiz}), L_{ui}v_i$ и D_i обеспечивают, что множество V_i выпукло и что $J_i(b_{uiz})$ выпукла на V_i , так что задача координации является задачей выпуклого программирования. Так как значения b_{uiz} мы выбираем сами с учетом выполнения глобального ограничения, то допустимость решения задачи (18) приводит к тому, что координирующая задача также имеет допустимое решение. Из компактности D_i вместе с непрерывностью $J_i(b_{uiz})$ и $L_{ui}v_i$ следует, что подзадача i имеет оптимальное решение, поскольку она имеет допустимое. Следовательно координирующая задача также будет иметь оптимальное решение.

Для решения локальных задач (18) могут быть использованы эффективные методы решения задач квадратичного программирования. Функции $J_i(b_{uiz})$ могут быть недифференцируемы в некоторых точках по переменным b_{iz} , и разрывы производных могут оказаться в критических точках, например в минимуме. Поэтому для решения координирующей задачи (19) могут использоваться методы недифференцируемой оптимизации, например, метод возможных направлений или Бандл-метод. Кроме того, для решения координирующей задачи могут использоваться любые численные методы решения задач нелинейного программирования, использующие только значения целевой функции, например, комплексный метод Бокса, генетические алгоритмы и т.д.

3. Проведение численных экспериментов

Авторами были исследованы различные варианты решения задачи оптимального управления, описываемой системой (10), (11), (12) и критерием (3) на интервале времени от 0 до 24 ч. Для реализации математических моделей и выполнения расчетов использовался пакет программ MATLAB.

Расчеты проводились для здания, состоящего из девяти зон общей площадью 1280 м². План здания показан на рис. 2. Из рис. 2 видно, что зона 1 имеет общие ограждающие конструкции почти со всеми другими зонами здания, за исключением зон 4 и 7. Остальные зоны также имеют общие ограждающие конструкции с соседними зонами.

В качестве источников тепла рассматривалось центральное водяное отопление и электрическое отопление с возможностью местного регулирования температуры. Значение T_{imin} принималось 20 °С, T_{imax} – 22 °С, φ_{imin} – 30%, φ_{imax} – 60%, C_{imax} – 800 ppm. Для водяного отопления значение q_{imax} принималось от 5 до 20 кВт, в зависимости от размеров зоны, для электрического отопления – от 2 до 5 кВт. Цены на энергоносители принимались по текущим тарифам на отопление и электроэнергию. Возмущения по параметрам внешней среды моделировались пе-

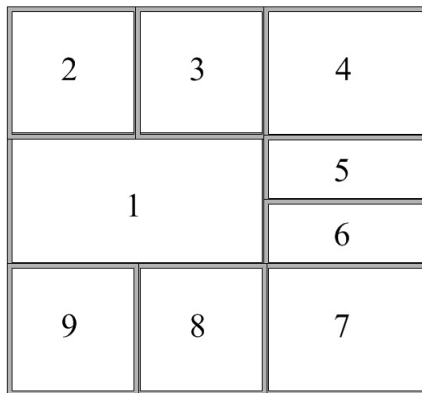


Рис. 2. План здания, состоящего из девяти зон

риодическими функциями на интервале времени от 0 до 24 ч. Эти возмущения учитываются в уравнениях модели и поэтому будут активно подавляться MPC-алгоритмом.

Влияние на параметры микроклимата таких не учитываемых случайных факторов, как инфильтрация через ограждающие конструкции, солнечная радиация, бытовые тепло-, влаго-, газовыделения и погрешности измерения моделировалось с помощью добавления случайной составляющей x_{wik} . В реальной системе текущие значения параметров микроклимата определяются в результате измерений. Полученные значения параметров микроклимата передаются по сети всем соседним зонам.

При выполнении численных экспериментов рассматривались следующие варианты организации оптимального управления: централизованное оптимальное управление, децентрализованное оптимальное управление, распределенное оптимальное управление, иерархическое оптимальное управление с использованием предложенного авторами метода. Полученные в результате проведения нескольких серий экспериментов результаты численных экспериментов для каждого варианта оптимального управления приведены в таблице 1.

Наибольшее энергосбережение достигается при децентрали-

Таблица 1. Результаты численных экспериментов

Используемый MPC-алгоритм	Потребление энергоресурсов, руб.
Централизованный	2425,5
Децентрализованный	2303,8
Распределенный	2533,1
Иерархический распределенный	2381,4

зованном управлении. Это объясняется тем, что в этом варианте подсистемы минимизируют свои затраты без учета существующих связей между подсистемами. Наименьшее энергосбережение достигается для распределенного варианта. Отсутствие координации для этого варианта приводит, как и ожидалось, к увеличению потребления энергоресурсов. Потребление энергоресурсов для централизованного и иерархически распределенного варианта практически одинаково.

Параметры микроклимата во всех зонах в течении всего времени эксперимента находились в пределах заданных ограничений. Нарушение ограничений может происходить, если общие ограничения на энергоресурсы примут такие значения, при которых количество подводимой энергии будет недостаточно для поддержания заданного микроклимата здания. Данные о средних за сутки расходах энергоносителей для зон 1–5 приведены в таблице 2, для зон 6–9 – в таблице 3.

В таблицах 2 и 3 рассматривается два режима потребления энергоносителей. Первый режим – когда общие ограничения потребления тепловой энергии по всему зданию всегда выполняются ($U_1 = 60$ кВт). Так как тариф на электроэнергию больше тарифа на отопление, то в этом режиме для поддержания микроклимата в помещении будет расходоваться только тепловая энергия. Однако такое положение может сильно измениться в часы пик. Дело в том, что тепловая энергия может тратиться на горячее водоснабжение, а электрическая энергия используется для

Таблица 2. Данные о расходах энергоносителей зон 1-5

Номер зоны	1	2	3	4	5
<i>Общее ограничение потребления тепловой энергии 60 кВт</i>					
Среднее потребление тепловой энергии, кВт	10	5,3	5,2	5,4	3,2
Среднее потребление электрической энергии, кВт	0	0	0	0	0
<i>Общее ограничение потребления тепловой энергии 45 кВт</i>					
Среднее потребление тепловой энергии, кВт	9,3	4,8	4,7	5,1	2,8
Среднее потребление электрической энергии, кВт	1	0,7	0,7	0,8	0,3

освещения и других бытовых нужд. Поэтому в часы пик значения ограничений U_1 для тепловой и U_2 для электрической энергии будут существенно снижаться. В крайнем случае в эти часы, ограничения на энергоресурсы могут привести к невозможности соблюдения требуемого микроклимата в помещениях.

Таблица 3. Данные о расходах энергоносителей зон 6-9

Номер зоны	6	7	8	9
<i>Общее ограничение потребления тепловой энергии 60 кВт</i>				
Среднее потребление тепловой энергии, кВт	3,2	5,4	5,1	5,5
Среднее потребление электрической энергии, кВт	0	0	0	0
<i>Общее ограничение потребления тепловой энергии 45 кВт</i>				
Среднее потребление тепловой энергии, кВт	2,8	5,1	4,6	4,8
Среднее потребление электрической энергии, кВт	0,4	0,8	0,6	0,7

Описанную ситуацию иллюстрирует второй режим, когда общие ограничения потребления тепловой энергии по всему зда-

нию снижаются до такого значения ($U_1 = 45$ кВт), при котором уже невозможно поддерживать микроклимат в помещении только за счет использования тепловой энергии. В этом случае включается электрическое отопление. При этом общее ограничение на потребление электрической энергии ($U_2 = 20$ кВт), расходуемой на отопление, не достигается.

На рис. 3 показаны графики изменения расхода тепла от водяного и электрического отопления для второй зоны при иерархическом оптимальном управлении. Сплошной линией показан расход тепла от водяного отопления, пунктирной – расход от электрического отопления. На рис. 3б показана ситуация, когда из-за необходимости учета глобального ограничения (12) в часы пик на короткое время становится невозможным использование одного вида энергоресурсов. Для поддержания требуемого микроклимата в такие моменты может резко возрасть расход альтернативного энергоносителя.

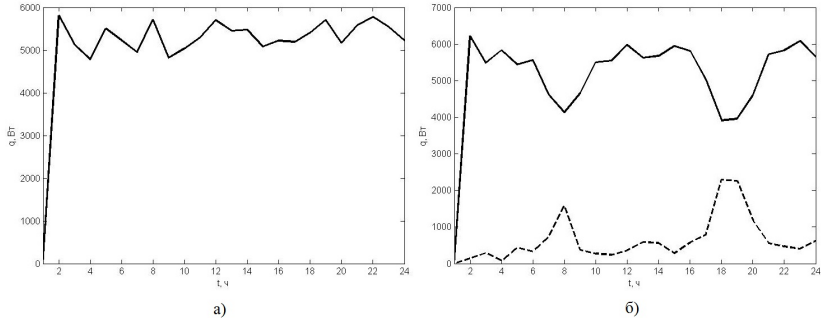


Рис. 3. Графики изменения расходов различных видов энергоресурсов: а) высокий уровень U_1 ; б) при снижении U_1 в часы пик

4. Выводы и перспективы

Таким образом, авторами предложен иерархический распределенный MPC-алгоритм, на каждом шаге которого реализуется процедура иерархической оптимизации, включающая итеративное решение локальных задач и задачи координации. Достоинством предложенного MPC-алгоритма является то, что, используя простые и доступные модели, он обеспечивает качественное управление микроклиматом в помещениях здания под действием случайных возмущений при минимизации энергозатрат и изменяющихся ограничениях на доступные энергоресурсы. Это позволяет более точно поддерживать заданный микроклимат в помещениях, существенно снизить расход энергоресурсов и исключить возникновение аварийных ситуаций, связанных с нарушением энергоснабжения. Результаты численных экспериментов подтвердили указанные положения.

Предложенный метод оптимального управления может применяться для управления энергопотреблением и микроклиматом больших многозонных зданий. Учитывая общую тенденцию широкого распространения распределенных систем управления (DCS-систем), такой подход к построению систем оптимального управления многозонными зданиями может стать все более перспективным.

Литература

1. АРТАМОНОВ А.Г., ВОЛОДИН В.М., АВДЕЕВ В.Г. *Математическое моделирование и оптимизация плазмохимических процессов*. – М.: Химия, 1989. – 224 с.
2. БАШКИН Б.В. *Энергосбережение в ЖКХ* / под. ред. Л.В. Примака, Л.Н. Чернышова. – М.: Академический Проект; Альма Матер, 2011. – 622 с.
3. БОГОСЛОВСКИЙ В.Н., СКАНАВИ А.Н. *Отопление*. – М.: Стройиздат, 1991. – 735 с.
4. ЛЭСДОН Л. *Оптимизация больших систем*. – М.: Наука, 1975. – 432 с.

5. МАРЬЯСИН О.Ю., КОЛОДКИНА А.С. *Управление тепловым режимом зданий с использованием прогнозирующих моделей* // Вестник СамГТУ. – 2017. – №1(53). – С. 122–132.
6. МЕССАРОВИЧ М., МАКО Д., ТАКАХАРА И. *Теория иерархических многоуровневых систем*. – М.: Мир, 1973. – 344 с.
7. СИНГХ М., ТИТЛИ А. *Системы: декомпозиция, оптимизация и управление*. – М.: Машиностроение, 1986. – 496 с.
8. VARATA F.A., IGREJA J.M., NEVES-SILVA R. *Model predictive control for thermal house comfort with limited energy resources* // Proc. of the 10th Portuguese Conference on Automatic Control, Funchal, Portugal, July 16–18, 2012. – P. 146–151.
9. DOUNIS A.I., CARAISCOS C. *Advanced control systems engineering for energy and comfort management in a building environment: A review* // Renewable and Sustainable Energy Reviews. – 2009. – No. 13(7). – P. 1246–1261.
10. FINDEISEN W. *Control and coordination in hierarchical system*. – N. Y.: Willey, 1980. – 480 p.
11. LAMOUDI M.Y., ALAMIR M. BEGUERY P. *Distributed constrained Model Predictive Control based on bundle method for building energy management* // Proc. of the 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC), Orlando, USA, December 12–15, 2011. – P. 8118–8124.
12. MORARI M., LEE J. *Model predictive control: past, present and future* // Computers and Chemical Engineering. – 1999. – Vol. 23. – P. 667–682.
13. MOROSAN P.D., BOURDAIS R., DUMUR D., BUISSON J. *Building temperature regulation using a distributed model predictive control* // Energy and Buildings. – 2010. – No. 42(9). – P. 1445–1452.

14. MOROSAN P.D., BUISSON J., BOURDAIS R. *Distributed MPC for Multi-zone Temperature Regulation with Coupled Constraints* // Proc. of the 18th IFAC World Congress, Milano, Italy, August 28 – September 2, 2011. – P. 1552–1557.
15. MOROSAN P.D., BUISSON J., BOURDAIS R. *A distributed MPC strategy based on Benders decomposition applied to multi-source multizone temperature regulation* // Journal of Process Control. – 2011. – No. 21(5). – P. 729–737.
16. SCATTOLINI R. *Architectures for distributed and hierarchical Model Predictive Control – A review* // Journal of Process Control. – 2009. – No. 19. – P. 723–731.
17. WANG L. *Model Predictive Control System Design and Implementation using MATLAB*. – Springer-Verlag London Limited, 2009. – 396 p.

OPTIMAL CONTROL FOR BUILDING ENERGY AND COMFORT MANAGEMENT OF LARGE MULTI-ZONE BUILDINGS

Anna Kolodkina, Yaroslavl Industrial and Economical College, Yaroslavl, Assistant (anuta__25@mail.ru).

Oleg Maryasin, Yaroslavl State Technical University, Yaroslavl, PhD in Technique, Associate Professor (maryasin2003@yandex.ru).

Andrey Ogarkov, Yaroslavl State Technical University, Yaroslavl, Postgraduate (drivemox@mail.ru).

Abstract: In this article, the problem of energy-efficient control of the indoor environment of large multi-zone buildings is considered. To solve this problem, it is proposed to use the hierarchical, distributed, predictive optimal control method (MPC approach). This method allows you to achieve a minimum of the global quality criterion and the implementation of restrictions for the whole system, taking into account the interrelationships between subsystems. When implementing a hierarchical distributed MPC-algorithm, there arises the problem of coordinated solution for the mathematical programming problems for each of the subsystems. To solve the global problem of mathematical programming, the authors propose a method based on the decomposition method via resource sharing. The authors prove that under certain assumptions on sets of admissible solutions for local problems, if the local optimization problems have a solution, then the coordination problem will have an admissible optimal solution. Numerical results are presented in order to illustrate the effectiveness of the proposed control strategy. As a result of comparison of various variants of the optimal control implementation, it is established that hierarchical distributed approach provides the best compliance with the limitations and the greatest energy saving. The availability of various types of energy resources allows, for example, in case of a sharp increase in the household needs consumption of thermal energy during peak hours, to increase the input of electricity to maintain the required microclimate.

Keywords: indoor environment, energy and resource efficiency, Model Predictive Control, hierarchical optimal control, MATLAB.

УДК 697.341

ББК 38.9

*Статья представлена к публикации
членом редакционной коллегии В.Н. Бурковым.*

Поступила в редакцию 20.11.2017.

Дата опубликования 30.09.2018.

НАЧАЛА ТЕОРИИ КОНЦЕПТУАЛЬНОГО ПРОЕКТИРОВАНИЯ С ПРИЛОЖЕНИЯМИ В ОБЛАСТИ АВИАЦИОННОЙ НАУКИ И ТЕХНОЛОГИЙ

Скворцов Е.Б.¹, Шелехова А.С.²

(«Центральный аэрогидродинамический институт
имени профессора Н.Е. Жуковского», г. Жуковский)

Представлены ключевые понятия концептуального проектирования, которые сформулированы в терминах теории множеств, векторного анализа и общей теории систем. Под концепцией понимается сочетание нескольких принципов, образующих информационно-логическую модель искусственного объекта. Принципы – это целенаправленные свертки связанной информации, векторы которых определяются идеями, а содержание – компонентами системы. Выделены понятия экосистемы (модель взаимодействия со средой) и эндосистемы (модель обособленного объекта). Единство моделей создаётся активной функциональной системой, которая в документированном виде образует проект. Разнообразие принципов расширено введением принципа «формы» наряду с принципами «действия» и «устройства» объекта. Его реализуемость обеспечивается технологиями. Процесс концептуального проектирования представлен как последовательная деятельность по устранению неопределённости, вызванных комплексным характером проектной информации с мнимой составляющей. Предложенная теория позволяет управлять концептуальным проектом, опираясь на законы математической логики. Формализованные благодаря этому проектные решения получают строгое обоснование.

Ключевые слова: концепция, принцип, идея, система, технология, проект, риск.

1. Постановка задачи

Концептуальное проектирование – это начальная стадия жизненного цикла создания уникального объекта (рис. 1), например, летательного аппарата в авиации. Существенным является акцент на уникальности, принципиальной новизне, отказе от готовых решений для будущего изделия, отличающегося нетрадиционными качествами. Поэтому для обеспечения

¹ Евгений Борисович Скворцов, заместитель начальника отделения, к.т.н. (skvortsov-tsagi@yandex.ru).

² Анна Сергеевна Шелехова, начальник сектора (anna.shelekhova@tsagi.ru).

новых потребностей необходимо формирование новой технической концепции и поиск нестандартных научных разработок, которые позволили бы осуществить проект.



Рис. 1. Жизненный цикл создания уникального объекта

Если идейная новизна отсутствует, то создание нового объекта может начаться с разработки прототипа на демонстрационной стадии (рис. 1) или даже на стадии производства изделия, в технологиях изготовления которого нет сомнений. Научный интерес к работе на этих стадиях жизненного цикла может сохраняться (методика летных испытаний, производственные технологии и т.п.), но исследования будут носить подчиненный характер, поскольку инициатива создания изделия в этих случаях находится вне научной среды. На демонстрационной стадии управление эскизным проектом объекта осуществляет, как правило, ОКБ, на дальнейших этапах жизненного цикла проектом управляют предприятия промышленности и затем — эксплуатационные предприятия.

Для данной работы представляет интерес случай, когда на основе достижений фундаментальных и поисковых исследований научным коллективом выполняется разработка концептуального проекта, а логическую основу управления концептуальным проектом составляют научные методы и, в частности, математические модели.

Известно, что сегодня не существует стандартов или регламентов концептуального проектирования. Теория управления

разработкой концепций создания перспективного изделия отсутствует. Парадигма концептуального проектирования еще не сложилась.

Из числа областей знаний, применяемых в области концептуального проектирования летательных аппаратов, можно упомянуть программы подготовки специалистов по проектированию самолетов, вертолетов, ракет, их двигателей и других агрегатов в высших учебных заведениях; военные и гражданские стандарты с описанием требований к содержанию конструкторской документации при выполнении аванпроекта (технического предложения) и эскизного проекта изделия авиационной техники; современные международные и российские стандарты по проектному менеджменту и управлению рисками; научные методы многодисциплинарной оптимизации и системы автоматизированного проектирования летательных аппаратов, в том числе разработанные в ЦАГИ. К этому направлению следует отнести методы теории систем и исследования операций, посвященные «внешнему проектированию» функциональных моделей, описывающих взаимодействие изделия с внешней средой и дающих оценку его эффективности.

Итак, на практике существует ряд образовательных, исследовательских и конструкторских методов, применяемых на концептуальной стадии жизненного цикла изделия. Однако они не связаны общей теорией, основанной на едином комплексе научных понятий. В настоящее время отсутствуют даже единообразные представления о концепции и системе, что позволило бы на основе общепризнанных понятий оперировать их моделями при управлении проектами. Не установлены логические связи с теориями, давно получившими широкое распространение в науке.

Современная организация авиационной науки требует преодоления этой разобщенности в связи с общей задачей формирования отраслевого научно-технического задела, необходимого для удовлетворения предметно-ориентированных потребностей авиационной техники. Выполнение продолжительных и затратных концептуальных проектов требует такого же, как в ОКБ, централизованного управления научным проектом на основе общей для всех научных школ теории концептуального проектирования.

Настоящая работа предлагает элементы такой теории, которые образуют универсальную логико-математическую модель многогранной проектной деятельности с согласованным понятийным аппаратом. Указаны причины, по которым управление концептуальным проектом необходимо проводить поэтапно в определенном порядке.

Научный подход к теории концептуального проектирования в данной работе формируется, прежде всего, в связи с логикой известных теорий – теории систем и теории множеств.

Системный метод получил всеобщее признание как метод решения сложных задач. К их числу, например, относятся организационные и экономические задачи, которым и были изначально адресованы математические методы моделирования систем. При всей обширности научных знаний в этой области они оставляют место для новых исследований.

2. Целеполагание

Управление любым проектом начинается с осознания его цели. Субъект, организующий проектную деятельность (проектант – руководитель проекта с проектным коллективом), заинтересованно воспринимает или сам формирует представления об ожидаемом результате проекта, к которому нужно стремиться. Но в этом случае конечной целью этой работы является, как правило, не новый материальный объект, а его потребительские ценности, т.е. функции, которыми можно будет воспользоваться на практике. При этом объект нужен лишь в качестве средства достижения цели. Собственно, для этого и организуется проект создания изделия с новыми свойствами. Цель инновации состоит в получении функций, соответствующих ожидаемым потребностям. Очевидно, что описание цели должно быть достаточно информативным, чтобы судить о достигнутом соответствии.

К определению цели проекта предъявляется ряд управленческих требований. Она должна быть:

- конкретной по сути, исчерпывающе определённой и ясно осознанной;
- структурированной «деревом цели» с подчинёнными задачами;

- измеримой с возможностью оценки результата;
- имеющей известную связь с ресурсами;
- привязанной ко времени осуществления проекта;
- достижимой в итоге проектной деятельности.

При выполнении этих условий проектное задание будет включать определённые требования к объекту в виде связей и ограничений, наложенных на целевые индикаторы, а сама цель может интерпретироваться как потребная (планируемая) система Z нормированных показателей объекта (рис. 2).



Рис. 2. Объект проектирования и система требований к нему

Тогда проектная деятельность будет направлена на то, чтобы конвертировать в проектируемый объект систему требований по множествам условий воздействия N_p среды обитания, используемых ресурсов N_r , ожидаемых полезных эффектов N_e :

$$(1) Z \subset N_p \times N_r \times N_e.$$

Возникает вопрос, как это сделать? Именно в этом состоит основная задача разработки теории проектирования, в том числе концептуального.

3. Объект и система

Задачи отыскания оптимального управления породили кибернетическое представление о *целенаправленно управляемом объекте*. Такая модель, выделяя из глобальной среды локаль-

ную систему, специализированную на решении одной задачи, включает информацию о внутренних связях объекта $\dot{x} = f_1(x, u, t)$, фазовом состоянии параметров $x \in V$ и процессе выработки управляющих сигналов $u \in Q$, изменяющих состояние объекта (рис. 3) в зависимости от времени t . В этой постановке описания способа функционирования Q и устройства V , а значит, и внешние связи объекта предполагаются известными.

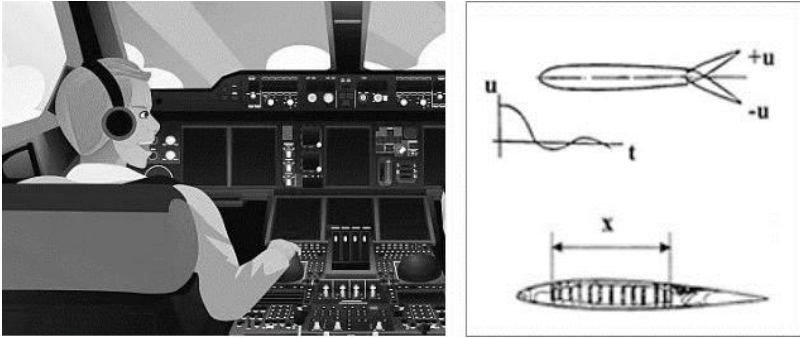


Рис. 3. Оператор и модель объекта управления

Объект находится под воздействием широкого поля внешних реакций $B_i (i = 1, 2, \dots, m)$, в том числе ресурсов среды R , получая из них средства $r \in R$ для покрытия издержек своего функционирования, и испытывает управляющие воздействия N_e поля заданий, нормирующих внешний спрос на продукцию E локальной системы с заданной целевой функцией (1). Её отклики на сигналы среды составляют, соответственно, различные полезные эффекты $e \in E$ и поле запросов на внешние ресурсы N_r для осуществления процесса, в котором участвует объект (рис. 4), управляемый оператором или автоматом.

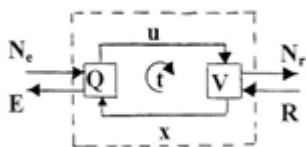


Рис. 4. Система «вход–выход»

Системный метод усматривает в предметах, прежде всего, признаки субъекта, абстрагируясь от его внутреннего содержания и уделяя основное внимание взаимным интересам субъекта и среды. Особый вклад системного метода в решение комплексных проблем обусловлен тем, что благодаря языку теории множеств он позволяет из всего поля реакций B_i , описывающих взаимодействия объекта и среды, выявить наиболее существенные взаимосвязи, чтобы включить их в исследование иерархических систем большой сложности.

Общая теория [5] под системой понимает отношение вида $S \subset X_{i=1}^m B_i$ на непустых множествах, образующих декартово произведение. Для рассмотренного выше субъекта управления целевая функция и ресурсы образуют непересекающиеся множества, которые создают систему типа «вход–выход»: $S \subset R \times E$. Элементы этого отношения образуют множество упорядоченных пар: $S \subset \{(r, e) \mid r \in R \wedge e \in E\}$.

Подобная модель системы получила широкое распространение, в частности, в задачах «исследования операций». Независимо от областей применения [2, 14] методы исследования операций включают выбор критерия эффективности объекта, построение модели его функционирования (развития) в среде обитания и введение исходных данных, применение разнообразных методов математического обеспечения и поиск оптимального управления объектом.

Описанный метод исследования применяют также для «внешнего проектирования», которое позволяет для моделей объектов с известными свойствами определять наилучшие условия применения или из нескольких конкурирующих объектов

рекомендовать наиболее успешный в моделируемых условиях применения, не вдаваясь в особенности его устройства.

Задача проектирования объекта, в том числе концептуального, принципиально отличается от задачи управления объектом. Как правило, *задача проектирования* объекта формулируется следующим образом: создать организацию или конструкцию, в частном случае их элементы, обеспечивающие наилучшее выполнение задачи при некоторых требованиях, накладываемых целеполаганием. Согласно этой задаче оператор-проектант лично не участвует в изменениях состояния объекта во время t его действия, как на рис. 3. Однако он знает, что возможны определенные программы управления $u(t) \in Q$ несуществующим объектом, так же как есть выбор из множества различных способов его функционирования, описанных моделями Q . Проектант знаком с разными моделями V возможного устройства будущего объекта, особенности которых могут быть описаны параметрами $x \in V$ (рис. 5) и оценены как потребными для их реализации ресурсами из числа возможных $r \in R$, так и локальными эффектами $e \in E$ из числа полезных. В число проектных данных входит также геометрическая информация, описывающая форму объекта $g \in F$.

В связи с этим в задаче проектирования преобразуется смысл обозначений, знакомых по теории управления. Теперь фазовые параметры не столько характеризуют текущее состояние объекта (эти детали нужны другим специалистам), сколько являются признаками особого типа программы действия и агрегатного состояния среди множества сведений о других способах и устройствах из того ассортимента, которым располагает и оперирует проектант во время t управления проектом.

Однако задача проектирования не может ограничиться только этим рассмотрением, поскольку задание на проект (1) должно содержать также характеристику условий его применения с указанием на различные внешние воздействия (например, рис. 6)

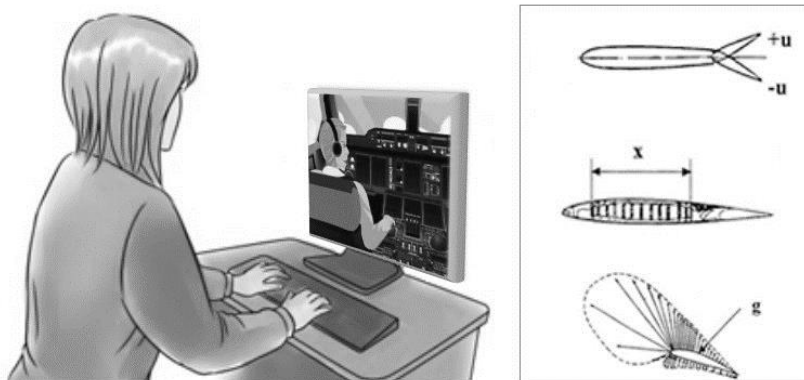


Рис. 5. Проектант и модель объекта проектирования

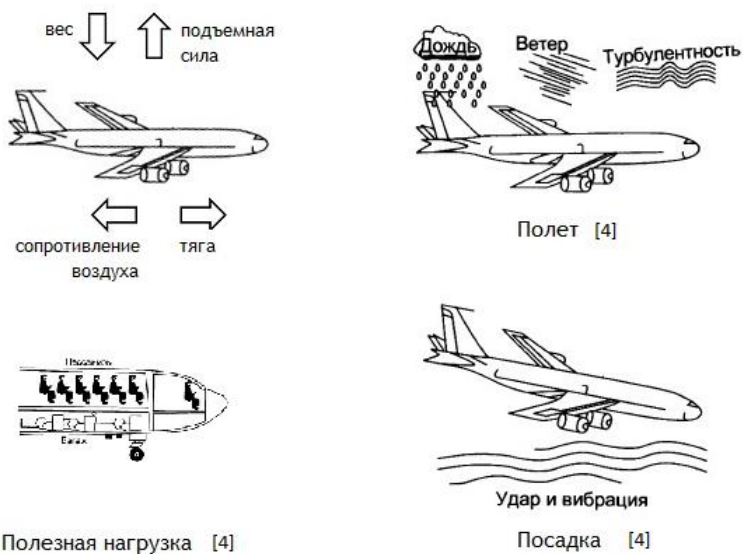


Рис. 6. Элементы внешнего воздействия и другие нагрузки на самолет

Схема системы проектирования может быть представлена в виде треугольной модели (рис. 7). В описание связей локальной и глобальной среды вошли понятия элементарных сил p взаимодействия, составляющих часть внешних нагрузок P различной физико-химической природы, в свою очередь включенных в поле допустимых условий эксплуатации N_p . Последние нормируются и определяют весь спектр воздействий среды. Следует специально отметить, что имеется в виду воздействие не только внешних, например, аэродинамических сил для летательного аппарата, но и внутренних, например, инерционных, в том числе со стороны топлива или груза летательного аппарата.

Управление проектом – это процесс, состоящий из последовательных действий, организованных проектантом в течение времени τ (например, рис. 1) и направленных на достижение цели проекта (1). Характер этой деятельности имеет смысл прямого поиска [15] удовлетворительных проектных решений.

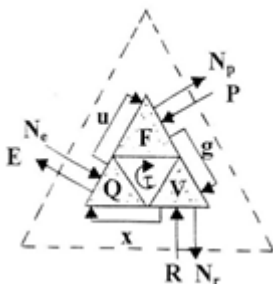


Рис. 7. Система «нагрузка-затраты-эффект»

Таким образом, модель системы проектирования сложнее, чем модель системы управления, для неё характерны тройственные отношения (степень сложности $n = 3$) вместо бинарных (степень сложности $n = 2$). Соответствующие оценки результатов проектирования представляются системой типа

$$(2) \quad S_z \subset R \times P \times E,$$

образованной множеством упорядоченных троек: $S_z \subset \{(r, p, e) \mid r \in R \wedge p \in P \wedge e \in E\}$. Из рис. 6 можно видеть, что факторы взаимодействия объекта с внешней средой (R, P, E)

и нормируемые проектным заданием требования (N_r, N_p, N_e) измеряются одинаковыми по своей природе величинами (например, располагаемый и потребный эффект). Так как системы (1) и (2) сопоставимы, это делает возможной верификацию разрабатываемой модели объекта.

В обоих случаях эти системы моделируют внешние связи. При этом (2) является *экзосистемой* (целевой системой) объекта, содержание которого оказывается вне этого понятия, оставаясь «черным ящиком». Очевидно, что задача проектирования не может ограничиваться рассмотренной системой как единственной моделью, поскольку при этом исчезает сам предмет исследования. Напрашивается вывод о целесообразности совместного применения взаимодополняющих моделей субъекта и объекта.

В модели объекта понятие организации M_u объединяет множество динамических элементов $u \in Q$, которые характеризуют модели взаимодействующих в общем технологическом процессе способов функционирования $Q \subset M_u$. Последние отличаются такими признаками, как приемы, режимы, последовательность и условия выполнения операций, реализующих различные программы.

Эти понятия соотносятся с интерпретацией структуры M_x как множества статических элементов $x \in V$ системы, которые характеризуют модели деталей, узлов, агрегатов, изделий и других орудий $V \subset M_x$, предназначенных для выполнения технологических процессов. Устройства как структурные единицы могут отличаться материалом, размерами, архитектурой и другими параметрами состояния.

То, что в задаче проектирования определено как способ и устройство, составляет, как известно, предмет патентно-изобретательской деятельности и входит в число основных инженерных понятий.

Форма – это особый фактор объекта. Достаточно сказать, что сам процесс проектирования сопровождается разработкой наглядной технической документации. При этом наиболее универсальной информацией – геометрической (g) – обладают схемы и чертежи, которые характеризуют плотность распределения материала в пространстве и обеспечивают единообразное понимание формы участниками проекта на всех стадиях его реализации. Понятием формы F в общем случае целесообразно харак-

теризовать любую информацию о границах объекта и его элементов. Множество её проявлений M_g составляет пограничную область взаимодействия с материальной средой: $g \in F \subset M_g$.

Согласование моделей объекта и субъекта проектирования [15] требует использования единого языка для их описания. Иерархически организованные субъекты получили обозначение «система» на языке теории множеств. Пользуясь этим же языком, представим информационно-логическую модель объекта как внутреннюю *эндосистему* (объектовую систему):

$$(3) S_o \subset V \times F \times Q$$

Описание (3) эндосистемы объекта S_o может быть представлено множеством упорядоченных троек: $S_o \subset \{(x, g, u) \mid x \in V \wedge g \in F \wedge u \in Q\}$. В таком виде представлено все возможное разнообразие компонентов эндосистемы, включающей модели устройства, формы и способа действия объекта. Однако действительное число локализованных подсистем определяется сложностью задачи проектирования, показателем которой может быть $n \geq 3$. Так, например, если требуется одним устройством произвести три различных действия (многоцелевая система), то сложность такой системы увеличится до $n = 5$ и она будет включать пять подсистем (V, F, Q_1, Q_2, Q_3). Наоборот, если способ функционирования облика задан и не требует проектирования, то сложность задачи ограничена ($n = 2$) поиском решающей системы вида (V, F, \emptyset) .

Таким образом, объектовая система (3) любой сложности n может быть представлена в виде: $S_o \subset X_{i=1}^n A_i$, где A_i – локальное множество подсистемы из общего поля взаимодействующих в составе эндосистемы.

Итак, сложилось два различных понятия «системы», которая в переводе с греческого означает «состав – целое, состоящее из соединенных частей». В обоих случаях это упорядоченные множества, в том числе экзосистема S_z – модель формализованных связей объекта со средой обитания, объединенных целью проекта. Вторая из рассмотренных – эндосистема S_o – модель обособленного от среды обитания объекта, состоящего из взаимосвязанных компонентов (подсистем), интегрированных в од-

но целое для достижения цели проекта. В данной работе оба понятия составляют единство субъективного и объективного, оба являются формализованными моделями и в этом качестве представляют интерес как предметы исследования проектантом, деятельность которого направлена на достижение цели проекта. Необходимость введения двух разных понятий в настоящей работе возникла впервые. Получив широкое распространение, до сих пор они развивались независимо друг от друга, используя один и тот же термин «система», например [4] и [5]. Для исследования двух разных моделей в рамках одной работы это неприемлемо, поэтому в их определения внесены уточнения.

Идеи, воспринятые в результате умственного напряжения, становятся элементами ментальной личности [14]. Они фиксируются в памяти и участвуют в информационном обмене, в том числе как множество элементов замысла $D(\tau)$, произведенных проектантом в стремлении постичь содержание объектовой системы $S_o \subset X_{i=1}^n A_i$. В то же творческое время τ сознание проектанта связано с множеством $S_z \subset X_{i=1}^m B_i$ реализующих систем, которые им измерены. Согласно теории множеств [10] подобные отношения называются **функциональной системой**:

$$(4) D: S_o \rightarrow S_z$$

Здесь элементы замысла из множества $D(\tau)$, возникающие благодаря интуиции и логике проектанта, в том числе динамические операторы $S_o(\tau) \in D(\tau)$, составляют информационную модель объекта как функцию сознания, определенную на множестве S_o со значениями на множестве S_z .

Понятие функциональной системы устанавливает адекватную связь моделирующего сознания проектанта с источником информации об объекте S_o , с одной стороны, и моделью его обитания S_z – с другой. В документальном виде такие функциональные системы образуют собственно **«проекты»**.

4. Концепция и принципы

В соответствии с теорией поля эндосистему можно рассматривать в качестве пространственного скалярного поля вида $S_o \subset X_{i=1}^n A_i$, например (3), в каждой точке которого взаимодей-

ствующие множества подсистем A_i производят функцию вида $S_z \subset X_{i=1}^n B_i$, например (2).

Ключевой характеристикой деятельности управляющего проектанта является его целеустремленность. При этом экзосистема S_z служит для измерения целевых факторов (2) и получения информации для принятия решений. Объектовая эндосистема S_o отображает результаты (3) целенаправленных усилий проектанта, однако собственно направление процесса проектирования заключено в более содержательной модели «концепции».

Целенаправленность проекта определяется на векторном поле $\vec{K} \subset X_{i=1}^n A_i$, в каждой точке которого взаимодействие векторов \vec{A}_i производит функцию (2). На этом поле задачей проекта является формирование его **концепции**:

$$(5) \vec{K} \subset \vec{V} \times \vec{F} \times \vec{Q},$$

определяющей весь дальнейший ход проектирования. Как известно, пространственный вектор полностью определяется его проекциями, в том числе принципами устройства, формы и действия соответственно:

$$(6) \vec{V} = V \cdot \vec{\lambda}_x; \vec{F} = F \cdot \vec{\lambda}_g; \vec{Q} = Q \cdot \vec{\lambda}_u,$$

где V, F, Q – скалярные компоненты эндосистемы; $\vec{\lambda}_x, \vec{\lambda}_g, \vec{\lambda}_u$, – векторы единичной длины (орты), задающие направление вектора концепции.

Здесь множества вида \vec{A} представляют собой направленные свертки логически связанной информации и именуются **принципами** [15]. Элементами вектора \vec{A} являются скалярные аргументы вида A (компоненты эндосистемы) и единичные векторы вида $\vec{\lambda}$, которые по смыслу являются **идеями**, придающими то или иное содержание модели объекта.

Таким образом, принципы в составе концепции являются векторами (рис. 8). Величина каждого из них определяется информационной сверткой множества компонентов эндосистемы S_o , а направление задано ортом-идеей, в котором заключена объединяющая логика принципа, выраженная его именем (назованим, формулой).



Рис. 8. Принципы формы, устройства и способа действия [15]

По сути обоснование и выбор концепции играет роль творческого зачатия – замысла, определяющего по какому пути пойдет содержательная работа над проектом. Момент выбора идеи аналогичен прохождению железнодорожной стрелки (рис. 9), движение за которой от начала до конца следует по направлению, отличному от отвергнутой альтернативы. Возможная ошибка в выборе пути может сделать проект ущербным, а затраты на него – бессмысленными. Предотвращение этих рисков – ключевое событие проекта.

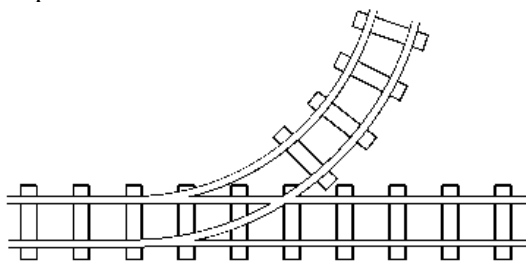


Рис. 9. Схема разветвления железнодорожных путей

5. Идея и принципы

Как можно понять из (6), между ортами-идеями и компонентами эндосистемы нет однозначной связи в составе принципов из-за их различной природы. Это обстоятельство вносит неопределенность в описание принципа (даже при наличии идеи), делая его нуменальной, т.е. непосредственно (в состоянии свертки) непознаваемой сущностью.

Например, идея «качения по поверхности» создает выбор из нескольких принципов для органов циклического движения, таких как цилиндр (колесо) с единственной осью вращения (рис. 10), сфера (шар) с неограниченным числом осей вращения и, наконец, шагающие опоры с дискретными контактными поверхностями.

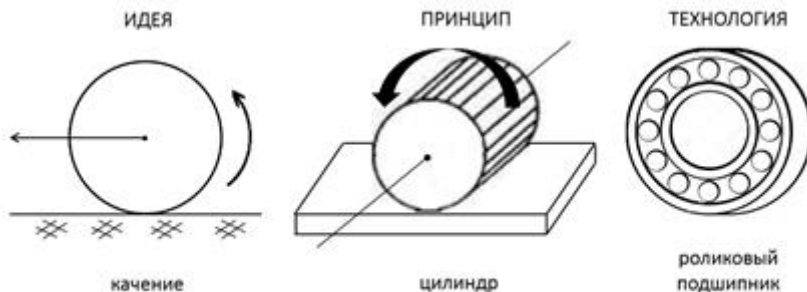


Рис.10 . Идея качения, принцип и технология на ее основе

Понятие идеи за тысячелетия развития гносеологии (от Аристотеля и Платона до Гегеля и современных философов) оставалось по сути неизменным. Все известные авторы сходятся в главном: «идея» – неделимая информационно-логическая единица, независимая от времени. В содержании идей заключена суть предметов и явлений материального мира. Идея представляет собой сущность не отдельного феномена действительности, но является прообразом некоторого типа/вида предметов и явлений (в переводе с греческого «идея» – вид или форма). Оставаясь неизменными, единичные идеи могут группироваться в различных сочетаниях в зависимости от объединяющей концепции. Это можно сравнить с участием атомов в строении различных молекул.

В познавательной деятельности идеи являются источником мыслей для человеческого сознания, отражаясь в нём в качестве представлений о возможном решении поставленной задачи. Таким образом, идеи не принадлежат субъекту мыслительной деятельности, и, кстати, авторское право на идеи (принципы и кон-

цепции) не распространяется (п. 5 ст.1259 ГК РФ). И хотя, как отмечал И. Кант, идеи и лежат в основании предметов и явлений природы, но с помощью органов чувств, восприняты быть не могут. Об их существовании мы узнаём только умозрительно благодаря человеческому сознанию.

Находясь вне сознания, идеи составляют трансцендентное поле (от лат. «transcendents» – запредельный, потусторонний). Это утверждение И. Кант выразил следующим образом: «основоположения, применение которых целиком остаётся в пределах возможного опыта, мы будем называть имманентными, а те основоположения, которые должны выходить за эти пределы, мы будем называть трансцендентными».

Отсюда следует, что сознание мыслящего субъекта включает знания двух видов: имманентное, которое отражает идеи, уже воплощённые в предметах и явлениях действительности, и трансцендентное знание, основанное на восприятии заранее неизвестных идей, возникающих в связи с целеустремлённой интеллектуальной деятельностью субъекта. И те, и другие могут составлять принципы.

Задачи научного исследования предусматривают численную характеристику информации, содержащейся в множествах вида A . В связи с этим будет уместно установить соответствие понятий «имманентный», т.е. известный из опыта, и «реальный» (Re), т.е. измеренный действительным числом. Другая аналогия – «трансцендентный», т.е. принципиально новый, неизвестный из опыта, – и «мнимый» (Im), т.е. воображаемый, существующий лишь в мысленном представлении и определённый мнимым числом согласно теории функций комплексного переменного:

$$(7) \quad A = ReA + i \cdot ImA .$$

Из комплексного представления принципов, а значит, и концепций из них построенных, следует, что если цель создания реального объекта содержит только реалистичную информацию, то проектная деятельность должна быть организована так, чтобы в процессе материализации идей преобразовать мнимые знания (мнения, гипотезы, прогнозы) в реальные, подтверждённые опытом и сопоставимые с целью.

Прибегая к комплексным числам для действий с принципами, необходимо понимать, какой смысл придается мнимому числу в отличие от действительного. Математик Валлис еще в 18 веке отметил [6]: «если в алгебре отрицательное число означает количество, меньшее, чем ничто, то в физике оно означает противоположность действия». В нашем случае, если положительное число говорит о приобретении, приросте содержательной информации (вектор положительного направления), то отрицательное – о её утрате или недостатке (вектор отрицательного направления). С этой точки зрения, корень квадратный из отрицательного числа (одномерная характеристика отрицательной величины) тоже является характеристикой дефицита информации, который должен быть покрыт заранее неизвестными идеями. Для теории концептуального проектирования существенной является не только принятая интерпретация мнимого числа. Теория функций комплексного переменного указывает на то, что существуют функции, благодаря которым действия с комплексными числами приводят к реальным результатам, например, в простейшем случае: $(a + bi)(a - bi) = a^2 + b^2$. Аналогичную цель преследуют функции проектанта.

Таким образом, математический аппарат ТФКП, с одной стороны, придает научную строгость рассмотренным понятиям гносеологии, а с другой стороны, дает теоретическое обоснование основному направлению проектной деятельности – от комплексной первичной информации о будущем изделии к реальному описанию объекта в завершении проекта.

Понятными становятся и задачи, которые должны быть решены при управлении проектом. Необходимо определить, какая именно функция или алгоритм управления позволяет завершить реальным результатом действия с комплексными концепциями (принципами, идеями) и соответствующими комплексными системами. В данном случае речь идет, конечно, не об алгебраической формуле, а о методологии концептуального проектирования. Очевидно, что ее особенностью должна быть обязательная оценка неопределенностей для целенаправленного сокращения мнимых представлений.

6. Принципы и технологии

Вернемся к обсуждению понятия «идея». Напомним, что идеи представляют собой частные информационно-логические модели не отдельных и конкретных феноменов действительности, а их разновидностей. Последние могут иметь множество конкретных реализаций, имеющих благодаря идее один общий признак.

В рассмотренном выше примере с идеей качения было названо несколько принципов ее осуществления, включая цилиндр с одной осью вращения. В реальности этот принцип нашел применение в огромном числе предметов, известных с древности до наших дней. Это колеса транспортных средств, подшипники качения, оси и валки станочного оборудования и т.п. Такой пример приведен на рис. 9.

Рассмотрим аналогичные по смыслу примеры из авиационной практики. Так, новым направлением создания боевых самолетов современного поколения стало выполнение требований малой заметности в различных диапазонах спектра, из которых важнейшим является радиолокационный с наибольшей дальностью обнаружения цели в зоне действия средств ПВО. В связи с этим при создании самолетов в соответствии с идеей минимальной «эффективной поверхности рассеяния» (ЭПР) появился ряд принципов проектирования основных отражателей в составе планера и силовой установки (например, рис. 11).

В число возможных мер включен принцип согласования углов стреловидности передних и задних кромок крыла, оперения и других элементов планера. В этом случае зондирующее излучение РЛС будет отражаться от всех кромок в одном узком диапазоне углов, и на экране оператора ПВО появится один сигнал отражения в этом ракурсе. Вероятность его обнаружения значительно меньше, чем вероятность обнаружения серии сигналов от разных частей самолета на круговой диаграмме. Далее возникает вопрос, как эту плодотворную идею реализовать в самолете? Оказывается (рис. 12), для этого есть ряд возможностей в связи с применением геометрии шеврона, пины или ромба в конструкции самолета, когда обеспечивается параллельность кромок, сводящая к минимуму число пиков отраженного сигнала

РЛС. Именно эти варианты исполнения крыла и нашли воплощение в различных проектах малозаметных военных самолетов.

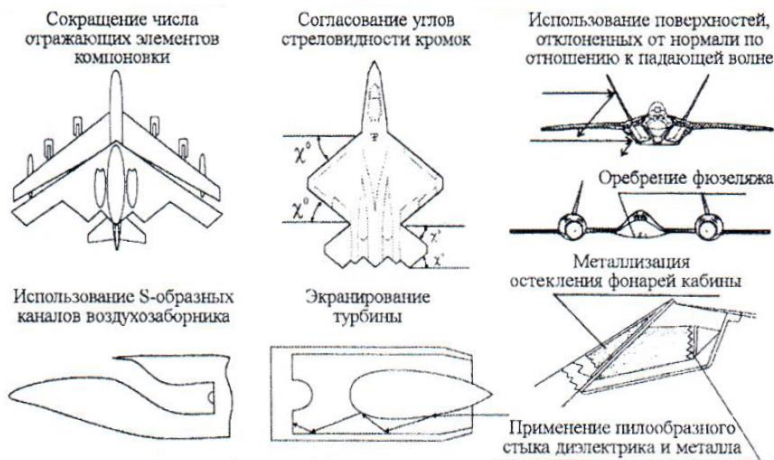


Рис. 11. Принципы уменьшения радиолокационной заметности самолётов [1]



Рис. 12. Формы крыла в плане с параллельными кромками [1]

Другой пример. В годы Великой Отечественной войны (1941–1945 гг.) наиболее известными советскими истребителями стали Як-3 и Ла-5 (рис. 13). Их сравнение показывает, что оба соответствуют одной и той же технической концепции. А именно, самолеты созданы в одинаковой нормальной схеме аэродинамической балансировки, что является определяющим

принципом в динамике полета. Существенным для динамики является также способ создания тяги винтовым двигателем, который ограничивает скорость полета. Принцип формы проявляется в смещенном к носу фюзеляжа положении крыла, что связано с большими инерционными нагрузками от двигателя внутреннего сгорания, определяющими переднее положение центра тяжести самолета. Принцип устройства является типичным для всех самолетов того времени. Крыло, оперение и фюзеляж образуют федеративную конструкцию, каждый элемент которой выполняет отдельную функцию без признаков взаимной интеграции.

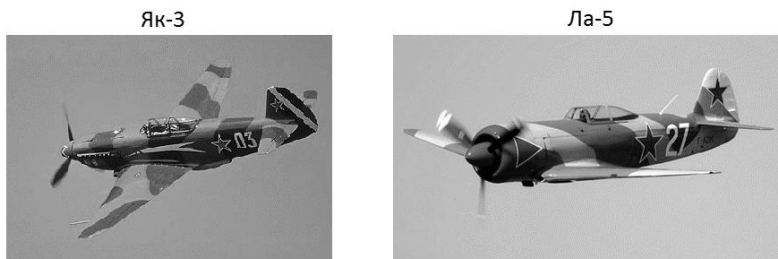


Рис.13. Советские истребители 1940-х годов

Итак, оба самолета в принципе одинаковы, но имеют очевидные технические особенности. Як-3 оснащен рядным поршневым двигателем жидкостного охлаждения, а Ла-5 имеет звездообразный ПД воздушного охлаждения. Силовые элементы конструкции крыла Як-3 были металлическими, а у Ла-5 даже лонжероны были сделаны из так называемой дельта-древесины.

Из представленных примеров следует очевидный вывод. Для достижения поставленной цели проектант свободен в выборе концепции будущего изделия, включающей различные идеи. Для каждой идеи проектант свободен в поиске адекватных принципов. И этим дело не ограничивается, потому что для материализации идей и принципов нужен выбор так называемых технологий.

При выполнении концептуальных исследований технологиями стали именоваться научные разработки, предназначенные

для практического применения. Такое понимание вошло в обиход проектной деятельности вместе со шкалой «уровней готовности технологии» [12], которая оценивает степень приближения разрабатываемого проекта к поставленной цели. Стоит напомнить, что цель проекта (1) нормирует свойства будущего натурального объекта. В связи с этим его экосистема должна быть показана в естественных условиях применения. Тогда для верификации модели (3) объекта и ее элементов по мере их разработки должны производиться ее материальные аналогии со своими характеристиками вида (2). Аналогии разной полноты могут обеспечить физические модели, функциональные демонстраторы и, наконец, полномасштабные изделия.

Для их создания необходимы производственные технологии [11], каждая из которых является совокупностью взаимодействующих операций, оборудования и условий для передела сырья в готовое изделие. Здесь «технология» – это индуктивная система (лат. «induction» – наведение), объединяющая в своем составе *понимание* (гр. – «логос») особенностей аналогии с моделью объекта (3) и *умение* (гр. – «техно») в располагаемых материальных условиях преобразовать ее в предметы/явления действительности.

Следовательно, технология предусматривает осуществление двух процессов, из которых один обуславливает второй. Первый направлен на подготовку проекта аналогии, сохраняющей особенности модели объекта (3). На практике условия аналогии и ее реализуемости соблюдаются благодаря участию автора-проектанта, конструктора, технолога и пользователя проекта.

Связанный с первым второй процесс направлен на практическое осуществление, например, конструкторского проекта аналога (рис. 14) в действующих условиях производства. С этой целью разрабатываются производственные технологии, в соответствии с которыми из подходящих материалов изготавливается конечный продукт, а на более ранних стадиях разработки – материальные аналогии информационной модели объекта или ее компонентов, которые делают возможной периодическую верификацию цели проекта.

Конечно, когда говорят о производственной технологии, то не обязательно имеют в виду преобразование научного знания и материальных ресурсов исключительно в физический объект (рис. 13). Точно так же реализуются энергетические, информационные и иные технологии, приносящие иные продукты производства, пригодные для практического применения.

Из приведенных рассуждений ясно, что участие многих субъектов и материальных факторов в разработке и осуществлении технологического проекта вносит значительное разнообразие и множество наблюдаемых отличий в практическое исполнение одинаковых идей, принципов, концепций. С этим приходится считаться даже на такой ранней стадии жизненного цикла, как концептуальный проект.

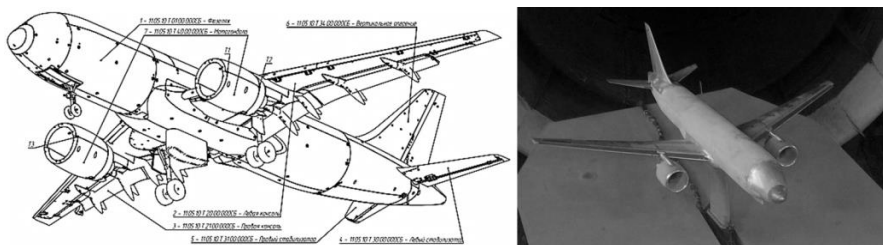


Рис. 14. Проект конструкции и изготовленная модель (аналог) самолёта для аэродинамических испытаний

Технология – сравнительно новое и многоликое понятие, которое используется разными участниками жизненного цикла (рис. 1) со своими представлениями о его смысле и значении. В современных условиях единой и однозначной парадигмы технологии не существует. В данном исследовании остается открытым вопрос о связи технологий, характерных для концептуальной, демонстрационной, производственной и т.д. стадий (рис. 1), о соотношении смыслов, которые вкладываются в этот термин различными специалистами, о необходимости внесения некоторых уточнений в его определение при выполнении обобщающего исследования.

Однако общим для технологий любой стадии разработки является то, что в результате их реализации происходит вопло-

щение ментальных идей в материальной среде. Именно посредством технологий мнимые идеи (принципы, концепции) становятся действительными и создаются практические возможности достижения цели проекта.

Итак, упомянутые выше цилиндрические ролики в подшипнике качения, ромбовидное крыло малозаметного самолета и, наконец, деревянные лонжероны крыла имеют технологические особенности. С одной стороны, они осуществляют породившие их принципы, являясь проводниками идей, а с другой – обеспечивают соответствие принципов тем условиям, которые предусмотрены целью проекта. Различные условия осуществления одной и той же идеи приводят к появлению разных технологий. При этом идея остается их общим прообразом.

7. Неопределённости и риски

Идеи и принципы познаются человеческим сознанием как интуитивно (мнения или гипотезы), так и конкретно-логическим путем через изучение явлений природы, выявление, сравнение и упорядочение фактов.

Среди идей, составляющих (6), можно выделить множества двух типов: множество реальных идей – фактов $Re(\lambda)$ и множество мнимых идей – гипотез $Im(\lambda)$. Вместе с названными идеями два типа принципов (реальных и мнимых) могут составлять комплексную модель концепции. Концептуальное проектирование на основе комплексной информации с мнимой составляющей из-за недостатка достоверной информации сопровождается множеством неопределённостей, которые должны быть исключены на пути к созданию реального объекта.

Последовательное устранение неопределенностей в содержании концепции составляет одну из основных проблем концептуального проектирования. Сам метод проектирования должен гарантировать при этом получение достоверного решения.

В начальный момент исследования область поиска новой концепции безгранична и неопределенна, а вероятность отыскания правильного решения ничтожна мала. Для принятия решения о запуске проекта необходима информация о его эффективности и технической реализуемости.

Для того чтобы получить такую информацию, необходимо выполнить несколько объективно необходимых этапов концептуального исследования (рис. 15). При этом каждый следующий этап осуществим при условии, что нужная для него исходная информация получена на предыдущем этапе.



Рис. 15. Процесс концептуального проектирования

В начале (Этап 0) создаётся информационное обеспечение - база данных из множества способов и устройств, которые могут быть полезны для решения новой задачи (рис. 16). Этот поиск сопровождается предварительной оценкой чувствительности системы характеристик будущего объекта к разнообразным технологиям (из числа известных), чтобы отказаться от очевидно бесполезных.

Применение апробированной, реалистичной информации способствует успеху концептуального проектирования. С начала проектирования в этом состоит особая роль известных по опыту практических технологий из множества $Re(\lambda)$. Их изучение составляет часть Этапа 0 и предвещает Этап 1.

Результатом начальных усилий становится множество вида $\{(x, g, u) \mid x \in M_x \wedge g \in M_g \wedge u \in M_u\}$, которое моделирует технологический базис для решения задач проекта, а также множество вида $\{(r, p, e) \mid r \in M_r \wedge p \in M_p \wedge e \in M_e\}$, модели-

рующее множество реакций внешней среды на применение технологий. Таким образом, устраняется *неопределённость* источника информации для проектных решений.

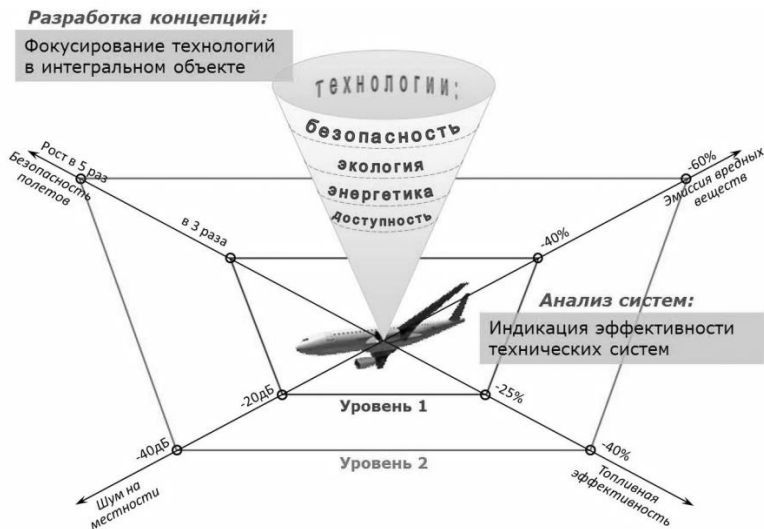


Рис. 16. Связь технологий с задачами развития
воздушного транспорта

К рискам, характерным для этого этапа, следует отнести неполноту информационного обеспечения, разрозненность фактических данных, возникновение мнимой информации в результате творческих усилий по решению уникальных задач проекта. Методом исключения части рисков и ограничения их влияния на результаты дальнейшего проектирования является разработка и использование правил, обобщающих известные факты, превращая общую для них закономерность в универсальную рекомендацию для разработки других проектных решений с аналогичными полезными свойствами, но в ином техническом исполнении. Это обеспечивает бóльшую свободу выбора проектных решений и позволяет избежать их противоречивости в составе единой концепции проекта.

На рис. 17 в систематизированном виде для примера представлены правила проектирования энергетических средств, которые включают несколько апробированных способов увеличения подъемной силы, в том числе увеличение эффективности аэродинамической механизации крыла при ее обдуве, создание дополнительной аэродинамической суперциркуляции крыла, отклонение вектора тяги с созданием ее вертикальной проекции. Точно также рекомендации, представленные на рис.11, следует понимать как правила проектирования малозаметных самолетов. Таким образом, правило – это апробированная идея $Re(\lambda)$.

Правила проектирования энергетических средств	Форма объекта	Способ увеличения подъемной силы	Устройство реактивной струи
обдув крыла с механизацией		реакция закрылка	винт
обдув крыла снизу		реакция закрылка	ТРДД
обдув крыла сверху		эффект Коанда	ТРДД
выдув на закрылки		суперциркуляция	ВСУ
эжекторный закрылок		суперциркуляция	эжектор
выдув на круглую кромку		суперциркуляция	ВСУ
реактивный закрылок		прямая реакция	ВСУ
отклонение вектора тяги		прямая реакция	ТРДД

Рис. 17. Опыт разработки энергетических технологий увеличения подъемной силы самолета

«Правило» рассматривается как устоявшаяся норма, предписание или частная рекомендация для проектирования объектов и их элементов. Разработка и освоение правил является началом самообучения проектанта на разрозненных фактах. Обычно этой цели в прикладных исследованиях служат учебные пособия, совместная работа со специалистами, а также специальные «Руководства для конструкторов».

На Этапе 1 анализируются различные комбинации из предложенных принципов (технологий), которые позволяют сформировать ряд возможных концепций изделия в целом (рис. 18). Характеристики вариантов (в данном случае самолётов) оцениваются и сравниваются, из них выбираются наиболее эффективные.

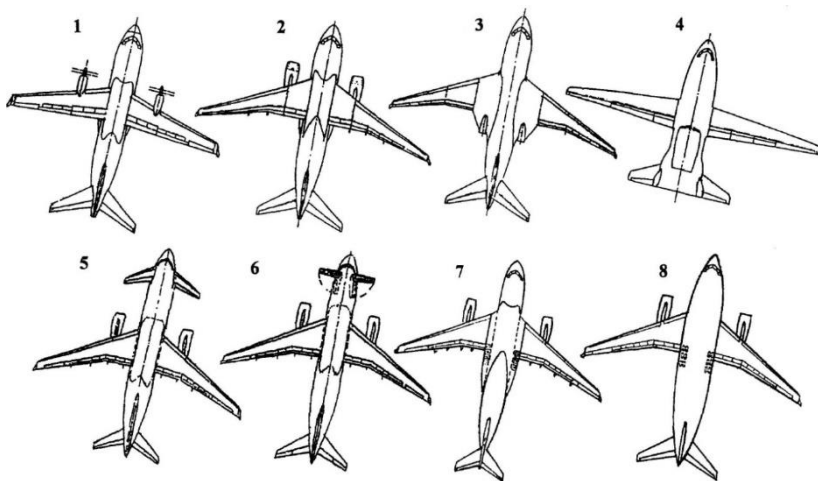


Рис. 18 Альтернативные варианты транспортного самолёта

Основным техническим риском данного этапа является выбор неправильной концепции. Ключевой причиной данного риска является построение недостаточного числа вариантов изделия и неправильный метод их формирования. Поскольку каждый из вариантов концепции (а это проектные эксперименты) является неповторяющимся соединением состояний в ряду возможных, то указанный выше риск может быть снижен с помощью полно-

факторного плана эксперимента 2^n для построения Парето-оптимальной траектории поиска наилучшей концепции, который обеспечивает 100%-ю информированность проектанта [15].

В результате Этапа 1 преодолевается принципиальная неопределённость расчетной концепции (5), которая в дальнейшем становится предметом глубокого изучения. При этом, хотя состав концепции и определён, остаются другие риски.

Так, возможное наличие нескольких принципов (технологий) в области определения одной идеи (например, цилиндра, сферы или шагающих опор для идеи качения) порождает некоторые неопределенности при формировании технической концепции, а также в системных оценках. Для их устранения существуют специальные методы. Однако по завершении этапа еще остается риск внутренней несогласованности между элементами концепции объекта (6), перечень которых был определен как относительно лучшее сочетание принципов в сравнении с другими вариантами. Но для совершенства концепции требуется еще и оптимальный баланс между ее разнородными элементами, тогда образуется и однозначная система характеристик объекта (2).

На Этапе 2 элементы концепции (принципы, идеи) согласуются между собой так, чтобы максимизировать эффективность в составе экзосистемы S_z . Основные технические риски данного этапа связаны с неопределенностью в отдельных характеристиках, что исключает возможность достоверной верификации концептуального проекта. Отдельные элементы концепции оказывают различное влияние на ее эффективность. Те из них, которые являются наименее изученными, а также те, к изменениям которых целевая функция особенно чувствительна, составляют перечень ключевых элементов выбранной концепции (рис. 19).

Характеристики элементов системы можно считать случайными величинами с заданными законами распределения (определяются экспертно), поэтому оценка степени влияния на интегральные критерии сводится к определению вероятностных характеристик выходных величин, характеризующих систему. Для решения этой задачи часто используются статистические методы [3].



Рис. 19. Ключевые элементы в расчётной концепции самолета

Примером может служить статистическая оценка дальности полета с помощью метода Монте-Карло. В рассматриваемом примере определяется функция нескольких случайных величин, среди которых: c_{x0} – коэффициент аэродинамического сопротивления при нулевой подъемной силе; A – коэффициент отвала крейсерской поляры; G_K – вес конструкции; $G_{об}$ – вес оборудования; $G_{с.у.}$ – вес силовой установки. Пример численной реализации заданных законов « $\alpha \times \beta$ » распределения входных параметров приведен на рис. 20.

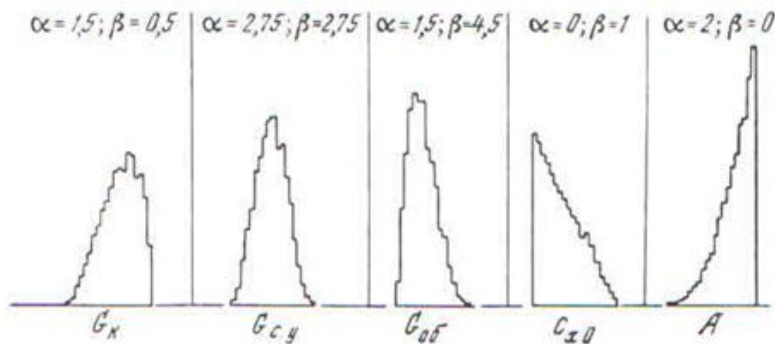


Рис. 20. Численная реализация законов распределения
входных параметров

На рис. 21 приведена гистограмма распределения дальности полета для числа испытаний $N = 9000$ с указанием математического ожидания (M) и среднеквадратического отклонения (σ).

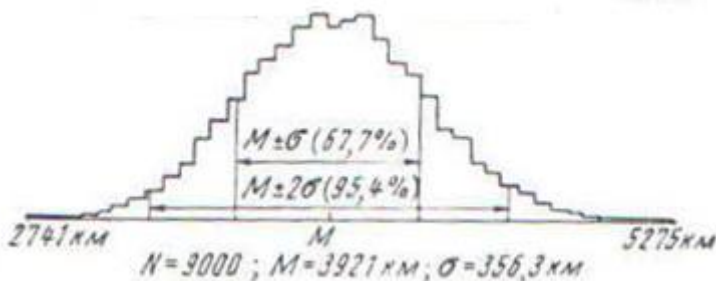


Рис. 21. Гистограмма распределения дальности полета

Экспертная неопределённость в отдельных характеристиках объекта создаёт для проектанта неопределённость его целевой функции и исключает возможность достоверной верификации концептуального проекта. Устранить этот риск расчётными исследованиями в большинстве случаев невозможно, поскольку они не избавляют от мнимого характера экспертных оценок. Необходима валидация критических характеристик. Таким образом, дальнейшие исследования должны быть направлены на уменьшение рисков обеспечения технической реализуемости и эффективности разрабатываемой концепции и ее технологий.

В связи с этим Этап 3 посвящается экспериментальным исследованиям (рис. 22) ключевых элементов (физическим, стендовым, имитационным), которые в лабораторных условиях дают первые доказательства их технической реализуемости, наличия эксплуатационных преимуществ [8] и позволяют предварительно установить соответствие характеристик требованиям к изделию. Техническими рисками данного этапа является риск несоответствия разработанной концепции критериям задания (1) и риски неполной аналогии расчетной концепции натуре. Создаются условия для перехода к демонстрационной стадии проекта, чтобы снизить указанные риски.



Рис. 22. Валидация технологий. Экспериментальные исследования модели самолета в аэродинамической трубе

Дело в том, что лабораторные условия для испытания множества моделей накладывают ограничения на их размеры, следуют различным критериям подобия [7] и выполняются на различных экспериментальных установках. Их единство в проекте обеспечивается универсальной моделью объекта, как правило, геометрической. Продемонстрировать физическое единство и согласованную эффективность может модель прототипа, действующая в натуральных условиях. При создании сложных технических объектов, таких как пилотируемый летательный аппарат, задача создания и испытаний демонстратора технологий выходит за пределы компетенций научных организаций и требует привлечения в проект конструкторской организации (ОКБ промышленности). В условиях демонстрационной стадии представляется возможность решить проблему трансфера научных разработок (технологий) по завершении концептуального проекта.

Таким образом, в процессе концептуального проектирования каждому этапу проектирования (рис. 15) присущи свои неопределённости и риски, связанные с уникальностью объекта. Для успешной разработки концептуального проекта необходимо их выявить, оценить и устранить. В настоящее время существуют универсальные методики качественного анализа и оценки рисков. Методы количественной оценки рисков предстоит создать для каждого этапа концептуального проектирования.

Анализ рисков предназначен не только для выбора концепции и разработки надёжных технологий, но и для выработки других мер по управлению проектом (финансовых, организационных, временных), которые позволят снизить вероятность возникновения негативной ситуации или минимизировать её последствия.

Управление рисками в целом является неотъемлемым атрибутом концептуального проектирования.

Заключение

В работе представлены основные аспекты методологии концептуального проектирования. Сформулировано понятие концепции в терминах теории множеств, теории поля и общей теории систем. В общем понятии системы выделено два частных: эндосистема (модель объекта) и экзосистема (модель связей объекта со средой обитания). Соответствующее понятие экзосистемы расширено включением фактора «нагрузка» от внешней среды, который дополняет известные факторы «эффект-затраты». Концепция определена как сочетание нескольких принципов. Принцип определен как свертка информации о компонентах эндосистемы, направление которой задано единичным вектором-идеей. В понятия концепции и эндосистемы введён принцип «формы», который вместе с принципами «действия» и «устройства» объекта, образует его трёхмерную модель.

Процесс концептуального проектирования представлен как последовательность объективно необходимых этапов, направленных на устранение неопределённостей, выявленных проектантом. Происхождение неопределённостей связано с ком-

плексным характером информации, насыщающей проект не только реальными, но и мнимыми данными, исходящими от проектанта. Это следствие уникальности проекта, требующей творческого воображения для решения новой задачи. С иллюстрациями из области авиации дан краткий обзор особенностей каждого этапа концептуального проектирования. Отмечено, что управление рисками неотделимо от управления проектом.

Предложенная теория позволяет оперировать элементами концепции создаваемого изделия, опираясь на законы математической логики. Формализованные проектные решения получают строгое обоснование благодаря интерфейсам (рис. 23), установленным между известными понятиями и различными теориями.



Рис. 23. Основные модели этапов концептуального проектирования (отмечены в скобках своими номерами)

Естественно, представленные материалы не исчерпывают всех знаний, которые нужны для решения задач концептуального проектирования во всей их сложности. В частности, предстоит показать идентичность предложенного описания этапов с описанием известных требований стандартов к другим этапам жизненного цикла и принятой системы оценки уровней готовности технологий [12], составить математические модели для описания этих этапов.

В данной работе смысл комплексного представления концепции и системы с мнимыми составляющими в их содержании

состоит в том, чтобы объяснить причины неопределенностей и рисков, которые сопровождают проект. Возможно, в дальнейшем будет построена общая математическая модель процесса концептуального проектирования, которая позволит управлять проектом с использованием практических методов теории функций комплексного переменного.

Литература

1. ВОЖДАЕВ В.В., ЛАЗАРЕВ В.В. *Конструктивно-компоновочные способы снижения радиолокационной заметности самолетов* // *Техника воздушного флота*. – 2011. – Т. LXXXV, №2(703). – С. 41–50.
2. ГОЛУБЕВ И.С., САКАЧ Р.В., ЛОГИНОВ Е.Л., ПИНАЕВ Е.Г. *Исследование операций в гражданской авиации*. – М: Транспорт, – 1980.
3. ДЕНИСОВ В.Е., ИСАЕВ В.К., РЯБОВ А.М., ШКАДОВ Л.И. *Статистическая оценка характеристик проектируемого самолета с помощью метода Монте-Карло* // *Ученые записки ЦАГИ*. – 1973. – Т. IV, №2. – С. 14–19.
4. КОСЯКОВ А., СВИТ У.Н., СЕЙМУР С.ДЖ., БИМЕР С.М. *Системная инженерия. Принципы и практика*. – М: ДМК, – 2014.
5. МЕСАРОВИЧ М., МАКО Д., ТАКАХАРА И. *Теория иерархических многоуровневых систем*. – М: Мир, – 1973.
6. МОЛОДШИЙ В.Н. *Основы учения о числе в XVIII веке*. – М.: Учпедгиз, – 1953.
7. СЕДОВ Л.И., *Методы подобия и размерности в механике*. – М.: Главная редакция физико-математической литературы издательства «Наука», – 1977.
8. СКВОРЦОВ Е.Б., ШЕЛЕХОВА А.С. *Верификация и валидация технологий в концептуальном проектировании авиационной техники* // XII Международная научно-практическая конференция «Перспективы развития науки и образования», Москва, 2016 – С. 143–149.

9. СКВОРЦОВ Е.Б., ШЕЛЕХОВА А.С. *Концептуальное проектирование и системная интеграция технологий* // XII Международная научно-практическая конференция «Перспективы развития науки и образования», Москва, 2016 – С. 150–158.
10. СТЕЛЛ Р.Р. *Множества. Логика. Аксиоматические теории*. – М: Просвещение, 1968.
11. *Технология самолетостроения* / Под. ред. А.Л. Абибова. – М.: Машиностроение, 1970.
12. *Трансфер технологий. Общие положения*. ГОСТ Р 57194.1-2016.
13. ХИЛЛСОН Д. *Управление рисками, пирамида Маслоу и меметика* // Управление проектами. – 2017, – №1(40). – С. 5–19.
14. ЧУЕВ Ю.В. *Исследование операций в военном деле*. – М: Воениздат, 1970.
15. SKVORTSOV E.B. *Direct Search in Conceptual Design* // Acta Politechnica Journal of Advanced Engineering, Prague: Czech Technical University (CTU). – 2000. – Vol. 40, No. 1. – P. 24–29.

BEGINNING OF THEORY OF CONCEPTUAL DESIGN WITH APPLICATIONS IN AERONAUTICAL SCIENCE AND TECHNOLOGY

Eugenii Skvortsov, Central Aerohydrodynamic Institute, Zhukovsky, Deputy Head of Department.

Anna Shelekhova, Central Aerohydrodynamic Institute, Zhukovsky, Head of Sector.

Abstract: The work presents key concepts of conceptual design, that are formulated in terms of set theory, vector analysis and general systems theory. The conception is understood as the combination of several principles that form the information-logical model of an artificial object. The principles are intentional convolutions of related information, the vectors of which are defined by ideas, and the content is defined by the components of the system. The concepts of the exosystem (the model of interaction with the environment) and the endosystem (the model of an separate object) are highlighted. Unity of the model is created by an active functional system, which in a documented form creates a project. The diversity of principles is expanded by the introduction of the principle of "form" along with the principles of «ac-

tion» and «structure» of the object. The realizability of the object is provided by the technologies. The process of conceptual design is presented as a sequential activity aimed at eliminating uncertainties, which are caused by the complex nature of project information with an imaginary component. The proposed theory allows to manage the conceptual project, relying on the laws of mathematical logic. Formalized through this design decisions are strictly justified.

Keywords: conception, principle, idea, system, technology, project, risk.

УДК 167/168

ББК 30.2

*Статья представлена к публикации
членом редакционной коллегии В.В. Клочковым.*

Поступила в редакцию 26.09.2017.

Опубликована 30.09.2018.