

ВЕСТНИК ТОМСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА

УПРАВЛЕНИЕ,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И ИНФОРМАТИКА

TOMSK STATE UNIVERSITY
JOURNAL OF CONTROL AND COMPUTER SCIENCE

Научный журнал

2016

№ 1 (34)

Свидетельство о регистрации: ПИ № ФС 77-29497 от 27 сентября 2007 г.



EDITORIAL BOARD

Alexander Gortsev – Editor-in-Chief, Doctor of Science, Prof., Head of the Operation Research Department, Dean of the Faculty of Applied Mathematics and Cybernetics.
Tel: +72822529599

Valery Smagin – Deputy Editor-in-Chief, Doctor of Science, Prof. of the Operation Research Department.
Tel: +72822529599

Lyudmila Nezhelskaya – Executive Editor, Cand. of Science, Associate Prof. of the Operation Research Department. E-mail: vestnik_uvti@mail.tsu.ru

Sergey Vorobeychikov – Doctor of Science, Prof. of the Higher Mathematics and Mathematical Modeling Department

Gennady Koshkin – Doctor of Science, Prof. of the Theoretical Cybernetics Department

Yury Kostyuk – Doctor of Science, Prof. of the Theoretical Informatics Department

Anjela Matrosova – Doctor of Science, Prof., Head of the Programming Department

Anatoly Nazarov – Doctor of Science, Prof., Head of the Probability Theory and Mathematical Statistics Department

Sergey Sushchenko – Doctor of Science, Prof., Head of the Applied of Information Department, Dean of the Faculty of Informatics

EDITORIAL COUNCIL

Ana Rosa Cavalli
PhD, Prof.
University VII
Paris, France

Vladimir Dombrovskii
Doctor of Science, Prof.
Tomsk State University
Russia

Alexander Dudin
Doctor of Science, Prof.
Belarusian State University
Minsk, Republic Belorussia

Enco Orzinger
PhD, Prof.
University of Rome
Italy

Paolo Prinetto
Prof.
Politechnic Institute
Torino, Italy

Gilbert Saporta
PhD, Prof.
Pierre and Marie Curie
University, Paris, France

Raimund Ubar
Doctor of Science, Prof.
University of Technology
Tallinn, Estonia

Nina Yevtushenko
Doctor of Science, Prof.
Tomsk State University
Russia

Yervant Zorian
PhD, Fellow & Chief Architect, Synopsys, Mountain View, CA, USA

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Горцев Александр Михайлович – гл. редактор, проф., д-р техн. наук, зав. каф. исследования операций, декан факультета прикладной математики и кибернетики ТГУ.
Тел. +72822529599

Смагин Валерий Иванович – зам. гл. редактора, проф., д-р техн. наук, проф. каф. исследования операций ТГУ.
Тел. +72822529599

Нежелская Людмила Алексеевна – отв. секретарь, доц., канд. техн. наук, доц. каф. исследования операций ТГУ.
E-mail: vestnik_uvti@mail.tsu.ru

Воробейчиков Сергей Эрикович – проф., д-р физ.-мат. наук, проф. каф. высшей математики и математического моделирования ТГУ

Коскин Геннадий Михайлович – проф., д-р физ.-мат. наук, проф. каф. теоретической кибернетики ТГУ

Костюк Юрий Леонидович – проф., д-р техн. наук, проф. каф. теоретической информатики ТГУ

Матросова Анжела Юрьевна – проф., д-р техн. наук, зав. каф. программирования ТГУ

Назаров Анатолий Андреевич – проф., д-р техн. наук, зав. каф. теории вероятностей и математической статистики ТГУ

Сущенко Сергей Петрович – проф., д-р техн. наук, зав. каф. прикладной информатики, декан факультета информатики ТГУ

РЕДАКЦИОННЫЙ СОВЕТ

Ана Роза Ковали
д-р философии, проф.
Университет VII
Париж, Франция

Владимир Домбровский
д-р техн. наук, проф.
ТГУ, Томск, Россия

Александр Дудин
д-р физ.-мат. наук, проф.
БГУ, Минск,
Республика Беларусь

Енцо Орзингер
д-р философии, проф.
Римский университет
Италия

Паоло Принетто
проф.
Политехнический институт
Турин, Италия

Жильберт Сапорта
д-р философии, проф.
Университет им. Пьера и Марии
Кюри, Париж, Франция

Раймонд Убар
д-р техн. наук, проф.
Технологический университет,
Таллинн, Эстония

Нина Евтушенко
д-р техн. наук, проф.
ТГУ, Томск, Россия

Ервант Зориан
д-р философии,
гл. науч. сотр. фирмы
«Синописис», США

JORNAL INFO

Tomsk State University Journal of Control and Computer Science is an independent peer-reviewed research journal that welcomes submissions from across the world.

Tomsk State University Journal of Control and Computer Science is issued four times per year, and can be subscribed to in the Russian Press Joint Catalogue (Subscription Index 44031).

The publication in the journal is free of charge and may be in Russian or in English.

The topics of the journal are the following:

- *control of dynamical systems,*
- *mathematical modeling,*
- *data processing,*
- *informatics and programming,*
- *discrete function and automation,*
- *designing and diagnostics of computer systems.*

Rules of registration articles are given in a site:

<http://journals.tsu.ru/informatics/>
ISSN 2311-2085 (Online), ISSN 1998-8605 (Print).

О ЖУРНАЛЕ

Журнал «Вестник Томского государственного университета. Управление, вычислительная техника и информатика» выходит ежеквартально и распространяется по подписке, его подписной индекс 44031 в объединённом каталоге «Пресса России». Статьи публикуются на русском и английском языках.

Тематика публикаций журнала:

- *управление динамическими системами,*
- *математическое моделирование,*
- *обработка информации,*
- *информатика и программирование,*
- *дискретные функции и автоматы,*
- *проектирование и диагностика вычислительных систем.*

Журнал «Вестник Томского государственного университета. Управление, вычислительная техника и информатика» входит в «Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук» Высшей аттестационной комиссии, а также входит в систему Российского Индекса Научного Цитирования (РИНЦ). Правила оформления статей приведены на сайте: <http://journals.tsu.ru/informatics/>

СОДЕРЖАНИЕ

УПРАВЛЕНИЕ ДИНАМИЧЕСКИМИ СИСТЕМАМИ

Домбровский В.В., Самородова М.В. Управление с прогнозированием по квадратичному критерию линейными дискретными системами с марковскими скачками при ограничениях	4
Ким К.С., Смагин В.И. Локально-оптимальное управление дискретными системами с запаздыванием в канале управления при неполной информации о состоянии и возмущениях	11

ОБРАБОТКА ИНФОРМАЦИИ

Бахолдина М.А. Вероятность ошибки при оценивании состояний модулированного обобщенного полусинхронного потока событий	18
Вожов С.С., Чимитова Е.В. Проверка гипотезы о виде распределения по интервальным данным	35
Назаров А.А., Бронер В.И. Система управления запасами с гиперэкспоненциальным распределением объемов потребления ресурсов	43
Сиротина М.Н., Горцев А.М. Оценка максимального правдоподобия длительности мертвого времени в модулированном синхронном дважды стохастическом потоке событий	50
Цициашвили Г.Ш. Синергетический эффект в сети с гиперэкспоненциальными распределениями времен обслуживания	65

ИНФОРМАТИКА И ПРОГРАММИРОВАНИЕ

Бабанов А.М., Квач Е.С. «IS-THE»-отношения в семантических моделях данных: основные понятия и разновидности	69
Царева Е.С. Разрешение конфликтных ситуаций при синхронизации многопользовательских online-приложений	79
Сведения об авторах	92

CONTENTS

CONTROL OF DYNAMICAL SYSTEMS

Dombrovskii V.V., Samorodova M.V. Model predictive control with quadratic criterion for jump Markov discrete linear systems under constraints	4
Kim K.S., Smagin V.I. Locally-optimal control of discrete delayed control systems with incomplete information about state and perturbations	11

DATA PROCESSING

Bakholdina M.A. Error probability when estimating the states of the modulated generalized semi-synchronous flow of events	18
Vozhov S.S., Chimitova E.V. Testing goodness-of-fit with interval data	35
Nazarov A.A., Broner V.I. Inventory model with hyperexponential distribution of demand's batch size	43
Sirotnina M.N., Gortzev A.M. Maximum likelihood estimator of dead time duration in modulated synchronous twice stochastic flow of events	50
Tsitsiashvili G.Sh. Synergetic effect in network with hyper exponential distributions of service times	65

INFORMATICS AND PROGRAMMING

Babanov A.M., Kvach E.S. «IS-THE»-relations in semantic data models: basic notions and variants	69
Tsareva E.V. Resolution of conflicts during synchronization multiplayer online-application	79
Brief information about the authors	92

УПРАВЛЕНИЕ ДИНАМИЧЕСКИМИ СИСТЕМАМИ

УДК 519.865.5

DOI: 10.17223/19988605/34/1

В.В. Домбровский, М.В. Самородова

УПРАВЛЕНИЕ С ПРОГНОЗИРОВАНИЕМ ПО КВАДРАТИЧНОМУ КРИТЕРИЮ ЛИНЕЙНЫМИ ДИСКРЕТНЫМИ СИСТЕМАМИ С МАРКОВСКИМИ СКАЧКАМИ ПРИ ОГРАНИЧЕНИЯХ

Рассматривается задача управления с прогнозированием по квадратичному критерию для линейных дискретных систем со скачкообразно меняющимися параметрами. Синтезированы стратегии управления при наличии явных ограничений на управляющие воздействия. Алгоритм синтеза прогнозирующей стратегии сводится к решению последовательности задач квадратичного программирования.

Ключевые слова: линейные системы с марковскими скачками; прогнозирующее управление; марковские скачки; ограничения.

Моделями с марковскими скачкообразными параметрами описывается широкий класс реальных систем [1]. В этих моделях предполагается, что смена структуры системы осуществляется в соответствии с эволюцией марковской цепи с конечным пространством состояний. Решению различных задач управления и оценивания для таких систем посвящено значительное количество работ [2–13].

Эффективным подходом к синтезу систем управления с ограничениями, получившим широкое признание и применение в практике управления сложными технологическими процессами, является метод управления с прогнозирующей моделью (управление со скользящим горизонтом) [14, 15]. Применению данного метода к управлению дискретными системами с марковскими скачками посвящены работы [3, 5, 11–13]. В работах [3, 11–13] рассматривается задача управления по квадратичному критерию дискретными системами при условии, что от состояния марковской цепи зависит только матрица управления системы при «жестких» ограничениях на управляющие переменные.

В настоящей работе рассматривается более общий случай, когда от состояния цепи зависит не только матрица управления, но и матрица динамики системы. Получены уравнения синтеза оптимальных стратегий управления с учетом «жестких» ограничений на управляющие переменные.

1. Постановка задачи

Пусть объект управления описывается уравнением

$$x(k+1) = A[\alpha(k+1)]x(k) + B[\alpha(k+1)]u(k), \quad (1)$$

где $x(k) \in \mathbb{R}^{n_x}$ – вектор состояния, $u(k) \in \mathbb{R}^{n_u}$ – вектор управления, $\alpha(k)$ ($k = 0, 1, 2, \dots, v$) – однородная дискретная марковская цепь с конечным множеством состояний $\{1, 2, \dots, v\}$, известной матрицей переходных вероятностей

$$P = [P_{i,j}] (i, j \in \{1, 2, \dots, v\}), \quad P_{j,i} = P\{\alpha(k+1) = \alpha_j | \alpha(k) = \alpha_i\}, \quad \sum_{j=1}^v P_{j,i} = 1,$$

и известным начальным распределением

$$p_i = P\{\alpha(0) = i\} (i = \overline{1, v}), \quad \sum_{i=1}^v p_i = 1.$$

Матрицы динамики $A[\alpha(k)]$ и управления $B[\alpha(k)]$ выбираются в соответствии с состоянием α_i марковской цепи $\alpha(k)$ из множеств $A = \{A^i \in \mathbb{R}^{n_x \times n_x} : i = \overline{1, v}\}$ и $B = \{B^i \in \mathbb{R}^{n_x \times n_u} : i = \overline{1, v}\}$ соответственно. Предполагается, что состояние марковской цепи в момент времени k доступно наблюдению.

На управляющие воздействия наложены ограничения:

$$u_{\min}(k) \leq S(k)u(k) \leq u_{\max}(k), \quad (2)$$

где $S(k) \in \mathbb{R}^{p \times n_u}$, $u_{\min}(k), u_{\max}(k) \in \mathbb{R}^p$.

На каждом шаге k будем определять закон управления системой (1) при ограничениях (2) из условия минимума критерия со скользящим горизонтом управления:

$$J(k+m/k) = E \left\{ \sum_{i=1}^m x^T(k+i) R_1 x(k+i) + u^T(k+i-1/k) R u(k+i-1/k) / x(k), \alpha(k) = \alpha_j \right\}, \quad (3)$$

где $E\{\dots/\dots\}$ – оператор условного математического ожидания; m – горизонт прогноза, $R_1 \geq 0$, $R > 0$ – весовые матрицы соответствующих размерностей.

2. Синтез стратегий прогнозирующего управления

Для решения сформулированной задачи используем методологию управления с прогнозирующей моделью. Данный подход позволяет получить стратегии управления с обратной связью с учетом явных ограничений на управляющие воздействия.

Стратегии управления с прогнозированием определяются по следующему правилу. На каждом шаге k минимизируем функционал (3) по последовательности прогнозирующих управлений $u(k/k), \dots, u(k+m-1/k)$, зависящих от состояния системы в момент времени k . В качестве управления в момент времени k берем $u(k) = u(k/k)$. Тем самым получаем управление $u(k)$ как функцию состояний $x(k)$ и $\alpha(k) = \alpha_j$, т.е. управление с обратной связью. Чтобы получить управление $u(k+1)$ на следующем шаге, процедура повторяется для следующего момента $k+1$ и т.д.

Дискретная марковская цепь с конечным множеством состояний $\{1, 2, \dots, v\}$ и матрицей переходных вероятностей P допускает следующее представление в пространстве состояний [9]:

$$\theta(k+1) = P\theta(k) + v(k+1), \quad (4)$$

где $\theta(k) = [\delta(\alpha(k), 1), \dots, \delta(\alpha(k), v)]^T$, $\delta(\alpha(k), j)$ – функция Кронекера; $\{v(k)\}$ – последовательность мартингал-разностей с условными моментами:

$$E\{v(k+1)/\theta(k)\} = 0, \quad (5)$$

$$E\{v(k+1)v^T(k+1)/\theta(k)\} = \text{diag}\{P\theta(k)\} - P\text{diag}\{\theta(k)\}P^T.$$

С учетом (4) систему (1) можно представить в следующем виде:

$$x(k+1) = A[\theta(k+1)]x(k) + B[\theta(k+1)]u(k), \quad (6)$$

где матрица динамики $A[\theta(k+1)]$ и матрица управления $B[\theta(k+1)]$ имеют вид

$$A[\theta(k+1)] = \sum_{i=1}^v \theta_i(k+1) A^i, \quad B[\theta(k+1)] = \sum_{i=1}^v \theta_i(k+1) B^i, \quad (7)$$

здесь $\theta_i(k+1)$ ($i = 1, 2, \dots, v$) – компоненты вектора $\theta(k+1)$.

Теорема 1. Вектор прогнозирующих управлений $U(k) = [u^T(k/k), \dots, u^T(k+m-1/k)]^T$, минимизирующий критерий (3) при ограничениях вида (2), на каждом шаге k определяется из решения задачи квадратичного программирования с критерием вида

$$Y(k+m/k) = 2x^T(k)G(k)U(k) + U^T(k)H(k)U(k),$$

при ограничениях

$$U_{\min}(k) \leq \bar{S}(k)U(k) \leq U_{\max}(k). \quad (8)$$

Оптимальное управление равно

$$u(k) = \begin{bmatrix} I_{n_u} & 0_{n_u} & \dots & 0_{n_u} \end{bmatrix} U(k),$$

где

$$\bar{S}(k) = \text{diag}(S(k), \dots, S(k+m-1)),$$

$$U_{\min}(k) = \left[u_{\min}^T(k), \dots, u_{\min}^T(k+m-1) \right]^T,$$

$$U_{\max}(k) = \left[u_{\max}^T(k), \dots, u_{\max}^T(k+m-1) \right]^T,$$

I_{n_u} – единичная матрица размерности n_u , 0_{n_u} – квадратная нулевая матрица размерности n_u , $H(k)$ и $G(k)$ – блочные матрицы, блоки которых равны:

$$H_{t,t}(k) = \sum_{i_t=1}^v \left(B^{i_t} \right)^T Q^{(i_t)}(k) B^{i_t} + R, \quad t = \overline{1, m}, \quad (9)$$

$$H_{t,s}(k) = \sum_{i_s=1}^v \dots \sum_{i_{t+1}=1}^v \sum_{i_t=1}^v \left(B^{i_t} \right)^T \left(A^{i_{t+1}} \right)^T \dots \left(A^{i_s} \right)^T Q^{(i_t, \dots, i_s)}(k) B^{i_s}, \quad s > t, \quad (10)$$

$$H_{s,t}(k) = H_{t,s}^T(k), \quad s < t, \quad (11)$$

$$G_t(k) = \sum_{i_t=1}^v \dots \sum_{i_1=1}^v \left(A^{i_1} \right)^T \dots \left(A^{i_t} \right)^T Q^{(i_1, \dots, i_t)}(k) B^{i_t}, \quad t = \overline{1, m}. \quad (12)$$

Последовательность матриц $Q^{(i_t, \dots, i_s)}(k)$ ($s, t = \overline{1, m}$) определяется рекуррентными уравнениями

$$Q^{(i_t, \dots, i_s)}(k) = \Theta_{i_t, \dots, i_s}(k) R_1 + \sum_{i_{s+1}=1}^v \left(A^{i_{s+1}} \right)^T Q^{(i_t, \dots, i_{s+1})}(k) A^{i_{s+1}}, \quad t = \overline{1, m-2}, \quad s > t, \quad (13)$$

$$Q^{(i_t)}(k) = E_{i_t} P^t \theta(k) R_1 + \sum_{i_{t+1}=1}^v \left(A^{i_{t+1}} \right)^T Q^{(i_t, i_{t+1})}(k) A^{i_{t+1}}, \quad t = \overline{1, m-1}, \quad (14)$$

$$Q^{(i_m)}(k) = E_{i_m} P^m \theta(k) R_1 \quad (15)$$

с граничными условиями

$$Q^{(i_t, \dots, i_m)}(k) = \Theta_{i_t, \dots, i_m}(k) R_1, \quad t = \overline{1, m-1}, \quad (16)$$

где

$$\Theta_{i_t, \dots, i_s}(k) = E_{i_s} P \text{diag} \left\{ P \text{diag} \left\{ \dots P \text{diag} \left\{ P \text{diag} \left\{ P^t \theta(k) \right\} E_{i_t}^T \right\} E_{i_{t+1}}^T \right\} E_{i_{t+2}}^T \dots \right\} E_{i_{s-2}}^T \right\} E_{i_{s-1}}^T, \quad (17)$$

$$E_{i_t} = [0, \dots, 0, 1, 0, \dots, 0]_{1 \times v}, \quad i_t = \overline{1, v}, \quad t = \overline{1, m}. \quad (18)$$

Доказательство. Критерий (3) можно записать следующим образом:

$$\begin{aligned} J(k+m/k) = & E \left\{ x^T(k+1) R_1 x(k+1) + u^T(k/k) R u(k/k) + \right. \\ & + E \left\{ x^T(k+2) R_1 x(k+2) + u^T(k+1/k) R u(k+1/k) + \dots + E \left\{ x^T(k+m) R_1 x(k+m) + \right. \right. \\ & \left. \left. + u^T(k+m-1/k) R u(k+m-1/k) / x(k+m-1), \theta(k+m-1) \right\} \dots / x(k+1), \theta(k+1) \right\} / x(k), \theta(k) \right\}. \end{aligned}$$

Введем обозначение

$$\begin{aligned} J_{k+s} = & E \left\{ x^T(k+s+1) R_1 x(k+s+1) + u^T(k+s/k) R u(k+s/k) + E \left\{ x^T(k+s+2) R_1 x(k+s+2) + \right. \right. \\ & \left. \left. + u^T(k+s+1/k) R u(k+s+1/k) + \dots + E \left\{ x^T(k+m) R_1 x(k+m) + u^T(k+m-1/k) \times \right. \right. \\ & \left. \left. \times R u(k+m-1/k) / x(k+m-1), \theta(k+m-1) \right\} / x(k+s+1), \theta(k+s+1) \right\} / x(k+s), \theta(k+s) \right\}. \end{aligned}$$

Очевидно, что

$$J_{k+s} = E \left\{ x^T(k+s+1) R_1 x(k+s+1) + u^T(k+s/k) R u(k+s/k) + J_{k+s+1} / x(k+s), \theta(k+s) \right\} \quad (19)$$

и

$$J(k+m/k) = J_k. \quad (20)$$

Рассмотрим

$$J_{k+m-1} = E \left\{ x^T(k+m) R_1 x(k+m) + u^T(k+m-1/k) R u(k+m-1/k) / x(k+m-1), \theta(k+m-1) \right\}. \quad (21)$$

Выражая $x(k+m)$ через $x(k+m-1)$ с учетом (6) и (7), будем иметь

$$x(k+m) = \sum_{i_m=1}^v \theta_{i_m}(k+m) \left[A^{i_m} x(k+m-1) + B^{i_m} u(k+m-1) \right]. \quad (22)$$

Подставляя (22) в (21) и взяв условное математическое ожидание с учетом (4) и (5), получим

$$\begin{aligned} J_{k+m-1} &= x^T(k+m-1) \sum_{i_m=1}^v \left(A^{i_m} \right)^T E_{i_m} P \theta(k+m-1) \times \\ &\times R_1 A^{i_m} x(k+m-1) + 2x^T(k+m-1) \sum_{i_m=1}^v \left(A^{i_m} \right)^T E_{i_m} P \theta(k+m-1) R_1 B^{i_m} u(k+m-1) + \\ &+ u^T(k+m-1) \left\{ \sum_{i_m=1}^v \left(B^{i_m} \right)^T E_{i_m} P \theta(k+m-1) R_1 B^{i_m} + R \right\} u(k+m-1). \end{aligned}$$

Предположим далее, что для некоторого q верно

$$\begin{aligned} J_{k+m-q} &= x^T(k+m-q) \sum_{i_{m-q+1}=1}^v \left(A^{i_{m-q+1}} \right)^T Q^{(i_{m-q+1})}(k+m-q) A^{i_{m-q+1}} x(k+m-q) + \\ &+ 2x^T(k+m-q) \sum_{t=m-q+1}^m \sum_{i_t=1}^v \dots \sum_{i_{m-q+1}=1}^v \left(A^{i_{m-q+1}} \right)^T \dots \left(A^{i_t} \right)^T Q^{(i_{m-q+1}, \dots, i_t)}(k+m-q) B^{i_t} u(k+t-1/k) + \\ &+ \sum_{t=m-q+1}^m u^T(k+t-1/k) \left[\sum_{i_t=1}^v \left(B^{i_t} \right)^T Q^{(i_t)}(k+m-q) B^{i_t} + R \right] u(k+t-1/k) + \\ &+ 2 \sum_{t=m-q+1}^{m-1} \sum_{s=t+1}^m u^T(k+t-1/k) \sum_{i_s=1}^v \dots \sum_{i_t=1}^v \left(B^{i_t} \right)^T \left(A^{i_{t+1}} \right)^T \dots \left(A^{i_s} \right)^T Q^{(i_t, \dots, i_s)}(k+m-q) B^{i_s} u(k+s-1/k), \quad (23) \end{aligned}$$

последовательность матриц $Q^{(i_t, \dots, i_s)}(k+m-q)$ ($s, t = \overline{m-q+1, m}$) определяется рекуррентными уравнениями

$$Q^{(i_t, \dots, i_s)}(k+m-q) = \Theta_{i_t, \dots, i_s}(k+m-q) R_1 + \sum_{i_{s+1}=1}^v \left(A^{i_{s+1}} \right)^T Q^{(i_t, \dots, i_{s+1})}(k+m-q) A^{i_{s+1}}, \quad t = \overline{m-q+1, m-2}, \quad s > t, \quad (24)$$

$$Q^{(i_t)}(k+m-q) = E_{i_t} P^{q-m+t} \theta(k+m-q) R_1 + \sum_{i_{t+1}=1}^v \left(A^{i_{t+1}} \right)^T Q^{(i_t, i_{t+1})}(k+m-q) A^{i_{t+1}}, \quad (25)$$

$$Q^{(i_m)}(k+m-q) = E_{i_m} P^q \theta(k+m-q) R_1, \quad (26)$$

с граничными условиями:

$$Q^{(i_t, \dots, i_m)}(k+m-q) = \Theta_{i_t, \dots, i_m}(k+m-q) R_1, \quad t = \overline{m-q+1, m-1}, \quad (27)$$

где

$$\begin{aligned} \Theta_{i_t, \dots, i_s}(k+m-q) &= E_{i_s} P \text{diag} \left\{ P \text{diag} \left\{ \dots P \text{diag} \left\{ P \text{diag} \left\{ P^{q-m+t} \theta(k+m-q) \right\} E_{i_t}^T \right\} E_{i_{t+1}}^T \right\} \times \right. \\ &\times E_{i_{t+2}}^T \left. \right\} \dots E_{i_{s-2}}^T \left. \right\} E_{i_{s-1}}^T. \end{aligned} \quad (28)$$

Покажем, что данная формула верна и для $q+1$. Действительно, из (19) следует, что

$$\begin{aligned} J_{k+m-(q+1)} &= E \left\{ x^T(k+m-q) R_1 x(k+m-q) + \right. \\ &+ u^T(k+m-(q+1)/k) R u(k+m-(q+1)/k) + J_{k+m-q} / x(k+m-(q+1)), \theta(k+m-(q+1)) \left. \right\}. \quad (29) \end{aligned}$$

Подставим в (29) вместо J_{k+m-q} его выражение через (23)–(28), вместо $x(k+m-q)$ – его выражение через $x(k+m-(q+1))$, используя (6) и (7); вместо $\theta(k+m-q)$ – его выражение через $\theta(k+m-(q+1))$, используя (4); возьмем условное математическое ожидание и, преобразовав выражение, получим, что

$$J_{k+m-(q+1)} = x^T(k+m-(q+1)) \sum_{i_{m-(q+1)+1}=1}^v \left(A^{i_{m-(q+1)+1}} \right)^T \times$$

$$\begin{aligned}
& \times Q^{(i_{m-(q+1)+1})}(k+m-(q+1))A^{i_{m-(q+1)+1}}x(k+m-(q+1)) + \\
& + 2x^T(k+m-(q+1)) \sum_{t=m-(q+1)+1}^m \sum_{i_t=1}^v \dots \sum_{i_{m-(q+1)+1}=1}^v \left(A^{i_{m-(q+1)+1}} \right)^T \dots \left(A^{i_t} \right)^T \times \\
& \times Q^{(i_{m-(q+1)+1}, \dots, i_t)}(k+m-(q+1))B^{i_t}u(k+t-1/k) + \\
& + \sum_{t=m-(q+1)+1}^m u^T(k+t-1/k) \left[\sum_{i_t=1}^v \left(B^{i_t} \right)^T Q^{(i_t)}(k+m-(q+1))B^{i_t} + R \right] u(k+t-1/k) + \\
& + 2 \sum_{t=1}^{m-1} \sum_{s=t+1}^m u^T(k+t-1/k) \sum_{i_s=1}^v \dots \sum_{i_t=1}^v \left(B^{i_t} \right)^T \left(A^{i_{t+1}} \right)^T \dots \left(A^{i_s} \right)^T Q^{(i_t, \dots, i_s)}(k+m-(q+1))B^{i_s}u(k+s-1/k), \quad (30)
\end{aligned}$$

последовательность матриц $Q^{(i_t, \dots, i_s)}(k+m-(q+1))$ ($s, t = \overline{m-(q+1)+1, m}$) определяется рекуррентными уравнениями

$$\begin{aligned}
Q^{(i_t, \dots, i_s)}(k+m-(q+1)) &= \Theta_{i_t, \dots, i_s}(k+m-(q+1))R_1 + \sum_{i_{s+1}=1}^v \left(A^{i_{s+1}} \right)^T Q^{(i_t, \dots, i_{s+1})}(k+m-(q+1))A^{i_{s+1}}, \quad (31) \\
t &= \overline{m-(q+1)+1, m-2}, s > t,
\end{aligned}$$

$$\begin{aligned}
Q^{(i_t)}(k+m-(q+1)) &= E_{i_t} P^{(q+1)-m+t} \theta(k+m-(q+1))R_1 + \sum_{i_{t+1}=1}^v \left(A^{i_{t+1}} \right)^T Q^{(i_t, i_{t+1})}(k+m-(q+1))A^{i_{t+1}}, \quad (32) \\
t &= \overline{m-(q+1)+1, m-1},
\end{aligned}$$

$$Q^{(i_m)}(k+m-(q+1)) = E_{i_m} P^{q+1} \theta(k+m-(q+1))R_1 \quad (33)$$

с граничными условиями

$$Q^{(i_t, \dots, i_m)}(k+m-(q+1)) = \Theta_{i_t, \dots, i_m}(k+m-(q+1))R_1, \quad t = \overline{m-(q+1)+1, m-1}, \quad (34)$$

где

$$\begin{aligned}
\Theta_{i_t, \dots, i_s}(k+m-(q+1)) &= E_{i_s} P \text{diag} \left\{ P \text{diag} \left\{ \dots P \text{diag} \left\{ P \text{diag} \left\{ P^{(q+1)-m+t} \theta(k+m-(q+1)) \right\} \times \right. \right. \right. \\
& \times E_{i_t}^T \left. \left. \left. E_{i_{t+1}}^T \right\} E_{i_{t+2}}^T \right\} \dots E_{i_{s-2}}^T \right\} E_{i_{s-1}}^T. \quad (35)
\end{aligned}$$

Формулы (30)–(35) совпадают с (23)–(28), если в (23)–(28) q заменить на $q+1$, а значит, согласно принципу математической индукции, формулы (23)–(28) верны для всех $q = \overline{1, m}$.

Из (23)–(28) и (20) следует, что

$$\begin{aligned}
J(k+m/k) &= x^T(k) \sum_{i_1=1}^v \left(A^{i_1} \right)^T Q^{(i_1)}(k)A^{i_1}x(k) + \\
& + 2x^T(k) \sum_{t=1}^m \sum_{i_t=1}^v \dots \sum_{i_1=1}^v \left(A^{i_t} \right)^T \dots \left(A^{i_1} \right)^T Q^{(i_1, \dots, i_t)}(k)B^{i_t}u(k+t-1/k) + \\
& + \sum_{t=1}^m u^T(k+t-1/k) \left[\sum_{i_t=1}^v \left(B^{i_t} \right)^T Q^{(i_t)}(k)B^{i_t} + R \right] u(k+t-1/k) + \\
& + 2 \sum_{t=1}^{m-1} \sum_{s=t+1}^m u^T(k+t-1/k) \sum_{i_s=1}^v \dots \sum_{i_t=1}^v \left(B^{i_t} \right)^T \left(A^{i_{t+1}} \right)^T \dots \left(A^{i_s} \right)^T Q^{(i_t, \dots, i_s)}(k)B^{i_s}u(k+s-1/k), \quad (36)
\end{aligned}$$

последовательность матриц $Q^{(i_t, \dots, i_s)}(k)$ ($s, t = \overline{1, m}$) определяется рекуррентными уравнениями (13)–(18).

Выражение (36) можно записать в матричной форме:

$$J(k+m/k) = x^T(k) \sum_{i_1=1}^v \left(A^{i_1} \right)^T Q^{(i_1)}(k)A^{i_1}x(k) +$$

$$+ 2x^T(k)G(k)U(k) + U^T(k)H(k)U(k), \quad (37)$$

где $G(k)$ и $H(k)$ определяются соотношениями (9)–(18).

Таким образом, имеем задачу минимизации критерия (37) при ограничениях (8), которая эквивалентна задаче квадратичного программирования с критерием (3) при ограничениях (2).

Заключение

В данной работе предложен метод синтеза стратегий прогнозирующего управления по квадратичному критерию для линейных дискретных систем со скачкообразно меняющимися параметрами в матрицах динамики и управления. Данный подход позволяет в явном виде учесть ограничения на управления. Алгоритм синтеза прогнозирующей стратегии включает решение последовательности задач квадратичного программирования. Синтезированы стратегии управления с учетом явных ограничений на управляющие воздействия.

ЛИТЕРАТУРА

1. Пакшин П.В. Дискретные системы со случайными параметрами и структурой. М.: Физматлит, 1994.
2. Пакшин П.В., Ретинский Д.М. Робастная стабилизация систем случайной структуры с переключаемой статической обратной связью по выходу // Автоматика и телемеханика. 2005. № 7. С. 135–147.
3. Домбровский В.В., Обьедко Т.Ю. Управление с прогнозированием системами с марковскими скачками при ограничениях и применение к оптимизации инвестиционного портфеля // Автоматика и телемеханика. 2011. № 5. С. 96–112.
4. Смагин В.И., Поползухина Е.В. Синтез следящих систем управления для объектов со случайными скачкообразными параметрами и мультипликативными возмущениями // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2000. № 271. С. 171–175.
5. Blackmore L., Bektassov A., Ono M., Williams B.C. Robust optimal predictive control of jump Markov linear systems using particles // Lecture Notes in Computer Science. 2007. V. 4416. P. 104–117.
6. Costa O.L.V., Okimura R.T. Discrete-time mean-variance optimal control of linear systems with Markovian jumps and multiplicative noise // International Journal of Control. 2009. V. 82, No. 2. P. 256–267.
7. Costa O.L.V., Oliveira A. Optimal mean-variance control for discrete-time linear systems with Markovian jumps and multiplicative noises // Automatica. 2012. V. 48, No. 2. P. 304–315.
8. Dragan V., Morozan T. The Linear Quadratic Optimization Problems for a Class of Linear Stochastic Systems With Multiplicative White Noise and Markovian Jumping // IEEE Transactions on Automatic Control. 2004. V. 49, No. 5. P. 665–675.
9. Elliott R.J., Aggoun L., Moore J.B. Hidden Markov Models: Estimation and Control. Berlin: Springer-Verlag, 1995.
10. Li X., Zhou X.Y. Indefinite stochastic LQ control with Markovian jumps in a finite time horizon // Communications in Information and Systems. 2002. No. 2. P. 265–282.
11. Домбровский В.В., Обьедко Т.Ю. Управление дискретными динамическими системами со случайными зависимыми параметрами при ограничениях // Вестник Томского государственного университета: управление, вычислительная техника и информатика. 2011. № 3. С. 5–12.
12. Домбровский В.В., Обьедко Т.Ю. Управление с прогнозированием взаимосвязанными гибридными системами с марковскими скачками при ограничениях // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2012. № 3. С. 5–12.
13. Домбровский В.В., Самородова М.В. Управление с прогнозированием нелинейными стохастическими системами с марковскими скачками при ограничениях // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2015. № 3. С. 14–22.
14. Rawlings J. Tutorial: Model Predictive Control Technology // Proc. Amer. Control Conf. San Diego. California. June 1999. P. 662–676.
15. Dombrovskii V., Obyedko T. Model predictive control for constrained systems with serially correlated stochastic parameters and portfolio optimization // Automatica. 2015. No. 54. P. 325–331.

Домбровский Владимир Валентинович, д-р техн. наук, профессор. E-mail: dombrovs@ef.tsu.ru

Самородова Мария Владимировна. E-mail: samorodova21@gmail.com

Томский государственный университет

Поступила в редакцию 15 ноября 2015 г.

Dombrovskii Vladimir V., Samorodova Mariya V. (Tomsk State University, Russian Federation).

Model predictive control with quadratic criterion for jump Markov discrete linear systems under constraints

Keywords: Markov linear systems; model predictive control; constraints.

DOI: 10.17223/19988605/34/1

Let the control object be described by the equation

$$x(k+1) = A[\alpha(k+1)]x(k) + B[\alpha(k+1)]u(k), \quad (1)$$

where $x(k) \in \mathbb{R}^{n_x}$ is the vector of state, $u(k) \in \mathbb{R}^{n_u}$ is the vector of control, $\alpha(k)$ ($k=0,1,2,\dots$) denotes a time-invariant Markov chain taking values in a finite set of observable states $\{1,2,\dots,v\}$ with the known transition probability matrix $P = [P_{i,j}]$ ($i,j \in \{1,2,\dots,v\}$),

$$P_{j,i} = P\{\alpha(k+1)=\alpha_j | \alpha(k)=\alpha_i\}, \quad \sum_{j=1}^v P_{j,i} = 1, \text{ and the initial distribution } p_i = P\{\alpha(0)=i\} \quad (i=1,2,\dots,v), \quad \sum_{i=1}^v p_i = 1.$$

The state α_i of the Markov chain $\alpha(k)$ selects the system matrices $A[\alpha(k)]$ and $B[\alpha(k)]$ from the sets $A = \{A^i \in \mathbb{R}^{n_x \times n_x} : i = \overline{1,v}\}$ и $B = \{B^i \in \mathbb{R}^{n_x \times n_u} : i = \overline{1,v}\}$ respectively. It is assumed that the state of the Markov chain is observable at the time instant k .

On control variables the following constraints are imposed:

$$u_{\min}(k) \leq S(k)u(k) \leq u_{\max}(k), \quad (2)$$

where $S(k) \in \mathbb{R}^{p \times n_u}$, $u_{\min}(k), u_{\max}(k) \in \mathbb{R}^p$.

To control system (1), we synthesize model predictive control strategies. At each step k we minimize the quadratic criterion with a receding horizon

$$J(k+m/k) = E \left\{ \sum_{i=1}^m x^T(k+i)R_1x(k+i) + u^T(k+i-1/k)Ru(k+i-1/k) / x(k), \alpha(k)=\alpha_j \right\} \quad (3)$$

on trajectories of system (1) over the sequence of predictive controls $u(k/k), \dots, u(k+m-1/k)$, which depend on system's state and on the state of Markov chain at the moment k under constraints (2), where m is a prediction horizon, $R_1 \geq 0, R > 0$ are weight matrices of corresponding dimensions, k is a current moment. The synthesis of predictive control strategies is reduced to the sequence of quadratic programming tasks.

REFERENCES

1. Pakshin, P.V. (1994) *Diskretnye sistemy so sluchaynymi parametrami i strukturoy* [Discrete-systems with stochastic parameters and structure]. Moscow: Fizmatlit.
2. Pakshin, P.V. & Retinskiy, D.M. (2005) Robust Stabilization of Random-Structure Systems via Switchable Static Output Feedback. *Automation and Remote Control*. 66(7). pp. 1153–1161. DOI: 10.1007/s10513-005-0155-5
3. Dombrovskii, V.V. & Obyedko, T.Yu. (2011) Predictive control of systems with Markovian jumps under constraints and its application to the investment portfolio optimization. *Automation and Remote Control*. 72(5). pp. 989–1003. DOI: 10.1134/S0005117911050079
4. Smagin, V.I. & Popolzhukhina, E.V. (2000) The synthesis of tracking control systems for objects with random switching parameters and multiplicative noises. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 271. pp. 171-175. (In Russian).
5. Blackmore, L., Bektassov, A., Ono, M. & Williams, B.C. (2007) Robust optimal predictive control of jump Markov linear systems using particles. *Lecture Notes in Computer Science*. 4416. pp. 104-117. DOI: 10.1007/978-3-540-71493-4_11
6. Costa, O.L.V. & Okimura, R.T. (2009) Discrete-time mean-variance optimal control of linear systems with Markovian jumps and multiplicative noise. *International Journal of Control*. 82(2). pp. 256-267. DOI: 10.1080/00207170802050825
7. Costa, O.L.V. & Oliveira, A. (2012) Optimal mean-variance control for discrete-time linear systems with Markovian jumps and multiplicative noises. *Automatica*. 48(2). pp. 304-315. DOI: 10.1016/j.automatica.2011.11.009
8. Dragan, V. & Morozan, T. (2004) The Linear Quadratic Optimization Problems for a Class of Linear Stochastic Systems With Multiplicative White Noise and Markovian Jumping. *IEEE Transactions on Automatic Control*. 49(5). pp. 665-675. DOI: 10.1109/TAC.2004.82671
9. Elliott, R.J., Aggoun, L. & Moore, J.B. (1995) *Hidden Markov Models: Estimation and Control*. Berlin: Springer-Verlag.
10. Li, X. & Zhou, X.Y. (2002) Indefinite stochastic LQ control with Markovian jumps in a finite time horizon. *Communications in Information and Systems*. 2. pp. 265-282. DOI: 10.4310/CIS.2002.v2.n3.a4
11. Dombrovskii, V.V. & Obyedko, T.Yu. (2011) Predictive control of discrete dynamic systems with stochastic dependent parameters under constraints. *Vestnik Tomskogo gosudarstvennogo universiteta: upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 3. pp. 5-12. (In Russian).
12. Dombrovskii, V.V. & Obyedko, T.Yu. (2012) Predictive control of interconnected hybrid systems with Markovian jumps under constraints. *Vestnik Tomskogo gosudarstvennogo universiteta: upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 3. pp. 5-12. (In Russian).
13. Dombrovskii, V.V. & Samorodova, M.V. (2015) Predictive control of nonlinear stochastic systems with Markovian jumps under constraints. *Vestnik Tomskogo gosudarstvennogo universiteta: upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 3(32). pp. 14-22. (In Russian). DOI: 10.17223/19988605/32/2
14. Rawlings, J. (1999) *Tutorial: Model Predictive Control Technology*. Proc. Amer. Control Conf. San Diego. California. June. pp. 662-676.
15. Dombrovskii, V. & Obyedko, T. (2015) Model predictive control for constrained systems with serially correlated stochastic parameters and portfolio optimization. *Automatica*. 54. pp. 325-331. DOI: 10.1016/j.automatica.2015.02.021

К.С. Ким, В.И. Смагин

ЛОКАЛЬНО-ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ ДИСКРЕТНЫМИ СИСТЕМАМИ С ЗАПАЗДЫВАНИЕМ В КАНАЛЕ УПРАВЛЕНИЯ ПРИ НЕПОЛНОЙ ИНФОРМАЦИИ О СОСТОЯНИИ И ВОЗМУЩЕНИЯХ

Работа выполнена в рамках государственного заказа Минобрнауки РФ на проведение научных исследований в Национальном исследовательском Томском государственном университете на 2014–2016 годы.

Рассматривается решение задачи управления в условиях неполной информации о состоянии объекта и модели возмущений с учетом запаздывания по управлению. Предполагается, что модель, описывающая возмущение, действующее на объект, содержит неизвестные параметры. Для определения оптимального управления используется метод локально-оптимального слежения, реализованный с использованием алгоритмов калмановской фильтрации и экстраполяции с неизвестным входом.

Ключевые слова: дискретные системы; локальный критерий; запаздывание по управлению; неполная информация.

Локально-оптимальные дискретные системы управления являются частным случаем дискретного прогнозирующего управления (Model predictive control) с прогнозом на один такт. Задачи управления для объектов с запаздываниями в канале управления исследовались в работах [1–7]. В [1–4] рассматривались задачи управления на основе метода расширения пространства состояний. В работах [5–7] изучались задачи синтеза управлений для объектов с запаздыванием в канале управления и с неполной информацией о возмущениях. В [5, 6] эта задача решалась на основе принципов адаптации, при этом в [6] модель объекта задавалась в непрерывном времени. В работе [7] рассматривалась задача синтеза управления в дискретных стохастических системах с использованием методов калмановской фильтрации с учетом оценок неизвестного входа (возмущений).

В настоящей работе решается задача управления в условиях неполной информации о состоянии объекта с запаздыванием по управлению при косвенных наблюдениях за возмущениями. Предполагается, что модель возмущений содержит неопределенные параметры.

1. Постановка задачи

Модель объекта с запаздыванием по управлению описывается дискретным уравнением

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k-h) + Fs(k), \\ x(0) &= x_0, u(j) = \psi(j), j = -h, -h+1, \dots, -1, \end{aligned} \quad (1)$$

где $x(k) \in \mathbb{R}^n$ – вектор состояния; $u(k-h) \in \mathbb{R}^m$ – вектор управления; h – количество тактов запаздывания; $s(k) \in \mathbb{R}^{n_1}$ – вектор возмущений, $\psi(j)$ ($j = -h, -h+1, \dots, -1$) – заданный вектор; A, B, F – заданные постоянные матрицы. Предполагается, что наблюдению доступен вектор $w_x(k) \in \mathbb{R}^l$:

$$w_x(k) = H_x x(k) + \tau_x(k), \quad (2)$$

где H_x – матрица канала наблюдений, $\tau_x(k)$ – гауссовская случайная последовательность.

Модель возмущений содержит неизвестные переменные параметры и определяется следующим разностным уравнением:

$$s(k+1) = (R(k) + \Delta R(k))s(k) + f(k) + \Delta f(k) + q(k), s(0) = s_0, \quad (3)$$

где $R(k)$ – известная матрица, $f(k)$ – известный вектор, $\Delta R(k)$ и $\Delta f(k)$ – некоторые неизвестные матрица и вектор, которые можно интерпретировать как ошибки определения параметров модели (3). Модель (3) представим как динамическую модель с неизвестным входом

$$s(k+1) = R(k)s(k) + f(k) + r(k) + q(k), \quad s(0) = s_0, \quad (4)$$

где $r(k) = \Delta R(k)s(k) + \Delta f(k)$ – вектор неизвестного входа.

Косвенные наблюдения за вектором возмущений описываются следующим соотношением:

$$\omega(k) = \Phi s(k) + \tau(k), \quad (5)$$

где $\omega(k) \in R^m$ – вектор наблюдений; Φ – $m_1 \times n$ -матрица; $\tau(k)$ – случайные ошибки наблюдений. В (1) и (3) x_0, s_0 – случайные векторы начальных условий, независимые от $q(k)$, $\tau(k)$ и $\tau_x(k)$ ($M\{x_0\} = \bar{x}_0$, $M\{(x_0 - \bar{x}_0)(x_0 - \bar{x}_0)^T\} = P_0$, $M\{s_0\} = \bar{s}_0$, $M\{(s_0 - \bar{s}_0)(s_0 - \bar{s}_0)^T\} = P_{s_0}$); $q(k)$, $\tau(k)$, $\tau_x(k)$ – независимые гауссовские случайные последовательности с характеристиками:

$$\begin{aligned} M\{q(k)\} &= 0, \quad M\{\tau(k)\} = 0, \quad M\{\tau_x(k)\} = 0, \\ M\{q(k)q^T(j)\} &= Q\delta_{kj}, \quad M\{\tau(k)\tau^T(j)\} = T\delta_{kj}, \quad M\{\tau_x(k)\tau_x^T(j)\} = T_x\delta_{kj}, \end{aligned} \quad (6)$$

δ_{kj} – символ Кронекера, T – символ операции транспонирования.

Требуется построить такое управление, чтобы вектор выхода системы $w(k) \in R^{n_2}$, $w(k) = Hx(k)$ отслеживал значение заданного вектора $z(k) \in R^{n_2}$.

2. Синтез локально-оптимального управления

Сначала определим управление, отслеживающее заданный вектор $z(k)$. Предположим, что все компоненты вектора $x(k)$ и $s(k)$ измеряются точно. Тогда оптимизируемый локальный критерий будет иметь вид

$$I(k) = M\{(w(k+1) - z(k))^T C(w(k+1) - z(k)) + u^T(k-h)Du(k-h)/S_0^k, X_0^k\}, \quad (7)$$

где $C > 0$, $D \geq 0$ – весовые матрицы; $z(k)$ – заданный отслеживаемый вектор; $S_0^k = \{s(0), s(1), \dots, s(k)\}$, $X_0^k = \{x(0), x(1), \dots, x(k)\}$.

Вычислим значение критерия (7):

$$\begin{aligned} I(k) &= u^T(k-h)(B^T H^T CHB + D)u(k-h) + u^T(k-h)B^T H^T C(HAx(k) + \\ &+ HFs(k) - z(k)) + (HAx(k) + HFs(k) - z(k))^T CHBu(k-h). \end{aligned} \quad (8)$$

Оптимальное управление определим из условия

$$\frac{dI(k)}{du(k-h)} = 0. \quad (9)$$

Тогда, в силу (9), получим уравнение

$$(B^T H^T CHB + D)u(k-h) + B^T H^T C(HAx(k) + HFs(k) - z(k)) = 0. \quad (10)$$

Выражая $u(k-h)$ из (10), получаем управление в следующем виде:

$$u(k-h) = -(B^T H^T CHB + D)^{-1} B^T H^T C(HAx(k) + HFs(k) - z(k)). \quad (11)$$

Далее, учитывая (1), имеем равенства

$$\begin{aligned} x(k) &= Ax(k-1) + Bu(k-h-1) + Fs(k-1), \\ x(k-1) &= Ax(k-2) + Bu(k-h-2) + Fs(k-2), \\ &\vdots \\ x(k-h+1) &= Ax(k-h) + Bu(k-2h) + Fs(k-h). \end{aligned} \quad (12)$$

Тогда для вычисления вектора $x(k)$ из системы (12) получим следующую формулу:

$$x(k) = A^h x(k-h) + \sum_{i=1}^h A^{i-1} Bu(k-h-i) + \sum_{i=1}^h A^{i-1} Fs(k-i). \quad (13)$$

Учитывая (13), локально-оптимальное управление (11) представим в виде

$$u(k-h) = -(B^T H^T CHB + D)^{-1} B^T H^T C(HA^{h+1}x(k-h) + \sum_{i=1}^h HA^i Bu(k-h-i) + \sum_{i=0}^h HA^i Fs(k-i) - z(k)). \quad (14)$$

Управление (14) формируется в момент времени $k-h$, и для его реализации необходимо знать состояние $x(k-h)$, возмущение $s(k-h)$ и прошлые значения управлений $u(k-h-i)$, а также необходимо вычислять прогноз возмущений для моментов времени $k, k-1, \dots, k-h+1$.

Построим управление для случая неполной информации об аддитивном возмущении $s(\cdot)$ и о состоянии объекта $x(\cdot)$. Управление в этом случае определим на основе принципа разделения, используя оценки фильтрации компонент $x(\cdot)$ и $s(\cdot)$ и оценки прогноза для вектора $s(\cdot)$. В результате для текущего времени $(k-h)$ получим

$$u(k-h) = -(B^T H^T CHB + D)^{-1} B^T H^T C(HA^{h+1}\hat{x}_f(k-h) + \sum_{i=1}^h HA^i Bu(k-h-i) + HA^h F\hat{s}_f(k-h) + \sum_{i=0}^{h-1} HA^i F\hat{s}_p(k-i) - z(k)), \quad (15)$$

где $\hat{s}_f(k-h)$ и $\hat{x}_f(k-h)$ – оценки фильтрации, которые определяются с помощью алгоритма оптимальной калмановской фильтрации:

$$\hat{s}_f(k-h) = R(k-h-1)\hat{s}_f(k-h-1) + f(k) + \hat{r}(k-h-1) + K_f(k-h)[\omega(k-h) - \Phi(R(k-h-1)\hat{s}_f(k-h-1) + f(k) + \hat{r}(k-h-1))], \quad \hat{s}_f(0) = \bar{s}_0, \quad (16)$$

$$K_f(k-h) = P(k-h/k-h-1)\Phi^T(\Phi P(k-h/k-h-1)\Phi^T + T)^{-1}, \quad (17)$$

$$P(k-h/k-h-1) = R(k-h-1)P(k-h-1)R(k-h-1)^T + Q, \quad (18)$$

$$P(k-h) = (E_{n_1} - K_f(k-h)\Phi)P(k-h/k-h-1), \quad P(0) = P_{s_0}. \quad (19)$$

$$\hat{x}_f(k-h) = A\hat{x}_f(k-h-1) + Bu(k-2h-1) + F\hat{s}_f(k-h-1) + \hat{r}_x(k-h-1) + K_x(k-h)[w(k-h) - H_x(A\hat{x}_f(k-h-1) + Bu(k-2h-1) + F\hat{s}_f(k-h-1) + \hat{r}_x(k-h-1))], \quad \hat{x}(0) = \bar{x}_0, \quad (20)$$

$$K_x(k-h) = P_x(k-h/k-h-1)H^T(HP_x(k-h/k-h-1)H^T + T_x)^{-1}, \quad (21)$$

$$P_x(k-h/k-h-1) = AP_x(k-h-1)A^T, \quad (22)$$

$$P_x(k-h) = (E_{n_1} - K_x(k-h)H)P_x(k-h/k-h-1), \quad P(0) = P_0, \quad (23)$$

где E_{n_1} – единичная матрица размерности n_1 . В (19) введена оценка вектора неизвестного входа $\hat{r}_x(k)$; так как в исходной модели (1) $s(k)$ точно не наблюдается, то введение в модели (1) аддитивного неизвестного вектора (вектора ошибок) $r_x(k)$ и последующая его оценка позволят обеспечить компенсацию ошибок, возникающих в модели объекта (1) из-за использования оценок $\hat{s}(k)$ при формировании управления. При определении управления (14) требуется вычислять также оценки и в моменты, большие, чем $(k-h)$ (оценки прогноза), поэтому здесь воспользуемся экстраполятором, который позволит найти оценку возмущения с прогнозом на один такт $\hat{s}_p(k-h+1)$:

$$\hat{s}_p(k-h+1) = R(k-h)\hat{s}_p(k-h) + f(k-h) + \hat{r}(k-h) + K_p(k-h)(\omega(k-h) - \Phi\hat{s}_p(k-h)), \quad \hat{s}_p(0) = \bar{s}_0, \quad (24)$$

$$K_p(k-h) = R(k-h)P_p(k-h)\Phi^T(\Phi P_p(k-h)\Phi^T + T)^{-1}, \quad (25)$$

$$P_p(k-h+1) = (R(k-h) - K_p(k-h)\Phi)P_p(k-h)(R(k-h) - K_p(k-h)\Phi)^T + Q + K_p(k-h)TK_p^T(k-h), \quad P_p(0) = P_{s_0}, \quad (26)$$

а оценки $\hat{s}_p(k-h+j)$ для $j \geq 2$ определяются по формулам

$$\hat{s}_p(k-h+j) = R(k-h+j-1)\hat{s}_p(k-h+j-1) + f(k-h+j-1) + \hat{r}(k-h+j-1). \quad (27)$$

В (16) и (24) оценка $\hat{r}(\cdot)$ вычисляется по методу наименьших квадратов на основе минимизации критерия [8]:

$$J_1 = \sum_{i=1}^p \left\{ \|\chi(i)\|_V^2 + \|r(i-1)\|_W^2 \right\}, \quad (28)$$

где $\chi(i) = \omega(i) - \Phi(R(i-1)\hat{s}_f(i-1) + f(i-1))$, $\rho = k-h-1$, $V > 0$, $W \geq 0$ – весовые матрицы соответствующих размерностей, $\|\chi(i)\|_V^2 = \chi^T(i)V\chi(i)$. В этом случае оценка имеет вид

$$\hat{r}(k-h-1) = [\Phi^T V \Phi + W]^{-1} \Phi^T V \{\omega(k-h) - \Phi[R(k-h-1)\hat{s}(k-h-1) + f(k-h-1)]\}, \quad (29)$$

которая учитывается при определении оценок $\hat{s}_f$ и $\hat{s}_p$, вычисляемых по формулам (16), (24).

Отметим, что в (24) используется оценка $\hat{r}(k-h)$, которая вычисляется по оценке (29) с использованием прогноза на один такт, здесь можно воспользоваться, например, методами прогнозирования временных рядов.

По аналогии с (29) находится оценка неизвестного входа $\hat{r}_x(k)$, минимизируя следующий критерий:

$$J_2 = \sum_{i=1}^p \left\{ \|\chi_x(i)\|_{V_x}^2 + \|r_x(i-1)\|_{W_x}^2 \right\}, \quad (30)$$

где $\chi_x(i) = w_x(i) - \Phi(A\hat{x}(i-1) + Bu(i-h-1) + F\hat{s}_f(i-1))$, $V_x > 0$, $W_x \geq 0$ – весовые матрицы, получим оценку

$$\hat{r}_x(k-h-1) = S_x [w_x(k-h) - H_x(A\hat{x}_f(k-h-1) + Bu(k-2h-1) + F\hat{s}_f(k-h-1))], \quad (31)$$

$$S_x = (H_x^T V_x H_x + W_x)^{-1} H_x^T V_x.$$

Отметим, что для построения оценок неизвестных входов можно также использовать методы, предложенные в работах [9–12].

3. Результаты моделирования

Рассмотрим модель объекта для следующих исходных данных:

$$A = \begin{pmatrix} 0,75 & 0 \\ 0,1 & 0,79 \end{pmatrix}, \quad R(k) = R = \begin{pmatrix} 0 & 1 \\ 0,2 & 0,6 \end{pmatrix}, \quad Q = \text{diag}\{0,05 \quad 0,02\}, \quad T = \text{diag}\{0,04 \quad 0,06\},$$

$$B = H = F = C = V = P_0 = V_x = E_2, \quad D = W = W_x = 0, \quad z = (20 \ 17)^T,$$

$$f_1(k) = \begin{cases} 0,4, & \text{если } 0 \leq k < 10, \\ -0,4, & \text{если } 10 \leq k < 20, \\ 0,4, & \text{если } 20 \leq k \leq 30, \end{cases} \quad f_2(k) = \begin{cases} 0,4, & \text{если } 0 \leq k < 10, \\ -0,4, & \text{если } 10 \leq k < 20, \\ 0,4, & \text{если } 20 \leq k \leq 30. \end{cases}$$

Алгоритм управления исследовался для следующей матрицы ΔR и компонент вектора $\Delta f(k)$:

$$\Delta R(k) = \Delta R = \begin{pmatrix} 0 & 0,03 \\ 0,04 & 0,05 \end{pmatrix}, \quad \Delta f_1(k) = \begin{cases} 0,1\sin(k) + 0,1, & \text{если } 0 \leq k < 10, \\ 0,1\sin(k) + 0,1, & \text{если } 10 \leq k < 20, \\ 0,1\sin(k) - 0,1, & \text{если } 20 \leq k \leq 30, \end{cases}$$

$$\Delta f_2(k) = \begin{cases} 0,1\sin(k) - 0,1, & \text{если } 0 \leq k < 10, \\ 0,1\sin(k) + 0,1, & \text{если } 10 \leq k < 20, \\ 0,1\sin(k) - 0,1, & \text{если } 20 \leq k \leq 30. \end{cases}$$

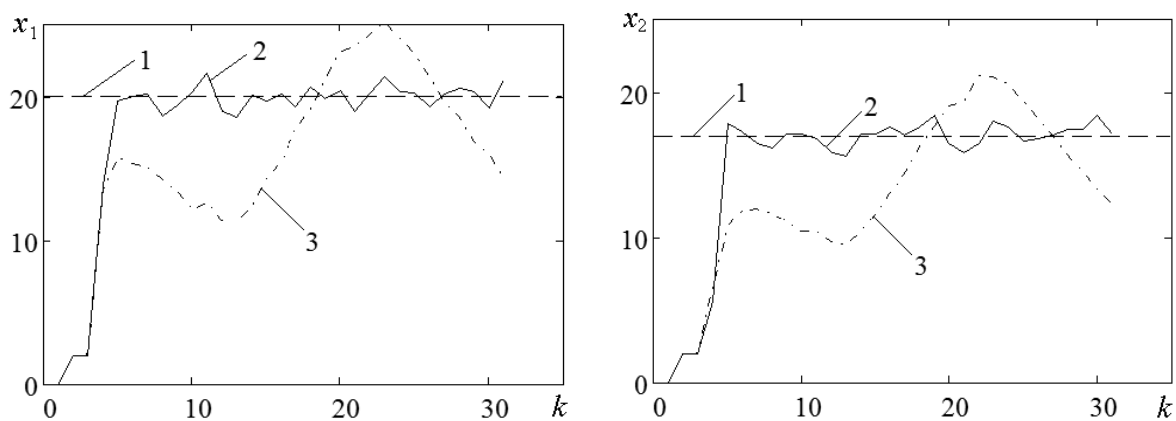


Рис. 1. Компоненты вектора состояния: 1 – компоненты отслеживаемого вектора; 2 – компоненты вектора состояния для управления (14)–(25); 3 – компоненты вектора состояния для управления (14), когда в фильтрах (14)–(25) не используются оценки неизвестных входов

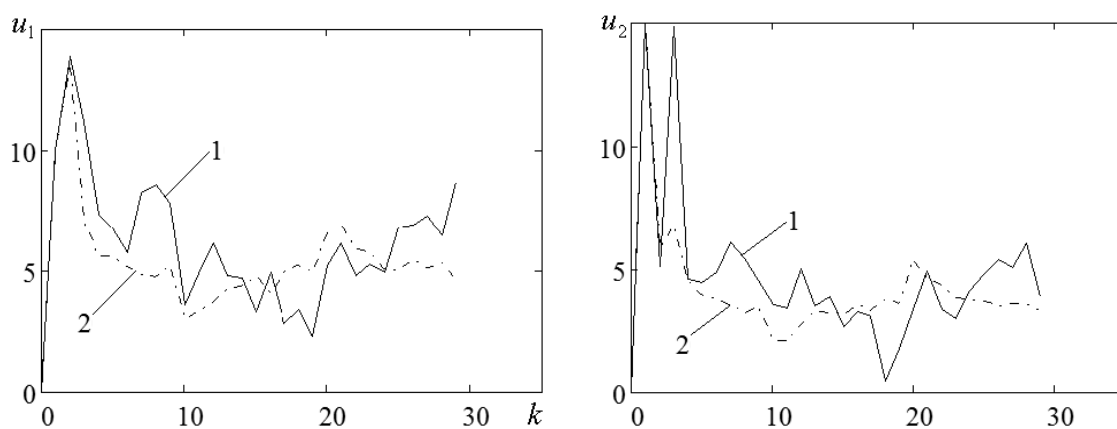


Рис. 2. Компоненты вектора управления: 1 – компоненты вектора управления для алгоритма, использующего оценки неизвестных входов (14)–(25); 2 – компоненты вектора управления для алгоритма (14), не использующего оценки неизвестных входов

Результаты моделирования показали, что исключение оценок неизвестных входов в используемых алгоритмах фильтрации и экстраполяции приводит к значительному снижению точности слежения или к срыву слежения.

Заключение

Предложен алгоритм локально-оптимального управления для дискретной стохастической системы с запаздыванием по управлению, функционирующей в условиях неполной информации о модели возмущений и компонентах вектора состояния. Показано, что применение в алгоритмах фильтрации и экстраполяции с учетом оценок неизвестного входа приводит к повышению точности отслеживания компонент заданного вектора.

ЛИТЕРАТУРА

1. Дегтярев Г.Л., Ризаев И.С. Синтез локально-оптимальных алгоритмов управления летательными аппаратами. М. : Машиностроение, 1991. 304 с.
2. Mohammad S., Modarres S., Karbassi S.M. Time-optimal control of discrete-time linear systems with state and input time-delays // Int. Journal of Innovative Computing, Information and Control. 2009. Vol. 5, No. 9. P. 2619–2625.
3. Tehrani H.A., Ramroodi N. Eigenvalue assignment of discrete-time linear systems with state and input time-delays // AIJ-MISC. 2013. Vol. 45, No. 2. P. 23–30.

4. Киселева М.Ю., Смагин В.И. Управление с прогнозирующей моделью с учетом запаздывания по управлению // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2010. № 2 (11). С. 5–12.
5. Смагин В.И., Смагин С.В. Адаптивное управление запасами с учетом ограничений и транспортных запаздываний // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2008. № 3 (4). С. 19–26.
6. Пыркин А.А. Адаптивный алгоритм компенсации параметрически неопределенного смещенного гармонического возмущения для линейного объекта с запаздыванием в канале управления // Автоматика и телемеханика. 2010. № 8. С. 62–78.
7. Kiseleva M.Yu., Smagin V.I. Model predictive control for linear discrete-time systems with time delays and unknown input // Communications in Computer and Information Science (CCIS-487). Springer International Publishing Switzerland, 2014. P. 181–188.
8. Janczak D., Grishin Yu. State estimation of linear dynamic system with unknown input and uncertain observation using dynamic programming // Control and Cybernetics. 2006. No. 4. P. 851–862.
9. Hsieh C.-S. On the optimality of two-stage Kalman filtering for systems with unknown inputs // Asian Journal of Control. 2010. No. 4. P. 510–523.
10. Witczak M. Fault diagnosis and fault-tolerant control strategies for non-linear systems. Chapter 2. Unknown input observers and filters // Lecture Notes in Electrical Engineering. Springer International Publishing, Switzerland, 2014. P. 19–56.
11. Смагин В.И., Смагин С.В. Фильтрация в линейных дискретных нестационарных системах с неизвестными возмущениями // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2011. № 3 (16). С. 43–51.
12. Смагин В.И. Оценивание состояний нестационарных дискретных систем при неизвестном входе с использованием компенсаций // Изв. вузов. Физика. 2015. Т. 58, № 7. С. 122–127.

Ким Константин Станиславович. E-mail: kks93@rambler.ru

Смагин Валерий Иванович, д-р техн. наук, профессор. E-mail: vsm@mail.tsu.ru

Томский государственный университет

Поступила в редакцию 2 ноября 2015 г.

Kim Konstantin S., Smagin Valery I. (Tomsk State University, Tomsk, Russian Federation).

Locally-optimal control of discrete delayed control systems with incomplete information about state and perturbations.

Keywords: discrete system; local criteria; delayed control; incomplete information.

DOI: 10.17223/19988605/34/2

Model of object with delayed control is described by equation

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k-h) + Fs(k), \\ x(0) &= x_0, u(j) = \psi(j), j = -h, -h+1, \dots, -1, \end{aligned}$$

where $x(k) \in \mathbb{R}^n$ is the state vector, $u(k-h) \in \mathbb{R}^m$ is the control vector, h is the time delay, $s(k) \in \mathbb{R}^{n_1}$ is the perturbation vector, x_0 and $\psi(j)$ ($j = -h, -h+1, \dots, -1$) are initial vector and initial function, A , B , and F are constant matrices. It is assumed that the observable vector $w_x(k) \in \mathbb{R}^l$, and

$$w_x(k) = H_x x(k) + \tau_x(k),$$

where H_x is the matrix of channel of observations, $\tau_x(k)$ is the Gaussian random sequence.

The perturbation model contains unknown parameters and is determined by the equation

$$s(k+1) = (R(k) + \Delta R(k))s(k) + f(k) + \Delta f(k) + q(k), s(0) = s_0,$$

where $R(k)$ is the known matrix, $f(k)$ is the known vector, $\Delta R(k)$ and $\Delta f(k)$ are some unknown matrix and vector, s_0 is the random vector of initial conditions independent of $q(k)$, $\tau(k)$ and $\tau_x(k)$; $q(k)$, $\tau(k)$, $\tau_x(k)$ are independent Gaussian random sequences with the known characteristics.

Indirect observations of the vector perturbations are described by the model

$$\omega(k) = \Phi s(k) + \tau(k),$$

where $\omega(k) \in \mathbb{R}^{m_1}$ is the vector of observations, Φ is $m_1 \times n$ -matrix, $\tau(k)$ are random errors of observations.

To solve the problem, we use the approach which is based on optimization of local criteria

$$I(k) = M \left\{ (w(k+1) - z(k))^T C (w(k+1) - z(k)) + u^T(k-h) D u(k-h) \right\},$$

where $w(k) = Hx(k)$ is the controlled output of the system, $C = C^T \geq 0$ and $D = D^T \geq 0$ are weight matrices, $z(k) \in \mathbb{R}^n$ is the tracking vector. The control is realized on the base of the Kalman filtering and extrapolation with considering the unknown input.

REFERENCES

1. Degtyarev, G.L. & Rizaev, I.S. (1991) *Sintez lokal'no-optimal'nykh algoritmov upravleniya letatel'nyimi apparatami* [Synthesis of locally optimal algorithms for flight vehicle control]. Moscow: Mashinostroenie.

2. Mohammad, S., Modarres, S. & Karbassi, S.M. (2009) Time-optimal control of discrete-time linear systems with state and input time-delays. *Int. Journal of Innovative Computing, Information and Control*. 5(9). pp. 2619-2625.
3. Tehrani, H.A. & Ramroodi, N. (2013) Eigenvalue assignment of discrete-time linear systems with state and input time-delays. *AIJ-MISC*. 45(2). pp. 23-30.
4. Kiseleva, M.Yu. & Smagin, V.I. (2010) Model predictive control with time-delay in control input. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2(11). pp. 5-12. (In Russian).
5. Smagin, V.I. & Smagin, S.V. (2008) Adaptive Inventory Control with Restrictions and Transport Delays. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 3(4). pp. 19-26. (In Russian)
6. Pyrkin, A.A. (2010) Adaptive algorithm to compensate parametrically uncertain biased disturbance of a linear plant with delay in the control channel. *Automation and Remote Control*. 71(8). pp. 1562-1577. DOI: 10.1134/S0005117910080060
7. Kiseleva, M.Yu. & Smagin, V.I. (2014) *Model predictive control for linear discrete-time systems with time delays and unknown input*. Communications in Computer and Information Science (CCIS–487). Springer International Publishing Switzerland. pp. 181-188.
8. Janczak, D. & Grishin, Yu. (2006) State estimation of linear dynamic system with unknown input and uncertain observation using dynamic programming. *Control and Cybernetics*. 4. pp. 851-862.
9. Hsieh, C-S. (2010) On the optimality of two-stage Kalman filtering for systems with unknown inputs. *Asian Journal of Control*. 4. pp. 510-523. DOI: 10.1002/asjc.205
10. Witczak, M. (2014) *Fault diagnosis and fault-tolerant control strategies for non-linear systems*. Springer International Publishing, Switzerland. pp. 19-56.
11. Smagin, V.I. & Smagin, S.V. (2011) Filtering for linear not stationary discrete system with unknown disturbances. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 3(16). pp. 43-51. (In Russian).
12. Smagin, V.I. (2015) State estimation for nonstationary discrete systems with unknown input using compensations. *Russian Physics Journal*. 58(7). pp. 122-127. (In Russian).

ОБРАБОТКА ИНФОРМАЦИИ

УДК 519.21

DOI: 10.17223/19988605/34/3

М.А. Бахолдина

ВЕРОЯТНОСТЬ ОШИБКИ ПРИ ОЦЕНИВАНИИ СОСТОЯНИЙ
МОДУЛИРОВАННОГО ОБОБЩЕННОГО ПОЛУСИНХРОННОГО
ПОТОКА СОБЫТИЙ

Работа выполнена в рамках государственного заказа Минобрнауки РФ на проведение научных исследований в Национальном исследовательском Томском государственном университете на 2014–2016 гг.

Рассматривается модулированный обобщенный полусинхронный поток событий, являющийся одной из математических моделей информационных потоков заявок, функционирующих в телекоммуникационных и информационно-вычислительных сетях связи, и относящийся к классу дважды стохастических потоков событий (DSPPs). В статье приводятся аналитические результаты по нахождению условной и безусловной вероятности ошибочного решения при оптимальном оценивании состояний потока событий.

Ключевые слова: модулированный обобщенный полусинхронный поток событий; дважды стохастический поток событий (DSPP); MAP (Markovian Arrival Process)-поток событий; апостериорная вероятность состояния; оценка состояния; критерий максимума апостериорной вероятности; вероятность ошибки при оценивании состояний потока.

Условия функционирования реальных систем массового обслуживания таковы, что если в отношении параметров обслуживающих устройств, как правило, можно утверждать, что они известны и с течением времени не меняются, то в отношении интенсивностей входящих потоков этого сказать во многих случаях нельзя. Более того, интенсивности входящих потоков заявок обычно меняются со временем, часто эти изменения носят случайный характер, что приводит к рассмотрению математических моделей дважды стохастических потоков событий (DSPPs). Данные потоки можно охарактеризовать двумя случайностями: первая случайность – это число событий на любом рассматриваемом интервале функционирования потока; вторая случайность – это случайный процесс $\lambda(t)$, называемый интенсивностью потока [1–8].

В настоящей работе рассматривается модулированный обобщенный полусинхронный поток событий, являющийся обобщением полусинхронного потока [9–12] и обобщенного полусинхронного потока событий [13–19] и относящийся к классу дважды стохастических потоков событий с кусочно-постоянной интенсивностью. Достаточно обширная литература по исследованию подобных потоков событий (асинхронных, синхронных и полусинхронных) приведена в [13, 20–33], при этом в [22] показано, что данные потоки могут быть представлены в виде моделей MAP-потоков событий с определенными ограничениями на параметры последних. Настоящая статья является непосредственным развитием работ [34, 35], где решается задача оптимального оценивания состояний модулированного обобщенного полусинхронного потока событий (задача фильтрации интенсивности потока) и в качестве решающего правила используется критерий максимума апостериорной вероятности, обеспечивающий минимум полной (безусловной) вероятности ошибки вынесения решения о состоянии потока [36]. В настоящей статье приводятся аналитические и численные результаты по нахождению условной и безусловной вероятности ошибочного решения при оптимальном оценивании состояний потока.

1. Постановка задачи

Рассматривается модулированный обобщенный полусинхронный поток событий (далее – поток или поток событий), интенсивность которого является кусочно-постоянным стационарным случайным процессом $\lambda(t)$ с двумя состояниями λ_1 и λ_2 ($\lambda_1 > \lambda_2 \geq 0$). Будем говорить, что имеет место первое состояние процесса (потока), если $\lambda(t) = \lambda_1$, и второе состояние процесса (потока), если $\lambda(t) = \lambda_2$. В течение временного интервала случайной длительности, когда процесс $\lambda(t)$ находится в состоянии λ_i ($\lambda(t) = \lambda_i$), имеет место пуассоновский поток событий с интенсивностью λ_i , $i = 1, 2$.

Переход из первого состояния процесса $\lambda(t)$ во второе возможен в момент наступления события пуассоновского потока интенсивности λ_1 , при этом переход осуществляется с вероятностью p ($0 \leq p \leq 1$); с вероятностью $1 - p$ процесс $\lambda(t)$ остается в первом состоянии (т.е. сначала наступает событие потока, затем происходит либо не происходит переход процесса $\lambda(t)$ из первого состояния во второе). Переход из первого состояния процесса $\lambda(t)$ во второе также возможен в произвольный момент времени, не совпадающий с моментом наступления события, при этом длительность пребывания процесса $\lambda(t)$ в первом состоянии распределена по экспоненциальному закону с параметром β : $F(\tau) = 1 - e^{-\beta\tau}$, $\tau \geq 0$. Тогда длительность пребывания процесса $\lambda(t)$ в первом состоянии есть случайная величина с экспоненциальной функцией распределения $F_1(\tau) = 1 - e^{-(p\lambda_1 + \beta)\tau}$, $\tau \geq 0$.

Переход из второго состояния процесса $\lambda(t)$ в первое в момент наступления события пуассоновского потока интенсивности λ_2 невозможен и может осуществляться только в произвольный момент времени. При этом длительность пребывания процесса $\lambda(t)$ во втором состоянии распределена по экспоненциальному закону с параметром α : $F_2(\tau) = 1 - e^{-\alpha\tau}$, $\tau \geq 0$. В момент окончания второго состояния процесса $\lambda(t)$ при его переходе из второго состояния в первое инициируется с вероятностью δ ($0 \leq \delta \leq 1$) дополнительное событие. Отметим, что события пуассоновских потоков и дополнительные события неразличимы для наблюдателя. В сделанных предположениях $\lambda(t)$ – скрытый марковский процесс. При этом матрицы инфинитезимальных характеристик принимают вид

$$D_0 = \begin{bmatrix} -(\lambda_1 + \beta) & \beta \\ \alpha(1 - \delta) & -(\lambda_2 + \alpha) \end{bmatrix}, \quad D_1 = \begin{bmatrix} (1 - p)\lambda_1 & p\lambda_1 \\ \alpha\delta & \lambda_2 \end{bmatrix}.$$

Элементами матрицы D_1 являются интенсивности переходов процесса $\lambda(t)$ из состояния в состояние с наступлением события. Недиagonальные элементы матрицы D_0 – это интенсивности переходов из состояния в состояние без наступления события. Диагональные элементы матрицы D_0 – это интенсивности выхода процесса $\lambda(t)$ из своих состояний, взятые с противоположным знаком. Отметим, что если $\beta = 0$, то имеет место обобщенный полусинхронный поток событий [15].

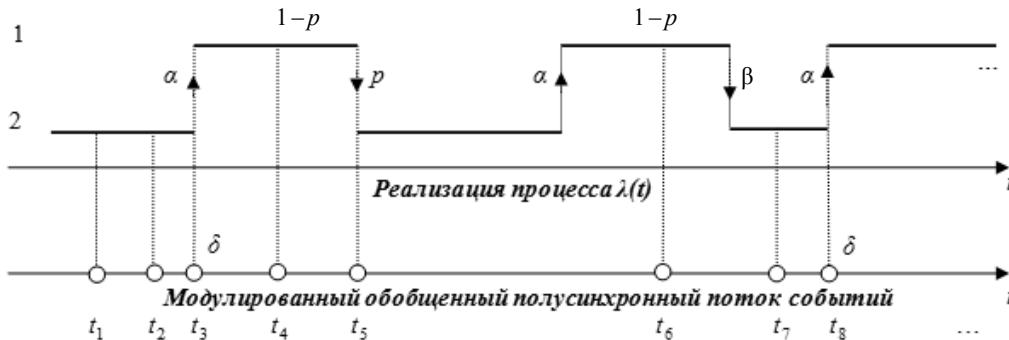


Рис. 1. Формирование модулированного обобщенного полусинхронного потока событий

Заметим, что в определении модулированного обобщенного полусинхронного потока событий в явном виде не оговаривается, в каком состоянии процесса $\lambda(t)$ наступает дополнительное событие потока при переходе процесса $\lambda(t)$ из второго состояния в первое. Данное обстоятельство при последующем получении аналитических результатов является несущественным, так как наступление дополнительного события и переход процесса $\lambda(t)$ из второго состояния в первое происходят мгновенно. В реальных ситуациях возможны два варианта, связанные с наступлением события и переходом процесса $\lambda(t)$ из второго состояния в первое: 1) первично наступление события во втором состоянии процесса $\lambda(t)$, затем его переход из второго состояния в первое; 2) первичен переход процесса $\lambda(t)$ из второго состояния в первое, затем наступление события в первом состоянии. Здесь принимается, что сначала происходит переход процесса $\lambda(t)$ из второго состояния в первое, затем дополнительное событие с вероятностью δ инициируется в первом состоянии.

Пример реализации модулированного обобщенного полусинхронного потока событий приведен на рис. 1, где 1, 2 – состояния процесса $\lambda(t)$; $t_1, t_2, \dots$ – моменты наступления событий потока; дополнительные события помечены буквами δ .

Рассматривается установившийся (стационарный) режим функционирования потока событий, поэтому переходными процессами на полуинтервале наблюдения $(t_0, t]$, где t_0 – начало наблюдений, t – окончание наблюдений (момент вынесения решения), пренебрегаем. Тогда без потери общности можно положить $t_0 = 0$. Поскольку процесс $\lambda(t)$ является принципиально ненаблюдаемым, то говорить о состоянии потока можно только в вероятностном смысле. Вся доступная информация о потоке – это моменты наступления событий $t_1, t_2, \dots, t_k$ с начала наблюдения t_0 до момента t .

Решение о состоянии потока выносится по критерию максимума апостериорной вероятности, который обеспечивает минимум полной вероятности ошибки вынесения решения [36]. Таким образом, для вынесения решения о состоянии ненаблюдаемого процесса $\lambda(t)$ в момент времени t необходимо определить апостериорные вероятности $w(\lambda_i | t) = w(\lambda_i | t_1, \dots, t_m, t) = P\{\lambda(t) = \lambda_i | t_1, \dots, t_m, t\}$, $i = 1, 2$, того, что в момент времени t значение процесса $\lambda(t)$ есть $\lambda(t) = \lambda_i$ (m – количество наблюдаемых событий за время t). При этом $w(\lambda_1 | t) + w(\lambda_2 | t) = 1$. Решение о состоянии процесса $\lambda(t)$ выносится путем сравнения апостериорных вероятностей: если $w(\lambda_j | t) \geq w(\lambda_i | t)$, $i, j = 1, 2$, $i \neq j$, то оценка состояния процесса $\lambda(t)$ есть $\hat{\lambda}(t) = \lambda_j$.

2. Условная вероятность ошибочного решения о состоянии потока событий в общем случае

Перейдем к выводу формул для вероятностей ошибок при вынесении решения о состоянии процесса $\lambda(t)$. Пусть полуинтервал наблюдения за потоком событий есть $(t_0, t]$. Момент времени t зафиксирован, т.е. $t - t_0$ есть длина полуинтервала наблюдения. В силу того что моменты наступления событий $t_1, t_2, \dots, t_k$ ($t_1 < t_2 < \dots < t_k < t$), попавшие в полуинтервал $(t_0, t]$, случайны, то случайна и разность $t - t_k$. Момент вынесения решения t лежит между моментами времени t_k и t_{k+1} ($t_k < t < t_{k+1}$), причем момент времени t_{k+1} может быть сколь угодно большим (разность $t_{k+1} - t$ также случайна).

Итак, вынесение решения о состоянии процесса $\lambda(t)$ привязано к интервалу между двумя соседними временными моментами наступления событий: (t_k, t_{k+1}) . Сам момент вынесения решения t можно трактовать как некоторую точку, случайным образом падающую на ось времени Ot , никак не связанную с потоком событий. Вследствие этого точка t может попасть в любой интервал между соседними событиями модулированного обобщенного полусинхронного потока событий. При этом начало наблюдений, т.е. точка $t_0 = 0$, однозначно определяется на оси времени Ot .

Рассмотрим апостериорную вероятность $w(\lambda_1 | t)$, $t_k \leq t < t_{k+1}$, полученную в [34, 35]. Обозначим через $\tau_k = t - t_k$, $k = 1, 2, \dots$. Следующее событие потока наступает в момент времени t_{k+1} ($t_{k+1} > t_k$). Тогда τ_k ограничено снизу нулем, сверху τ_k может быть в принципе неограниченным, т.е. $\tau_k \geq 0$. С уче-

том введенного обозначения $w(\lambda_1 | t) = w(\lambda_1 | t_k + \tau_k)$, $\tau_k \geq 0$. В момент $t = t_k$ имеем $w(\lambda_1 | t = t_k) = w(\lambda_1 | t_k + 0)$. Так как τ_k привязано к моменту времени t_k наступления k -го события, то для простоты обозначим $w(\lambda_1 | t_k + \tau_k) = w(\lambda_1 | \tau_k)$, $\tau_k \geq 0$.

Остановимся более подробно на алгоритме принятия решения. Процесс $\lambda(\tau_k)$, $\tau_k \geq 0$, является ненаблюдаемым. В момент t_k наступления события значение процесса $\lambda(\tau_k = 0)$ может быть равным либо λ_1 , либо λ_2 . Вследствие этого в момент τ_k вынесения решения процесс $\lambda(\tau_k)$ может также принимать любое значение: либо λ_1 ($\lambda(\tau_k) = \lambda_1$), либо λ_2 ($\lambda(\tau_k) = \lambda_2$). Тогда оценка $\hat{\lambda}(\tau_k)$ значения процесса $\lambda(\tau_k)$ в момент времени τ_k , получаемая по критерию максимума апостериорной вероятности, может принимать либо значение λ_1 ($\hat{\lambda}(\tau_k) = \lambda_1$), либо значение λ_2 ($\hat{\lambda}(\tau_k) = \lambda_2$). При этом возможны следующие варианты: 1) если в момент времени τ_k значение процесса $\lambda(\tau_k) = \lambda_1$, то правильное решение ($\hat{\lambda}(\tau_k) = \lambda_1$) будет приниматься, если $w(\lambda_1 | \tau_k) \geq w(\lambda_2 | \tau_k)$; если же $w(\lambda_1 | \tau_k) < w(\lambda_2 | \tau_k)$, то будет приниматься ошибочное решение (совершаться ошибка): $\hat{\lambda}(\tau_k) = \lambda_2$; 2) если в момент времени τ_k значение процесса $\lambda(\tau_k) = \lambda_2$, то правильное решение ($\hat{\lambda}(\tau_k) = \lambda_2$) будет приниматься, если $w(\lambda_1 | \tau_k) < w(\lambda_2 | \tau_k)$; если же $w(\lambda_1 | \tau_k) \geq w(\lambda_2 | \tau_k)$, то будет приниматься ошибочное решение (совершаться ошибка): $\hat{\lambda}(\tau_k) = \lambda_1$.

Обозначим далее $w(\lambda(\tau_k), \tau_k)$ – распределение вероятностей значений двумерной смешанной случайной величины $(\lambda(\tau_k), \tau_k)$, здесь $\lambda(\tau_k)$ – значение дискретной случайной величины ($\lambda(\tau_k) = \lambda_1$ либо $\lambda(\tau_k) = \lambda_2$); τ_k – значение непрерывной случайной величины ($\tau_k \geq 0$). Тогда уравнение $w(\lambda(\tau_k) = \lambda_1, \tau_k) = w(\lambda(\tau_k) = \lambda_2, \tau_k)$ определяет границу τ_k^0 критической области, в которой отклоняется гипотеза $\hat{\lambda}(\tau_k) = \lambda_2$ и принимается гипотеза $\hat{\lambda}(\tau_k) = \lambda_1$ (либо наоборот, отклоняется гипотеза $\hat{\lambda}(\tau_k) = \lambda_1$ и принимается гипотеза $\hat{\lambda}(\tau_k) = \lambda_2$). Сам корень данного уравнения (если он существует и единствен) может быть меньше нуля ($\tau_k^0 < 0$), равен нулю ($\tau_k^0 = 0$) и может быть больше нуля ($\tau_k^0 > 0$). Кроме того, в принципе возможны ситуации, когда данное уравнение определяет некоторое множество корней либо корней не имеет (корни не существуют).

Расписывая в данном уравнении $w(\lambda(\tau_k) = \lambda_j, \tau_k)$, $j = 1, 2$, через безусловную плотность $w(\tau_k)$ и апостериорную вероятность $w(\lambda(\tau_k) = \lambda_j | \tau_k) = w(\lambda_j | \tau_k)$, приходим к следующему виду уравнения для границы критической области τ_k^0 :

$$w(\lambda_1 | \tau_k) = w(\lambda_2 | \tau_k), \quad k = 1, 2, \dots \quad (1)$$

Тогда если $w(\lambda_1 | \tau_k) \geq w(\lambda_2 | \tau_k)$, то апостериорную вероятность $w(\lambda_1 | \tau_k)$ можно интерпретировать как условную вероятность вынесения правильного решения: $\hat{\lambda}(\tau_k) = \lambda_1$ при условии, что вынесение решения произведено в момент времени τ_k ($\tau_k \geq 0$); апостериорную же вероятность $w(\lambda_2 | \tau_k) = 1 - w(\lambda_1 | \tau_k)$ – как условную вероятность вынесения ошибочного решения (условную вероятность ошибки): решение выносится в пользу $\hat{\lambda}(\tau_k) = \lambda_1$, хотя на самом деле имеет место $\lambda(\tau_k) = \lambda_2$. Аналогичная трактовка имеет место для случая $w(\lambda_1 | \tau_k) < w(\lambda_2 | \tau_k)$.

Во введенных обозначениях согласно формуле, полученной для расчета апостериорной вероятности в [34, 35], поведение вероятности $w(\lambda_1 | \tau_k)$ на полуинтервале $[t_k, t_{k+1})$ между соседними событиями потока определяется выражением

$$w(\lambda_1 | \tau_k) = \frac{w_1 [w_2 - w(\lambda_1 | t_k + 0)] - w_2 [w_1 - w(\lambda_1 | t_k + 0)] e^{-b\tau_k}}{w_2 - w(\lambda_1 | t_k + 0) - [w_1 - w(\lambda_1 | t_k + 0)] e^{-b\tau_k}}, \quad \tau_k = t - t_k \geq 0, \quad k = 0, 1, \dots, \quad (2)$$

где $w_1 = \frac{\lambda_1 - \lambda_2 + \alpha + \beta - 2\alpha\delta - b}{2(\lambda_1 - \lambda_2 - \alpha\delta)}$, $w_2 = \frac{\lambda_1 - \lambda_2 + \alpha + \beta - 2\alpha\delta + b}{2(\lambda_1 - \lambda_2 - \alpha\delta)}$, $b = \sqrt{(\lambda_1 - \lambda_2 - \alpha + \beta)^2 + 4\alpha\beta(1 - \delta)}$.

В момент времени $\tau_k = t - t_k = 0$ (т.е. тогда, когда момент вынесения решения t совпадает с моментом t_k наступления события) апостериорная вероятность (2) претерпевает разрыв 1-го рода ($k = 1, 2, \dots$), поэтому в момент времени $\tau_k = 0$ согласно формуле пересчета, полученной в [34, 35], имеет место

$$w(\lambda_1 | t_k + 0) = \frac{\alpha\delta + [\lambda_1(1-p) - \alpha\delta]w(\lambda_1 | t_k - 0)}{\lambda_2 + \alpha\delta + (\lambda_1 - \lambda_2 - \alpha\delta)w(\lambda_1 | t_k - 0)}, \quad k = 1, 2, \dots, \quad (3)$$

где $w(\lambda_1 | t_k - 0)$ вычисляется по формуле (2), в которой, во-первых, вместо τ_k нужно подставить τ_{k-1} и, во-вторых, вычисления производить для $\tau_{k-1} = t_k - t_{k-1}$, $k = 1, 2, \dots$. Последнее реализует вычисление предела слева апостериорной вероятности $w(\lambda_1 | \tau_{k-1})$ в момент времени t_k (в момент наступления события). В качестве начального значения $w(\lambda_1 | t_0 + 0) = w(\lambda_1 | t_0 = 0)$ в (2) выбирается априорная финальная вероятность первого состояния процесса $\lambda(t)$: $\pi_1 = \alpha / (p\lambda_1 + \beta + \alpha)$.

Изучим поведение апостериорной вероятности $w(\lambda_1 | \tau_k)$ как функции τ_k ($\tau_k \geq 0$). Производная функции (2) по τ_k принимает вид

$$\frac{dw(\lambda_1 | \tau_k)}{d\tau_k} = \frac{a(w_2 - w_1)^2 [w_1 - w(\lambda_1 | t_k + 0)] [w_2 - w(\lambda_1 | t_k + 0)] e^{-a(w_2 - w_1)\tau_k}}{[w_2 - w(\lambda_1 | t_k + 0) - [w_1 - w(\lambda_1 | t_k + 0)] e^{-a(w_2 - w_1)\tau_k}]^2}, \quad \tau_k \geq 0, \quad k = 0, 1, \dots, \quad (4)$$

где $a = \lambda_1 - \lambda_2 - \alpha\delta$; апостериорная вероятность $w(\lambda_1 | t_k + 0)$ определена в (3), $k = 1, 2, \dots$; $w(\lambda_1 | t_0 + 0) = \pi_1 = \alpha / (p\lambda_1 + \beta + \alpha)$.

Рассмотрим поведение производной (4) в зависимости от τ_k ($\tau_k \geq 0$). Из (2) вытекает, что $\lim_{\tau_k \rightarrow \infty} w(\lambda_1 | \tau_k) = w_1$. При этом знак производной (4) определяется знаком выражения $a[w_1 - w(\lambda_1 | t_k + 0)][w_2 - w(\lambda_1 | t_k + 0)]$, где a определена в (4).

Можно показать, что знак производной (4) для любого a ($a > 0$ либо $a < 0$) определяется знаком разности $w_1 - w(\lambda_1 | t_k + 0)$, тогда: 1) если $0 \leq w(\lambda_1 | t_k + 0) < w_1$, то $dw(\lambda_1 | \tau_k) / d\tau_k > 0$ и $w(\lambda_1 | \tau_k)$ является возрастающей функцией переменной τ_k , стремящейся к w_1 снизу при $\tau_k \rightarrow \infty$; 2) если $w_1 < w(\lambda_1 | t_k + 0) \leq 1$, то $dw(\lambda_1 | \tau_k) / d\tau_k < 0$ и $w(\lambda_1 | \tau_k)$ является убывающей функцией переменной τ_k , стремящейся к w_1 сверху при $\tau_k \rightarrow \infty$; 3) если $w(\lambda_1 | t_k + 0) = w_1$, то $w(\lambda_1 | \tau_k) = w_1$ для $\tau_k \geq 0$. Тогда уравнение (1) имеет либо единственный корень τ_k^0 ($\tau_k^0 < 0$ или $\tau_k^0 \geq 0$), либо корень τ_k^0 не существует, так как апостериорная вероятность $w(\lambda_1 | \tau_k)$ есть монотонная функция переменной τ_k . Подставляя (2) в (1) и решая полученное уравнение относительно τ_k , находим

$$\tau_k^0 = \frac{1}{a(w_2 - w_1)} \ln \frac{(w_2 - 1/2)[w_1 - w(\lambda_1 | t_k + 0)]}{(w_1 - 1/2)[w_2 - w(\lambda_1 | t_k + 0)]}, \quad k = 0, 1, \dots \quad (5)$$

Выражение (5) определяет границу критической области τ_k^0 и в зависимости от соотношения величин w_1 и $w(\lambda_1 | t_k + 0)$ для любого a ($a > 0$ либо $a < 0$) возможны различные варианты положения τ_k^0 на временной оси:

1) если $1/2 \leq w_1 \leq w(\lambda_1 | t_k + 0) \leq 1$, то корень τ_k^0 не существует, при этом условная вероятность ошибки определяется в виде $P_0(w(\lambda_1 | t_k + 0), \tau_k) = 1 - w(\lambda_1 | \tau_k)$, $\tau_k \geq 0$;

2) если $1/2 < w(\lambda_1 | t_k + 0) < w_1 < 1$, то корень τ_k^0 существует и $\tau_k^0 < 0$, при этом условная вероятность ошибки определяется в виде $P_0(w(\lambda_1 | t_k + 0), \tau_k) = 1 - w(\lambda_1 | \tau_k)$, $\tau_k \geq 0$;

3) если $0 \leq w(\lambda_1 | t_k + 0) \leq 1/2 < w_1 < 1$, то корень τ_k^0 существует и $\tau_k^0 \geq 0$, при этом условная вероятность ошибки определяется в виде

$$P_0(w(\lambda_1 | t_k + 0), \tau_k) = \begin{cases} w(\lambda_1 | \tau_k), & 0 \leq \tau_k < \tau_k^0, \\ 1 - w(\lambda_1 | \tau_k), & \tau_k \geq \tau_k^0; \end{cases}$$

4) если $w_1 = 1/2$, $w(\lambda_1 | t_k + 0) < w_1$, то корень τ_k^0 не существует, при этом условная вероятность ошибки определяется в виде $P_0(w(\lambda_1 | t_k + 0), \tau_k) = w(\lambda_1 | \tau_k)$, $\tau_k \geq 0$;

5) если $0 \leq w(\lambda_1 | t_k + 0) \leq w_1 \leq 1/2$, то корень τ_k^0 не существует, при этом условная вероятность ошибки определяется в виде $P_0(w(\lambda_1 | t_k + 0), \tau_k) = w(\lambda_1 | \tau_k)$, $\tau_k \geq 0$;

6) если $0 \leq w_1 < w(\lambda_1 | t_k + 0) < 1/2$, то корень τ_k^0 существует и $\tau_k^0 < 0$, при этом условная вероятность ошибки определяется в виде $P_0(w(\lambda_1 | t_k + 0), \tau_k) = w(\lambda_1 | \tau_k)$, $\tau_k \geq 0$;

7) если $0 \leq w_1 < 1/2 \leq w(\lambda_1 | t_k + 0) \leq 1$, то корень τ_k^0 существует и $\tau_k^0 \geq 0$, при этом условная вероятность ошибки определяется в виде

$$P_0(w(\lambda_1 | t_k + 0), \tau_k) = \begin{cases} 1 - w(\lambda_1 | \tau_k), & 0 \leq \tau_k < \tau_k^0, \\ w(\lambda_1 | \tau_k), & \tau_k \geq \tau_k^0; \end{cases}$$

8) если $w_1 = 1/2$, $w_1 \leq w(\lambda_1 | t_k + 0)$, то корень τ_k^0 не существует, при этом условная вероятность ошибки определяется в виде $P_0(w(\lambda_1 | t_k + 0), \tau_k) = 1 - w(\lambda_1 | \tau_k)$, $\tau_k \geq 0$.

Полученные формулы позволяют сформулировать алгоритм расчета условной вероятности вынесения ошибочного решения $P_0(w(\lambda_1 | t_k + 0), \tau_k)$ в любой момент времени $\tau_k \geq 0$, $k = 0, 1, \dots, 1$) в момент времени $t_0 = 0$ задается $w(\lambda_1 | t_0 + 0) = w(\lambda_1 | t_0 = 0) = \pi_1$; 2) по формуле (5) для $k = 0$ рассчитывается τ_k^0 , тем самым устанавливается положение границы критической области на временной оси; 3) находится один из восьми возможных вариантов соотношения величин w_1 и $w(\lambda_1 | t_k + 0)$; 4) для найденного варианта рассчитывается (с использованием формулы (2)) вероятность $P_0(w(\lambda_1 | t_k + 0), \tau_k)$ в любой момент времени $0 \leq \tau_k < t_{k+1} - t_k$; 5) рассчитывается вероятность $w(\lambda_1 | \tau_k)$ в момент времени $\tau_k = t_{k+1} - t_k$, т.е. $w(\lambda_1 | t_{k+1} - 0)$, по формуле (2); затем производится пересчет апостериорной вероятности в момент времени t_{k+1} , т.е. находится $w(\lambda_1 | t_{k+1} + 0)$ по формуле пересчета (3); 6) k увеличивается на 1, алгоритм переходит на шаг 2 и т.д.

Замечание. В силу формулы пересчета (3) значение $w(\lambda_1 | t_{k+1} + 0)$ зависит от всех моментов $t_1, t_2, \dots, t_k$ наступления событий в потоке, т.е. вся предыдущая информация «сосредоточена» в вероятности $w(\lambda_1 | t_{k+1} + 0)$. Вследствие этого для определения безусловной вероятности ошибки необходимо усреднить условную вероятность ошибки $P_0(w(\lambda_1 | t_k + 0), \tau_k)$ по моментам наступления событий $t_1, t_2, \dots, t_k$. Однако найти функцию распределения вероятностей моментов $t_1, t_2, \dots, t_k$ наступления событий в модулированном обобщенном полусинхронном потоке в явном виде представляется затруднительным или вообще невозможным. Определить безусловную вероятность ошибки возможно только для некоторых случаев соотношения параметров потока.

3. Условная и безусловная вероятность ошибочного решения о состоянии потока для частных и особых случаев

Представляет интерес рассмотреть частные и особые случаи соотношения параметров потока, для которых возможно вычисление безусловной вероятности ошибки.

3.1. Частный случай: $p = 1$, $\delta = 0$

В данном случае $w_1 = \frac{\lambda_1 - \lambda_2 + \alpha + \beta - b}{2(\lambda_1 - \lambda_2)}$, $w_2 = \frac{\lambda_1 - \lambda_2 + \alpha + \beta + b}{2(\lambda_1 - \lambda_2)}$, $b = \sqrt{(\lambda_1 - \lambda_2 - \alpha + \beta)^2 + 4\alpha\beta}$, при этом формула пересчета (3) принимает вид

$$w(\lambda_1 | t_k + 0) = 0, \quad k = 1, 2, \dots \quad (6)$$

Отметим тот факт, что равенство (6) означает, что значение апостериорной вероятности $w(\lambda_1 | t_k + \tau_k)$ (или $w(\lambda_1 | t)$) после любого момента t_k не зависит от моментов наступлений событий до t_k , т.е. от $t_1, t_2, \dots, t_{k-1}$. Вследствие этого в обозначении апостериорной вероятности индекс k можно опустить: $w(\lambda_1 | t) = w(\lambda_1 | t_k + \tau_k) = w(\lambda_1 | \tau_k) = w(\lambda_1 | \tau)$, $\tau \geq 0$. Значение условной вероятности ошибки $P_0(w(\lambda_1 | t_k + 0), \tau_k)$ также не зависит от предыстории, вследствие этого $P_0(w(\lambda_1 | t_k + 0), \tau_k) = P_0(\tau)$.

Стоит отметить, что для рассматриваемого частного случая модулированный обобщенный полусинхронный поток событий является рекуррентным потоком. Действительно, можно показать, что при значении параметров $p = 1$, $\delta = 0$ совместная плотность вероятностей длительностей двух смежных интервалов $p(\tau_1, \tau_2)$ факторизуется: $p(\tau_1, \tau_2) = p(\tau_1)p(\tau_2)$. Общий вид плотности вероятности и совместной плотности вероятности получен в статье [37]:

$$p(\tau) = \gamma z_1 e^{-z_1 \tau} + (1 - \gamma) z_2 e^{-z_2 \tau}, \quad \tau \geq 0, \quad (7)$$

$$p(\tau_1, \tau_2) = p(\tau_1)p(\tau_2) + \gamma(1 - \gamma) \frac{\lambda_1 [\lambda_2 - p(\lambda_2 + \alpha\delta)]}{z_1 z_2} [z_1 e^{-z_1 \tau_1} - z_2 e^{-z_2 \tau_1}] [z_1 e^{-z_1 \tau_2} - z_2 e^{-z_2 \tau_2}], \quad \tau_1 \geq 0, \quad \tau_2 \geq 0, \quad (8)$$

$$\gamma = \frac{1}{z_2 - z_1} [z_2 - \lambda_1 \pi_1(0) - (\lambda_2 + \alpha\delta) \pi_2(0)], \quad \pi_1(0) = \alpha \frac{\lambda_1(1 - p + p\delta) + \delta\beta}{\lambda_1 \alpha + (p\lambda_1 + \beta)(\lambda_2 + \alpha\delta)}, \quad \pi_2(0) = \frac{p\lambda_1(\lambda_2 + \alpha) + \lambda_2 \beta}{\lambda_1 \alpha + (p\lambda_1 + \beta)(\lambda_2 + \alpha\delta)},$$

$$z_{1,2} = \frac{1}{2} [\lambda_1 + \lambda_2 + \alpha + \beta \mp b], \quad b = \sqrt{(\lambda_1 - \lambda_2 - \alpha + \beta)^2 + 4\alpha\beta(1 - \delta)}, \quad z_1 z_2 = \lambda_1 \lambda_2 + \lambda_1 \alpha + \lambda_2 \beta + \alpha\delta\beta.$$

В выражении для совместной плотности (8) $p(\tau_k)$ определены в (7) для $\tau = \tau_k$, $k = 1, 2$.

Согласно постановке задачи рассматривается установившийся (стационарный) режим функционирования потока событий. Это означает, что для частного случая рассматриваемый поток событий является потоком Пальма, или рекуррентным потоком, и интервалы между соседними по времени событиями потока представляют собой независимые случайные величины $\tau_1, \tau_2, \dots, \tau_k, \dots$, распределенные одинаково с плотностью $p(\tau)$, определенной в (7).

Формула (2) в данном случае принимает вид

$$w(\lambda_1 | \tau_k) = w_1 w_2 (1 - e^{-b\tau_k}) / (w_2 - w_1 e^{-b\tau_k}), \quad \tau_k = t - t_k \geq 0, \quad k = 0, 1, \dots$$

Так как $w(\lambda_1 | \tau_k)$, $k = 1, 2, \dots$, не зависит от предыстории, то для любых $k = 1, 2, \dots$

$$w(\lambda_1 | \tau) = w_1 w_2 (1 - e^{-b\tau}) / (w_2 - w_1 e^{-b\tau}), \quad \tau \geq 0. \quad (9)$$

Для начального интервала изменения τ ($t_0 \leq \tau < t_1$) справедлива формула (2), в которой $k = 0$ и $w(\lambda_1 | t_0 + 0) = \pi_1 = \alpha / (p\lambda_1 + \beta + \alpha)$.

Подставляя (9) в уравнение (1) и решая полученное уравнение относительно τ , находим границу критической области

$$\tau^0 = \frac{1}{a(w_2 - w_1)} \ln \frac{w_1(w_2 - 1/2)}{w_2(w_1 - 1/2)} \quad (10)$$

для любого полуинтервала $[t_k, t_{k+1})$, $k = 1, 2, \dots$. Для начального полуинтервала $[t_0, t_1)$ граница критической области τ_0^0 определяется выражением (5), где $k = 0$ и $w(\lambda_1 | t_0 + 0) = \pi_1 = \alpha / (p\lambda_1 + \beta + \alpha)$.

Из (10) следует, что вне зависимости от знака a ($a > 0$ либо $a < 0$) для любого полуинтервала $[t_k, t_{k+1})$, $k = 1, 2, \dots$, справедливо: 1) если $0 \leq w_1 \leq 1/2$, то корень τ^0 не существует, при этом условная вероятность ошибки определяется в виде $P_0(\tau) = w(\lambda_1 | \tau)$, $\tau \geq 0$; 2) если $1/2 < w_1 < 1$, то корень τ^0 существует и $\tau^0 > 0$, при этом условная вероятность ошибки определяется в виде

$$P_0(\tau) = \begin{cases} w(\lambda_1 | \tau), & 0 \leq \tau < \tau^0, \\ 1 - w(\lambda_1 | \tau), & \tau \geq \tau^0. \end{cases}$$

Для полуинтервала $[t_0, t_1)$ условная вероятность ошибки $P_0(\pi_1, \tau_0)$ в зависимости от значений величин π_1 , w_1 определится одной из формул, соответствующих вариантам 1–8 раздела 2.

В силу того что момент вынесения решения t есть некоторая точка, случайным образом падающая на ось времени Ot , плотность вероятности $w(\tau)$ длительности интервала (t_k, t_{k+1}) , в который попала точка t , для рекуррентных потоков определяется в виде [38]:

$$w(\tau) = \frac{\tau p(\tau)}{M\tau}, \quad M\tau = \int_0^{\infty} \tau p(\tau) d\tau, \quad (11)$$

где $p(\tau)$ – плотность вероятности длительности интервала между соседними событиями рекуррентного потока. Тогда безусловная вероятность ошибки P_0 вычисляется по формуле

$$P_0 = \int_0^{\infty} w(\tau) P_0(\tau) d\tau. \quad (12)$$

Для рассматриваемого частного случая $p(\tau)$ определяется в виде (7), где

$$\gamma = \frac{z_2 - \lambda_2}{z_2 - z_1}; \quad \pi_1(0) = 0, \quad \pi_2(0) = 1; \quad z_{1,2} = \frac{1}{2}[\lambda_1 + \lambda_2 + \alpha + \beta \mp b], \quad b = \sqrt{(\lambda_1 - \lambda_2 - \alpha + \beta)^2 + 4\alpha\beta}, \quad 0 < z_1 < z_2.$$

Подставляя (7) в (11), получаем

$$M\tau = \int_0^{\infty} \tau p(\tau) d\tau = [(1 - \gamma)z_1 + \gamma z_2] / z_1 z_2;$$

$$w(\tau) = \tau (A_1 e^{-z_1 \tau} + A_2 e^{-z_2 \tau}), \quad \tau \geq 0, \quad \text{где } A_1 = \gamma z_1 / M\tau \text{ и } A_2 = (1 - \gamma) z_2 / M\tau. \quad (13)$$

Используя формулу (12), найдем явный вид безусловной вероятности ошибки P_0 в зависимости от значения величины w_1 для любого полуинтервала $[t_k, t_{k+1})$, $k = 1, 2, \dots$:

1) если $0 \leq w_1 \leq 1/2$, то корень τ^0 не существует, при этом безусловная вероятность ошибки определяется в виде

$$P_0^{(1)} = \int_0^{\infty} w(\tau) P_0(\tau) d\tau = \int_0^{\infty} w(\tau) w(\lambda_1 | \tau) d\tau = \int_0^{\infty} f(\tau) d\tau, \quad (14)$$

$$f(\tau) = w(\tau) w(\lambda_1 | \tau) = \frac{w_1 w_2 \tau (1 - e^{-b\tau})}{w_2 - w_1 e^{-b\tau}} [A_1 e^{-z_1 \tau} + A_2 e^{-z_2 \tau}], \quad \text{где } A_1, A_2 \text{ определены в (13)}. \quad (15)$$

2) если $1/2 < w_1 < 1$, то корень τ^0 существует и $\tau^0 > 0$, при этом безусловная вероятность ошибки определяется в виде

$$P_0^{(2)} = \int_0^{\infty} w(\tau) P_0(\tau) d\tau = \int_0^{\tau^0} w(\tau) w(\lambda_1 | \tau) d\tau + \int_{\tau^0}^{\infty} w(\tau) (1 - w(\lambda_1 | \tau)) d\tau = \int_{\tau^0}^{\infty} w(\tau) d\tau + \int_0^{\tau^0} f(\tau) d\tau - \int_{\tau^0}^{\infty} f(\tau) d\tau, \quad (16)$$

где $f(\tau)$ определена в (15),

$$\int_{\tau^0}^{\infty} w(\tau) d\tau = A_1 \frac{e^{-z_1 \tau^0}}{z_1} \left(\tau^0 + \frac{1}{z_1} \right) + A_2 \frac{e^{-z_2 \tau^0}}{z_2} \left(\tau^0 + \frac{1}{z_2} \right). \quad (17)$$

Подставляя (17) в (16), окончательно получаем

$$P_0^{(2)} = A_1 \frac{e^{-z_1 \tau^0}}{z_1} \left(\tau^0 + \frac{1}{z_1} \right) + A_2 \frac{e^{-z_2 \tau^0}}{z_2} \left(\tau^0 + \frac{1}{z_2} \right) + \int_0^{\tau^0} f(\tau) d\tau - \int_{\tau^0}^{\infty} f(\tau) d\tau, \quad (18)$$

где $f(\tau)$ определена в (15); A_1, A_2 определены в (13).

Интегралы, входящие в (14), (18), могут быть вычислены только численно.

Для начального полуинтервала $[t_0, t_1)$ безусловная вероятность ошибки P_0 находится аналогично нахождению безусловных вероятностей ошибок $P_0^{(1)}$ и $P_0^{(2)}$ в зависимости от значений величин π_1 , w_1 , соответствующих вариантам 1–8 раздела 2.

3.2. Особый случай: $\lambda_1 - \lambda_2 - \alpha\delta = 0$

В данном случае поведение апостериорной вероятности $w(\lambda_1 | \tau_k)$ на полуинтервале $[t_k, t_{k+1})$ между соседними событиями потока определяется выражением

$$w(\lambda_1 | \tau_k) = \frac{\alpha(1-\delta)}{\beta + \alpha(1-\delta)} + \left[w(\lambda_1 | t_k + 0) - \frac{\alpha(1-\delta)}{\beta + \alpha(1-\delta)} \right] e^{-(\beta + \alpha(1-\delta))\tau_k}, \quad \tau_k = t - t_k \geq 0, \quad k = 0, 1, \dots \quad (19)$$

При этом формула пересчета (3) принимает вид

$$w(\lambda_1 | t_k + 0) = \frac{\alpha\delta + [\lambda_1(1-p) - \alpha\delta]w(\lambda_1 | t_k - 0)}{\lambda_2 + \alpha\delta}, \quad k = 1, 2, \dots$$

Подставляя (19) в уравнение (1) и решая полученное уравнение относительно τ_k , находим границу критической области

$$\tau_k^0 = \frac{1}{\beta + \alpha(1-\delta)} \ln \frac{2[(\beta + \alpha(1-\delta))w(\lambda_1 | t_k + 0) - \alpha(1-\delta)]}{\beta - \alpha(1-\delta)} \quad (20)$$

для любого полуинтервала $[t_k, t_{k+1})$, $k = 1, 2, \dots$. Для начального полуинтервала $[t_0, t_1)$ граница критической области τ_0^0 определяется выражением (20), где $k = 0$ и $w(\lambda_1 | t_0 + 0) = \pi_1 = \alpha / (p\lambda_1 + \beta + \alpha)$.

Если к ограничению $\lambda_1 - \lambda_2 - \alpha\delta = 0$ добавить дополнительное ограничение $\lambda_1(1-p) - \alpha\delta = 0$, то формула пересчета (3) выпишется в виде

$$w(\lambda_1 | t_k + 0) = \alpha\delta / (\lambda_2 + \alpha\delta), \quad k = 1, 2, \dots \quad (21)$$

Равенство (21) означает, что значение апостериорной вероятности $w(\lambda_1 | \tau_k)$ после любого момента t_k не зависит от моментов наступления событий $t_1, t_2, \dots, t_{k-1}$. Далее, подставляя (21) в (20), находим границу критической области

$$\tau_k^0 = \frac{1}{\beta + \alpha(1-\delta)} \ln \frac{2\alpha[\delta\beta - \lambda_2(1-\delta)]}{\lambda_1(\beta - \alpha(1-\delta))} \quad (22)$$

для любого полуинтервала $[t_k, t_{k+1})$, $k = 1, 2, \dots$. Для начального полуинтервала $[t_0, t_1)$ граница критической области τ_0^0 определяется в виде (20), где $k = 0$ и $w(\lambda_1 | t_0 + 0) = \pi_1 = \alpha / (p\lambda_1 + \beta + \alpha)$:

$$\tau_0^0 = \frac{1}{\beta + \alpha(1-\delta)} \ln \frac{2\alpha[\delta\beta - p\lambda_1(1-\delta)]}{(p\lambda_1 + \beta + \alpha)(\beta - \alpha(1-\delta))}. \quad (23)$$

Подчеркнем, что для рассматриваемого особого случая модулированный обобщенный полусинхронный поток событий является рекуррентным потоком. Действительно, можно показать [37], что при реализации условия $\lambda_1 - \lambda_2 - \alpha\delta = 0$ совместная плотность вероятностей длительностей двух смежных интервалов $p(\tau_1, \tau_2)$ факторизуется: $p(\tau_1, \tau_2) = p(\tau_1)p(\tau_2)$.

Обозначим через L выражение под знаком логарифма в (22), (23). Тогда для случая соотношения параметров $\lambda_1 - \lambda_2 - \alpha\delta = 0$, $\lambda_1(1-p) - \alpha\delta = 0$ имеем

1) для начального полуинтервала $[t_0, t_1)$

$$L = \frac{2\alpha[\delta\beta - p\lambda_1(1-\delta)]}{(p\lambda_1 + \beta + \alpha)(\beta - \alpha(1-\delta))}, \quad w(\lambda_1 | t_k + 0) = w(\lambda_1 | t_0 + 0) = \pi_1 = \alpha / (p\lambda_1 + \beta + \alpha);$$

2) для любого полуинтервала $[t_k, t_{k+1})$, $k = 1, 2, \dots$,

$$L = \frac{2\alpha[\delta\beta - \lambda_2(1-\delta)]}{\lambda_1(\beta - \alpha(1-\delta))}, \quad w(\lambda_1 | t_k + 0) = \alpha\delta / (\lambda_2 + \alpha\delta).$$

Тогда в зависимости от значений величин L , $w(\lambda_1 | t_k + 0)$ условная вероятность ошибки $P_0(\tau)$ для любого полуинтервала $[t_k, t_{k+1})$, $k = 0, 1, \dots$, определяется в виде

1) $0 \leq w(\lambda_1 | t_k + 0) < 1/2$:

1.1) если $L \leq 0$, то корень τ^0 не существует, при этом условная вероятность ошибки определяется в виде $P_0(\tau) = w(\lambda_1 | \tau)$, $\tau \geq 0$;

1.2) если $0 < L < 1$, то корень τ^0 существует и $\tau^0 < 0$, при этом условная вероятность ошибки определяется в виде $P_0(\tau) = w(\lambda_1 | \tau)$, $\tau \geq 0$;

1.3) если $L \geq 1$, то корень τ^0 существует и $\tau^0 > 0$, при этом условная вероятность ошибки определяется в виде

$$P_0(\tau) = \begin{cases} w(\lambda_1 | \tau), & 0 \leq \tau < \tau^0, \\ 1 - w(\lambda_1 | \tau), & \tau \geq \tau^0; \end{cases}$$

2) $1/2 \leq w(\lambda_1 | t_k + 0) \leq 1$:

2.1) если $L \leq 0$, то корень τ^0 не существует, при этом условная вероятность ошибки определяется в виде $P_0(\tau) = 1 - w(\lambda_1 | \tau)$, $\tau \geq 0$;

2.2) если $0 < L < 1$, то корень τ^0 существует и $\tau^0 < 0$, при этом условная вероятность ошибки определяется в виде $P_0(\tau) = 1 - w(\lambda_1 | \tau)$, $\tau \geq 0$;

2.3) если $L \geq 1$, то корень τ^0 существует и $\tau^0 > 0$, при этом условная вероятность ошибки определяется в виде

$$P_0(\tau) = \begin{cases} 1 - w(\lambda_1 | \tau), & 0 \leq \tau < \tau^0, \\ w(\lambda_1 | \tau), & \tau \geq \tau^0. \end{cases}$$

Поскольку в рассматриваемом случае поток событий представляет собой рекуррентный поток, то возможно вычисление безусловной вероятности ошибки P_0 по формуле (12). Для рассматриваемого особого случая $p(\tau)$ определяется в виде (7), где

$$\gamma = (z_2 - \lambda_1) / (z_2 - z_1) = 1, \quad 1 - \gamma = -(z_1 - \lambda_1) / (z_2 - z_1) = 0; \quad z_1 = \lambda_1, \quad z_2 = \lambda_2 + \alpha + \beta, \quad 0 < z_1 < z_2.$$

Подставляя (7) в (11), получаем

$$M\tau = \int_0^\infty \tau p(\tau) d\tau = \frac{1}{\lambda_1}; \quad w(\tau) = \lambda_1^2 \tau e^{-z_1 \tau}, \quad \tau \geq 0.$$

Тогда в зависимости от значений величин L , $w(\lambda_1 | t_k + 0)$ безусловная вероятность ошибки P_0 для любого полуинтервала $[t_k, t_{k+1})$, $k = 0, 1, \dots$, определяется в виде

1) $0 \leq w(\lambda_1 | t_k + 0) < 1/2$:

1.1) если $L \leq 0$, то корень τ^0 не существует, при этом безусловная вероятность ошибки определяется в виде

$$P_0^{(1)} = \int_0^\infty w(\tau) P_0(\tau) d\tau = \int_0^\infty w(\tau) w(\lambda_1 | \tau) d\tau = \int_0^\infty f(\tau) d\tau,$$

$$f(\tau) = w(\tau) w(\lambda_1 | \tau) = \lambda_1^2 \tau e^{-z_1 \tau} \left\{ \frac{\alpha(1-\delta)}{\beta + \alpha(1-\delta)} + \left[w(\lambda_1 | t_k + 0) - \frac{\alpha(1-\delta)}{\beta + \alpha(1-\delta)} \right] e^{-(\beta + \alpha(1-\delta))\tau} \right\}.$$

Обозначая через

$$b_1 = \frac{\alpha(1-\delta)}{\beta + \alpha(1-\delta)}, \quad b_2 = w(\lambda_1 | t_k + 0) - \frac{\alpha(1-\delta)}{\beta + \alpha(1-\delta)}, \quad (24)$$

получаем $f(\tau)$ в виде

$$f(\tau) = \lambda_1^2 \tau e^{-z_1 \tau} [b_1 + b_2 e^{-(\beta + \alpha(1-\delta))\tau}]. \quad (25)$$

Учитывая, что $\int_0^\infty \tau e^{-z\tau} d\tau = 1/z^2$, окончательно получаем

$$P_0^{(1)} = \int_0^\infty f(\tau) d\tau = b_1 + \frac{\lambda_1^2 b_2}{(z_1 + \beta + \alpha(1-\delta))^2}, \quad (26)$$

где b_1, b_2 определены в (24);

1.2) если $0 < L < 1$, то корень τ^0 существует и $\tau^0 < 0$, при этом безусловная вероятность ошибки определяется выражением (26);

1.3) если $L \geq 1$, то корень τ^0 существует и $\tau^0 > 0$, при этом безусловная вероятность ошибки определяется в виде

$$\begin{aligned} P_0^{(1)} &= \int_0^\infty w(\tau) P_0(\tau) d\tau = \int_0^{\tau^0} w(\tau) w(\lambda_1 | \tau) d\tau + \int_{\tau^0}^\infty w(\tau) (1 - w(\lambda_1 | \tau)) d\tau = \int_{\tau^0}^\infty w(\tau) d\tau + \int_0^{\tau^0} f(\tau) d\tau - \\ &- \int_{\tau^0}^\infty f(\tau) d\tau \doteq D_1 + D_2; \quad D_1 = \int_{\tau^0}^\infty w(\tau) d\tau = e^{-z_1 \tau^0} (1 + z_1 \tau^0); \quad D_2 = \int_{\tau^0}^\infty f(\tau) d\tau - \int_{\tau^0}^\infty f(\tau) d\tau = \\ &= b_1 \left[1 - 2e^{-z_1 \tau^0} (1 + z_1 \tau^0) \right] + \frac{\lambda_1^2 b_2}{(z_1 + \beta + \alpha(1-\delta))^2} \left[1 - 2e^{-(z_1 + \beta + \alpha(1-\delta)) \tau^0} (1 + (z_1 + \beta + \alpha(1-\delta)) \tau^0) \right], \end{aligned} \quad (27)$$

где b_1, b_2 определены в (24);

2) $1/2 \leq w(\lambda_1 | t_k + 0) \leq 1$:

2.1) если $L \leq 0$, то корень τ^0 не существует, при этом безусловная вероятность ошибки определяется в виде

$$P_0^{(2)} = \int_0^\infty w(\tau) P_0(\tau) d\tau = \int_0^\infty w(\tau) [1 - w(\lambda_1 | \tau)] d\tau = \int_0^\infty w(\tau) d\tau - \int_0^\infty f(\tau) d\tau,$$

где $f(\tau) = w(\tau) w(\lambda_1 | \tau)$ и определена в (25), $\int_0^\infty w(\tau) d\tau = 1$. Тогда

$$P_0^{(2)} = 1 - b_1 - \frac{\lambda_1^2 b_2}{(z_1 + \beta + \alpha(1-\delta))^2}; \quad (28)$$

2.2) если $0 < L < 1$, то корень τ^0 существует и $\tau^0 < 0$, при этом безусловная вероятность ошибки определяется выражением (28);

2.3) если $L \geq 1$, то корень τ^0 существует и $\tau^0 > 0$, при этом безусловная вероятность ошибки определяется в виде

$$\begin{aligned} P_0^{(2)} &= \int_0^\infty w(\tau) P_0(\tau) d\tau = \int_0^{\tau^0} w(\tau) [1 - w(\lambda_1 | \tau)] d\tau + \int_{\tau^0}^\infty w(\tau) w(\lambda_1 | \tau) d\tau = \int_0^{\tau^0} w(\tau) d\tau - \\ &- \int_0^{\tau^0} f(\tau) d\tau + \int_{\tau^0}^\infty f(\tau) d\tau \doteq D_3 - D_2, \end{aligned}$$

где D_2 определено в (27), $D_3 = \int_0^{\tau^0} w(\tau) d\tau = 1 - e^{-z_1 \tau^0} (1 + z_1 \tau^0)$.

4. Результаты численных расчетов

В настоящем разделе приводятся результаты статистических экспериментов по вычислению условной вероятности вынесения ошибочного решения $P_0(w(\lambda_1 | t_k + 0), \tau_k)$ в любой момент времени $\tau_k \geq 0$, $k = 0, 1, \dots$, для общего случая и безусловной вероятности ошибки P_0 для особого случая соотношения параметров ($\lambda_1 - \lambda_2 - \alpha\delta = 0$, $\lambda_1(1-p) - \alpha\delta = 0$) для различных значений параметров модулированного обобщенного полусинхронного потока событий.

Приведенный в разделе 2 алгоритм расчета условной вероятности $P_0(w(\lambda_1 | t_k + 0), \tau_k)$ в любой момент времени $\tau_k \geq 0$, $k = 0, 1, \dots$, положен в основу реализации программы расчета для проведения статистиче-

ских экспериментов с целью вычисления условной вероятности ошибки $P_0(w(\lambda_1 | t_k + 0), \tau_k)$ для общего случая. Программа расчета реализована на языке программирования C# в интегрированной среде разработки Microsoft Visual Studio 2015 в виде пользовательского приложения и интерфейса командной строки.

На первом этапе эксперимента осуществляется имитационное моделирование потока при заданных значениях параметров потока и заданном времени моделирования и как результат получение истинной траектории интенсивности процесса $\lambda(t)$ и временных моментов $t_1, t_2, \dots$ наступления событий потока. Описание алгоритма имитационного моделирования здесь не приводится, так как никаких принципиальных трудностей он не содержит. На втором этапе производится непосредственное вычисление апостериорных вероятностей $w(\lambda_1 | \tau_0)$, $\tau_0 = t - t_0 \geq 0$; $w(\lambda_1 | t_k + 0)$, $k = 1, 2, \dots$; $w(\lambda_1 | \tau_k)$, $\tau_k = t - t_k \geq 0$, $k = 1, 2, \dots$, по формулам (2), (3); построение оценки $\hat{\lambda}(t)$ процесса $\lambda(t)$ и вычисление условной вероятности ошибки $P_0(w(\lambda_1 | t_k + 0), \tau_k)$, $k = 0, 1, \dots$, для общего случая.

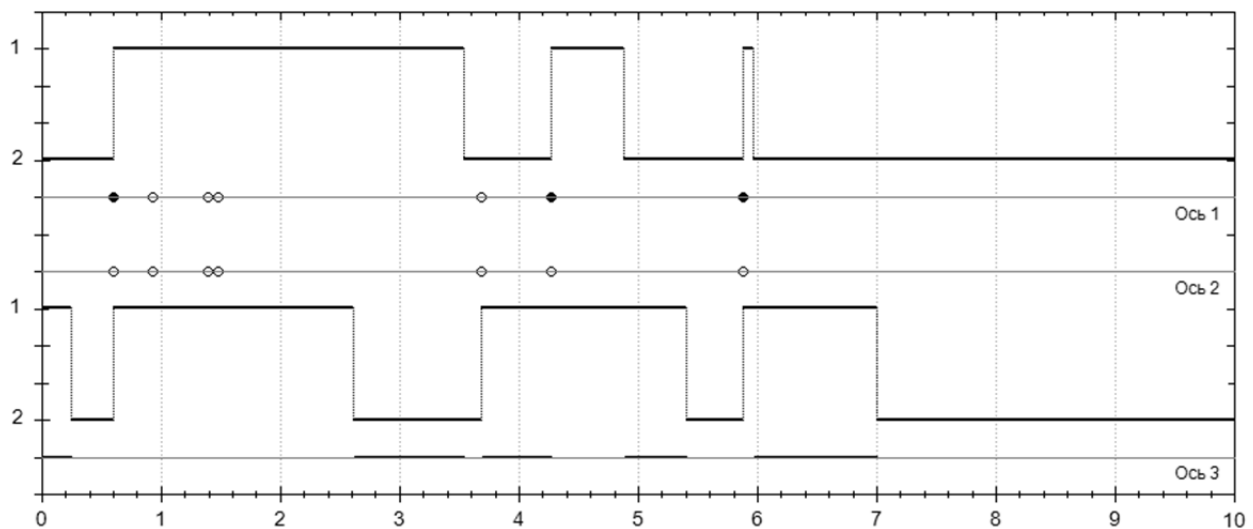


Рис. 2. Траектории процесса $\lambda(t)$ (верхняя часть рисунка) и оценки $\hat{\lambda}(t)$ (нижняя часть рисунка)

В качестве иллюстрации на рис. 2 приведен пример реализации потока событий при следующих значениях параметров: $\lambda_1 = 0,8$, $\lambda_2 = 0,2$, $p = 0,2$, $\beta = 0,5$, $\alpha = 0,8$, $\delta = 0,9$ и общем времени моделирования $T_m = 1000$. В верхней части рис. 2 приведена траектория случайного процесса $\lambda(t)$ (истинная траектория процесса $\lambda(t)$), полученная путем имитационного моделирования. Цифрами 1, 2 обозначены первое и второе состояния процесса $\lambda(t)$ соответственно. На оси 1 отображены события модулированного обобщенного полусинхронного потока событий: белыми кружками показаны события пуассоновских потоков, черными кружками – дополнительные события, которые могут наступить в первом состоянии потока при переходе процесса $\lambda(t)$ из второго состояния в первое. Напомним, что события пуассоновских потоков и дополнительные события неразличимы для наблюдателя, поэтому на оси 2 все события потока отображены одинаково (белыми кружками). В нижней части рис. 2 приведена траектория оценки $\hat{\lambda}(t)$ процесса $\lambda(t)$. Цифрами 1, 2 обозначены первое и второе состояния оценки $\hat{\lambda}(t)$ соответственно. На оси 3 жирными линиями отмечены промежутки, на которых оценка $\hat{\lambda}(t)$ не совпадает с истинным значением процесса $\lambda(t)$ (область ошибочных решений). Расчет апостериорной вероятности $w(\lambda_1 | \tau_k)$, $k = 0, 1, \dots$, и вынесение решения о состоянии процесса $\lambda(t)$ производились с шагом $\Delta t = 0,001$.

На рис. 3 приведена траектория поведения апостериорной вероятности $w(\lambda_1 | \tau_k)$, $k = 0, 1, \dots$, первого состояния процесса $\lambda(t)$, соответствующая полученной при имитационном моделировании последовательности наступления событий $t_1, t_2, \dots$ модулированного обобщенного полусинхронного потока

событий (см. рис. 2, ось 2). Цифрами 1; 0,5; 0 обозначены возможные значения апостериорной вероятности по оси ординат. На рис. 4 приведена траектория условной вероятности ошибки $P_0(w(\lambda_1 | t_k + 0), \tau_k)$, $k = 0, 1, \dots$, соответствующая той же последовательности наступления событий. Цифрами 1, 0,5, 0 обозначены возможные значения условной вероятности по оси ординат.

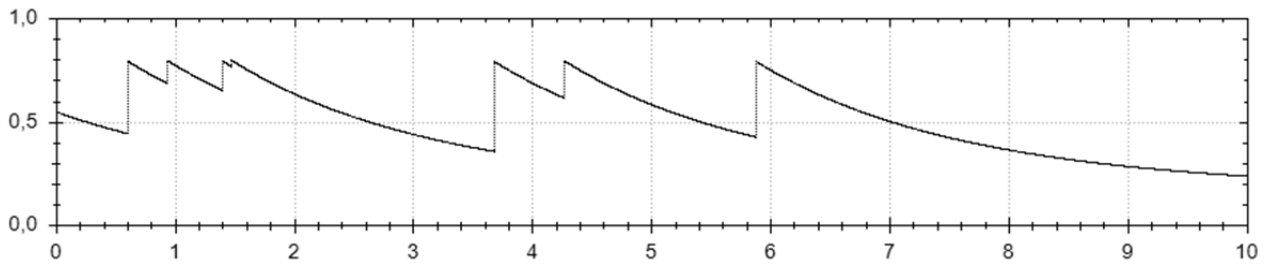


Рис. 3. Траектория апостериорной вероятности $w(\lambda_1 | \tau_k)$, $k = 0, 1, \dots$

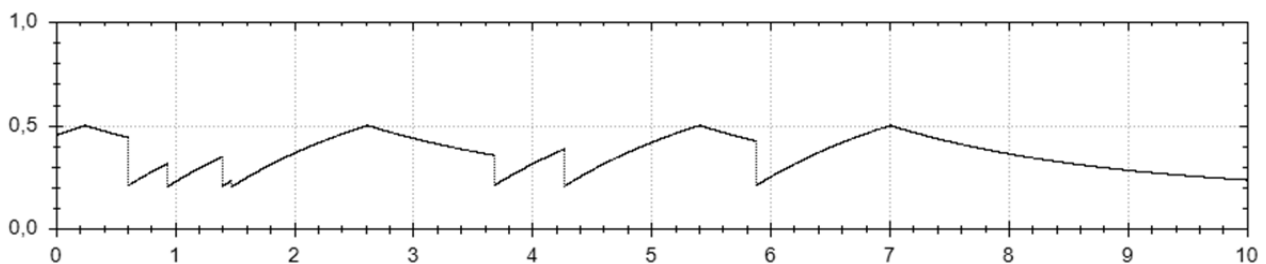


Рис. 4. Траектория условной вероятности ошибки $P_0(w(\lambda_1 | t_k + 0), \tau_k)$, $k = 0, 1, \dots$

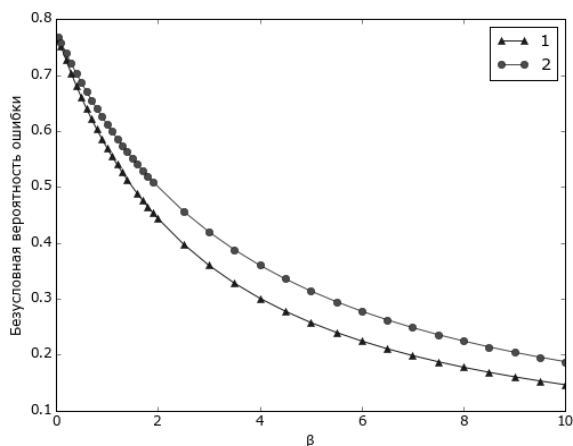


Рис. 5. Графики изменения значений безусловной вероятности ошибки P_0

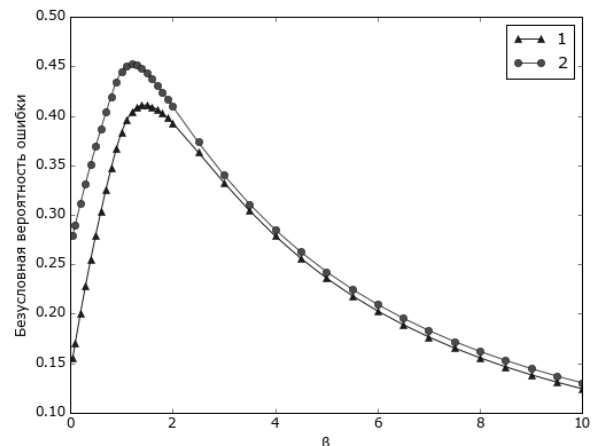


Рис. 6. Графики изменения значений безусловной вероятности ошибки P_0

Таким образом, при реализации предложенного алгоритма осуществляется оценка состояний процесса $\lambda(t)$ в любой момент времени t , и в этот же момент времени вычисляется условная вероятность ошибки вынесения решения.

Для особого случая соотношения параметров ($\lambda_1 - \lambda_2 - \alpha\delta = 0$, $\lambda_1(1 - p) - \alpha\delta = 0$) построены графики безусловной вероятности ошибки P_0 для различных значений параметров модулированного обобщенного полусинхронного потока событий и для любого полуинтервала $[t_k, t_{k+1})$, $k = 1, 2, \dots$. Алгоритм расчета состоит из следующих этапов: 1) задается β ; 2) вычисляется L , $w(\lambda_1 | t_k + 0)$; 3) определяется один из шести возможных вариантов (см. подраздел 3.2) и вычисляется значение P_0 для этого варианта; 4) значение β увеличивается; 5) алгоритм переходит на шаг 1 и т.д.

На рис. 5 представлены графики изменения значений безусловной вероятности ошибки P_0 в зависимости от значений параметров потока: 1 – при $\lambda_1 = 3$, $\lambda_2 = 1,5$, $p = 0,5$, $\alpha = 3$, $\delta = 0,5$; 2 – при $\lambda_1 = 4$, $\lambda_2 = 2$, $p = 0,5$, $\alpha = 4$, $\delta = 0,5$. На рис. 6 приведены графики изменения значений безусловной вероятности ошибки P_0 в зависимости от значений параметров потока: 1 – при $\lambda_1 = 4$, $\lambda_2 = 0,8$, $p = 0,2$, $\alpha = 4$, $\delta = 0,8$; 2 – при $\lambda_1 = 5$, $\lambda_2 = 1,8$, $p = 0,36$, $\alpha = 4$, $\delta = 0,8$. Параметр β изменяется по оси абсцисс и принимает значения 0,05; 0,1; 0,2 и так далее до 10.

Заключение

Алгоритм оптимального оценивания состояний модулированного обобщенного полусинхронного потока событий, предложенный в работах [34, 35], позволяет осуществлять оценку состояний процесса $\lambda(t)$ по результатам наблюдений за моментами наступлений событий в потоке. При реализации предложенного алгоритма осуществляется оценка состояний процесса $\lambda(t)$ в любой момент времени t , и в этот же момент времени вычисляется условная вероятность сделанной при вынесении решения ошибки. Для частного и особого случаев (при некоторых дополнительных ограничениях на параметры потока) вычисление безусловной вероятности ошибки возможно в явном виде, что позволяет при заданном наборе параметров определить значение безусловной вероятности ошибки до начала наблюдений за потоком событий и не привлекая методы имитационного моделирования.

ЛИТЕРАТУРА

1. Cox D.R. Some Statistical Methods Connected with Series of Events // J. Royal Statistical Society B. 1955. V. 17. P. 129–164.
2. Kingman Y.F.C. On doubly stochastic Poisson process // Proceedings of Cambridge Philosophical Society. 1964. V. 60, No. 4. P. 923–930.
3. Basharin G.P., Kokotushkin V.A., Naumov V.A. Method of equivalent substitutions for calculating fragments of communication networks for digital computer // Engineering cybernetics. 1979. V. 17 (6). P. 66–73.
4. Башарин Г.П., Кокотушкин В.А., Наумов В.А. О методе эквивалентных замен расчета фрагментов сетей связи // Изв. АН СССР. Техн. кибернетика. 1980. № 1. С. 55–61.
5. Neuts M.F. A versatile Markov point process // Journal of Applied Probability. 1979. V.16. P. 764–779.
6. Cox D. R., Isham V. Point Processes. London : Chapman & Hall, 1980.
7. Bremaud P. Point Processes and Queues: Martingale Dynamics. New York : Springer-Verlag, 1981.
8. Last G., Brandt A. Marked Point Process on the Real Line: The Dynamic Approach. New York : Springer-Verlag, 1995.
9. Нежелская Л.А. Оптимальное оценивание состояний полусинхронного потока событий в условиях его частичной наблюдаемости // Вестник Томского государственного университета. 2000. № 269. С. 95–98.
10. Горцев А.М., Нежелская Л.А. Оценивание параметров полусинхронного дважды стохастического потока событий методом моментов // Вестник Томского государственного университета. 2002. № 1. С. 18–23.
11. Gortsev A.M., Nezhelskaya L.A. Estimation of the dead-time period and parameters of a semi-synchronous double-stochastic stream of events // Measurement Techniques. 2003. V. 46, No. 6. P. 536–545.
12. Горцев А.М., Нежелская Л.А. Полусинхронный дважды стохастический поток событий при продлеваемом мертвом времени // Вычислительные технологии. 2008. Т. 13, № 1. С. 31–41.
13. Горцев А.М., Калягин А.А., Нежелская Л.А. Оптимальная оценка состояний обобщенного полусинхронного потока событий // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2010. № 2 (11). С. 66–81.
14. Горцев А.М., Калягин А.А. Оптимальная оценка состояний обобщенного полусинхронного потока событий в условиях непродлеваемого мертвого времени // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2010. № 4 (13). С. 50–60.
15. Горцев А.М., Калягин А.А. Вероятность ошибочных решений при оценивании состояний обобщенного полусинхронного потока событий // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2012. № 1 (18). С. 58–70.
16. Горцев А.М., Калягин А.А. Совместная плотность вероятностей длительности интервалов обобщенного полусинхронного потока // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2012. № 2 (19). С. 80–87.
17. Горцев А.М., Калягин А.А., Нежелская Л.А. Совместная плотность вероятностей длительности интервалов обобщенного полусинхронного потока событий при непродлеваемом мертвом времени // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2014. № 2 (27). С. 19–29.

18. Горцев А.М., Калягин А.А., Нежелская Л.А. Оценка максимального правдоподобия длительности мертвого времени в обобщенном полусинхронном потоке // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2015. № 1 (30). С. 27–37.
19. Калягин А.А., Нежелская Л.А. Сравнение МП- и ММ-оценок длительности мертвого времени в обобщенном полусинхронном потоке событий // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2015. № 3 (32). С. 23–32.
20. Горцев А.М., Нежелская Л.А., Шевченко Т.И. Оценивание состояний МС-потока событий при наличии ошибок измерений // Известия высших учебных заведений. Физика. 1993. № 12. С. 67–85.
21. Горцев А.М., Зуевич В.Л. Оптимальная оценка состояний асинхронного дважды стохастического потока событий с произвольным числом состояний // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2010. № 2 (11). С. 44–65.
22. Горцев А.М., Нежелская Л.А. О связи МС-потоков и МАР-потоков событий // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2011. № 1 (14). С. 13–21.
23. Горцев А.М., Леонова М.А. Оптимальная оценка состояний обобщенного асинхронного дважды стохастического потока // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2010. № 1 (10). С. 33–47.
24. Горцев А.М., Леонова М.А., Нежелская Л.А. Совместная плотность вероятностей длительности интервалов обобщенного асинхронного потока событий при непродлеваемом мертвом времени // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2012. № 4 (21). С. 14–25.
25. Леонова М.А., Нежелская Л.А. Оценка максимального правдоподобия длительности мертвого времени в обобщенном асинхронном потоке событий // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2013. № 2 (23). С. 54–63.
26. Горцев А.М., Нежелская Л.А. Оценивание длительности «мертвого времени» и интенсивностей синхронного дважды стохастического потока событий // Радиотехника. 2004. № 10. С. 8–16.
27. Васильева Л.А., Горцев А.М. Оценивание параметров дважды стохастического потока событий в условиях его неполной наблюдаемости // Автоматика и телемеханика. 2002. № 3. С. 179–184.
28. Горцев А.М., Шмырин И.С. Оптимальная оценка состояний дважды стохастического потока событий при наличии ошибок в измерениях моментов времени // Автоматика и телемеханика. 1999. № 1. С. 52–66.
29. Горцев А.М., Паршина М.Е. Оценивание параметров альтернирующего потока событий в условиях «мертвого времени» // Изв. вузов. Физика. 1999. № 4. С. 8–13.
30. Горцев А.М., Завгородняя М.Е. Оценка параметров альтернирующего потока событий при условии его частичной наблюдаемости // Оптика атмосферы и океана. 1997. Т. 10, № 3. С. 273–280.
31. Горцев А.М., Климов И.С. Оценка интенсивности пуассоновского потока событий в условиях частичной его ненаблюдаемости // Радиотехника. 1991. № 12. С. 3–7.
32. Горцев А.М., Нежелская Л.А. Оценка параметров синхронного альтернирующего пуассоновского потока событий методом моментов // Радиотехника. 1995. № 7–8. С. 6–10.
33. Бахолдина М.А. Совместная плотность вероятностей длительности интервалов модулированного обобщенного полусинхронного потока событий при непродлеваемом мертвом времени и условия его рекуррентности // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2015. № 2 (31). С. 4–17.
34. Бахолдина М.А. Оптимальная оценка состояний модулированного обобщенного полусинхронного потока событий // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2013. № 2 (23). С. 10–21.
35. Bakholdina M.A., Gortsev A.M. Optimal estimation of the states of modulated semi-synchronous integrated flow of events in condition of its incomplete observability // Applied Mathematical Sciences. 2015. V. 9, No. 29. P. 1433–1451.
36. Левин Б.Р. Теоретические основы статистической радиотехники. М. : Советское радио, 1968.
37. Bakholdina M., Gortsev A. Joint probability density of the intervals length of the modulated semi-synchronous integrated flow of events and its recurrence conditions // Communications in Computer and Information Science. 2014. V. 487. P. 18–25.
38. Вентцель Е.С., Овчаров Л.А. Теория случайных процессов и ее инженерные приложения. М. : Высшая школа, 2000. 383 с.

Бахолдина Мария Алексеевна. E-mail: maria.bakholdina@gmail.com
Томский государственный университет

Поступила в редакцию 19 октября 2015 г.

Bakholdina Maria A. (Tomsk State University, Russian Federation). **Probability of error on the states estimation of the modulated generalized semi-synchronous flow of events.**

Keywords: modulated semi-synchronous generalized flow of events; doubly stochastic Poisson process (DSPP); Markovian arrival process (MAP); a posteriori probability of the flow state; state estimation; criterion of a posteriori probability maximum; probability of error on estimation of the flow states.

DOI: 10.17223/19988605/34/3

In this paper, we consider the modulated semi-synchronous generalized flow of events, which is one of the mathematical models for incoming streams of events in computer communication networks and which is related to the class of doubly stochastic Poisson process-

es (DSPPs). The flow intensity process is a piecewise constant stationary random process $\lambda(t)$ with two states λ_1 and λ_2 ($\lambda_1 > \lambda_2 \geq 0$) (first and second states correspondingly). During the time interval with $\lambda(t) = \lambda_i$, a Poisson flow of events with intensity λ_i , $i = 1, 2$, arrives. At any moment of an event occurrence in state 1 of the process $\lambda(t)$, the process $\lambda(t)$ can change its state to state 2 with probability p ($0 \leq p \leq 1$) or continue to stay in state 1 with complementary probability $1 - p$ (i.e., after an event occurrence the process $\lambda(t)$ can change or not change its state from state 1 to state 2). The transition of the process $\lambda(t)$ from the first state to the second one is also possible at any moment that does not coincide with the moment of an event occurrence, in this case the duration of the process $\lambda(t)$ staying in the first state is distributed according to the exponential law with parameter β : $F(\tau) = 1 - e^{-\beta\tau}$, $\tau \geq 0$. Then the duration of the process $\lambda(t)$ staying in the first state is distributed according to the exponential law with distribution function $F_1(\tau) = 1 - e^{-(p\lambda_1 + \beta)\tau}$, $\tau \geq 0$. The transition of the process $\lambda(t)$ from state 2 to state 1 at the moment of an event occurrence in the second state is impossible. The duration of the process $\lambda(t)$ staying in the second state is distributed according to the exponential law with the parameter α : $F_2(\tau) = 1 - e^{-\alpha\tau}$, $\tau \geq 0$. Also, at the moment when the state changes from the second to the first state, an additional event is assumed to be initiated with probability δ ($0 \leq \delta \leq 1$).

So, we present analytical and numerical results of obtaining the conditional and unconditional error probability on the estimation of the flow states. For the general case the algorithm of calculating the conditional error probability $P_0(w(\lambda_1 | t_k + 0) | \tau_k)$ at any moment $\tau_k \geq 0$, $k = 0, 1, \dots$, is proposed. Also, for the special cases of the flow parameters relation expressions for unconditional error probability calculation are obtained explicitly.

REFERENCES

1. Cox, D.R. (1955) Some Statistical Methods Connected with Series of Events. *J. Royal Statistical Society B.* 17. pp. 129-164. DOI: 10.2307/2983950
2. Kingman, Y.F.C. (1964) On doubly stochastic Poisson process. *Proceedings of Cambridge Philosophical Society.* 60(4). pp. 923-930.
3. Basharin, G.P., Kokotushkin, V.A. & Naumov, V.A. (1979) Method of equivalent substitutions for calculating fragments of communication networks for digital computer. *Engineering cybernetics.* 17(6). pp. 66-73. DOI: 10.1016/0166-5316(84)90009-9
4. Basharin, G.P., Kokotushkin, V.A. & Naumov, V.A. (1980) O metode ekvivalentnykh zamen rascheta fragmentov setey svyazi. *Izv. AN SSSR. Tekhn. kibernetika.* 1. pp. 55-61.
5. Neuts, M.F. (1979) A versatile Markov point process. *Journal of Applied Probability.* 16. pp. 764-779. DOI: 10.2307/3213143
6. Cox, D.R. & Isham, V. (1980) *Point Processes*. London: Chapman & Hall.
7. Bremaud, P. (1981) *Point Processes and Queues: Martingale Dynamics*. New York: Springer-Verlag.
8. Last, G. & Brandt, A. (1995) *Marked Point Process on the Real Line: The Dynamic Approach*. New York: Springer-Verlag.
9. Nezhelskaya, L.A. (2000) Optimal'noe otsenivanie sostoyaniy polusinkhronnogo potoka sobyitiy v usloviyakh ego chastichnoy nablyudaemosti [Optimal state estimation of semi-synchronous flow in conditions of its incomplete observability]. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal.* 269. pp. 95-98.
10. Gortsev, A.M. & Nezhelskaya, L.A. (2002) Otsenivanie parametrov polusinkhronnogo dvazhdy stokhasticheskogo potoka sobyitiy metodom momentov [Parameters estimation of a semi-synchronous doubly stochastic flow of events using method of moments]. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal.* 1. pp. 18-23.
11. Gortsev, A.M. & Nezhelskaya, L.A. (2003) Estimation of the dead-time period and parameters of a semi-synchronous double-stochastic stream of events. *Measurement Techniques.* 46(6). pp. 536-545. DOI: 10.1023/A:1025499509015
12. Gortsev, A.M. & Nezhelskaya, L.A. (2008) Semi-synchronous twice-stochastic event flow in conditions of prolonged dead time. *Vychislitel'nye tekhnologii – Computational Technologies.* 13(1). pp. 31-41. (In Russian).
13. Gortsev, A.M., Kalyagin, A.A. & Nezhelskaya, L.A. (2010) Optimal state estimation of generic semi-synchronous flow of events. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science.* 2(11). pp. 66-81. (In Russian).
14. Gortsev, A.M. & Kalyagin, A.A. (2010) Optimal states estimation of generalized semi-synchronous flow of events in conditions of constant dead time. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science.* 4(13). pp. 50-60. (In Russian).
15. Gortsev, A.M. & Kalyagin, A.A. (2012) The probability of wrong decisions in the estimation of states of a generalized semi-synchronous flow of events. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science.* 1(18). pp. 58-70. (In Russian).
16. Gortsev, A.M. & Kalyagin, A.A. (2012) The joint probability density of duration of the intervals in a generalized semi-synchronous flow. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science.* 2(19). pp. 80-87. (In Russian).
17. Gortsev, A.M., Kalyagin, A.A. & Nezhelskaya, L.A. (2014) The joint probability density of duration of the intervals in a generalized semi-synchronous flow of events with unprolonging dead time. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science.* 2(27). pp. 19-29. (In Russian).

18. Gortsev, A.M., Kalyagin, A.A. & Nezhelskaya, L.A. (2015) Maximum likelihood estimation of dead time value at a generalized semi-synchronous flow of events. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 1(30). pp. 27-37. (In Russian).
19. Kalyagin, A.A. & Nezhelskaya, L.A. (2015) Comparison of MP- and MM-estimations of dead time in generic semisynchronous flow of events. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 3(32). pp. 23-32. (In Russian).
20. Gortsev, A.M., Nezhelskaya, L.A. & Shevchenko, T.I. (1993) Optimal states estimation of asynchronous doubly stochastic flow of events with arbitrary number of states. *Russian Physics Journal*. 12. pp. 67-85. (In Russian).
21. Gortsev, A.M. & Zuevich, V.L. (2010) Optimal state estimation of generic asynchronous doubly stochastic flow with random number of states. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2(11). pp. 44-65. (In Russian).
22. Gortsev, A.M. & Nezhelskaya, L.A. (2011) About connectivity of MC-flows and MAP-flows of events. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 1(14). pp. 13-21. (In Russian).
23. Gortsev, A.M. & Leonova, M.A. (2010) Optimal state estimation of generic asynchronous doubly stochastic flow. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 1(10). pp. 33-47. (In Russian).
24. Gortsev, A.M., Leonova, M.A. & Nezhelskaya, L.A. (2012) The joint probability density of duration of the intervals in generalized asynchronous flow of events with unprolonging dead time. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 4(21). pp. 14-25. (In Russian).
25. Leonova, M.A. & Nezhelskaya, L.A. (2013) Maximum likelihood estimation of dead time value at a generalized asynchronous flow of events. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2(23). pp. 54-63. (In Russian).
26. Gortsev, A.M. & Nezhelskaya, L.A. (2004) Estimation of the dead time period and intensities of the synchronous double stochastic event flow. *Radiotekhnics*. 10. pp. 8-16.
27. Vasilyeva, L.A. & Gortsev, A.M. (2002) Parameter estimation of a doubly stochastic flow of events under incomplete observability. *Avtomatika i Telemekhanika*. 3. pp. 179-184.
28. Gortsev, A.M. & Shmyrin, I.S. (1999) Optimal states estimation of the doubly stochastic flow of events in the presence of errors in time moments measurements. *Automation and Remote Control*. 1. pp. 52-66.
29. Gortsev, A.M. & Parshina, M.E. (1999) Otsenivanie parametrov al'terniruyushchego potoka sobytiy v usloviyakh "mertvogo vremeni" [Parameter estimation of alternating flow of events under conditions of dead time]. *Russian Physics Journal*. 4. pp. 8-13.
30. Gortsev, A.M. & Zavgorodnyaya, M.E. (1997) Otsenka parametrov al'terniruyushchego potoka sobytiy pri uslovii ego chastichnoy nablyudaemosti [Parameter estimation of alternating flow of events under conditions of particulate observability]. *Optika atmosfery i okeana – Atmospheric and Oceanic Optics*. 10(3). pp. 273-280.
31. Gortsev, A.M. & Klimov, I.S. (1991) Intensity estimation of the Poisson flow of events in conditions of its incomplete observability. *Radiotekhnika – Radioengineering*. 12. pp. 3-7. (In Russian).
32. Gortsev, A.M. & Nezhelskaya, L.A. (1995) Parameters estimation of the synchronous alternating Poisson flow of events with the method of moments. *Radiotekhnika – Radioengineering*. 7-8. pp. 6-10. (In Russian).
33. Bakholdina, M.A. (2015) Joint probability density of the intervals length of the modulated semi-synchronous integrated flow of events in conditions of constant dead time and flow recurrence conditions. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2(31). pp. 4-17. (In Russian).
34. Bakholdina, M.A. (2013) Optimal states estimation of the modulated semi-synchronous integrated flow of events in conditions of constant dead time. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2(23). pp. 10-21. (In Russian).
35. Bakholdina, M.A. & Gortsev, A.M. (2015) Optimal estimation of the states of modulated semi-synchronous integrated flow of events in condition of its incomplete observability. *Applied Mathematical Sciences*. 9(29). pp. 1433-1451. DOI: 10.12988/ams.2015.5135
36. Levin, B.R. (1968) *Teoreticheskie osnovy statisticheskoy radiotekhniki* [The theory of statistical radio engineering]. Moscow: Sovetskoe radio.
37. Bakholdina, M. & Gortsev, A. (2014) Joint probability density of the intervals length of the modulated semi-synchronous integrated flow of events and its recurrence conditions. *Communications in Computer and Information Science*. 487. pp. 18-25. DOI: 10.1007/978-3-319-13671-4_3
38. Ventcel, E.S. & Ovcharov, L.A. (2000) *Teoriya sluchaynykh protsessov i ee inzhenernye prilozheniya* [The theory of random processes and its engineering applications]. Moscow: Vysshaya shkola.

С.С. Вожов, Е.В. Чимитова

**ПРОВЕРКА ГИПОТЕЗЫ О ВИДЕ РАСПРЕДЕЛЕНИЯ
ПО ИНТЕРВАЛЬНЫМ ДАННЫМ**

*Работа выполнена при поддержке Министерства образования и науки РФ
в рамках проектной части государственного задания (№ 2.541.2014К).*

Рассматривается ICM (Iterative Convex Minorant)-алгоритм построения непараметрической оценки функции распределения по интервальным данным. Предлагаются критерии согласия, основанные на расстоянии между предполагаемой функцией распределения и её непараметрической оценкой. Применение критериев согласия опирается на статистическое моделирование условных распределений статистик данных критериев в интерактивном режиме проверки гипотезы о виде распределения.

Ключевые слова: интервальные данные, алгоритм Тёрнбулла, ICM-алгоритм, критерии согласия.

Основы анализа интервальных данных были заложены в метрологии, где интервал неопределённости вводится естественным образом. Предполагается, что каждое наблюдение – это величина, измеренная прибором, который имеет абсолютную ошибку измерения Δ . Таким образом, если точное значение наблюдаемого параметра объекта равно $\dot{x}$, ошибка измерения равна $e \in [-\Delta, \Delta]$, то точечное наблюдение выглядит как $x = \dot{x} + e$. В этом случае мы имеем дело с точечной выборкой $\mathbb{X}_n = \{X_1, \dots, X_n\}$. При проведении статистического анализа на основе точечных выборок мы, по сути, не учитываем величину погрешности измерения Δ . Однако существует и другой подход, который заключается в представлении наблюдения в виде интервала $(x - \Delta, x + \Delta) = (L, R)$, тогда для выборки из n наблюдений получаем интервальную выборку вида

$$\mathbb{I}_n = \{(L_1, R_1), \dots, (L_n, R_n)\}.$$

Интервальные наблюдения рассматриваются во многих современных исследованиях, например в [1–8]. В каждой из них аргументируется необходимость развития методов математической статистики, когда результаты наблюдений – не числа, а интервалы.

1. Непараметрическая оценка функции распределения. ICM-алгоритм

Основная идея построения непараметрической оценки функции распределения по интервальным данным заключается в том, что находится максимум логарифма функции правдоподобия

$$\ln L(\mathbb{I}_n) = \sum_{i=1}^n \ln(F(R_i) - F(L_i))$$

по значениям функции распределения в граничных точках интервалов наблюдений при соблюдении условия монотонности функции распределения. Однако решение данной оптимизационной задачи методом штрафных функций требует больших вычислительных ресурсов. Вместо этого целесообразно использование специальных алгоритмов. Одним из таких алгоритмов является алгоритм Тёрнбулла [9], подробное описание которого представлено в [5, 10, 11]. Однако данный алгоритм требует значительных вычислительных затрат, поэтому рекомендуется использовать ICM-алгоритм, предложенный в [12–14]. В [10, 11] показано, что время вычисления оценки по ICM-алгоритму существенно меньше, чем по алгоритму Тёрнбулла.

Идея ИСМ-алгоритма заключается в том, чтобы свести задачу максимизации функции правдоподобия к задаче построения изотонической регрессии [12]. Для оценивания функции распределения $F(x)$ вводится разбиение $0 < \tau_0 < \tau_1 < \dots < \tau_m$, состоящее из всех неповторяющихся упорядоченных границ интервалов L_i и R_i , $i = \overline{1, n}$.

В ИСМ-алгоритме для всех $j = \overline{1, m}$ и $i = \overline{1, n}$ определяются веса:

$$\alpha_{ji} = \begin{cases} 1, & \text{если } \tau_j = L_i, \\ -1, & \text{если } \tau_j = R_i, \\ 0, & \text{если } \tau_j \notin (L_i, R_i]. \end{cases}$$

Итерационный процесс ИСМ-алгоритма для произвольного шага алгоритма $k \geq 1$:

1. Для всех $j = \overline{1, m}$ найти точки $(G_j^{(k)}, V_j^{(k)})$ в соответствии со следующими выражениями:

$$\begin{aligned} G_j^{(k)} &= G_{j-1}^{(k)} + \sum_{i=1}^n \frac{1}{|\alpha_{ji}| \cdot \left(\hat{F}^{(k)}(\tau_j) - \hat{F}^{(k)}(\tau_{j-1}) \right)^2}, \\ W_j^{(k)} &= W_{j-1}^{(k)} + \sum_{i=1}^n \frac{1}{\alpha_{ji} \cdot \left(\hat{F}^{(k)}(L_i) - \hat{F}^{(k)}(R_i) \right)}, \\ D_j^{(k)} &= D_{j-1}^{(k)} + \sum_{i=1}^n \frac{\hat{F}^{(k)}(\tau_j)}{|\alpha_{ji}| \cdot \left(\hat{F}^{(k)}(\tau_j) - \hat{F}^{(k)}(\tau_{j-1}) \right)^2}, \\ V_j^{(k)} &= W_j^{(k)} + D_j^{(k)}. \end{aligned}$$

2. Установить $l = 0$.

3. Оценка функции распределения равна левой производной выпуклой миноранты:

$$\hat{F}^{(k)}(\tau_j) = \min_{l+1 \leq s \leq m} \left(\frac{V_s^{(k)} - V_l^{(k)}}{G_s^{(k)} - G_l^{(k)}} \right),$$

где $j = \overline{l+1, s}$, s – индекс угловой точки выпуклой миноранты (функция, значение которой не больше соответствующих значений данной функции).

4. Изменить $l = s$. Если $l < m$ переходим к пункту 3.

Алгоритм повторяем до тех пор, пока для всех $j = \overline{1, m}$ не выполнится условие:

$$\left| \hat{F}^{(k)}(\tau_j) - \hat{F}^{(k-1)}(\tau_j) \right| \leq 10^{-7}.$$

2. Критерии согласия для проверки сложной гипотезы о виде распределения по интервальным данным

Пусть имеется априорная информация о том, что выборка наблюдений принадлежит некоторому параметрическому семейству распределений $F_0(x; \theta)$, $\theta \in \Theta$. Для оценивания неизвестных параметров распределения $F_0(x; \theta)$ можно использовать метод максимального правдоподобия, который заключается в максимизации логарифма функции правдоподобия:

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \ln L(\mathcal{I}_n; \theta) = \arg \max_{\theta \in \Theta} \sum_{i=1}^n \ln (F(R_i; \theta) - F(L_i; \theta)).$$

Прежде чем делать выводы на основе полученного распределения, необходимо убедиться в адекватности полученной вероятностной модели. Для этого проверяется сложная гипотеза о виде распределения

$$H_0 : F(t) \in \{F_0(t; \theta), \theta \in \Theta\}.$$

Сложность гипотезы определяется тем, что неизвестные параметры предполагаемого распределения $F_0(x; \theta)$ оцениваются по той же выборке, по которой проверяется гипотеза.

Для проверки гипотезы H_0 по интервальной выборке рассмотрим модификации известных критериев согласия типа Колмогорова, Крамера–Мизеса–Смирнова и Андерсона–Дарлингга. Суть модификации заключается в том, что вместо эмпирической функции распределения в выражениях статистик данных критериев будем использовать непараметрическую оценку функции распределения по интервальной выборке $\mathbb{I}_n$. В этом случае статистика критерия типа Колмогорова вычисляется следующим образом:

$$D_n = \sup_{0 < t < \tau_m} |\hat{F}_n(t) - F_0(t; \hat{\theta})|,$$

статистика критерия типа Крамера–Мизеса–Смирнова вычисляется по формуле

$$S_{\omega^2} = \int_0^{\tau_m} \left(\hat{F}_n(t) - F_0(t; \hat{\theta}) \right)^2 dF_0(t; \hat{\theta})$$

и статистика Андерсона–Дарлингга имеет вид

$$S_W = \int_0^{\tau_m} \left(\hat{F}_n(t) - F_0(t; \hat{\theta}) \right)^2 \frac{dF_0(t; \hat{\theta})}{F_0(t; \hat{\theta})(1 - F_0(t; \hat{\theta}))},$$

где $\hat{F}_n(t)$ – непараметрическая оценка функции распределения по интервальным данным, которая вычисляется с помощью ИСМ-алгоритма.

Гипотеза H_0 отвергается при больших значениях статистик. Аналитический вид распределений рассматриваемых статистик при справедливости нулевой гипотезы неизвестен. При этом достигнутые уровни значимости (p -value) можно оценить с использованием методов статистического моделирования. Для этого необходимо задать количество повторений N . Порядок проверки гипотезы о виде распределения в этом случае имеет следующий вид:

1. Установить уровень значимости α .
2. По исходной интервальной выборке вычислить оценку неизвестного параметра θ распределения $F_0(t; \theta)$.
3. В соответствии с выбранным критерием согласия вычислить значение статистики критерия $S(\mathbb{I}_n)$.
4. Положить $M = 0$ и повторить следующую последовательность действий N раз:
 - 4.1) сгенерировать интервальную выборку $\mathbb{Z}_n = \{(L_1, R_1), \dots, (L_n, R_n)\}$:

$$(L_i, R_i) = \left(X_i - \frac{\Delta}{2}, X_i + \frac{\Delta}{2} \right), X_i = F_0^{-1}(\xi_i; \hat{\theta}), \xi_i \in \text{Uniform}(0, 1),$$

где $\hat{\theta}$ – оценка параметра θ , полученная в п. 2, Δ – абсолютная ошибка измерения в исходной выборке $\mathbb{I}_n$;

- 4.2) по полученной выборке $\mathbb{Z}_n$ оценить параметры распределения $F_0(t; \theta)$;
- 4.3) вычислить значение статистики критерия по смоделированной выборке;
- 4.4) если $S(\mathbb{Z}_n) > S(\mathbb{I}_n)$, то увеличить M на 1.

5. Вычислить оценку достигнутого уровня значимости $p = \frac{M}{N}$.

6. Если $p > \alpha$, то гипотеза H_0 не отвергается.

Важнейшим свойством критериев проверки статистических гипотез является мощность критерия. Чем выше мощность критерия, тем меньше вероятность допустить ошибку второго рода (вероятность принять гипотезу H_0 , когда она не верна). Целью исследования мощности критериев согласия может быть не только выявление наиболее мощного среди них, но и сравнение величины мощности критериев

по интервальным данным с аналогичной величиной мощности критериев по точечным данным. Точечная выборка содержит больше информации о распределении, чем интервальная, так как точечное наблюдение – это, по сути, измерение с нулевой погрешностью Δ (что на практике, как правило, невозможно). Это означает, что мощность критериев согласия для точечных выборок будет выше мощности соответствующих модифицированных критериев для интервальных данных. Однако то, насколько велики эти потери, неизвестно.

Для исследования мощности непараметрических критериев согласия по точечным и интервальным выборкам рассматривалась пара близких конкурирующих гипотез: H_0 , соответствующая нормальному распределению (сложная гипотеза), против H_1 , соответствующей логистическому распределению с параметрами $\theta_1 = 5$, $\theta_2 = \frac{\sqrt{3}}{\pi}$. В табл. 1 представлены оценки мощности критериев типа Колмогорова, Крамера–Мизеса–Смирнова и Андерсона–Дарлинга для точечных и интервальных выборок наблюдений объёмом $n = 50, 100, 200, 300, 500$. Оценки мощности вычислялись по выборкам статистик объёма $N = 20\,000$ при уровне значимости $\alpha = 0,1$. Функции плотности рассматриваемых распределений представлены в табл. 2.

Т а б л и ц а 1

Оценки мощности критериев согласия по точечным и интервальным данным

n	D_n		S_{ω^2}		S_W	
	X_n	I_n	X_n	I_n	X_n	I_n
50	0,18	0,17	0,21	0,21	0,24	0,23
100	0,23	0,21	0,29	0,29	0,33	0,32
200	0,36	0,31	0,46	0,45	0,49	0,49
300	0,45	0,36	0,58	0,58	0,67	0,62
500	0,65	0,50	0,79	0,78	0,89	0,82

Из табл. 1 видно, что среди рассмотренных критериев наиболее мощным является критерий Андерсона–Дарлинга как для точечных, так и для интервальных выборок. Как и ожидалось, мощность для всех видов критериев выше для точечных выборок. При этом эта разница увеличивается с ростом объёма выборки.

3. Пример проверки гипотезы о виде распределения по интервальной выборке

Рассмотрим выборку цен потребительского спроса на биоэнергетический напиток НПК «САВА», которые были получены путём анкетирования потенциальных потребителей продукции НПК «САВА» [5]. Участникам предлагалось указать диапазон для цены, которую они готовы заплатить за биоэнергетический напиток. В результате были получены интервальные данные по ценовым предпочтениям. В [5] по полученной выборке с помощью алгоритма Тёрнбулла была вычислена оценка функции распределения. В данной работе подберем параметрическое распределение, которое достаточно хорошо описывает распределение наблюдаемой случайной величины. В табл. 2 представлены результаты оценки параметров ряда распределений методом максимального правдоподобия и соответствующие значения информационных критериев Акаике (AIC) и Шварца (BIC), позволяющих сравнивать полученные модели:

$$AIC = 2k - 2l, \quad BIC = k \ln n - 2l,$$

где k – количество оцененных параметров, l – значение логарифмической функции правдоподобия для построенной модели. Чем меньше значение критерия, тем лучше модель.

На основе полученных результатов по критериям Акаике и Шварца в качестве распределения, описывающего цену потребительского спроса, выбрано нормальное распределение. На рис. 1 представлены графики непараметрической оценки функции распределения по ICM-алгоритму и функции распределения нормального закона с оцененными параметрами.

Выбор распределения

Распределение	Функция плотности	Оценки параметров	l	BIC	AIC
Нормальное	$\frac{1}{\theta_2 \sqrt{2\pi}} \exp\left(-\frac{(x-\theta_1)^2}{2\theta_2^2}\right)$	(38,88; 12,48)	-66,59	139,69	137,18
Логистическое	$\exp\left(-\frac{x-\theta_1}{\theta_2}\right) / \theta_2 \left(\exp\left(-\frac{x-\theta_1}{\theta_2}\right) + 1\right)^2$	(38,99; 7,69)	-67,91	142,34	139,82
Логнормальное	$\frac{1}{\sqrt{2\pi}(x-\theta_1) \cdot \theta_3} \exp\left(-\frac{1}{2}\left(\frac{1}{\theta_3} \ln\left(\frac{x-\theta_1}{\theta_2}\right)\right)^2\right)$	(16; 18,50; 0,74)	-69,72	149,22	145,44
Гамма	$\frac{1}{\theta_2 \cdot \Gamma(\theta_3)} \left(\frac{x-\theta_1}{\theta_2}\right)^{\theta_3-1} \exp\left(-\frac{x-\theta_1}{\theta_2}\right)$	(16; 9,78; 2,35)	-67,29	144,37	140,59
Экспоненциальное	$\frac{1}{\theta_2} \exp\left(-\frac{x-\theta_1}{\theta_2}\right)$	(16; 22,83)	-71,32	149,15	146,63
Вейбулла	$\frac{\theta_3}{x-\theta_1} \left(\frac{x-\theta_1}{\theta_2}\right)^{\theta_3} \exp\left(-\left(\frac{x-\theta_1}{\theta_2}\right)^{\theta_3}\right)$	(16; 25,65; 1,85)	-66,15	142,07	138,29
Логлогистическое	$\theta_3 \left(\frac{x-\theta_1}{\theta_2}\right)^{\theta_3-1} / \theta_2 \left(1 + \left(\frac{x-\theta_1}{\theta_2}\right)^{\theta_3}\right)^2$	(16; 20,08; 2,36)	-69,66	149,09	145,31

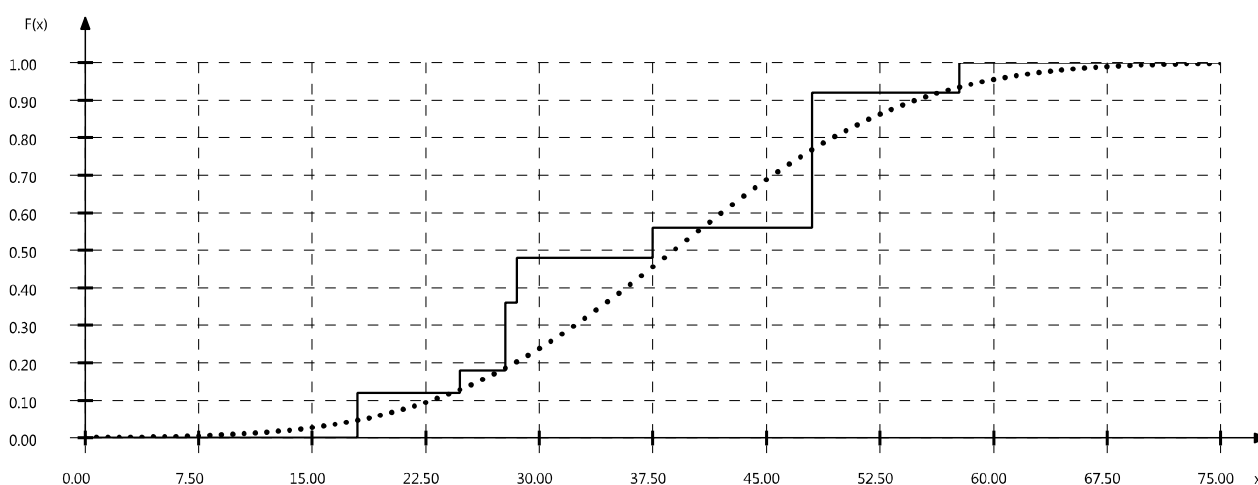


Рис. 1. Непараметрическая оценка функции распределения и нормальное распределение (пунктирная линия — нормальное распределение, сплошная — оценка функции распределения)

Информационные критерии позволяют сравнивать различные вероятностные модели, однако они не гарантируют адекватность выбранной модели. Для обоснования выбора модели необходима проверка статистической гипотезы о виде распределения.

Для проверки гипотезы о принадлежности рассматриваемой выборки нормальному распределению используем предложенные в данной работе модификации критериев согласия. Зададимся уровнем значимости $\alpha = 0,05$. В соответствии с представленным выше порядком проверки гипотезы о виде распределения при $N = 20\,000$ получены следующие оценки достигнутого уровня значимости (p -value): для критерия типа Колмогорова $p = 0,1853$, для критерия типа Крамера–Мизеса–Смирнова $p = 0,2786$ и для критерия типа Андерсона–Дарлингса $p = 0,3423$. Поскольку по всем критериям p -value больше

заданного уровня значимости, то нет оснований для отвержения гипотезы о принадлежности выборки нормальному распределению.

Заключение

В данной работе предложены модификации критериев согласия Колмогорова, Крамера–Мизеса–Смирнова и Андерсона–Дарлинга для проверки гипотез о виде распределения по интервальным данным. Применение предложенных критериев предполагает использование методов статистического моделирования для оценки достигнутого уровня значимости (p -value) для конкретной гипотезы при заданных значениях объема выборки и длины интервалов наблюдений. Сформулирован порядок проверки гипотезы о виде распределения по интервальным данным.

В результате исследования мощности предложенных критериев в качестве наиболее предпочтительного можно рекомендовать критерий согласия типа Андерсона–Дарлинга.

На основе данных о ценах потребительского спроса на биоэнергетический напиток НПК «САВА» проверена гипотеза о принадлежности данной выборки нормальному распределению.

ЛИТЕРАТУРА

1. Kreinovich V. Interval computations and interval-related statistical techniques: estimating uncertainty of the results of data processing and indirect measurements // *Advanced Mathematical and Computational Tools in Metrology and Testing X*. Singapore : World Scientific, 2015. P. 38–49. (Book series: *Advances in Mathematics for Applied Sciences*. V. 86). DOI: 10.1142/9789814678629_0014.
2. Вошинин А.П. Интервальный анализ данных: развитие и перспективы // *Заводская лаборатория*. 2002. Т. 68, № 1. С. 118–126.
3. Вошинин А.П. Метод анализа данных с интервальными ошибками в задачах проверки гипотез и оценивания параметров неявных линейно параметризованных функций // *Заводская лаборатория*. 2000. Т. 66, № 3. С. 51–64.
4. Орлов А.И. Основные идеи статистики интервальных данных // *Научный журнал КубГАУ*. 2013. № 94 (10). С. 1–26.
5. Зенкова Ж.Н., Краковецкая И.В. Непараметрическая оценка Тёрнбулла для интервально-цензурированных данных в маркетинговом исследовании спроса на биоэнергетические напитки // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. 2013. № 3 (24). С. 64–69.
6. Лемешко Б.Ю., Постовалов С.Н. Об оценивании параметров распределений по интервальным наблюдениям // *Вычислительные технологии*. 1998. Т. 3, № 2. С. 31–38.
7. Лемешко Б.Ю., Постовалов С.Н. О решении задач статистического анализа интервальных наблюдений // *Вычислительные технологии*. 1997. Т. 2, № 1. С. 28–36.
8. Лемешко Б.Ю., Постовалов С.Н. Статистический анализ наблюдений, имеющих интервальное представление // *Сборник научных трудов НГТУ*. Новосибирск : Изд-во НГТУ, 1996. № 1. С. 3–12.
9. Turnbull B.W. Nonparametric estimation of a survivorship function with doubly-censored data // *J. Am. Statist. Assoc.* 1974. V. 69. P. 169–73.
10. Вожов С.С. Исследование свойств непараметрической оценки функции распределения по интервальным данным // *Сборник научных трудов НГТУ*. Новосибирск : Изд-во НГТУ, 2015. № 1 (79). С. 33–44.
11. Vozhov S., Chimitova E. Investigation of Maximum Likelihood Estimates and Goodness-of-Fit Tests for Data with Known Measurement Error // *Applied methods of statistical analysis. Applications in survival analysis, reliability and quality control. AMSA'2015, Novosibirsk, 14–19 Sept. 2015 : proc. of the intern. workshop. Novosibirsk : NSTU publ., 2015. P. 124–130.*
12. Groeneboom P. Asymptotics for interval censored observations // *Technical Report 87-18*. Department of Mathematics, University of Amsterdam, 1987. 69 p.
13. Groeneboom P. Nonparametric maximum likelihood estimation for interval censored data // *Technical Report*, Statistics Department, Stanford University, 1991. 87 p.
14. Groeneboom P., Wellner J.A. *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Basel : Birkhauser Verlag, 1992. 126 p.

Вожов Станислав Сергеевич. E-mail: vss920414@gmail.com

Чимитова Екатерина Владимировна, канд. техн. наук, доцент. E-mail: chimitova@corp.nstu.ru
Новосибирский государственный технический университет

Поступила в редакцию 3 декабря 2015 г.

Vozhov Stanislav S., Chimitova Ekaterina V. (Novosibirsk State Technical University, Russian Federation).

Testing goodness-of-fit with interval data.

Keywords: interval data; nonparametric estimation of distribution function; Turnbull algorithm; ICM-algorithm.

DOI: 10.17223/19988605/34/4

The main terms of interval data analysis was initially founded the measurement theory in metrology where an interval uncertainty is naturally introduced. It is expected that every observation is measured by an instrument with absolute error Δ . Thus, if the precise value of an observed response is $\dot{x}$, measurement error is $e \in [-\Delta, \Delta]$, then the measurement is equal to $x = \dot{x} + e$. In this case, we deal with a usual complete sample $\mathcal{X}_n = \{X_1, \dots, X_n\}$. Nevertheless, the measurement can be represented as an interval $(x - \Delta, x + \Delta) = (L, R)$. In this case, for the sample of n observations we obtain an interval sample of the form

$$\mathcal{I}_n = \{(L_1, R_1), \dots, (L_n, R_n)\}.$$

The main idea of nonparametric estimation of the distribution function with interval data is based on maximization of the loglikelihood function

$$\ln L(\mathcal{I}_n) = \sum_{i=1}^n \ln(F(R_i) - F(L_i))$$

at the boundary points of observations $L_i, R_i, i = \overline{1, n}$, under condition of monotonicity of the distribution function.

The Turnbull and ICM algorithms are used for calculation of the nonparametric estimate of the distribution function with interval data. The accuracy of the estimates calculation is the same for both algorithm, but the computing time is less for the ICM algorithm.

Unknown distribution parameters can be estimated by the maximum likelihood method, which is based on the maximization of likelihood function by parameter θ :

$$L(\mathcal{I}_n | \theta) = \prod_{i=1}^n (F(R_i | \theta) - F(L_i | \theta)).$$

Thus, the maximum likelihood estimates can be written as

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \ln L(\mathcal{I}_n | \theta).$$

In this paper, the modifications of the classical goodness-of-fit tests for composite hypothesis $H_0: F(t) \in \{F_0(t; \theta), \theta \in \Theta\}$ have been proposed. The main idea of this modification is based on the usage of nonparametric estimate of the distribution function, obtained by the ICM algorithm, instead of the empirical distribution function. In this case, we have the test statistic of the Kolmogorov type as

$$D_n = \sup_{0 < t < \tau_m} |\hat{F}_n(t) - F_0(t, \hat{\theta})|,$$

the statistic of Cramer-von Mises-Smirnov type test as

$$S_{\omega^2} = \int_0^{\tau_m} (\hat{F}_n(t) - F_0(t, \hat{\theta}))^2 dF_0(t, \hat{\theta}),$$

and the statistic of Anderson-Darling type test as

$$S_W = \int_0^{\tau_m} (\hat{F}_n(t) - F_0(t, \hat{\theta}))^2 \frac{dF_0(t, \hat{\theta})}{F_0(t, \hat{\theta})(1 - F_0(t, \hat{\theta}))},$$

where $\hat{F}_n(t)$ is the nonparametric estimate of the distribution function by the interval data, $0 < \tau_0 < \tau_1 < \dots < \tau_m$ are ordered different values L_i and $R_i, i = \overline{1, n}$.

We have formulated the sequence of steps for estimation of the p -value for the proposed tests. The hypothesis is not rejected if the obtained p -value is larger than the significance level α .

As an example, we have tested the normality hypothesis by the interval sample of consumer demand prices for bio-energy drink SPC "SAVA". It has been shown that there is no reason for rejecting the hypothesis of normality of consumer demand prices.

REFERENCES

1. Kreinovich, V. (2015) Interval computations and interval-related statistical techniques: estimating uncertainty of the results of data processing and indirect measurements. In: Pavese, P., Bremser, W., Chunovkina, A., Fischer, N. & Forbes, A. (eds) *Advanced Mathematical and Computational Tools in Metrology and Testing X*. Singapore: World Scientific, pp. 38–49. (Book series: Advances in Mathematics for Applied Sciences, vol. 86). DOI 10.1142/9789814678629_0014.
2. Voshchinin, A.P. (2002) Interval'nyy analiz dannykh: razvitie i perspektivy [Interval data analysis: development and prospects]. *Zavodskaya laboratoriya*. 68(1). pp. 118-126.

3. Voshchinin, A.P. (2000) Metod analiza dannykh s interval'nymi oshibkami v zadachakh proverki gipotez i otsenivaniya parametrov neyavnykh lineynoy parametrizovannykh funktsiy [A method for data analysis with interval errors in problems of hypothesis testing and parameter estimation of fussy linear parameterized functions]. *Zavodskaya laboratoriya*. 66(3). pp. 51–64.
4. Orlov, A.I. (2013) Osnovnye idei statistiki interval'nykh dannykh [Basic ideas of interval data statistics]. *Nauchnyy zhurnal KubGAU – Scientific Journal of KubSAU*. 94(10). pp. 1-26.
5. Zenkova, Zh.N. & Krakovetskaya, I.V. (2013) Nonparametric Turnbull estimator for interval censored data in the marketing research of the demand of bio-energy drinks. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 3(24). pp. 64-69. (In Russian).
6. Lemeshko, B.Yu. & Postovalov, S.N. (1998) Ob otsenivanii parametrov raspredeleniy po interval'nym nablyudeniyam [On estimation of distribution parameters by interval observations]. *Vychislitel'nye tekhnologii*. 3(2). pp. 31-38.
7. Lemeshko, B.Yu. & Postovalov, S.N. (1997) O reshenii zadach statisticheskogo analiza interval'nykh nablyudeniy [On solving the problems of statistical analysis of interval data]. *Vychislitel'nye tekhnologii*. 2(1). pp. 28-36.
8. Lemeshko, B.Yu. & Postovalov, S.N. (1996) Statisticheskiy analiz nablyudeniy, imeyushchikh interval'noe predstavlenie [Statistical analysis of interval observations]. *Sbornik nauchnykh trudov NGTU*. 1. pp. 3-12.
9. Turnbull, B.W. (1974) Nonparametric estimation of a survivorship function with doubly-censored data. *J. Am. Statist. Assoc.* 69. pp. 169-73. DOI: 10.1080/01621459.1974.10480146
10. Vozhov, S.S. (2015) Issledovanie svoystv neparametricheskoy otsenki funktsii raspredeleniya po interval'nym dannym [Investigation of the properties of nonparametric estimate for distribution function with interval data]. *Sbornik nauchnykh trudov NGTU*. 1(79). pp. 33-44.
11. Vozhov, S. & Chimitova, E. (2015) Investigation of Maximum Likelihood Estimates and Goodness-of-Fit Tests for Data with Known Measurement Error. *Applied methods of statistical analysis. Applications in survival analysis, reliability and quality control*. AMSA'2015, Novosibirsk, 14–19 Sept. 2015. Proc. of the Intern. Workshop. Novosibirsk: NSTU. pp. 124-130.
12. Groeneboom, P. (1987) *Asymptotics for interval censored observations*. Technical Report 87-18. Department of Mathematics, University of Amsterdam.
13. Groeneboom, P. (1991) *Nonparametric maximum likelihood estimation for interval censored data*. Technical Report, Statistics Department, Stanford University, 1991. 87 p.
14. Groeneboom, P. & Wellner, J.A. (1992) *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Basel: Birkhauser Verlag.

А.А. Назаров, В.И. Бронер**СИСТЕМА УПРАВЛЕНИЯ ЗАПАСАМИ С ГИПЕРЭКСПОНЕНЦИАЛЬНЫМ
РАСПРЕДЕЛЕНИЕМ ОБЪЕМОВ ПОТРЕБЛЕНИЯ РЕСУРСОВ**

Проводится исследование математической модели системы управления запасами с релейным управлением объемом накопленных запасов. Рассмотрен случай гиперэкспоненциального распределения объемов потребления ресурсов. Найдено явное выражение для стационарной плотности распределения значения запасов в системе. Приводятся результаты численного эксперимента.

Ключевые слова: управление запасами; релейное управление; гиперэкспоненциальное распределение; математическое моделирование.

В последние десятилетия к математическим моделям управления запасами проявляют большой интерес. В качестве таковых в работах [1–5] рассматриваются математические модели деятельности фонда социального страхования с релейным управлением капиталом фонда. В [1] исследуются основные характеристики деятельности фонда социального страхования в случае непрерывной скорости поступления денежных средств и экспоненциально распределённых страховых выплат. В [2, 5] рассматриваются и исследуются модели фонда социального страхования при релейном управлении (в [2] также рассмотрено релейно-гистерезисное управление) капиталом такого фонда, когда выплаты по страховым случаям и выплаты на финансирование социальных программ образуют пуассоновские потоки событий с постоянной и переменной интенсивностями соответственно, а величины выплат являются одинаково распределёнными независимыми случайными величинами с экспоненциальной функцией распределения.

В [3] построена и исследована математическая модель деятельности некоммерческого фонда в случае пуассоновского потока поступающих платежей постоянной интенсивности при экспоненциальном распределении страховых премий и релейного управления капиталом. А в [6] на основе диффузионного приближения исследуется аналогичная [3] модель.

В [7] находится выражение для функции скорости выделения средств на социальные программы в диффузионном приближении для процесса изменения капитала фонда в условиях математической модели [1].

В работах [8–11] рассматриваются различные математические модели управления запасами. Например, в [9] предполагается, что продавец приобретает ресурс в фиксированном объеме, который потребляется в течение торговой сессии. Так как спрос не определен, то целью исследования в аналогичных работах, как правило, ставится задача нахождения объема запасов, такого, чтобы спрос был удовлетворен и в конце торговой сессии не оставалось нереализованной продукции.

В данной работе исследуется модель, аналогичная [2, 5], в случае, когда объемы запроса на расходование имеют гиперэкспоненциальное распределение.

1. Математическая модель

В качестве математической модели управления запасами рассмотрим систему (рис. 1), на вход которой непрерывно поступают некоторые ресурсы с постоянной скоростью $v = 1$.

Обозначим через $s(t)$ объем накопленных ресурсов в системе к моменту времени t . Будем считать, что потребление ресурса осуществляется в случайные моменты времени партиями случайного объема.

Моменты потребления образуют пуассоновский поток с кусочно-постоянной интенсивностью $\lambda(s)$, зависящей от значений $s(t) = s$ величин накопленных запасов к моменту времени t поступления заявки на расходование ресурса, здесь

$$\lambda(s) = \begin{cases} \lambda_1, & s < S, \\ \lambda_2, & s \geq S, \end{cases} \quad (1)$$

где S – некоторое пороговое значение уровня запасов $s(t)$.

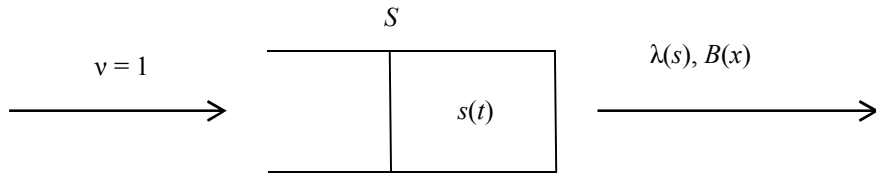


Рис. 1. Система управления запасами

Будем полагать, что объемы потребления ресурсов имеют гиперэкспоненциальную функцию распределения

$$B(x) = \sum_{k=1}^n b_k (1 - e^{-\mu_k x}) \quad (2)$$

n -го порядка с параметрами $\mu_k > 0$ и $b_k > 0$, причем

$$\sum_{k=1}^n b_k = 1. \quad (3)$$

Заметим, что процесс $s(t)$ может принимать отрицательные значения $s(t) < 0$ и система продолжает функционировать, откладывая исполнение заявки на потребление ресурсов.

Условие существования стационарного режима в рассматриваемой системе имеет вид

$$\lambda_1 b < 1 < \lambda_2 b, \quad (4)$$

где b – среднее значение объема одной партии на потребление ресурсов.

Таким образом, при $\lambda_1 < \lambda_2$ и $s(t) < S$ объем ресурса в системе будет увеличиваться в среднем, а при достижении уровня S и его превышении, т.е. $s(t) \geq S$, в связи с возрастанием интенсивности потребления объем ресурса будет уменьшаться.

В силу (2) величина b может быть представлена следующим образом

$$b = \sum_{k=1}^n \frac{b_k}{\mu_k}. \quad (5)$$

Из описания математической модели следует, что случайный процесс $s(t)$ является марковским с непрерывным временем t и непрерывным множеством значений $-\infty < s < \infty$.

Обозначим его плотность распределения

$$P(s, t) = \frac{\partial P\{s(t) < s\}}{\partial s}$$

и запишем следующее равенство:

$$P(s + \Delta t, t + \Delta t) = P(s, t)(1 - \lambda(s)\Delta t) + \Delta t \int_0^\infty \lambda(s + x)P(s + x, t)dB(x) + o(\Delta t),$$

из которого для стационарного распределения $P(s) = \lim_{t \rightarrow \infty} P(s, t)$ получим уравнение

$$P'(s) + \lambda(s)P(s) = \int_0^\infty \lambda(s + x)P(s + x)dB(x), \quad (6)$$

решение $P(s)$ которого удовлетворяет краевым условиям

$$P(-\infty) = P(\infty) = 0. \quad (7)$$

Отметим, что уравнение (6) является основным при исследовании математических моделей систем управления запасами.

Найдем решение $P(s)$ уравнения (6) в явном виде, взяв в качестве функции распределения $B(x)$ объемов партий потребления гиперэкспоненциальную функцию распределения.

Обозначив

$$P(s) = \begin{cases} P_1(s), & s < S, \\ P_2(s), & s > S, \end{cases} \quad (8)$$

можем записать уравнение (7) в виде двух уравнений

$$P_2'(s) + \lambda_2 P_2(s) = \lambda_2 \int_0^{\infty} P_2(s+x) dB(x), \quad s > S, \quad (9)$$

$$P_1'(s) + \lambda_1 P_1(s) = \lambda_1 \int_0^{S-s} P_1(s+x) dB(x) + \lambda_2 \int_{S-s}^{\infty} P_2(s+x) dB(x), \quad s < S. \quad (10)$$

Найдем решения уравнений (9) и (10), удовлетворяющие краевым условиям

$$P_1(-\infty) = 0, \quad P_2(\infty) = 0. \quad (11)$$

2. Решение уравнения для $P_2(s)$

Решение $P_2(s)$, $s > S$, уравнения (9) будем искать в виде

$$P_2(s) = C e^{-\gamma(s-S)}, \quad s > S. \quad (12)$$

Подставляя (12) в (9), получим равенство

$$\lambda_2 - \gamma = \lambda_2 \int_0^{\infty} e^{-\gamma x} dB(x), \quad (13)$$

которое является нелинейным уравнением относительно величины γ .

Очевидно, что уравнение (13) имеет нулевой корень $\gamma = 0$, но в силу краевого условия (11) $P_2(\infty) = 0$ он является посторонним в рассматриваемой задаче.

Нетрудно показать, что при выполнении условия (4) $\lambda_2 b > 1$ уравнение (13) кроме нулевого решения имеет единственный положительный корень $\gamma > 0$ для любой функции распределения $B(x)$, поэтому решением уравнения (9) является функция (12), определяемая с точностью до мультипликативной постоянной C , значение которой найдем ниже.

3. Решение уравнения для $P_1(s)$

В силу (12) представим уравнение (10) в виде

$$P_1'(s) + \lambda_1 P_1(s) = \lambda_1 \int_0^{S-s} P_1(s+x) dB(x) + \lambda_2 C e^{-\gamma(s-S)} \int_{S-s}^{\infty} e^{-\gamma x} dB(x). \quad (14)$$

Принимая во внимание (2), получим

$$\int_{S-s}^{\infty} e^{-\gamma x} dB(x) = \int_{S-s}^{\infty} e^{-\gamma x} \sum_{k=1}^n b_k \mu_k e^{-\mu_k x} dx = \sum_{k=1}^n b_k \mu_k \int_{S-s}^{\infty} e^{-(\mu_k + \gamma)x} dx = \sum_{k=1}^n b_k \frac{\mu_k}{\mu_k + \gamma} e^{-(\mu_k + \gamma)(S-s)},$$

поэтому (14) можем записать следующим образом:

$$P_1'(s) + \lambda_1 P_1(s) = \lambda_1 \int_0^{S-s} P_1(s+x) dB(x) + \lambda_2 C \sum_{k=1}^n b_k \frac{\mu_k}{\mu_k + \gamma} e^{\mu_k(s-S)}.$$

Подставляя в это равенство выражение (2) для функции распределения $B(x)$, получим уравнение для $P_1(s)$

$$P_1'(s) + \lambda_1 P_1(s) = \sum_{k=1}^n b_k \mu_k \left\{ \lambda_1 \int_0^{S-s} P_1(s+x) e^{-\mu_k x} dx + C e^{\mu_k s} \frac{\lambda_2}{\mu_k + \gamma} e^{-\mu_k S} \right\}. \quad (15)$$

Прежде чем сформулировать теорему о виде функции $P_1(s)$, рассмотрим уравнение

$$z + \lambda_1 = \lambda_1 \sum_{k=1}^n b_k \frac{\mu_k}{\mu_k - z}, \quad (16)$$

которое нетрудно преобразовать к алгебраическому уравнению степени $n + 1$, откуда следует, что уравнение (16) имеет $n + 1$ корней. Достаточно очевидно, что $z = 0$ является корнем этого уравнения.

Для остальных корней $z = z_v$, $v = \overline{1, n}$, уравнения (16) докажем следующее утверждение.

Лемма 1. При выполнении условия (4)

$$\lambda_1 b < 1$$

все корни $z = z_v$, $v = \overline{1, n}$, уравнения (16) действительные и положительные.

Доказательство. Будем полагать, что значения μ_k упорядочены по возрастанию, т.е. $\mu_1 < \mu_2 < \dots < \mu_n$. Рассмотрим функцию

$$f(z) = \lambda_1 \sum_{k=1}^n b_k \frac{\mu_k}{\mu_k - z},$$

совпадающую с правой частью уравнения (16). Так как

$$f'(z) = \lambda_1 \sum_{k=1}^n b_k \frac{\mu_k}{(\mu_k - z)^2} > 0,$$

то в интервале $0 < z < \mu_1$ функция $f(z)$ непрерывна, монотонно возрастает и принимает значения $\lambda_1 < f(z) < \infty$. При выполнении условия $\lambda_1 b < 1$ на интервале $0 < z < \mu_1$ уравнение (16) имеет по крайней мере один корень $z_1 > 0$.

Далее рассмотрим функцию $f(z)$ на интервале $\mu_{v-1} < z < \mu_v$, где $f(z)$ непрерывна, монотонно возрастает и принимает значения $-\infty < f(z) < \infty$, поэтому уравнение (16) на интервале $\mu_{v-1} < z < \mu_v$ также имеет по крайней мере один корень $z_v > 0$.

Количество рассматриваемых интервалов равно n , совпадающее с числом положительных корней $z = z_v$, $v = \overline{1, n}$.

Лемма доказана.

Следствие 1. Корень $0 < z_1 < \mu_1$, а для любого $v = \overline{2, n}$, $\mu_{v-1} < z_v < \mu_v$.

Сформулированное следствие существенно упрощает численное нахождение всех положительных корней уравнения (16).

Докажем следующее утверждение.

Теорема 1. Решение $P_1(s)$ уравнения (15) имеет вид

$$P_1(s) = C \sum_{v=1}^n x_v e^{z_v(s-S)}, \quad s < S, \quad (17)$$

где z_v – положительные корни уравнения (16), параметры x_v распределения (17) являются компонентами вектора $\mathbf{X}$ – решения системы линейных алгебраических уравнений

$$\mathbf{A}\mathbf{X} = \mathbf{h}, \quad (18)$$

где элементы A_{kv} матрицы $\mathbf{A}$ и компоненты h_k вектора $\mathbf{h}$ имеют вид

$$A_{kv} = \frac{\lambda_1}{\mu_k - z_v}, \quad h_k = \frac{\lambda_2}{\mu_k + \gamma}, \quad (19)$$

нормирующая константа C определяется равенством

$$C = \left(\frac{1}{\gamma} + \sum_{v=1}^n \frac{x_v}{z_v} \right)^{-1}. \quad (20)$$

Доказательство. Решение $P_1(s)$ уравнения (15) будем искать в виде (17).

Подставляя выражение (17) в (15) и выполняя несложные преобразования, получим равенство

$$\sum_{v=1}^n x_v e^{z_v(s-S)} \left\{ z_v + \lambda_1 + \lambda_1 \sum_{k=1}^n b_k \frac{\mu_k}{z_v - \mu_k} \right\} = \sum_{k=1}^n b_k \mu_k e^{\mu_k(s-S)} \left\{ \sum_{v=1}^n x_v \frac{\lambda_1}{z_v - \mu_k} + \frac{\lambda_2}{\mu_k + \gamma} \right\}.$$

Приравнявая в полученном выражении коэффициенты в линейной комбинации экспонент $e^{z_v s}$ к нулю, получим равенства

$$z_v + \lambda_1 + \lambda_1 \sum_{k=1}^n b_k \frac{\mu_k}{z_v - \mu_k} = 0, \quad v = \overline{1, m},$$

которые при всех $v = \overline{1, n}$ совпадают с (16), следовательно, $z = z_v$ являются корнями уравнения (16).

Аналогично для экспонент $e^{\mu_k s}$ получим равенства

$$\sum_{v=1}^n x_v \frac{\lambda_1}{\mu_k - z_v} = \frac{\lambda_2}{\mu_k + \gamma}, \quad k = \overline{1, n},$$

которые составляют неоднородную систему линейных алгебраических уравнений относительно x_v , совпадающую с системой (18), в которой элементы A_{kv} матрицы $\mathbf{A}$ и компоненты h_k вектора $\mathbf{h}$ определяются равенствами (19).

Значение константы C найдем из условия нормировки

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} P(s) ds = \int_{-\infty}^S P_1(s) ds + \int_S^{\infty} P_2(s) ds = C \sum_{v=1}^n x_v \int_{-\infty}^S e^{z_v(s-S)} ds + C \int_S^{\infty} e^{-\gamma(s-S)} ds = \\ &= C \sum_{v=1}^n x_v \int_{-\infty}^0 e^{z_v x} dx + C \int_0^{\infty} e^{-\gamma x} dx = C \left\{ \sum_{v=1}^n \frac{x_v}{z_v} + \frac{1}{\gamma} \right\}. \end{aligned}$$

Отсюда следует равенство

$$C = \left(\sum_{v=1}^n \frac{x_v}{z_v} + \frac{1}{\gamma} \right)^{-1},$$

которое совпадает с (20). Теорема доказана.

В силу (17) и (12) распределение $P(s)$ из (8) имеет вид

$$P(s) = \left(\sum_{v=1}^n \frac{x_v}{z_v} + \frac{1}{\gamma} \right)^{-1} \cdot \begin{cases} \sum_{v=1}^n x_v e^{z_v(s-S)}, & s < S, \\ e^{-\gamma(s-S)}, & s > S, \end{cases} \quad (21)$$

где параметры γ и z_v этого распределения являются положительными корнями уравнений (13) и (16), параметры x_v являются компонентами вектора $\mathbf{X}$ – решения системы линейных алгебраических уравнений (18).

Явное выражение (21) для решения $P(s)$ уравнения (6) полностью решает проблему исследования математической модели управления запасами при выполнении указанных ограничений: релейное управление и гиперэкспоненциальное распределение объемов партий потребления ресурсов.

4. Пример

Рассмотрим в качестве закона распределения объемов потребления гиперэкспоненциальное распределение третьего порядка

$$B(x) = \sum_{k=1}^3 b_k (1 - e^{-\mu_k x}),$$

где значения μ_k и b_k определяются вектор-строками $\mathbf{\mu}$ и $\mathbf{b}$ соответственно:

$$\mathbf{\mu} = (1 \quad 0,4 \quad 10), \quad \mathbf{b} = (0,2 \quad 0,3 \quad 0,5), \quad (22)$$

при которых средняя величина объемов потребления $b = 1$.

Для заданных значений параметров $\lambda_1 = 0,8$ и $\lambda_2 = 1,2$ найдены положительные корни уравнений (13) и (16)

$$\gamma = 0,099; \quad z_1 = 0,094; \quad z_2 = 0,899; \quad z_3 = 9,617.$$

Таким образом, уравнение (13) имеет единственное решение, а уравнение (16) имеет три положительных корня. Оба уравнения имеют нулевые корни, которые, как было показано выше, являются посторонними.

Найдем плотность распределения вероятностей значений объема запасов при заданных параметрах (22) и $S = 10$.

Параметры $x_v, v = \overline{1,3}$, распределения (16), являющиеся компонентами вектора $\mathbf{X}$ – решения системы (18) линейных алгебраических уравнений, имеют вид

$$x_1 = 0,945; \quad x_2 = 0,036; \quad x_3 = 0,019,$$

а нормирующая константа $C = 0,049$, тогда имеет место график, представленный на рис. 2.

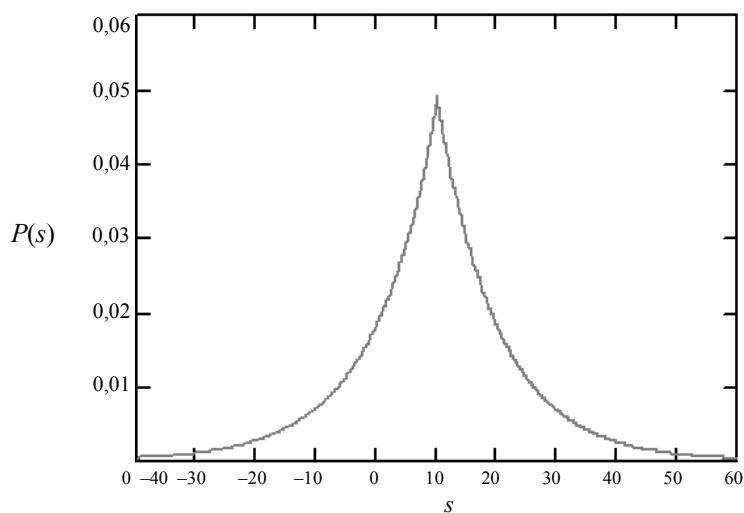


Рис. 2. Плотность распределения вероятностей значений процесса $P(s)$

Следует отметить, что плотность распределения вероятностей $P(s)$ значений процесса $s(t)$ является непрерывной для всех значений s , что естественно, но также и в точке $s = S$, что не является очевидным.

Заключение

В данной работе построена математическая модель системы управления запасами. Получено аналитическое выражение для стационарной плотности распределения значений объема запасов при гиперэкспоненциальном распределении объемов потребления и релейном управлении объемом запасов. Предложенный подход может быть применен к аналогичным задачам при различных распределениях объемов расходования ресурсов.

ЛИТЕРАТУРА

1. Змеев О.А. Математическая модель деятельности фонда социального страхования при экспоненциальных страховых выплатах // Вестник Томского государственного университета. 2003. № 280. С. 130–135.
2. Вальц О.В., Змеев О.А. Математическая модель деятельности фонда социального страхования при экспоненциальных страховых выплатах и со случайными расходами на социальные программы // Вестник Томского государственного университета. 2004. № 284. С. 37–41.
3. Лившиц К.И., Шифердекер И.Ю. Математическая модель деятельности некоммерческого фонда при релейном управлении капиталом // Вестник Томского государственного университета. Приложение. 2006. № 18. С. 302–308.
4. Лившиц К.И., Сухотина Л.Ю., Шифердекер И.Ю. Пуассоновская модель деятельности некоммерческого фонда при релейном управлении капиталом // Вестник Томского государственного университета. Приложение. 2006. № 19. С. 302–312.
5. Китаева А.В., Терпугов А.Ф. Модель фонда социального страхования при релейном управлении капиталом и экспоненциально распределенных страховых выплатах и выплатах по социальным программам // Вестник Томского государственного университета. 2006. № 293. С. 35–37.
6. Лившиц К.И., Шифердекер И.Ю. Диффузионная аппроксимация математической модели деятельности некоммерческого фонда при релейном управлении капиталом // Вестник Томского государственного университета. 2006. № 293. С. 38–44.
7. Китаева А.В., Терпугов А.Ф. Управление капиталом фонда социального страхования // Вестник Томского государственного университета. 2006. № 290. С. 167–168.
8. Arrow K.J., Harris Th.E., Marschak J. Optimal Inventory Policy // Econometrica. 1951. V. 19, is. 3. P. 205–272.
9. Khouja M. The single-period (newsvendor) problem: Literature review and suggestions for future research // Omega. 2000. V. 27. P. 537–553.
10. Nahmias S. Demand estimation in lost sales inventory systems // Naval Research Logistics. 1994. V. 41. P. 739–757.
11. Gallego G., Moon I. The distribution free newsboy problem: Review and extensions // The Journal of the Operational Research Society. 1993. V. 44. P. 825–834.

Назаров Анатолий Андреевич, д-р техн. наук, профессор. E-mail: nazarov.tsu@gmail.com

Бронер Валентина Игоревна. E-mail: valsubbotina@mail.ru

Томский государственный университет

Поступила в редакцию 12 декабря 2015 г.

Consider a mathematical model of inventory control. Let the product flow be continuous with the fixed rate $v = 1$. Let $s(t)$ be an inventory level at the moment t . The demand occurs according to a Poisson process with piecewise constant intensity $\lambda(s)$,

$$\lambda(s) = \begin{cases} \lambda_1, & s < S, \\ \lambda_2, & s \geq S, \end{cases}$$

where S is the threshold inventory level of $s(t)$. The values of purchases are independent and identically distributed random variables from the n -th order hyperexponential distribution with the first moment equal to 1.

For stationary distribution probability density function $P(s)$ we obtained the equation

$$P'(s) + \lambda(s)P(s) = \int_0^\infty \lambda(s+x)P(s+x)dB(x),$$

where $P(s)$ satisfies the boundary conditions

$$P(-\infty) = P(\infty) = 0.$$

Then, the expression for the probability density function $P(s)$ is derived:

$$P(s) = \left(\sum_{v=1}^n \frac{x_v}{z_v} + \frac{1}{\gamma} \right)^{-1} \cdot \begin{cases} \sum_{v=1}^n x_v e^{z_v(s-S)}, & s < S, \\ e^{-\gamma(s-S)}, & s \geq S, \end{cases}$$

where z_v and γ are positive roots of equations

$$z + \lambda_1 = \lambda_1 \sum_{k=1}^n b_k \frac{\mu_k}{\mu_k - z}, \quad \lambda_2 - \gamma = \lambda_2 \int_0^\infty e^{-\gamma x} dB(x).$$

Here x_v are components of the vector $\mathbf{X}$, which is a solution of a system of linear algebraic equations

$$\mathbf{A}\mathbf{X} = \mathbf{h},$$

where A_{kv} are elements of the matrix $\mathbf{A}$, h_k are elements of the vector $\mathbf{h}$. The elements A_{kv} and h_k have the form

$$A_{kv} = \frac{\lambda_1}{\mu_k - z_v}, \quad h_k = \frac{\lambda_2}{\mu_k + \gamma},$$

and normalizing constant C is determined by the equation

$$C = \left(\frac{1}{\gamma} + \sum_{v=1}^n \frac{x_v}{z_v} \right)^{-1}.$$

REFERENCES

1. Zmeyer, O.A. (2003) The model of the social insurance fund with exponential distributed insurance payments. *Vestnik Tomskogo gosudarstvennogo universiteta – Journal of the Tomsk State University*. 280. pp. 130-135. (In Russian).
2. Valts, O.V. & Zmeyer, O.A. (2004) Mathematical model of advertising campaign taking into account the effect of “boring” of advertisement. *Vestnik Tomskogo gosudarstvennogo universiteta – Journal of the Tomsk State University*. 284. pp. 37-41. (In Russian).
3. Livshits, K.I. & Shiferdeker, I.Yu. (2006) Mathematical model of incommercial fund functioning under the relay control of its capital. *Vestnik Tomskogo gosudarstvennogo universiteta – Journal of the Tomsk State University*. 18. pp. 302-308. (In Russian).
4. Livshits, K.I., Suhotina, L.Yu. & Shiferdeker, I.Yu. (2006) Puasson model of incommercial fund functioning under the relay control of its capital. *Vestnik Tomskogo gosudarstvennogo universiteta – Journal of the Tomsk State University*. 19. pp. 302-312. (In Russian).
5. Kitaeva, A.V. & Terpugov, A.F. (2006) The model of the social insurance fund on the relay management of capital and exponential distributed insurance payments and payments on social programs. *Vestnik Tomskogo gosudarstvennogo universiteta – Journal of the Tomsk State University*. 293. pp. 35-37. (In Russian).
6. Livshits, K.I. & Shiferdeker, I.Yu. (2006) Diffusion approximation of the mathematical model of the non-profit foundation with the relay money management. *Vestnik Tomskogo gosudarstvennogo universiteta – Journal of the Tomsk State University*. 293. pp. 38-44.
7. Kitaeva, A.V. & Terpugov, A.F. (2006) Control of the social insurance fund's surplus. *Vestnik Tomskogo gosudarstvennogo universiteta – Journal of the Tomsk State University*. 290. pp. 167-168. (In Russian).
8. Arrow, K.J., Harris, Th.E. & Marschak, J. (1951) Optimal Inventory Policy. *Econometrica*. 19(3). pp. 205-272. DOI: 10.2307/1906813
9. Khouja, M. (2000) The single-period (newsvendor) problem: Literature review and suggestions for future research. *Omega*. 27. pp. 537-553.
10. Nahmias, S. (1994) Demand estimation in lost sales inventory systems. *Naval Research Logistics*. 41. pp. 739-757. DOI: 10.1002/1520-6750(199410)41:6<739::AID-NAV3220410605>3.0.CO;2-A
11. Gallego, G. & Moon, I. (1993) The distribution free newsboy problem: Review and extensions. *The Journal of the Operational Research Society*. 44. pp. 825-834. DOI: 10.1038/sj/jors/0440809

М.Н. Сиротина, А.М. Горцев

ОЦЕНКА МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ ДЛИТЕЛЬНОСТИ МЕРТВОГО ВРЕМЕНИ В МОДУЛИРОВАННОМ СИНХРОННОМ ДВАЖДЫ СТОХАСТИЧЕСКОМ ПОТОКЕ СОБЫТИЙ

Работа выполнена в рамках государственного заказа Минобрнауки РФ на проведение научных исследований в Национальном исследовательском Томском государственном университете на 2014–2016 гг.

Рассматривается модулированный синхронный дважды стохастический поток событий. Поток функционирует в условиях непродлевающегося мертвого времени, т.е. после каждого зарегистрированного события наступает время фиксированной длительности, в течение которого другие события исходного модулированного синхронного потока недоступны наблюдению. Полагается, что длительность мертвого времени – неизвестная величина. Методом максимального правдоподобия решается задача об оценке длительности мертвого времени по наблюдениям за моментами наступления событий рассматриваемого потока.

Ключевые слова: модулированный синхронный поток событий; непродлевающееся мертвое время; функция правдоподобия; оценка максимального правдоподобия; длительность мертвого времени.

Настоящая работа является непосредственным продолжением исследований модулированного синхронного потока событий, начатых в статьях [1–4].

Математические модели систем массового обслуживания широко применяются при описании реальных физических, технических и других процессов и систем. В связи с бурным развитием компьютерной техники и информационных технологий появилась важная сфера приложений теории массового обслуживания – проектирование и создание информационно-вычислительных сетей, компьютерных сетей связи, спутниковых сетей, телекоммуникационных сетей, объединенных термином «цифровые сети интегрального обслуживания» (ЦСИО) [5, 6].

На практике параметры, определяющие входящий поток событий, изменяются со временем, при этом изменения часто носят случайный характер, последнее приводит к рассмотрению дважды стохастических потоков событий. По-видимому, одной из первых работ в этом направлении явилась статья [7], в которой дважды стохастический поток определяется как поток, интенсивность которого есть случайный процесс. Дважды стохастические потоки можно разделить на два класса: к первому классу относятся потоки, интенсивность которых есть непрерывный случайный процесс; ко второму классу относятся потоки, интенсивность которых есть кусочно-постоянный случайный процесс с конечным числом состояний. Подчеркнем, что потоки второго класса впервые введены в рассмотрение практически одновременно в 1979 г. в [8, 9]. В [8] введенные потоки названы МС (Markov chain)-потоками; в [9] – MVP (Markov versatile processes)-потоками. Отечественные и зарубежные авторы в своих работах, начиная с начала 1990-х гг. [10–15], называют введенные в [8, 9] потоки событий либо дважды стохастическими потоками событий, либо МАР-потоками, либо МС-потоками. В свою очередь, в зависимости от того, каким образом происходит переход из состояния в состояние, МС-потоки можно разделить на три типа: 1) синхронные потоки событий [16–21]; 2) асинхронные и обобщенные асинхронные потоки событий [22–27]; 3) полусинхронные и обобщенные полусинхронные потоки событий [28–33]. В [34] введены в рассмотрение МАР-потоки событий первого порядка (собственно МАР-потоки, введенные в [9]) и МАР-потоки событий второго порядка (суперпозиция двух синхронизированных МАР-потоков первого порядка, отличающихся друг от друга исходными параметрами). В [34] показывается, что син-

хронный МС-поток является частным случаем МАР-потока первого порядка, асинхронный и полусинхронный МС-потоки являются частными случаями МАР-потока второго порядка.

В реальных ситуациях параметры, задающие входящий поток событий, известны либо частично, либо вообще неизвестны, либо (что еще более ухудшает ситуацию) изменяются со временем. Вследствие этого возникают задачи: 1) оценки состояний потока (задача фильтрации интенсивности потока) по наблюдениям за моментами наступления событий [24, 27, 29, 33, 35–38]; 2) оценка параметров потока по наблюдениям за моментами наступления событий [16–22, 26, 28, 30–32].

Одним из искажающих факторов при оценке состояний и параметров потока событий выступает мертвое время регистрирующих приборов [39–46], которое порождается зарегистрированным событием. Другие же события, наступившие в течение периода мертвого времени, недоступны наблюдению (теряются). Можно считать, что этот период продолжается некоторое фиксированное время (непродлевающееся мертвое время). В качестве примера приведем протокол CSMA/CD – протокол случайного множественного доступа с обнаружением конфликта, широко используемый в компьютерных сетях. В момент регистрации (обнаружения) конфликта на входе некоторого узла сети по сети рассылается сигнал «заглушки» («пробки»); в течение времени рассылки сигнала «заглушки» заявки, поступившие в данный узел сети, получают отказ в обслуживании и направляются в источник повторных вызовов. Здесь время, в течение которого узел сети закрыт для обслуживания заявок, поступающих в него после обнаружения конфликта, можно трактовать как мертвое время прибора, регистрирующего конфликт в узле сети. Для того чтобы оценить потери заявок в узле сети, необходимо оценить длительность мертвого времени, которым выступает в данном случае длительность сигнала «заглушки».

В работах [1–4] введен в рассмотрение модулированный синхронный поток событий, являющийся обобщением синхронного потока и относящийся к классу МАР-потоков второго порядка. Синхронный поток событий систематически исследовался в работах [16–21, 47–52]. В настоящей статье, являющейся непосредственным развитием работ [1–4], для решения задачи оценивания длительности мертвого времени применяется метод максимального правдоподобия [53–59], так как оценки, построенные при использовании данного метода, как правило, обладают привлекательными свойствами.

1. Постановка задачи

Рассматривается модулированный синхронный поток событий (далее поток), интенсивность которого есть кусочно-постоянный стационарный случайный процесс $\lambda(t)$ с двумя состояниями: $\lambda_1, \lambda_2 (\lambda_1 > \lambda_2 \geq 0)$. Длительность пребывания процесса $\lambda(t)$ (потока) в i -м состоянии распределена по экспоненциальному закону с параметром $\alpha_i, i = 1, 2$. Если процесс $\lambda(t)$ в момент времени t находится в i -м состоянии, то на полуинтервале $[t, t + \Delta t)$, где Δt – достаточно малая величина, с вероятностью $\alpha_i \Delta t + o(\Delta t)$ пребывание процесса $\lambda(t)$ в i -м состоянии закончится и процесс $\lambda(t)$ с вероятностью, равной единице, перейдет из i -го состояния в j -е ($i, j = 1, 2, i \neq j$). В течение временного интервала случайной длительности, когда $\lambda(t) = \lambda_i$, имеет место пуассоновский поток событий с интенсивностью $\lambda_i, i = 1, 2$. Кроме того, переход из первого состояния процесса $\lambda(t)$ во второе возможен в момент наступления события пуассоновского потока интенсивности λ_1 ; переход осуществляется с вероятностью p ($0 \leq p \leq 1$); с вероятностью $1 - p$ процесс $\lambda(t)$ остается в первом состоянии. Переход из второго состояния процесса $\lambda(t)$ в первое возможен также в момент наступления события пуассоновского потока интенсивности λ_2 ; переход осуществляется с вероятностью q ($0 \leq q \leq 1$); с вероятностью $1 - q$ процесс $\lambda(t)$ остается во втором состоянии. В сделанных предположениях $\lambda(t)$ – марковский процесс.

Матрицы инфинитезимальных коэффициентов рассматриваемого потока примут вид

$$D_0 = \begin{pmatrix} -(\lambda_1 + \alpha_1) & \alpha_1 \\ \alpha_2 & -(\lambda_2 + \alpha_2) \end{pmatrix}, \quad D_1 = \begin{pmatrix} (1-p)\lambda_1 & p\lambda_1 \\ q\lambda_2 & (1-q)\lambda_2 \end{pmatrix}.$$

Элементами матрицы D_1 являются интенсивности переходов процесса $\lambda(t)$ из состояния в состояние с наступлением события. Недиagonальные элементы матрицы D_0 – интенсивности переходов из состояния в состояние без наступления события. Диагональные элементы матрицы D_0 – интенсивности выхода процесса $\lambda(t)$ из своих состояний, взятые с противоположным знаком. Отметим, что если $\alpha_i = 0, i = 1, 2$, то имеет место обычный синхронный поток событий [50].

После каждого зарегистрированного события наступает время фиксированной длительности T (далее мертвое время), в течение которого другие события потока недоступны наблюдению. Рассматривается непродлевающееся мертвое время, т.е. события, наступившие в течение интервала мертвого времени не вызывают его продления. По окончании длительности периода мертвого времени первое наступившее событие вновь порождает период мертвого времени длительности T и т.д. Вариант возникающей ситуации представлен на рис. 1, где λ_1, λ_2 – состояния процесса $\lambda(t)$, $t_1, t_2, \dots$ – моменты наступления наблюдаемых событий потока, штриховка – периоды мертвого времени длительности T , ось 1 отображает исходный модулированный поток событий, 2 – схему создания мертвого времени, 3 – наблюдаемые события модулированного синхронного потока.

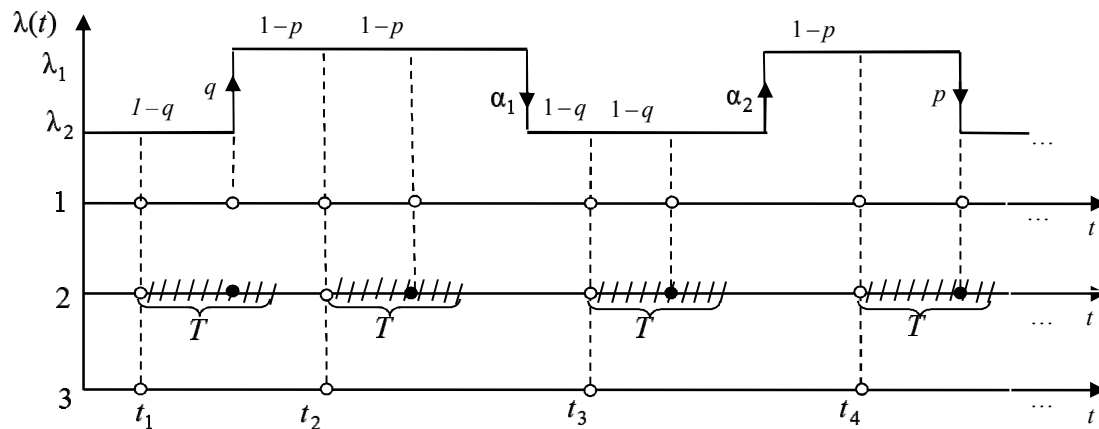


Рис. 1. Формирование наблюдаемого потока событий

Процесс $\lambda(t)$ и типы событий (события пуассоновских потоков с интенсивностями λ_1 либо λ_2) являются принципиально ненаблюдаемыми (в сделанных предположениях $\lambda(t)$ – скрытый марковский процесс), а наблюдаемыми являются только моменты наступления наблюдаемых событий на временной оси $t_1, t_2, \dots$. Рассматривается стационарный режим функционирования потока. Последовательность моментов $t_1, t_2, \dots, t_k, \dots$ наступления событий наблюдаемого потока порождает вложенную цепь Маркова $\{\lambda(t_k)\}$, т.е. поток обладает марковским свойством, если его эволюцию рассматривать с момента t_k (момент наступления события потока), $k = 1, 2, \dots$.

Основной задачей работы является построение оценки $\hat{T}$ длительности мертвого времени методом максимального правдоподобия (в предположении, что остальные параметры потока $\lambda_1, \lambda_2, \alpha_1, \alpha_2, p, q$ известны).

2. Построение функции правдоподобия

Обозначим $\tau_k = t_{k+1} - t_k, k = 1, 2, \dots$, – значение длительности k -го интервала между моментами наступления соседних событий наблюдаемого потока (рис. 2).

Так как поток функционирует в стационарном режиме, то плотность вероятности значений длительности интервала между моментами наступления соседних событий наблюдаемого потока

$p_T(\tau_k) = p_T(\tau)$, $\tau \geq 0$, для любого k (индекс T подчеркивает, что плотность вероятности зависит от длительности мертвого времени). Таким образом, без потери общности момент наступления события t_k можно положить равным нулю, т.е. $\tau = 0$.

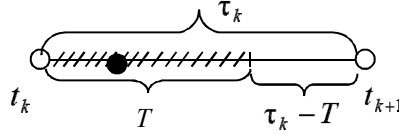


Рис. 2. Интервал между соседними наблюдаемыми событиями потока

Тогда плотность вероятностей $p_T(\tau)$ значений длительности интервала между наступлениями соседних событий модулированного синхронного потока, функционирующего в условиях мертвого времени, запишется в виде [4]:

$$p_T(\tau) = \begin{cases} 0, & 0 \leq \tau < T, \\ \gamma(T)z_1 e^{-z_1(\tau-T)} + (1-\gamma(T))z_2 e^{-z_2(\tau-T)}, & \tau \geq T, \end{cases} \quad (1)$$

$$\gamma(T) = \frac{1}{z_2 - z_1} (z_2 - \lambda_1 \pi_1(T) - \lambda_2 \pi_2(T)),$$

$$z_1 = (\alpha_1 + \lambda_1 + \alpha_2 + \lambda_2 - \sqrt{(\alpha_1 + \lambda_1 - \alpha_2 - \lambda_2)^2 + 4\alpha_1 \alpha_2}),$$

$$z_2 = (\alpha_1 + \lambda_1 + \alpha_2 + \lambda_2) + \sqrt{(\alpha_1 + \lambda_1 - \alpha_2 - \lambda_2)^2 + 4\alpha_1 \alpha_2}, \quad z_1 < z_2,$$

$$\pi_1(T) = \pi_1 - (\pi_1 - \pi_1(0|T))e^{-(\alpha_1 + \alpha_2 + p\lambda_1 + q\lambda_2)T}, \quad \pi_2(T) = \pi_2 - (\pi_2 - \pi_2(0|T))e^{-(\alpha_1 + \alpha_2 + p\lambda_1 + q\lambda_2)T},$$

$$\pi_1 = \frac{\alpha_2 + q\lambda_2}{\alpha_1 + p\lambda_1 + \alpha_2 + q\lambda_2}, \quad \pi_2 = \frac{\alpha_1 + p\lambda_1}{\alpha_1 + p\lambda_1 + \alpha_2 + q\lambda_2}.$$

Пусть $\tau_1 = t_2 - t_1, \tau_2 = t_3 - t_2, \dots, \tau_k = t_{k+1} - t_k, k = 1, 2, \dots$, – последовательность измеренных в результате наблюдения за потоком на интервале наблюдения $(0, t)$ значений длительностей интервалов между соседними событиями наблюдаемого потока. Упорядочим величины $\tau_1, \tau_2, \dots, \tau_k$ по возрастанию: $\tau_{\min} = \tau^{(1)} < \tau^{(2)} < \dots < \tau^{(k)}$. Тогда функция правдоподобия, с учетом (1), примет вид

$$L(\gamma_i, \lambda_i, \alpha_i, p, q, T | \tau^{(1)}, \dots, \tau^{(k)}) = 0 < \tau_{\min} < T,$$

$$L(\lambda_i, \alpha_i, p, q, T | \tau^{(1)}, \dots, \tau^{(k)}) = \prod_{j=1}^k p_T(\tau^{(j)}), \quad T \leq \tau_{\min}.$$

Так как основная задача исследования заключается в построении оценки длительности мертвого времени $\hat{T}$, то согласно методу максимального правдоподобия ее решение есть нахождение максимума функции правдоподобия по неизвестному параметру T :

$$L(T | \tau^{(1)}, \dots, \tau^{(k)}) = \prod_{j=1}^k p_T(\tau^{(j)}) = \prod_{j=1}^k \left\{ \gamma(T)z_1 e^{-z_1(\tau^{(j)}-T)} + (1-\gamma(T))z_2 e^{-z_2(\tau^{(j)}-T)} \right\} \Rightarrow \max_T, \quad 0 \leq T \leq \tau_{\min}, \quad (2)$$

где $z_1, z_2, \gamma(T)$ определены в (1).

Значение T , при котором (2) достигает своего глобального максимума, есть оценка $\hat{T}$ длительности мертвого времени.

3. Решение оптимизационной задачи

Сделаем переобозначение: $\tau_{\min} = \tau_m$. Так как функция правдоподобия (2) отлична от нуля при $0 \leq T \leq \tau_m$, то положим $p_T(\tau^{(j)}) = 0, j = \overline{2, k}$, при $T > \tau_m$ ($\tau_m > 0$). Изучим поведение функции $p_T(\tau_m), 0 \leq T \leq \tau_m$ как функции переменной T . В последующем изложении ситуация, когда принимается

$\tau_m = 0$, означает доопределение изучаемых функций в граничной точке. Рассмотрим производную $p'_T(\tau_m)$ функции $p_T(\tau_m)$ по переменной T . Производная примет вид

$$p'_T(\tau_m) = \frac{F_1(T)e^{-z_1(\tau_m-T)} - F_2(T)e^{-z_2(\tau_m-T)}}{(z_2 - z_1)(\beta_1 + \beta_2)},$$

$$F_1(T) = z_1[(\beta_1 + \beta_2)z_1z_2 - z_1f(T) - f'(T)], \quad F_2(T) = z_2[(\beta_1 + \beta_2)z_1z_2 - z_2f(T) - f'(T)], \quad (3)$$

$$f(T) = P_1 + P_2 + (\lambda_1 - \lambda_2) \frac{(\beta_1 P_1 - \beta_2 P_2)e^{-(\beta_1 + \beta_2)T}}{F(T)}, \quad f'(T) = -(\lambda_1 - \lambda_2)(\beta_1 + \beta_2)z_1z_2 \frac{(\beta_1 P_1 - \beta_2 P_2)e^{-(\beta_1 + \beta_2)T}}{F^2(T)},$$

$$F(T) = z_1z_2 - Pe^{-(\beta_1 + \beta_2)T} > 0, \quad 0 \leq T \leq \tau_m, \quad \tau_m \geq 0,$$

где $\beta_1 = \alpha_1 + p\lambda_1$, $\beta_2 = \alpha_2 + q\lambda_2$, $P_1 = \lambda_1\alpha_2(1-p) + q\lambda_2(\lambda_1 + \alpha_1)$, $P_2 = \lambda_2\alpha_1(1-q) + p\lambda_1(\lambda_2 + \alpha_2)$, $P = \lambda_1\lambda_2(1-p-q)$, $z_1z_2 = \lambda_1\lambda_2 + \lambda_1\alpha_2 + \lambda_2\alpha_1$; z_1, z_2 определены в (1).

Лемма 1. Производная $p'_T(\tau_m)$ – положительная функция переменной τ_m при $T = 0$ ($p'_0(\tau_m) > 0$).

Доказательство. Так как τ_m – любое неотрицательное число ($\tau_m \geq 0$), то $p'_0(\tau_m)$ можно рассматривать как функцию переменной τ_m . Подставляя $T = 0$ в (3), получаем

$$p'_0(\tau_m) = \frac{C}{(z_2 - z_1)A^2} \left\{ z_1 e^{-z_1\tau_m} (z_2 A - C) - z_2 e^{-z_2\tau_m} (z_1 A - C) \right\}, \quad \tau_m \geq 0, \quad (4)$$

где $A = P_1 + P_2$, $C = \lambda_1 P_1 + \lambda_2 P_2$; P_1, P_2 определены в (3); z_1, z_2 определены в (1).

Значение производной (4) в точке $\tau_m = 0$ примет вид

$$p'_0(0) = p'_0(\tau_m = 0) = (C/A)^2 > 0, \quad (5)$$

где A, C определены в (4).

Предел функции $p'_0(\tau_m)$ при $\tau_m \rightarrow \infty$ определится в виде

$$p'_0(\infty) = \lim_{\tau_m \rightarrow \infty} p'_0(\tau_m) = \pm 0. \quad (6)$$

Исследуем функцию $p'_0(\tau_m)$ ($\tau_m > 0$) на присутствие нуля. Рассмотрим уравнение $p'_0(\tau_m) = 0$, которое приводится к виду

$$e^{-(z_2 - z_1)\tau_m} = \frac{z_1(z_2 A - C)}{z_2(z_1 A - C)}, \quad \tau_m > 0. \quad (7)$$

Обозначим $B = B_1 / B_2$; $B_1 = z_1(z_2 A - C)$, $B_2 = z_2(z_1 A - C)$. Подчеркнем, что если $B_2 = 0$, то $z_1 A - C = 0$, т.е. $z_1 A = C$. Тогда $z_2 A - C = (z_2 - z_1)A > 0$ и из (4) следует $p'_0(\tau_m) > 0, \tau_m \geq 0$.

Преобразовывая B_1 , получаем

$$B_1 = \frac{1}{2} z_1^2 \left\{ -2A + (\beta_1 + \beta_2)(z_1 + z_2) + (\beta_1 + \beta_2) \sqrt{(\lambda_1 - \lambda_2 + \alpha_1 - \alpha_2)^2 + 4\alpha_1\alpha_2} \right\} > \\ > \frac{1}{2} z_1^2 \left\{ -2A + (\beta_1 + \beta_2)(z_1 + z_2) + (\beta_1 + \beta_2) |\lambda_1 - \lambda_2 + \alpha_1 - \alpha_2| \right\}.$$

Нетрудно показать, что правая часть последнего неравенства всегда строго больше нуля. Тогда $B_1 > 0$. Таким образом, в уравнении (7) имеют место два варианта: 1) $B_1 > 0$, $B_2 < 0$; 2) $B_1 > 0$, $B_2 > 0$. Для первого варианта $B < 0$, и уравнение (7) не имеет решения. Для второго варианта $B > 1$, так как $B_1 - B_2 = (z_2 - z_1)C > 0$, и уравнение (7) также не имеет решения. Отсюда следует, что производная $p'_0(\tau_m)$ не достигает нуля для $\tau_m > 0$. Тогда (6) принимает вид $p'_0(\infty) = +0$. Объединение (5), (6) и утверждения $p'_0(\tau_m) \neq 0, \tau_m > 0$, доказывает лемму 1.

Из леммы 1 следует, что плотность вероятности $p_T(\tau_m)$ – возрастающая функция переменной T в точке $T = 0$ ($\tau_m \geq 0$).

Лемма 2. Производная $p'_T(\tau_m)$ при $T = \tau_m$ ($\tau_m \geq 0$) строго больше нуля ($p'_{T=\tau_m}(\tau_m) = p'_0(\tau_m) > 0$).

Доказательство. Подставляя $T = \tau_m$ в (3), получаем, что

$$p'(\tau_m) = \frac{1}{\beta_1 + \beta_2} \left\{ C + (\lambda_1 - \lambda_2)(\beta_1 P_1 - \beta_2 P_2) \frac{e^{-(\beta_1 + \beta_2)\tau_m}}{(z_1 z_2 - P e^{-(\beta_1 + \beta_2)\tau_m})^2} X(\tau_m) \right\}, \quad (8)$$

$$X(\tau_m) = C + P(z_1 + z_2)(1 - e^{-(\beta_1 + \beta_2)\tau_m}), \tau_m \geq 0,$$

где C определена в (4); $\beta_1, \beta_2, P_1, P_2, P$ определены в (3); z_1, z_2 определены в (1).

Нетрудно показать, что $X(\tau_m) > 0$ для $\tau_m \geq 0$. Последнее означает, что знак производной $p'(\tau_m)$ определяется множителем $(\beta_1 P_1 - \beta_2 P_2)$.

Рассмотрим $p'(\tau_m)$ как функцию τ_m ($\tau_m \geq 0$). Имеем

$$p'(\tau_m = 0) = (C/A)^2 > 0; \quad p'(\tau_m = \infty) = \lim_{\tau_m \rightarrow \infty} p'(\tau_m) = C/(\beta_1 + \beta_2) > 0.$$

Знак разности $p'(\tau_m = 0) - p'(\tau_m = \infty) = C(\lambda_1 - \lambda_2)(\beta_1 P_1 - \beta_2 P_2)/(A^2(\beta_1 + \beta_2))$ определяется знаком множителя $(\beta_1 P_1 - \beta_2 P_2)$: если $(\beta_1 P_1 - \beta_2 P_2) > 0$, то $p'(\tau_m = 0) > p'(\tau_m = \infty)$; если $(\beta_1 P_1 - \beta_2 P_2) < 0$, то $p'(\tau_m = 0) < p'(\tau_m = \infty)$; если $(\beta_1 P_1 - \beta_2 P_2) = 0$, то $p'(\tau_m = 0) = p'(\tau_m = \infty)$.

Из вида производной $p'(\tau_m)$ следует, что в случае, когда $(\beta_1 P_1 - \beta_2 P_2) \geq 0$, функция $p'(\tau_m) > 0$, т.е. функция $p_T(\tau_m)$ есть возрастающая функция переменной T в точке $T = \tau_m$ ($\tau_m \geq 0$). Случай $(\beta_1 P_1 - \beta_2 P_2) < 0$ требует отдельного рассмотрения. Для этого исследуем вторую производную $p''(\tau_m)$ функции $p(\tau_m)$. Производная $p''(\tau_m)$, с учетом (8), примет вид

$$p''(\tau_m) = -(\lambda_1 - \lambda_2)(\beta_1 P_1 - \beta_2 P_2) z_1 z_2 \frac{e^{-(\beta_1 + \beta_2)\tau_m}}{F^3(\tau_m)} Y(\tau_m), \quad (9)$$

$$Y(\tau_m) = \left\{ C + (z_1 + z_2)P - (z_1 + z_2 + \beta_1 + \beta_2)P e^{-(\beta_1 + \beta_2)\tau_m} \right\}, \tau_m \geq 0,$$

где C определена в (4); $\beta_1, \beta_2, P_1, P_2, P, F(T = \tau_m)$ определены в (3); z_1, z_2 определены в (1).

Так как рассматривается случай $(\beta_1 P_1 - \beta_2 P_2) < 0$, то множитель в (9) $-(\lambda_1 - \lambda_2)(\beta_1 P_1 - \beta_2 P_2) z_1 z_2 e^{-(\beta_1 + \beta_2)\tau_m} / F^3(\tau_m) > 0$; тогда знак производной $p''(\tau_m)$ определяется знаком выражения $Y(\tau_m)$. Исследуем функцию $Y(\tau_m)$. Имеем $Y(\infty) = C + (z_1 + z_2)P \geq 0$, $Y(0) = C - P(\beta_1 + \beta_2)$, при этом $Y(0) > 0$, либо $Y(0) < 0$, либо $Y(0) = 0$. Производная $Y'(\tau_m)$ примет вид

$$Y'(\tau_m) = (\beta_1 + \beta_2)(z_1 + z_2 + \beta_1 + \beta_2)P e^{-(\beta_1 + \beta_2)\tau_m}, \tau_m \geq 0.$$

Знак производной $Y'(\tau_m)$ зависит от множителя P : 1) если $P < 0$, то $Y'(\tau_m) < 0$ ($\tau_m \geq 0$) и тогда $p''(\tau_m) > 0$ ($\tau_m \geq 0$); 2) если $P = 0$, то $Y(\tau_m) = C > 0$ ($\tau_m \geq 0$) и тогда $p''(\tau_m) > 0$ ($\tau_m \geq 0$); 3) если $P > 0$, то $Y'(\tau_m) > 0$ ($\tau_m \geq 0$) и тогда $p''(\tau_m) > 0$, либо $p''(\tau_m) < 0$, либо $p''(\tau_m) = 0$ ($\tau_m \geq 0$).

Для первых двух случаев производная $p'(\tau_m) > 0$ ($\tau_m \geq 0$). Если реализуется третий случай, то возможны варианты поведения функции $p'(\tau_m)$ ($\tau_m \geq 0$): а) если $Y(0) = C - P(\beta_1 + \beta_2) > 0$, то $p''(\tau_m) > 0$ ($\tau_m \geq 0$); тогда $p'(\tau_m) > 0$ ($\tau_m \geq 0$); б) если $Y(0) = C - P(\beta_1 + \beta_2) = 0$, то $p''(\tau_m) = 0$ ($\tau_m = 0$), $p''(\tau_m) > 0$ ($\tau_m > 0$); тогда $p'(\tau_m) > 0$ ($\tau_m \geq 0$); в) если $Y(0) = C - P(\beta_1 + \beta_2) < 0$, то поведение $p''(\tau_m)$ определяется тремя ситуациями: 1) $0 \leq \tau_m < \tau_m^*$, тогда $p''(\tau_m) < 0$; 2) $\tau_m = \tau_m^*$, тогда $p''(\tau_m) = 0$; 3) $\tau_m^* < \tau_m$, тогда $p''(\tau_m) > 0$.

Таким образом, при реализации варианта $Y(0) = C - P(\beta_1 + \beta_2) < 0$ производная $p'(\tau_m)$ ($\tau_m > 0$) достигает своего глобального минимума в точке $\tau_m = \tau_m^*$. При этом точка τ_m^* определяется выражением

$$\tau_m^* = -\frac{1}{\beta_1 + \beta_2} \ln \frac{z_1 z_2 (z_1 + z_2 - \beta_1 - \beta_2)}{(z_1 + z_2 + \beta_1 + \beta_2)P}, \quad P > 0. \quad (10)$$

Подставляя (10) в (8), находим выражение для производной $p'(\tau_m)$ в точке минимума $\tau_m = \tau_m^*$:

$$p'(\tau_m^*) = \frac{1}{\beta_1 + \beta_2} \left\{ C + (\lambda_1 - \lambda_2)(\beta_1 P_1 - \beta_2 P_2) \frac{(z_1 + z_2 - \beta_1 - \beta_2)^2}{4P(\beta_1 + \beta_2)} \right\}, P > 0. \quad (11)$$

Можно показать, что производная, определяемая в (11), строго больше нуля ($p'(\tau_m^*) > 0$, $P > 0$). Отсюда следует, что при реализации варианта $Y(0) = C - P(\beta_1 + \beta_2) < 0$ выполняется $p'(\tau_m) > 0$ ($\tau_m \geq 0$). Лемма 2 доказана.

Далее изучим поведение производной $p'_T(\tau_m)$ как функции переменной T на интервале $(0, \tau_m)$.

Рассмотрим на предмет существования корней уравнение $p'_T(\tau_m) = 0$, которое, с учетом (3), преобразуется в выражение

$$\Psi(T) = e^{-(z_2 - z_1)(\tau_m - T)}, \Psi(T) = \frac{F_1(T)}{F_2(T)}, 0 \leq T \leq \tau_m, \quad (12)$$

где $F_1(T), F_2(T)$ определены в (3). Так как, в принципе, τ_m может быть сколь угодно большим числом, то изучим поведение функции $\Psi(T)$ при $T \geq 0$.

Отметим, что $e^{-(z_2 - z_1)\tau_m} \leq e^{-(z_2 - z_1)(\tau_m - T)} \leq 1, 0 \leq T \leq \tau_m, \tau_m \geq 0$.

Обозначим $h = 1 - p - q$, тогда $P = \lambda_1 \lambda_2 h$. Величину $(\lambda_1 - \lambda_2)(\beta_1 P_1 - \beta_2 P_2)$ представим в виде

$$(\lambda_1 - \lambda_2)(\beta_1 P_1 - \beta_2 P_2) = -(\lambda_1 \lambda_2)^2 h^2 + \lambda_1 \lambda_2 [2z_1 z_2 - (\beta_1 + \beta_2)(z_1 + z_2)]h - z_1 z_2 (z_1 - \beta_1 - \beta_2)(z_2 - \beta_1 - \beta_2) = x(h), -1 \leq h \leq 1. \quad (13)$$

Функция (13) достигает нуля в точках $h = h_1$ и $h = h_2$; $h_1 = z_2(z_1 - \beta_1 - \beta_2)/\lambda_1 \lambda_2$, $h_2 = z_1(z_2 - \beta_1 - \beta_2)/\lambda_1 \lambda_2$, $h_1 < h_2$.

Исследуем функцию $F_1(T)$, $T \geq 0$, из (12) и определенную в (3).

Лемма 3. Функция $F_1(T) > 0$ при $T \geq 0$.

Доказательство. Имеем

$$F_1(0) = z_1(\beta_1 + \beta_2)(z_2 A - C)C/A^2 > 0, F_1(\infty) = \lim_{T \rightarrow \infty} F_1(T) = z_1(z_2 A - C) > 0,$$

так как $(z_2 A - C) > 0$ (лемма 1).

Тогда $(z_2 A - C) = \lambda_1 \lambda_2 z_1 (h - h_1) > 0$. Отсюда следует, что всегда $h > h_1$.

Для дальнейшего исследования $F_1(T)$ на знак представим ее как функцию, зависящую от параметра h ($F_1(T) = F_1(T, h)$):

$$F_1(T, h) = z_1 [(\beta_1 + \beta_2)z_1 z_2 - z_1 f(T, h) - f'(T, h)], T \geq 0, -1 \leq h \leq 1, \quad (14)$$

$$f(T, h) = A + x(h) \frac{e^{-(\beta_1 + \beta_2)T}}{F(T, h)}, f'(T, h) = -z_1 z_2 (\beta_1 + \beta_2) x(h) \frac{e^{-(\beta_1 + \beta_2)T}}{F^2(T, h)},$$

где $x(h)$ определена в (13); A – в (4); β_1, β_2 ; $F(T, h)$ – в (3); z_1, z_2 – в (1).

Знак функции $F_1(T, h)$ зависит от знака функции $x(h)$. Рассмотрим все возможные варианты.

Пусть $x(h) = 0$. Это возможно, как следует из (13), если $h = h_1$ либо $h = h_2$. Так как всегда $h > h_1$, то случай $h = h_1$ исключается. Тогда для случая $h = h_2$ имеем $F_1(T, h = h_2) = z_1^2 (z_2 - z_1)(\beta_1 + \beta_2) > 0$, $T \geq 0$.

Пусть $x(h) > 0$. Последнее возможно, если реализуется один из следующих вариантов: $0 < h_1 < h < h_2$; $0 = h_1 < h < h_2$; $h_1 < 0 < h < h_2$; $h_1 < h = 0 < h_2$; $h_1 < h < 0 < h_2$; $h_1 < h < h_2 = 0$; $h_1 < h < h_2 < 0$.

Рассмотрим вариант $0 < h_1 < h < h_2$. Имеем

$$\frac{\partial F_1(T, h)}{\partial T} = \frac{z_1^2 z_2 (\beta_1 + \beta_2) x(h) e^{-(\beta_1 + \beta_2)T}}{F^3(T, h)} \left\{ z_1 z_2 (z_1 - \beta_1 - \beta_2) - \lambda_1 \lambda_2 (z_1 + \beta_1 + \beta_2) h e^{-(\beta_1 + \beta_2)T} \right\} \quad (15)$$

$$T \geq 0, 0 < h_1 < h < h_2.$$

Знак производной (15) определяется знаком функции:

$$y_1(T, h) = z_1 z_2 (z_1 - \beta_1 - \beta_2) - \lambda_1 \lambda_2 (z_1 + \beta_1 + \beta_2) h e^{-(\beta_1 + \beta_2)T}, \quad T \geq 0, 0 < h_1 < h < h_2. \quad (16)$$

Тогда, поскольку $h > 0$, функция (15) – возрастающая функция переменной T (возрастает от $y_1(T=0, h) = z_1 z_2 (z_1 - \beta_1 - \beta_2) - \lambda_1 \lambda_2 (z_1 + \beta_1 + \beta_2) h$ до $y_1(T=\infty) = z_1 z_2 (z_1 - \beta_1 - \beta_2) > 0$). Тогда $y_1(T=0, h) < 0$ для $0 < h_1 < h < h_2$. Отсюда следует, что функция (16) проходит через ноль в точке

$$T_1(h) = -\frac{1}{(\beta_1 + \beta_2)} \ln \frac{z_1 z_2 (z_1 - \beta_1 - \beta_2)}{\lambda_1 \lambda_2 (z_1 + \beta_1 + \beta_2) h}, \quad 0 < h_1 < h < h_2.$$

В силу этого поведение производной $F'_1(T, h)$ (определяемой в (15)) в зависимости от T будет выглядеть следующим образом: $F'_1(T, h) < 0, 0 \leq T < T_1(h)$; $F'_1(T, h) = 0, T = T_1(h)$; $F'_1(T, h) > 0, T > T_1(h)$. Таким образом, функция $F_1(T, h)$ в точке $T = T_1(h)$ достигает своего глобального минимума, при этом $F_1(T=0, h) > F_1(T=\infty, h)$, $0 < h_1 < h < h_2$. Покажем, что в точке $T = T_1(h)$ функция $F_1(T, h)$ строго больше нуля ($F_1(T = T_1, h) > 0, 0 < h_1 < h < h_2$). Имеем

$$F_1(T = T_1(h), h) = z_1 \left\{ (\lambda_1 \lambda_2)^2 (z_1 + \beta_1 + \beta_2)^2 h^2 - \lambda_1 \lambda_2 (z_1 - \beta_1 - \beta_2) [z_1 (z_1 - \beta_1 - \beta_2)(z_2 - \beta_1 - \beta_2) + \right. \\ \left. + z_2 (z_1 + \beta_1 + \beta_2)^2] h + z_1 z_2 (z_1 - \beta_1 - \beta_2)^3 (z_2 - \beta_1 - \beta_2) \right\}, \quad 0 < h_1 < h < h_2. \quad (17)$$

Функция (17), в частности, достигает нуля в точке $h = h_1$. Тогда при $h > h_1$ имеет место $F_1(T = T_1, h) > 0$. Используя формулы (15), (16), показывается, что $F_1(T, h) > 0, T \geq 0$, для оставшихся вариантов, реализующихся для $x(h) > 0$.

Пусть $x(h) < 0$. Это возможно, если реализуется один из следующих вариантов: $0 < h_1 < h_2 < h \leq 1$; $0 = h_1 < h_2 < h \leq 1$; $h_1 < 0 < h_2 < h \leq 1$; $h_1 < h_2 = 0 < h \leq 1$; $h_1 < h_2 < 0 < h \leq 1$; $h_1 < h_2 < h = 0$; $h_1 < h_2 < h < 0$.

Используя формулы (15)–(17), показывается, что для всех приведенных вариантов функция $F_1(T, h) > 0, T \geq 0, h > h_2$. Лемма 3 доказана.

Перейдем к рассмотрению функции $F_2(T)$, $T \geq 0$, из (12) и определенной в (3).

Лемма 4. 1) Функция $F_2(T) = 0$ ($T \geq 0$), если $x(h) = 0$; 2) функция $F_2(T) < 0$ ($T \geq 0$), если $x(h) > 0$.

Доказательство. Имеем

$$F_2(0) = \lambda_1 \lambda_2 z_2^2 (\beta_1 + \beta_2) (h - h_2) C / A^2, \quad F_2(\infty) = \lim_{T \rightarrow \infty} F_2(T) = \lambda_1 \lambda_2 z_2^2 (h - h_2).$$

Для дальнейшего исследования функции $F_2(T)$ на знак представим ее как функцию, зависящую от параметра h ($F_2(T) = F_2(T, h)$):

$$F_2(T, h) = z_2 [(\beta_1 + \beta_2) z_1 z_2 - z_2 f(T, h) - f'(T, h)], \quad T \geq 0, -1 \leq h \leq 1, \quad (18)$$

где $f(T, h)$, $f'(T, h)$ определены в (14).

Знак функции $F_2(T, h)$ зависит от знака функции $x(h)$, определенной в (13).

Пусть $x(h) = 0$. Это возможно, как следует из (13), если $h = h_1$ либо $h = h_2$. Так как всегда $h > h_1$ (лемма 3), то случай $h = h_1$ исключается. Тогда для случая $h = h_2$, подставляя h_2 в (18), имеем $F_2(T, h = h_2) = 0, T \geq 0$.

Пусть $x(h) > 0$. Неравенство выполняется, если реализуется один из вариантов, приведенных в лемме 3 для $x(h) > 0$.

Рассмотрим вариант $0 < h_1 < h < h_2$. Имеем

$$\frac{\partial F_2(T, h)}{\partial T} = \frac{z_1 z_2^2 (\beta_1 + \beta_2) x(h) e^{-(\beta_1 + \beta_2)T}}{F^3(T, h)} \left\{ z_1 z_2 (z_2 - \beta_1 - \beta_2) - \lambda_1 \lambda_2 (z_2 + \beta_1 + \beta_2) h e^{-(\beta_1 + \beta_2)T} \right\}, \quad (19)$$

$$T \geq 0, 0 < h_1 < h < h_2.$$

Знак производной (19) определяется знаком функции:

$$y_2(T, h) = z_1 z_2 (z_2 - \beta_1 - \beta_2) - \lambda_1 \lambda_2 (z_2 + \beta_1 + \beta_2) h e^{-(\beta_1 + \beta_2)T}, \quad T \geq 0, \quad 0 < h_1 < h < h_2. \quad (20)$$

Тогда, поскольку $h > 0$, функция (20) – возрастающая функция переменной T (возрастает от $y_2(T=0, h) = z_1 z_2 (z_2 - \beta_1 - \beta_2) - \lambda_1 \lambda_2 (z_2 + \beta_1 + \beta_2) h$ до $y_2(T=\infty, h) = z_1 z_2 (z_2 - \beta_1 - \beta_2) > 0$). Здесь возможны две ситуации: а) $y_2(T=0, h) < 0$, $0 < h_1 < h < h_2$; б) $y_2(T=0, h) > 0$, $0 < h_1 < h < h_2^*$; $y_2(T=0, h) = 0$, $h = h_2^*$, $y_2(T=0, h) < 0$, $h_2^* < h < h_2$, где $h_2^* = z_1 z_2 (z_2 - \beta_1 - \beta_2) [\lambda_1 \lambda_2 (z_2 + \beta_1 + \beta_2)]^{-1}$.

Рассмотрим ситуацию а). Тогда функция (20) проходит через ноль в точке

$$T_2(h) = -\frac{1}{(\beta_1 + \beta_2)} \ln \frac{z_1 z_2 (z_2 - \beta_1 - \beta_2)}{\lambda_1 \lambda_2 (z_2 + \beta_1 + \beta_2) h}, \quad 0 < h_1 < h < h_2.$$

В силу этого поведение производной $F_2'(T, h)$ (определяемой в (19)) в зависимости от T будет выглядеть следующим образом: $F_2'(T, h) < 0$, $0 \leq T < T_2(h)$; $F_2'(T, h) = 0$, $T = T_2(h)$; $F_2'(T, h) > 0$, $T > T_2(h)$. Таким образом, функция $F_2(T, h)$ в точке $T = T_2(h)$ достигает своего глобального минимума, при этом $F_2(0) = F_2(T=0, h) < 0$, $F_2(\infty) = F_2(T=\infty, h) < 0$, $F_2(T=0, h) < F_2(T=\infty, h)$, $0 < h_1 < h < h_2$. Все это доказывает, что $F_2(T, h) < 0$, $T \geq 0$, для ситуации а). Аналогично доказывается, что для ситуации б) функция $F_2(T, h) < 0$, $T \geq 0$.

Используя формулы (19), (20), показывается, что для оставшихся вариантов, приведенных в лемме 3, $F_2(T, h) < 0$, $T \geq 0$. Лемма 4 доказана.

Для случая $x(h) < 0$ имеет место лемма.

Лемма 5. Функция $F_2(T) > 0$ ($T \geq 0$), если $x(h) < 0$.

Доказательство осуществляется применением формул (19), (20) (аналогично применению этих формул в лемме 4) для вариантов изменения параметра h , приведенных в лемме 3 для случая $x(h) < 0$. Лемма 5 доказана.

Введем в рассмотрение функцию

$$\Phi(T, h) = F_1(T, h) - F_2(T, h) = (z_2 - z_1) [-z_1 z_2 (\beta_1 + \beta_2) + (z_1 + z_2) f(T, h) + f'(T, h)], \quad T \geq 0, \quad -1 \leq h \leq 1, \quad (21)$$

где $F_1(T, h)$, $f(T, h)$, $f'(T, h)$ определены в (14); $F_2(T, h)$ – в (18).

Лемма 6. Функция $\Phi(T, h) > 0$, $T \geq 0$, если $x(h) \geq 0$.

Доказательство. Имеем

$$\Phi(T=0, h) = (z_2 - z_1)(\beta_1 + \beta_2)(C/A)^2 > 0, \quad \Phi(T=\infty, h) = \lim_{T \rightarrow \infty} \Phi(T, h) = (z_2 - z_1)C > 0.$$

Знак функции $\Phi(T, h)$ зависит от знака функции $x(h)$, определенной в (13).

Пусть $x(h) = 0$. Это возможно, как следует из (13), если $h = h_1$ либо $h = h_2$. Так как всегда $h > h_1$ (лемма 3), то случай $h = h_1$ исключается. Тогда для случая $h = h_2$, подставляя h_2 в (21), получаем $\Phi_2(T, h = h_2) = (z_2 - z_1)(\beta_1 + \beta_2)z_1^2 > 0$, $T \geq 0$.

Пусть $x(h) > 0$. Неравенство выполняется, если реализуется один из вариантов, приведенных в лемме 3 для $x(h) > 0$.

Рассмотрим вариант $0 < h_1 < h < h_2$. Имеем

$$\begin{aligned} \partial \Phi(T, h) / \partial T = & -(1/F^3(T, h)) z_1 z_2 (z_2 - z_1)(\beta_1 + \beta_2) x(h) e^{-(\beta_1 + \beta_2)T} \times \\ & \times \left\{ z_1 z_2 (z_1 + z_2 - \beta_1 - \beta_2) - \lambda_1 \lambda_2 (z_1 + z_2 + \beta_1 + \beta_2) h e^{-(\beta_1 + \beta_2)T} \right\}, \quad T \geq 0, \quad 0 < h_1 < h < h_2. \end{aligned} \quad (22)$$

Знак производной (22) определяется знаком функции:

$$y(T, h) = z_1 z_2 (z_1 + z_2 - \beta_1 - \beta_2) - \lambda_1 \lambda_2 (z_1 + z_2 + \beta_1 + \beta_2) h e^{-(\beta_1 + \beta_2)T}, \quad T \geq 0, \quad 0 < h_1 < h < h_2. \quad (23)$$

Тогда, поскольку $h > 0$, функция (23) – возрастающая функция переменной T (возрастает от $y(T=0, h) = z_1 z_2 (z_1 + z_2 - \beta_1 - \beta_2) - \lambda_1 \lambda_2 (z_1 + z_2 + \beta_1 + \beta_2) h$ до $y(T=\infty, h) = z_1 z_2 (z_1 + z_2 - \beta_1 - \beta_2) > 0$).

Здесь возможны три ситуации: а) $y(T=0, h) > 0$, $0 < h_1 < h < h_2$; б) $y(T=0, h=h_2) = 0$; в) $y(T=0, h) > 0$, $0 < h_1 < h^*$; $y(T=0, h) = 0$, $h = h^*$; $y(T=0, h) < 0$, $h^* < h < h_2$, где $h^* = z_1 z_2 (z_1 + z_2 - \beta_1 - \beta_2) \times [\lambda_1 \lambda_2 (z_1 + z_2 + \beta_1 + \beta_2)]^{-1}$.

Рассмотрим ситуацию а). Тогда функция (23) – положительная функция переменной T ($y(T, h) > 0, T \geq 0$). В силу этого производная (22) – отрицательная функция переменной T ($T \geq 0$). Все это означает, что функция (21) убывает от $\Phi(T=0, h) > 0$ до $\Phi(T=\infty, h) > 0$. Отсюда следует, что для ситуации а) функция $\Phi(T, h) > 0, T \geq 0$. Аналогично доказывается, что для ситуаций б), в) функция $\Phi(T, h) > 0, T \geq 0$.

Используя формулы (22), (23), показывается, что для оставшихся вариантов, приведенных в лемме 3 для $x(h) > 0$, функция $\Phi(T, h) > 0, T \geq 0$. Лемма 6 доказана.

Для случая $x(h) < 0$ имеет место лемма.

Лемма 7. Функция $\Phi(T, h) > 0, T \geq 0$, если $x(h) < 0$.

Доказательство осуществляется применением формул (22), (23) (аналогично применению этих формул в лемме 6) для вариантов изменения параметра h , приведенных в лемме 3 для случая $x(h) < 0$. Лемма 7 доказана.

Лемма 8. Функции $F_1(T, h)$, $F_2(T, h)$ подчиняются неравенству: $F_1(T, h) > F_2(T, h)$, $T \geq 0$, $-1 \leq h \leq 1$.

Доказательство осуществляется объединением результатов лемм 6, 7. Лемма 8 доказана.

Лемма 9. Уравнение (12) решения не имеет.

Доказательство вытекает из последовательного применения лемм 3–5 и леммы 8. Лемма 9 доказана.

Лемма 10. Производная $p'(\tau_m)$, определяемая формулой (3), является положительной функцией переменной T ($p'(\tau_m) > 0, 0 \leq T \leq \tau_m, \tau_m \geq 0$).

Доказательство осуществляется применением результатов лемм 1, 2, 9. Лемма 10 доказана.

Теорема 1. Плотность вероятности $p_T(\tau_m)$ – возрастающая функция переменной T ($0 \leq T \leq \tau_m, \tau_m \geq 0$).

Доказательство осуществляется применением результата леммы 10. Теорема 1 доказана.

Теорема 2. Плотность вероятности $p_T(\tau_m)$ достигает своего глобального максимума в точке $T = \tau_m$ ($0 \leq T \leq \tau_m, \tau_m \geq 0$).

Доказательство вытекает из результата теоремы 1.

Следствие 1. Из теоремы 1 следует, что функции $p_T(\tau^{(j)})$, $j = \overline{1, k}$, из (2) – возрастающие функции переменной T ($0 \leq T \leq \tau_m, \tau_m \geq 0$).

Следствие 2. Из теоремы 2 следует, что функция правдоподобия $L(T | \tau^{(1)}, \dots, \tau^{(k)})$, определяемая в (2), достигает своего глобального максимума в точке $T = \tau_m$, т.е. решением оптимизационной задачи (2) является оценка длительности мертвого времени $\hat{T} = \tau_m$.

Заключение

Полученный результат делает возможным решение задачи оценки длительности мертвого времени без привлечения численных методов: в процессе наблюдения (в течение временного интервала (t_0, t)) потока событий вычисляются величины τ_k , $k = \overline{1, n}$, после чего находится $\tau_m = \min \tau_k$ ($k = \overline{1, n}$) и полагается $\hat{T} = \tau_m$.

ЛИТЕРАТУРА

1. Горцев А.М., Голофастова М.Н. Оптимальная оценка состояний модулированного синхронного дважды стохастического потока событий // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2013. № 2 (23). С. 42–53.

2. Сиротина М.Н. Оптимальная оценка состояний модулированного синхронного дважды стохастического потока событий в условиях непродлевающегося мертвого времени // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2014. № 1 (26). С. 63–74.
3. Gortsev A., Sirotina M. Joint Probability Density Function of Modulated Synchronous Flow Interval Duration // Communications in Computer and Information Science. 2014. V. 487. P. 145–152.
4. Сиротина М.Н., Горцев А.М. Плотность вероятностей длительности интервала между соседними событиями в модулированном синхронном потоке при непродлевающемся мертвом времени // Материалы Международной научной конференции, посвящ. 80-летию проф., д-ра физ.-мат. наук Геннадия Алексеевича Медведева. Минск. 23–26 февраля 2015. Минск : Изд-во РИВШ, 2015. С. 299–304.
5. Дудин А.Н., Клименок В.Н. Системы массового обслуживания с коррелированными потоками. Минск: Изд-во БГУ, 2000. 175 с.
6. Башарин Г.П., Гайдамака Ю.В., Самуйлов К.Е. Математическая теория телетрафика и ее применения к анализу мультисервисных сетей связи следующих поколений // Автоматика и вычислительная техника. 2013. № 2. С. 11–21.
7. Kingman J.F.C. On doubly stochastic Poisson process // Proceedings Cambridge Philosophical Society. 1964. V. 60, No. 4. P. 923–930.
8. Башарин Г.П., Кокотушкин В.А., Наумов В.А. О методе эквивалентных замен расчета фрагментов сетей связи // Изв. АН СССР. Техн. кибернетика. 1979. № 6. С. 92–99.
9. Neuts M.F. A versatile Markov point process // Journal of Applied Probability. 1979. V. 16. P. 764–779.
10. Lucantoni D.M. New results on the single server queue with a batch markovian arrival process // Communication in Statistics Stochastic Models. 1991. V. 7. P. 1–46.
11. Lucantoni D.M., Neuts, M.F. Some steady-state distributions for the MAP/SM/1 queue // Communication in Statistics Stochastic Models. 1994. V. 10. P. 575–598.
12. Лившиц К.Н., Бублик Я.С. Вероятность разорения страховой компании при дважды стохастическом потоке страховых выплат // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2010. № 1 (10). С. 66–77.
13. Yang Y.W. Woo Shin. BMAP/G/1 queue with correlated arrivals of customers and disasters // Operation Research Letters. 2004. V. 32, Issue 4. P. 364–373.
14. Artalejo Jesus R., Chakravarthy Srinivas R. Computational analysis of the maximal queue length in the MAP/M/c retrieval queue // Applied Mathematics and Computation. 2006. V. 183, Issue 2. P. 1399–1409.
15. Best J. Doubly Stochastic Processes: an Approach for Understanding Central Nervous System Activity // Selected Topics on Applied Mathematics, Circuits, Systems and Signals. WSEAS Press, 2009. P. 155–158.
16. Bushalnov I.V., Gortsev A.M., Nezhel'skaya L.A. Estimating parameters of the synchronous twofold-stochastic flow of events // Automation and Remote Control. 2008. V. 69, No. 9. P. 1517–1533.
17. Горцев А.М., Нежелская Л.А. Оценивание длительности «мертвого времени» и интенсивностей синхронного дважды стохастического потока событий // Радиотехника. 2004. № 10. С. 8–16.
18. Василевская Т.П., Горцев А.М., Нежелская Л.А. Оценивание длительности мертвого времени и параметров синхронного альтернирующего потока с проявлением либо непроявлением событий // Вестник Томского государственного университета. 2004. № 9. С. 129–138.
19. Горцев А.М., Нежелская Л.А. Оценивание длительности мертвого времени и параметров синхронного альтернирующего потока событий // Вестник Томского государственного университета. 2003. № 6. С. 232–239.
20. Горцев А.М., Нежелская Л.А. Оценивание параметров синхронного дважды стохастического потока событий методом моментов // Вестник Томского государственного университета. 2002. № 1. С. 24–29.
21. Горцев А.М., Нежелская Л.А. Оценка параметров синхронного альтернирующего пуассоновского потока событий методом моментов // Радиотехника. 1995. № 7–8. С. 6–10.
22. Леонова М.А., Нежелская Л.А. Оценка максимального правдоподобия длительности мертвого времени в обобщенном асинхронном потоке событий // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2013. № 2 (23). С. 54–63.
23. Горцев А.М., Леонова М.А., Нежелская Л.А. Совместная плотность вероятностей длительности интервалов обобщенного асинхронного потока событий при непродлевающемся мертвом времени // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2012. № 4 (21). С. 14–25.
24. Леонова М.А., Нежелская Л.А. Вероятность ошибки при оценивании состояний обобщенного синхронного потока событий // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2012. № 2 (19). С. 88–101.
25. Gortsev A.M., Nezhel'skaya L.A. An asynchronous double stochastic flow with initiation of superfluous events // Discrete Mathematics and Applications. 2011. V. 21, No. 3. P. 283–290.
26. Горцев А.М., Нежелская Л.А. Оценивание параметров асинхронного потока с иницированием лишних событий методом моментов // Вестник Томского государственного университета. 2006. № 18. С. 267–273.
27. Горцев А.М., Нежелская Л.А., Шевченко Т.И. Оценивание состояний МС-потока событий при наличии ошибок измерений // Известия высших учебных заведений. Физика. 1993. № 12. С. 67–85.
28. Калягин А.А., Нежелская Л.А. Оценка длительности мертвого времени в обобщенном полусинхронном потоке событий // Новые информационные технологии в исследовании сложных структур : материалы Десятой рос. конф. с междунар. участием (9–13 июня 2014 г.). Томск : Издательский Дом Томского государственного университета, 2014. С. 96–97.

29. Горцев А.М., Калягин А.А., Нежелская Л.А. Оптимальная оценка состояний обобщенного полусинхронного потока событий // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2010. № 2 (11). С. 66–81.
30. Горцев А.М., Нежелская Л.А. Полусинхронный дважды стохастический поток событий при продлеваемом мертвом времени // Вычислительные технологии. 2008. Т. 13, № 1. С. 31–41.
31. Gortsev A.M., Nezhel'skaya L.A. Estimation of the dead-time period and parameters of a semi-synchronous double-stochastic stream of events // Measurement Techniques. 2003. V. 46, No. 6. P. 536–545.
32. Горцев А.М., Нежелская Л.А. Оценивание параметров полусинхронного дважды стохастического потока событий методом моментов // Вестник Томского государственного университета. 2002. № 1. С. 18–23.
33. Нежелская Л.А. Оптимальное оценивание состояний полусинхронного потока событий в условиях его частичной наблюдаемости // Вестник Томского государственного университета. 2000. № 269. С. 95–98.
34. Горцев А.М., Нежелская Л.А. О связи МС-потоков и MAP-потоков событий // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2011. № 1 (14). С. 13–21.
35. Gortsev A.M., Nezhel'skaya L.A., Solov'ev A.A. Optimal State Estimation in MAP Event Flows with Unextendable Died Time // Automation and Remote Control. 2012. V. 73, No. 8. P. 1316–1326.
36. Горцев А.М., Шмырин И.С. Оптимальная оценка состояний дважды стохастического потока событий при наличии ошибок в измерениях моментов времени // Автоматика и телемеханика. 1999. № 1. С. 52–66.
37. Горцев А.М., Леонова М.А. Оптимальная оценка состояний обобщенного асинхронного дважды стохастического потока // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2010. № 1 (10). С. 33–47.
38. Горцев А.М., Зуевич В.Л. Оптимальная оценка состояний асинхронного дважды стохастического потока событий с произвольным числом состояний // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2010. № 2 (11). С. 44–65.
39. Апанасович В.В., Коляда А.А., Чернявский А.Ф. Статистический анализ случайных потоков в физическом эксперименте. Минск : Университетское, 1988. 254 с.
40. Горцев А.М., Васильева Л.А. Оценивание параметров дважды стохастического потока событий в условиях его неполной наблюдаемости // Автоматика и телемеханика. 2002. № 3. С. 179–184.
41. Горцев А.М., Завгородняя М.Е. Оценка параметров альтернирующего потока событий при условии его частичной наблюдаемости // Оптика атмосферы и океана. 1997. Т. 10, № 3. С. 273–280.
42. Горцев А.М., Климов И.С. Оценка интенсивности пуассоновского потока событий в условиях частичной его ненаблюдаемости // Радиотехника. 1991. № 12. С. 3–7.
43. Горцев А.М., Паршина М.Е. Оценивание параметров альтернирующего потока событий в условиях «мертвого времени» // Изв. вузов. Физика. 1999. № 4. С. 8–13.
44. Горцев А.М., Ниссенбаум О.В. Оценивание длительности мертвого времени и параметров асинхронного альтернирующего потока событий при непродлеваемом мертвом времени // Изв. вузов. Физика. 2005. № 10. С. 35–40.
45. Горцев А.М., Ниссенбаум О.В. Оценивание длительности мертвого времени и параметров асинхронного альтернирующего потока событий с иницированием лишнего события // Вестник Томского государственного университета. 2004. № 284. С. 137–145.
46. Горцев А.М., Климов И.С. Оценивание периода ненаблюдаемости и интенсивности пуассоновского потока событий // Радиотехника. 1996. № 2. С. 8–11.
47. Gortsev A.M., Nezhel'skaya L.A. Estimate of parameters of synchronously alternating Poisson stream of events by the moment method // Telecommunications and Radio Engineering (English translation of Elektrosvyaz and Radiotekhnika). 1996. V. 50, No. 1. P. 56–63.
48. Нежелская Л.А. Оптимальная оценка состояний синхронного МС-потока событий // Избранные доклады международной конференции «Всесибирские чтения по математике и механике». Томск : Изд-во Том. гос. ун-та, 1997. Т. 1. С. 97–102.
49. Bushlanov I.V., Gortsev A.M. Optimal estimation of the states of a synchronous double stochastic flow of events // Automation and Remote Control. 2004. V. 65, No. 9. P. 1389–1399.
50. Gortsev A.M., Nezhel'skaya L.A. Estimation of the dead time period and intensities of the synchronous double stochastic event flow // Radiotekhnika. 2004. No. 10. P. 8–16.
51. Горцев А.М., Нежелская Л.А. Синхронный дважды стохастический поток событий при продлеваемом мертвом времени // Теория вероятностей, случайные процессы, математическая статистика и приложения : материалы международной конференции. Минск : Изд-во БГУ, 2005. С. 60–69.
52. Бушланов И.В., Горцев А.М. Оптимальная оценка состояний синхронного дважды стохастического потока событий // Автоматика и телемеханика. 2004. № 9. С. 40–51.
53. Горцев А.М., Калягин А.А., Нежелская Л.А. Оценка максимального правдоподобия длительности мертвого времени в обобщенном полусинхронном потоке // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2015. № 1 (30). С. 27–37.
54. Горцев А.М., Соловьев А.А. Совместная плотность вероятностей длительности интервалов потока физических событий при непродлеваемом мертвом времени // Известия высших учебных заведений. Физика. 2014. Т. 57, № 7. С. 103–111.
55. Горцев А.М., Леонова М.А., Нежелская Л.А. Сравнение МП- и ММ-оценок длительности мертвого времени в обобщенном асинхронном потоке событий // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2013. № 4 (25). С. 32–42.

56. Калягин А.А., Нежелская Л.А. Сравнение МП- и ММ-оценок длительности мертвого времени в обобщенном полусинхронном потоке событий // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2015. № 3 (32). С. 23–32.
57. Леонова М.А., Нежелская Л.А. Оценка максимального правдоподобия длительности мертвого времени в обобщенном асинхронном потоке событий // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2013. № 2 (23). С. 54–63.
58. Горцев А.М., Соловьев А.А. Оценка максимального правдоподобия длительности непродлевающегося мертвого времени в потоке физических событий // Известия высших учебных заведений. Физика. 2015. Т. 58, № 11. С. 141–149.
59. Горцев А.М., Соловьев А.А. Сравнение МП- и ММ-оценок длительности непродлевающегося мертвого времени в МАР-потоке событий // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2015. № 4 (33). С. 13–22.

Сиротина Мария Николаевна. E-mail: mashuliagol@mail.ru

Горцев Александр Михайлович, д-р техн. наук, профессор. E-mail: gam@fpmk.tsu.ru

Томский государственный университет

Поступила в редакцию 12 ноября 2015 г.

Sirotina Maria N., Gortsev Alexander M. (Tomsk State University, Russian Federation).

Maximum likelihood estimator of dead time duration in modulated synchronous twice stochastic flow of events.

Keywords: modulated synchronous flow; dead time; the likelihood function; the maximum likelihood function estimation; dead time duration.

DOI: 10.17223/19988605/34/6

Consider the modulated synchronous twice stochastic flow of events, which rate is piecewise constant random process $\lambda(t)$ with two states: λ_1, λ_2 ($\lambda_1 > \lambda_2$). The time when process $\lambda(t)$ is staying in the i -th state has exponential probability density function with a parameter $\alpha_i, i = 1, 2$. During the time random interval when $\lambda(t) = \lambda_i$ there is a Poisson flow with the rate $\lambda_i, i = 1, 2$. A state transition of process $\lambda(t)$ occurs at the moment of arrival of a Poisson flow event, moreover, the passing from the first to the second state is realized with probability p , the passing from the second to the first state is realized with probability q .

The flow functions in conditions of dead time, which is one of the falsifying factors of the state and the parameter estimation of the flow. After each registered event there is some time of fixed duration T (dead time), during which another flow events is inaccessible for observation. Consider non-extendable dead time. Such time is all events, which happen during the dead time interval without its prolongation. When the duration of the dead time period finishes, the first happened event creates the dead time period of duration T again, and etc. Note that for the flow, which functions in dead time conditions, events are observable if they did not get into the dead time interval. For the flow, which functions in dead time absence, all events are observable.

One of confounding factor by the flow state and parameter estimation is the dead time of recording device, which is generated by the observable flow of events. All other events occurred during the dead time interval are not accessible for observation.

The main purpose of this paper is to estimate the dead time T . On the base of the maximum likelihood function method, the solution of the problem is obtained. It is shown that the likelihood function $L(T | \tau^{(1)}, \dots, \tau^{(k)})$ reaches its maximum at the point $\hat{T} = \tau_m = \min \tau_k (k = \overline{1, n})$. So, the solution of optimization problem is the dead time estimator $\hat{T} = \tau_m$.

REFERENCES

1. Gortsev, A.M. & Golofastova, M.N. (2013) Optimal state estimation of modulated synchronous twice-stochastic flow of events. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2(23). pp. 42–53. (In Russian).
2. Sirotina, M.N. (2014) Optimal state estimation of modulated synchronous twice-stochastic flow of events in conditions of fixed dead time. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 1(26). pp. 63–74. (In Russian).
3. Gortsev, A. & Sirotina, M. (2014) Joint Probability Density Function of Modulated Synchronous Flow Interval Duration. *Communications in Computer and Information Science*. 487. pp. 145–152. (In Russian). DOI: 10.1007/978-3-319-13671-4_17
4. Sirotina, M.N. & Gortsev, A.M. (2015) [Probability density function of interval duration between two adjacent events of modulated synchronous flow in conditions of fixed dead time]. *Proceedings International Scientific Conference devoted to the 80th anniversary of Prof. G.A. Medvedev*. Belorussia, 23–26 February, 2015. Minsk: RIVSH. pp. 299–304.
5. Dudin, A.N. & Klimenok, V.N. (2000) *Sistemy massovogo obsluzhivaniya s korrelirovannymi potokami* [Queueing systems with correlated flows]. Minsk: BSU.
6. Basharin, G.P., Gajdamaka, U.V. & Samyjlov, K.E. (2013) Mathematical theory of teletraffic and its applications to analysis of multi-service network systems of the nextages. *Automation and computation*. 2. pp. 11–21.
7. Kingman, J.F.C. (1964) On doubly stochastic Poisson process. *Proceedings Cambridge Philosophical Society*. 60(4). pp. 923–930.
8. Basharin, G.P., Kokotushkin, V.A. & Naumov, V.A. (1979) O metode ekvivalentnykh zamen rascheta fragmentov setey svyazi [About the method of renewal of subnetwork computation]. *Izv. AN SSSR. Tekhn. kibernetika*. 6. pp. 92–99.
9. Neuts, M.F. (1979) A versatile Markov point process. *Journal of Applied Probability*. 16. pp. 764–779. DOI: 10.2307/3213143

10. Lucantoni, D.M. (1991) New results on the single server queue with a batch markovian arrival process. *Communication in Statistics Stochastic Models*. 7. pp. 1-46. DOI: 10.1080/15326349108807174
11. Lucantoni, D.M. & Neuts, M.F. (1994) Some steady-state distributions for the MAP/SM/1 queue. *Communication in Statistics Stochastic Models*. 10. pp. 575-598. DOI: 10.1080/15326349408807311
12. Livshits, K.N. & Bublik, Ya.S. (2010) The probability of insurance company collapse with twice stochastic flow of insurance payouts. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 1(10). pp. 66-77. (In Russian).
13. YangWoo Shin, Y.W. (2004). BMAP/G/1 queue with correlated arrivals of customers and disasters. *Operation Research Letters*. 32(4). pp. 364-373. DOI: 10.1016/j.orl.2003.09.005
14. Artalejo, J.R. & Chakravarthy, S.R. (2006) Computational analysis of the maximal queue length in the MAP/M/c retrival queue. *Applied Mathematics and Computation*. 183(2). pp. 1399-1409.
15. Best, J. (2009) Doubly Stochastic Processes: an Approach for Understanding Central Nervous System Activity. In: Pardalos, P. et al. (eds) *Selected Topics on Applied Mathematics, Circuits, Systems and Signals*. WSEAS Press. pp. 155-158.
16. Bushalnov, I.V., Gortsev, A.M. & Nezhelskaya, L.A. (2008) Estimating parameters of the synchronous twofold-stochastic flow of events. *Automation and Remote Control*. 69(9). pp. 1517-1533. DOI: 10.1134/S0005117908090075
17. Gortsev, A.M. & Nezhelskaya, L.A. Estimation of the dead time period and intensities of the synchronous double stochastic event flow. *Radioengineering*. 10. pp. 8-16.
18. Vasilevskaya, T.P., Gortsev, A.M. & Nezhelskaya, L.A. (2004) Otsenivanie dlitel'nosti mertvogo vremeni i parametrov sinkhronnogo al'terniruyushchego potoka s proyavleniem libo neproyavleniem sobyitiy [Dead time and parameters estimation of synchronous alternating flow with or without event manifestation]. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 9. pp. 129-138.
19. Gortsev, A.M. & Nezhel'skaya, L.A. (2003) Otsenivanie dlitel'nosti mertvogo vremeni i parametrov sinkhronnogo al'terniruyushchego potoka sobyitiy [Dead time period and parameter estimation of synchronous alternating flow of events]. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 6. pp. 232-239.
20. Gortsev, A.M. & Nezhelskaya, L.A. (2002) Otsenivanie parametrov sinkhronnogo dvazhdy stokhasticheskogo potoka sobyitiy metodom momentov [Parameter estimation of synchronous twice-stochastic flow of events using the method of moments]. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 1. pp. 24-29.
21. Gortsev, A.M. & Nezhelskaya, L.A. (1995) Estimation of the parameters of a synchro-alternating Poisson event flow by the method of moments. *Radiotechnics*. 7–8. pp. 6-10.
22. Leonova, M.A. & Nezhelskaya, L.A. (2013) Maximum likelihood estimation of dead time value at a generalized asynchronous flow of events. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2(23). pp. 54-63. (In Russian).
23. Gortsev, A.M., Leonova, M.A. & Nezhelskaya, L.A. (2012) Joint probability density function of interval duration of generic asynchronous event flow in conditions of fixed dead time. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 4(21). pp. 14-25. (In Russian).
24. Leonova, M.A. & Nezhelskaya, L.A. (2012) Error probability of state estimation of generic synchronous flow. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2(19). pp. 88-101. (In Russian).
25. Gortsev, A.M. & Nezhelskaya, L.A. (2011) An asynchronous double stochastic flow with initiation of superfluous events. *Discrete Mathematics and Applications*. 21(3). pp. 283-290. DOI: 10.4213/dm1141.
26. Gortsev, A.M. & Nezhelskaya, L.A. (2006) Otsenivanie parametrov asinkhronnogo potoka s initsirovaniem lishnikh sobyitiy metodom momentov [Parameter estimation of asynchronous flow with additional event initiation using the method of moments]. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 18. pp. 267-273.
27. Gortsev, A.M., Nezhelskaya, L.A. & Shevchenko, T.I. (1993) Estimation of the states of an MC-stream of events in the presence of measurement errors. *Russian Physics Journal*. 36(12). pp. 1153-1167.
28. Kalyagin, A.A. & Nezhelskaya, L.A. (2014) [Dead time period estimation of generic semi-synchronous flow of events]. *Novye informatsionnye tekhnologii v issledovanii slozhnykh struktur* [New information technologies in complex structure investigation]. *Proceedings of 10th Russian conference with international participation* (9-13 of June, 2014). Tomsk: Tomsk State University. pp. 96-97.
29. Gortsev, A.M., Kalyagin, A.A. & Nezhelskaya, L.A. (2010) Optimal state estimation of generic semi-synchronous flow of events. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2(11). pp. 66-81. (In Russian).
30. Gortsev, A.M. & Nezhelskaya, L.A. (2008) Semi-synchronous twice-stochastic event flow in conditions of prolonged dead time. *Vychislitel'nye tekhnologii – Computational Technologies*. 13(1). pp. 31-41. (In Russian).
31. Gortsev, A.M. & Nezhelskaya, L.A. (2003) Estimation of the dead-time period and parameters of a semi-synchronous double-stochastic stream of events. *Measurement Techniques*. 46(6). pp. 536-545. DOI: 10.1023/A:1025499509015
32. Gortsev, A.M. & Nezhelskaya, L.A. (2002) Otsenivanie parametrov polusinkhronnogo dvazhdy stokhasticheskogo potoka sobyitiy metodom momentov [Parameter estimation of semi-synchronous twice-stochastic event flow using the method of moments]. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 1. pp. 18-23.
33. Nezhelskaya, L.A. (2000) Optimal'noe otsenivanie sostoyaniy polusinkhronnogo potoka sobyitiy v usloviyakh ego chastichnoy nablyudaemosti [Optimal state estimation of semi-synchronous flow in conditions of its incomplete observability]. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 269. pp. 95-98.

34. Gortsev, A.M. & Nezhelskaya, L.A. (2011) About connectivity of MC-flows and MAP-flows of events. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 1(14). pp. 13-21. (In Russian).
35. Gortsev, A.M., Nezhelskaya, L.A. & Solovov, A.A. (2012) Optimal State Estimation in MAP Event Flows with Unextendable Dead Time. *Automation and Remote Control*. 73(8). pp. 1316-1326. DOI: 10.1134/S000511791208005X
36. Gortsev, A.M. & Shmyrin, I.S. (1999) Optimal estimation of states of a double stochastic flow of events under conditions of time moments measurement errors. *Automation and Remote Control*. 60(1). pp. 41-51.
37. Gortsev, A.M. & Leonova, M.A. (2010) Optimal state estimation of generic asynchronous doubly stochastic flow. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 1(10). pp. 33-47. (In Russian).
38. Gortsev, A.M. & Zuevich, V.L. (2010) Optimal state estimation of generic asynchronous doubly stochastic flow with random number of states. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2(11). pp. 44-65. (In Russian).
39. Apanasovich, V.V., Kolyada, A.A. & Chernyavskiy, A.F. (1988) *Statisticheskii analiz sluchaynykh potokov v fizicheskom eksperimente* [Statistical analysis of stochastic flows in physical experiment]. Minsk: Universitetskoe.
40. Gortsev, A.M. & Vasileva, L.A. (2002) Parameter estimation of a doubly stochastic flow of events under incomplete observability. *Avtomatika i Telemekhanika*. 3. pp. 179-184.
41. Gortsev, A.M. & Zavgorodnyaya, M.E. (1997) Otsenka parametrov al'terniruyushchego potoka sobyitiy pri uslovii ego chastichnoy nablyudaemosti [Parameter estimation of alternating flow of events under conditions of particulate observability]. *Optika atmosfery i okeana – Atmospheric and Oceanic Optics*. 10(3). pp. 273-280.
42. Gortsev, A.M. & Klimov, I.S. (1991) An estimate for intensity of Poisson flow of events under the condition of its partial missing. *Radiotekhnika*. 12. pp. 3-7.
43. Gortsev, A.M. & Parshina, M.E. (1999) Otsenivanie parametrov al'terniruyushchego potoka sobyitiy v usloviyakh "mertvogo vremeni" [Parameter estimation of alternating flow of events under conditions of dead time]. *Russian Physics Journal*. 4. pp. 8-13.
44. Gortsev, A.M. & Nissenbaum, O.V. (2005) Estimation of the dead time period and parameters of an asynchronous alternative flow of events with unextendable dead time period. *Russian Physics Journal*. 48(10). pp. 1039-1054. DOI: 10.1007/s11182-006-0023-y.
45. Gortsev, A.M. & Nissenbaum, O.V. (2004) Dead time and parameter estimation of asynchronous alternating flow with additional event initiation. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 284. pp. 137-145. (In Russian).
46. Gortsev, A.M. & Klimov, I.S. (1996) Estimation of the non-observability period and intensity of Poisson event flow. *Radiotekhnika*. 2. pp. 8-11.
47. Gortsev, A.M. & Nezhelskaya, L.A. (1996) Estimate of parameters of synchronously alternating Poisson stream of events by the moment method. *Telecommunications and Radio Engineering*. 50(1). pp. 56-63.
48. Nezhelskaya, L.A. (1997) [Optimal state estimation of synchronous MC-flow of events]. *Vsesibirskie chteniya po matematike i mekhanike* [Siberian readings on mathematics and mechanics]. Proc. of the International Conference. Vol. 1. Tomsk: Tomsk state University. pp. 97-102.
49. Bushlanov, I.V. & Gortsev, A.M. (2004) Optimal estimation of the states of a synchronous double stochastic flow of events. *Automation and Remote Control*. 65(9). pp. 1389-1399. DOI: 10.1023/B:AURC.0000041418.09187.63
50. Gortsev, A.M. & Nezhelskaya, L.A. (2004) Estimation of the dead time period and intensities of the synchronous double stochastic event flow. *Radiotekhnika*. 10. pp. 8-16.
51. Gortsev, A.M. & Nezhelskaya, L.A. (2005) [Synchronous twice-stochastic flow of events in conditions of fixed dead time]. *Teoriya veroyatnostey, sluchaynye protsessy, matematicheskaya statistika i prilozheniya* [Probability theory, stochastic processes, mathematic statistics and applications]. Proc. of the International Conference. Minsk: BSU. pp. 60-69. (In Russian).
52. Bushlanov, I.V. & Gortsev, A.M. (2004) Optimal estimation of the states of a synchronous double stochastic flow of events. *Automation and Remote Control*. 65(9). pp. 1389-1399. DOI: 10.1023/B:AURC.0000041418.09187.63
53. Gortsev, A.M., Kalyagin, A.A. & Nezhelskaya, L.A. (2015) Maximum likelihood estimation of dead time of generic semichronous flow. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 1(30). pp. 27-37. (In Russian).
54. Gortsev, A.M. & Solovyov, A.A. (2014) Joint Probability Density of Interarrival Interval of a Flow of Physical Events with Unextendable Dead Time Period. *Russian Physics Journal*. 57(7). pp. 973-983. DOI: 10.1007/s11182-014-0333-4
55. Gortsev, A.M., Leonova, M.A. & Nezhelskaya, L.A. (2013) Comparison of MP- and MM-estimations of dead time in generic asynchronous flow. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 4(25). pp. 32-42. (In Russian).
56. Kalyagin, A.A. & Nezhelskaya, L.A. (2015) Comparison of MP- and MM-estimations of dead time in generic semisynchronous flow of events. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 3(32). pp. 23-32. (In Russian).
57. Leonova, M.A. & Nezhelskaya, L.A. (2013) Maximum likelihood estimation of dead time in generic asynchronous flow of events. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2(23). pp. 54-63. (In Russian).
58. Gortsev, A.M. & Solovyov, A.A. (2015) Maximum likelihood estimation of fixed dead time in physical flow of events. *Russian Physics Journal*. 58(11). pp. 141-149.
59. Gortsev, A.M. & Solovyov, A.A. (2015) Comparison of MP- and MM-estimations of fixed dead time in MAP-flow of events. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 4(33). pp. 13-22. (In Russian).

Г.Ш. Цициашвили

СИНЕРГЕТИЧЕСКИЙ ЭФФЕКТ В СЕТИ С ГИПЕРЭКСПОНЕНЦИАЛЬНЫМИ РАСПРЕДЕЛЕНИЯМИ ВРЕМЕН ОБСЛУЖИВАНИЯ

В статье сеть массового обслуживания с многоканальными узлами и гиперэкспоненциальными распределениями времен обслуживания преобразована в сеть Джексоновского типа. С помощью теоремы мультипликативности установлен синергетический эффект, состоящий в исчезновении очередей при увеличении числа приборов в узлах сети.

Ключевые слова: гиперэкспоненциальное распределение; сеть массового обслуживания; теорема мультипликативности.

В настоящей работе решается задача конструирования сети массового обслуживания с многоканальными узлами и исчезающей при большом числе каналов очередью. В системах массового обслуживания такие синергетические эффекты достаточно детально изучены. Однако в сетях обслуживания подобное явление установить сложнее из-за перемешивания потоков заявок, выходящих из узлов. Подобная постановка вопроса связана с интенсивно развивающимися работами по исследованию бесконечнолинейных сетей массового обслуживания [1, 2], в которых отсутствуют очереди.

В статье делается предположение, что времена обслуживания в каналах отдельных узлов гиперэкспоненциальные, т.е. являются вероятностными смесями конечного числа экспоненциальных распределений. Каждый узел разбивается на определенное число многоканальных систем типа $M | M | n | \infty$ со специально подобранным числом каналов. В результате исходная сеть преобразуется в открытую сеть Джексона, для которой мультипликативная теорема позволяет редуцировать первоначальную задачу конструирования сети, обладающей синергетическим эффектом, в отдельные многоканальные системы обслуживания (синергетический эффект для них уже установлен). Таким образом, расчетные методы сочетаются с элементами конструирования сетей обслуживания, обладающими синергетическим эффектом.

Основной результат

Рассмотрим открытую сеть массового обслуживания G , состоящую из m узлов, между которыми циркулируют заявки и времена обслуживания в которых являются независимыми и одинаково распределенными случайными величинами с функциями распределения (ф.р.)

$$F_i(t) = \sum_{k=1}^{r_i} p_{ik} (1 - \exp(-\mu_{ik}t)), \quad a_i = \int_0^\infty \overline{F}_i(x) dx = \sum_{k=1}^{r_i} \frac{p_{ik}}{\mu_{ik}}, \quad i = 1, \dots, m,$$

являющимися смесями r_i показательных распределений: $p_{ik} > 0$, $\sum_{k=1}^{r_i} p_{ik} = 1$. Пусть все положительные

числа $R_{ik} = \frac{p_{ik}}{\mu_{ik} a_i}$, $k = 1, \dots, r_i$, $i = 1, \dots, m$, являются рациональными и представляются равенствами

$R_{ik} = \frac{A_{ik}}{B_{ik}}$, где числитель и знаменатель – взаимно простые числа. Обозначим N наименьшее общее

кратное чисел B_{ik} , $k = 1, \dots, r_i$, $i = 1, \dots, m$, и положим, что в каждом из узлов сети G содержится nN приборов.

Динамика перемещения заявки в сети G задается неразложимой маршрутной матрицей $\Theta = \|\theta_{ij}\|_{i,j=0}^m$, где θ_{ij} – вероятность перехода заявки после обслуживания из узла i в узел j , θ_{i0} – вероятность ухода заявки из сети, θ_{0i} – вероятность поступления заявки входного потока в узел i , $\theta_{00} = 0$. Обозначим $(\lambda_1, \dots, \lambda_m)$ единственное решение системы линейных алгебраических уравнений

$$(1, \lambda_1, \dots, \lambda_m) = (1, \lambda_1, \dots, \lambda_m) \Theta, \lambda_1, \dots, \lambda_m > 0. \quad (1)$$

Предположим, что входной поток в сети G является пуассоновским с интенсивностью nN и выполнены неравенства

$$\rho_i = \lambda_i a_i < 1, i = 1, \dots, m. \quad (2)$$

Для получения синергетического эффекта в явном виде преобразуем сеть G с m узлами в сеть R с $\sum_{i=1}^m r_i$ узлами. Обозначим

$$n_{ik} = nNR_{ik}, k = 1, \dots, r_i, \sum_{k=1}^{r_i} n_{ik} = Nn, i = 1, \dots, m, \quad (3)$$

и заменим узел i сети G блоком (i) , состоящим из r_i узлов (i, k) с n_{ik} приборами, имеющими показательные распределения времен обслуживания $1 - \exp(-\mu_{ik}t)$, $k = 1, \dots, r_i$. Заявка входного потока поступает в узел (i, k) с вероятностью $\theta_{0i} p_{ik}$, а после окончания обслуживания уходит в узел (j, l) с вероятностью $\theta_{ij} p_{jl}$ или уходит из сети с вероятностью θ_{i0} .

Очевидно, что ф.р. времени обслуживания заявки, поступающей в блок (i) , совпадает с $F_i(t)$, а число приборов в блоке (i) в силу (3) равно nN , как и в сети G . Маршрутная матрица сети R также является неразложимой. Таким образом, сеть R является сетью Джексона. Докажем, что вероятность $P_R(n)$ существования очереди хотя бы в одном из узлов (i, k) сети R удовлетворяет соотношению

$$P_R(n) \rightarrow 0, n \rightarrow \infty. \quad (4)$$

Во всех узлах сети R в силу условий (2) выполняются неравенства

$$\frac{nN\lambda_i p_{ik}}{n_{ik}\mu_{ik}} = \rho_i < 1, k = 1, \dots, r_i, i = 1, \dots, m. \quad (5)$$

Тогда из мультипликативной теоремы Джексона [3] следует, что предельное распределение числа заявок в узлах (i, k) , $k = 1, \dots, r_i$, $i = 1, \dots, m$ сети R удовлетворяет равенству

$$P(t_{ik}, k = 1, \dots, r_i, i = 1, \dots, m) = \prod_{i=1}^m \prod_{k=1}^{r_i} P_{ik}(t_{ik}). \quad (6)$$

Здесь $P_{ik}(t_{ik})$ предельное распределение числа заявок в n_{ik} – канальной системе массового обслуживания типа $M|M|n| \infty$ с интенсивностью входного потока $nN\lambda_i p_{ik}$ и интенсивностью обслуживания на отдельном приборе μ_{ik} .

Известно, что для системы массового обслуживания $M|M|n| \infty$ с интенсивностью входного потока $n\lambda$ и интенсивностью обслуживания μ на отдельном приборе стационарная вероятность $p_n(k)$ нахождения в системе k заявок вычисляется по формулам [4]:

$$p_n(k) = p_n(0) \frac{n^k \rho^k}{\prod_{j=1}^k \min(j, n)}, \rho = \frac{\lambda}{\mu}, \quad (7)$$

$$p_n(0) = \left[1 + \sum_{k=1}^{\infty} \frac{n^k \rho^k}{\prod_{j=1}^k \min(j, n)} \right]^{-1} \leq \left[1 + \sum_{k=1}^{\infty} \frac{n^k \rho^k}{k!} \right]^{-1} = e^{-n\rho}.$$

В силу формулы Стирлинга [5. Глава I, § 2]:

$$n! = n^n e^{-n} \sqrt{2\pi n} \exp\left(\frac{\psi(n)}{12n}\right), \quad 0 < \psi(n) < 1,$$

и неравенства $f(\rho) = 1 + \ln \rho - \rho < 0$, $0 < \rho < 1$, стационарная вероятность $P(n)$ существования очереди в такой системе удовлетворяет соотношению

$$P(n) = p_n(0) \sum_{k>n} \frac{n^k \rho^k}{n! n^{k-n}} \leq e^n \frac{p_n(0)}{\sqrt{2\pi n}} \sum_{k>n} \rho^k \leq \frac{\exp(nf(\rho))}{(1-\rho)\sqrt{2\pi n}} \rightarrow 0, \quad n \rightarrow \infty. \quad (8)$$

Обозначим $P_{ik}(n)$ стационарную вероятность существования очереди в узле (i, k) сети R . В силу условия (5) и соотношений (6), (8) получим

$$P_{ik}(n) \rightarrow 0, \quad n \rightarrow \infty, \quad k = 1, \dots, r_i, \quad i = 1, \dots, m.$$

Отсюда следует соотношение (4), характеризующее синергетический эффект в сети R .

Замечание. Нетрудно доказать, что в случае большой загрузки при $\rho = \rho(n) = 1 - a n^{-\gamma}$, $0 < a < 1$, если $0 < \gamma < 1/2$, то $P(n) \rightarrow 0$, $n \rightarrow \infty$, если $\gamma > 1/2$, то $P(n) \rightarrow 1$, $n \rightarrow \infty$.

Заключение

Рассмотренная в настоящей работе модель сети может быть распространена на случай вполне монотонных (complete monotone) [6] распределений времен обслуживания. Под вполне монотонными понимаются распределения с плотностью $f(t)$, удовлетворяющей неравенствам

$$(-1)^k \frac{d^k f(t)}{dt^k} \geq 0, \quad t > 0, \quad k \geq 1.$$

С помощью [7. Theorem 1, Lemma 2] и ее обобщений доказывается, что для любой вполне монотонной ф.р. $F(t)$ и для любого $\varepsilon > 0$ существует гиперэкспоненциальная ф.р. $G(t)$, удовлетворяющая неравенству $\sup_{t \geq 0} |F(t) - G(t)| \leq \varepsilon$. Важными для приложений в теории массового обслуживания вполне монотонными ф.р. являются распределения Вейбулла, Парето и другие распределения с тяжелыми хвостами [8, 9]. В свою очередь, аппроксимировать вероятность существования очереди в системе обслуживания с вполне монотонными ф.р. времен обслуживания можно с помощью теорем устойчивости [10] применительно к рассмотренной в работе сети G .

Пользуясь замечанием, нетрудно исследовать синергетические эффекты в режиме большой загрузки в отдельных узлах сети.

ЛИТЕРАТУРА

1. Назаров А.А., Моисеева С.П. Метод асимптотического анализа в теории массового обслуживания. Томск : Изд-во НТЛ, 2006.
2. Моисеев А.Н., Назаров А.А. Бесконечнолинейные системы и сети массового обслуживания. Томск : Изд-во НТЛ, 2015.
3. Jackson J.R. Networks of Waiting Lines // Oper. Res. 1957. V. 5, No. 4. P. 518–521.
4. Ивченко Г.И., Каштанов В.А., Коваленко И.Н. Теория массового обслуживания. М. : Высшая школа, 1982.
5. Ширяев А.Н. Вероятность. М. : Наука, 1989.
6. Feldmann A., Whitt W. Fitting mixtures of exponentials to long tailed distributions to analyze network perfomance models// Perfomance Evaluation. 1998. V. 31. P. 245–279.
7. Vatamidou E [et al.]. On the accuracy of phase-type approximations of heavy-tailed risk models // Scandinavian Actuarial Journal. 2014. V. 6. P. 510–534.
8. Embrechts P., Cluppelberg C., Mikosch T. Modelling Extremal Events: for Insurance and Finance. Berlin : Springer-Verlag 1997.
9. Asmussen S. Ruin Probabilities. Singapore: World Scientific Publishing Co Inc, 2000.
10. Боровков А.А. Предельные теоремы для сетей обслуживания // Теория вероятностей и ее применения. 1986. Т. XXXI, вып. 3. С. 474–490.

Цициашвили Гурами Шалвович, д-р физ.-мат. наук, профессор. E-mail: guram@iam.dvo.ru
Институт прикладной математики ДВО РАН,
Дальневосточный федеральный университет, г. Владивосток

Поступила в редакцию 23 декабря 2015 г.

Synergetic effect in network with hyperexponential distributions of service times.

Keywords: hyperexponential distribution; queuing network; product theorem.

DOI: 10.17223/19988605/34/7

In this paper, a problem of construction of queuing network with multiserver nodes and vanishing queues with large numbers of servers is solved. In queuing systems such synergetic effects are analyzed sufficiently detailed. But in queuing networks an analysis of these effects are more complicated because of mixing of outgoing flows. This problem is closely connected with intensively developing researches of networks with infinite numbers of servers in which there are no queues. An interest to these networks is arisen by a simplicity of probability distributions in the networks so by manifold applications like cloud computing and systems of public service.

The networks with service times which have hyperexponential distributions are analyzed. That is service times distributions are probability mixtures of finite numbers of exponential distributions. Any network node is divided into few $M|M|n|\infty$ systems with special numbers of servers. As a result initial queuing network is transformed into Jackson type network. So, to analyze synergetic effect it is enough to consider separate multiserver systems of $M|M|n|\infty$ type for which this effects may be analyzed sufficiently simple. In such a way calculational and constructive methods are combined.

Main element of this construction is an algorithm of a transformation of queuing network with hyperexponential distributions of service times into network with exponentially distributed service times. Main property of this transformation is a retention of servers numbers in different nodes. This algorithm is based on ergodicity condition essentially. Then product theorem for the Jackson type queuing network is used. This theorem allows to calculate limit distribution of customers numbers in the network by limit distribution of customers numbers in different nodes. So, to obtain synergetic effect in the network it is possible to use synergetic effect in the nodes.

Considered model of queuing network may be spread onto networks with completely monotone distributions of service times. Here complete monotone distribution is distribution with probability density $f(t)$, satisfying the inequalities $(-1)^k \frac{d^k f(t)}{dt^k} \geq 0, t > 0, k \geq 1$.

Well known results allow to approximate any completely monotone distribution $F(t)$ by hyperexponential distribution $G(t)$, satisfying for any $\varepsilon > 0$ the inequality $\sup_{t \geq 0} |F(t) - G(t)| \leq \varepsilon$. In queuing theory completely monotone distributions are the Weibull, Pareto, and other subexponential distributions. To approximate probability of queue existence in networks with completely monotone distribution it is possible using stability theorems. A possibility to consider synergetic effects in separate nodes of the network in the regime of heavy traffic is discussed also.

REFERENCES

1. Nazarov, A.A. & Moiseeva, S.P. (2006) *Metod asimptoticheskogo analiza v teorii massovogo obsluzhivaniya* [Method of asymptotic analysis in queuing theory]. Tomsk: NTL.
2. Moiseev, A.N. & Nazarov, A.A. (2015) *Beskonechnolinyne sistemy i seti massovogo obsluzhivaniya* [Queuing systems and networks with infinite number of servers]. Tomsk: NTL.
3. Jackson, J.R. (1957) Networks of Waiting Lines. *Oper. Res.* 5(4). pp. 518-521. DOI: 10.1287/opre.5.4.518
4. Ivchenko, G.I., Kashtanov, V.A. & Kovalenko, I.N. (1982) *Teoriya massovogo obsluzhivaniya* [Queuing theory]. Moscow: Vysshaya shkola.
5. Shiriaev, A.N. (1989) *Veroyatnost'* [Probability]. Moscow: Nauka.
6. Feldmann, A. & Whitt, W. (1998) Fitting mixtures of exponentials to long tailed distributions to analyze network performance models. *Performance Evaluation*. 31. pp. 245-279. DOI: 10.1109/INFCOM.1997.631130
7. Vatamidou, E et al. (2014) On the accuracy of phase-type approximations of heavy-tailed risk models. *Scandinavian Actuarial Journal*. 6. pp. 510-534. DOI: 10.1080/03461238.2012.729154
8. Embrechts, P., Kluppelberg, C. & Mikosch, T. (1997) *Modelling Extremal Events: for Insurance and Finance*. Springer-Verlag: Berlin.
9. Asmussen, S. (2000) *Ruin Probabilities*. Singapore: World Scientific Publishing Co Inc.
10. Borovkov, A.A. (1986) Predel'nye teoremy dlya setey obsluzhivaniya [Limit theorems for queuing networks]. *Teoriya veroyatnostey i ee primeneniya – Probability theory and its applications*. 31(3). pp. 474-490.

ИНФОРМАТИКА И ПРОГРАММИРОВАНИЕ

УДК 004.652.8

DOI: 10.17223/19988605/34/8

А.М. Бабанов, Е.С. Квач

**«IS-THE»-ОТНОШЕНИЯ В СЕМАНТИЧЕСКИХ МОДЕЛЯХ ДАННЫХ:
ОСНОВНЫЕ ПОНЯТИЯ И РАЗНОВИДНОСТИ**

Статья открывает серию публикаций, посвященных исследованиям и результатам, которые получены в такой редко упоминаемой области моделирования данных, как «IS-A»-отношения. В первой статье вводится понятие «IS-THE»-отношения, более подходящее для работы с данными, даются основополагающие определения и рассматриваются различные виды «IS-THE»-отношений. Предлагаемый подход отличается следованием логическим канонам, но без ущерба функциональности, исторически сложившейся в этом разделе моделирования данных.

Ключевые слова: «IS-THE»-отношение; «IS-THE»-кластер; «IS-THE»-категоризация; семантическая модель данных; ERM-модель.

«IS-A»-отношения (отношения «ЕСТЬ-НЕКОТОРЫЙ») являются важным элементом человеческой умственной деятельности (не только научной). И на бытовом уровне мы часто используем обобщенные и специализированные понятия: «Пушкин А.С. есть некоторый поэт», «Поэт есть некоторый человек».

Что же касается науки, то в ней операции обобщения и специализации являются важными мыслительными инструментами, позволяющими переходить от частных явлений к их абстракциям различных уровней и наоборот. Многочисленные классификации в различных отраслях науки тому подтверждение. Это важное для человека явление давно изучается исследователями, правда, задачи конкретной науки накладывают свой отпечаток на представление об этом предмете [1, 2]. Соответственно, используются и различные термины для обозначения «IS-A»-отношений: собственно «IS-A»-отношение, «IS-IN»-отношение (семантические сети), отношение наследования (системы представления знаний и системы объектно-ориентированного программирования), отношения принадлежности и включения (теория множеств), операции обобщения и ограничения понятий (логика), отношения «суперкласс-подкласс», «супертип-подтип» и т.п. Интуитивно все эти термины имеют близкие смыслы, но есть и существенные отличия.

В [2, 3] выделяют две группы смысловых значений «IS-A»-отношений – класс / объект (англ. generic / individual) и класс / класс (англ. generic / generic). Первую группу составляют трактовки этого отношения, близкие теоретико-множественному отношению принадлежности элемента множеству. Вторую группу составляют трактовки «IS-A»-отношения, близкие теоретико-множественному отношению включения одного множества в другое.

В моделировании данных эта естественная для людей форма представления информации была предложена не сразу. В семантических моделях она появилась вместе с расширенной моделью «Сущность–Связь» (Enhanced / Extended Entity-Relationship Model – EERM) в 80-х гг. прошлого века [4, 5]. В этой модели для представления «IS-A»-отношений введены такие структурные понятия, как специализация и категоризация.

Однако даже в моделировании данных нет единого мнения во взглядах на «IS-A»-структуры, что уж говорить об отличии этих представлений от логических основ. В статье будет предложено новое по-

нятие «IS-THE»-отношения (отношения «ЕСТЬ-ЭТОТ»), ориентированного на моделирование данных, и предпринята попытка синтеза логически непротиворечивого понятийного базиса рассматриваемого явления. Далее в работе будут определены разновидности структур данных для представления «IS-THE»-отношений при моделировании данных. Среди них будут и уже традиционные специализация и категоризация.

Приведенные определения положены в основу правил структуризации данных и определения ограничений целостности семантической модели «Сущность–Связь–Отображение» (Entity–Relationship–Mapping Model – ERMM) [6], касающихся представления в ней «IS-THE»-отношений.

1. Базовые понятия «IS-A»-отношений

Чтобы получить полную и непротиворечивую картину представлений об «IS-A»-отношении и окружающих его понятиях, начнем рассмотрение с позиции современной логики по этому вопросу.

«Обобщение некоторого понятия есть операция образования из этого понятия некоторого нового с более широким объемом. Обратная операция перехода от некоторого понятия к понятию с меньшим объемом называется ограничением понятия. Обобщение понятия осуществляется за счет определенного ослабления содержания первого. Ограничение понятия осуществляется за счет расширения содержания исходного. Каждый акт обобщения должен быть переходом от вида к некоторому ближайшему роду, а ограничения – от рода к некоторому ближайшему виду.

Деление понятий – это операция разбиения объема понятия на подклассы, представляющие собой виды предметов, мыслимых в этом понятии. То, что мы называем точкой зрения или аспектом различения предметов, называют основанием деления понятий. Цель деления понятия состоит в том, чтобы выделить все возможные виды предметов каждый раз по некоторому определенному основанию. В составе каждого деления можно выделить делимое понятие, основание деления и члены деления.

Деление является правильным, если оно удовлетворяет следующим условиям:

1. Деление должно происходить по одному определенному основанию.
2. Полученные при делении понятия должны быть попарно несовместимы.
3. Члены деления как классы должны исчерпывать объем исходного понятия.
4. Никакой из членов деления не должен быть пустым классом.
5. Деление должно быть непрерывным, т.е. все его члены являются ближайшими видами исходного понятия, выделяемыми по выбранному основанию» [7].

В.А. Бочаров и В.И. Маркин дополняют эту картину рассмотрением возможных вариантов деления понятий.

«С каждым делением связывается основание деления – такая характеристика предметов, входящих в объем делимого понятия, модификация которой и порождает систему членов деления. В зависимости от того, что берется в качестве основания деления, различают два их вида – дихотомическое деление и деление по видоизменению основания. В случае дихотомического деления родового понятия основанием деления является некоторый признак, присущий лишь части предметов, входящих в объем исходного понятия. Деление осуществляется по наличию или отсутствию этого признака у предметов делимого понятия. В случае деления по видоизменению основания в качестве основания деления используются предметно-функциональные характеристики элементов объема делимого понятия. Необходимо только, чтобы предметная функция, берущаяся в качестве основания, была всюду определенной на этом объеме. При таком делении объемы видовых понятий представляют собой собрания тех и только тех предметов a и b , для которых выполняется условие $f(a) = f(b)$.

Всякое деление должно осуществляться по одному основанию» [8].

Насколько представления логиков, формировавшиеся веками, точны и понятны, настолько размыты и неглубоки определения, которыми пользуются в моделировании данных. Там принято то, что логики называют делением понятия, обозначать термином «специализация». Он же используется для именования нисходящего процесса образования «IS-A»-иерархий с полным наследованием признаков. Обратный процесс, при котором эти «IS-A»-иерархии образуются с помощью восходящего процесса, принято называть генерализацией.

По мнению специалистов по моделированию данных, специализация – это:

- «процесс выделения подклассов в суперклассе, который основан на группировке исключительных характеристик и связей подклассов» [9];
- «процесс выявления различий между экземплярами типа сущностей» [10];
- «процесс определения набора подклассов для типа сущностей на основе отличительных характеристик» [11];
- «процедура, в ходе которой вводятся подклассы более общих типов сущностей, чтобы объявить, что определенные роли зарегистрированы только для этих подклассов» [12];
- «процесс концептуального уточнения суперкласса путем выделения подклассов» [13].

Как видим, если логики обращают внимание, прежде всего, на содержание (признак) понятий деления, то специалисты по моделированию данных видят преимущественно структурные и ограничительные особенности классов данных. Главное назначение специализаций в моделировании данных – возможность группирования признаков (атрибутов и типов связей): обобщенные поднимаются на уровень суперкласса, специализированные опускаются до соответствующего подкласса. Т. Халпин, кроме этого, замечает, что специализации позволяют делать опциональные роли суперкласса обязательными для подкласса и за счет этого вводить дополнительные ограничения целостности [12. Р. 238].

И только Дж. Смит и Д. Смит подходят к концепции специализации аналогично логикам.

«Наш метод представления родовой иерархии требует, чтобы непосредственные потомки любого узла были разбиты на группы. Каждая группа должна содержать родовые объекты, классы которых являются взаимно исключающими. На практике такое группирование обычно может быть весьма легко сделано из семантических соображений. Например, потомки класса Транспорт могли бы быть сгруппированы следующим образом: {Транспортные средства, приводимые в движение ветром; Моторизованные транспортные средства; Транспортные средства, приводимые в движение человеком} и {Воздушный транспорт; Водный транспорт; Наземный транспорт}. Первая группа содержит взаимно-исключающие классы, которые соответствуют альтернативным типам “Системы приведения в движение”. Вторая группа, в свою очередь, содержит взаимно исключающие классы, которые соответствуют различным типам “Среды передвижения”».

Назовем взаимно исключающие группы родовых объектов, имеющие общие родительские объекты, кластерами. Будем говорить, что кластер относится к его родительскому родовому объекту. Например, мы можем говорить о двух кластерах, относящихся к типу объектов Транспорт. Кластер, относящийся к “Моторизованным транспортным средствам”, и кластер, относящийся к “Воздушному транспорту”, могут иметь общий элемент – Вертолет» [14].

Дж. Смит и Д. Смит предлагают, как и логики, сводить иерархии обобщения к отдельным правильным делениям родительских понятий, правда, они их называют кластерами. Но по-прежнему налицо одно основание деления и непересекаемость подклассов одного кластера.

2. Дополнительные понятия «IS-A»-отношений

Важными понятиями, сопутствующими рассмотрению «IS-A»-отношений, являются понятия, применяемые к элементам этих отношений.

«Если экземпляры типа объектов классифицируются в более специфичный тип, то этот специализированный тип называется подтипом» [12. Р. 260]. «Супертип – это обобщенная форма его подтипов, в таком случае подтип – специализированная форма супертипа» [Там же].

«Суперкласс – тип сущностей, включающий одну или несколько различных вспомогательных группировок его экземпляров, которые должны быть представлены в модели данных. Подкласс – различимая вспомогательная группировка экземпляров типа сущностей, которая должна быть представлена в модели данных» [4. С. 430]. «Связь между суперклассом и подклассом является связью “один-к-одному” (1 : 1) и называется связью суперкласс / подкласс» [Там же. С. 431].

Одни авторы (Халпин) предпочитают рассматривать иерархии обобщения типов данных, другие, такие как Т. Коннолли и К. Бэгг, строят «IS-A»-иерархии на классах данных. Логика же говорит об

иерархиях понятий. И эта позиция выглядит более универсальной по сравнению с узкими трактовками специалистов по моделированию данных.

Действительно, «понятие – это мысль, которая посредством указания на некоторый признак (содержание понятия) выделяет из универсума и собирает в класс (обобщает) все предметы, обладающие этим признаком (объем понятия) [8]». И далее: «Содержание представляет собой интенциональную характеристику понятия; экстенциональной характеристикой понятия является его объем» [8]. Таким образом, говоря об иерархии понятий, мы предполагаем и иерархию признаков, и иерархию объемов. Оба эти аспекта исчерпывают вариации других авторов.

В таком случае, рассматривая двухуровневую «IS-A»-иерархию, корректнее говорить о родовом (аналог супертипа или суперкласса) и видовых (аналоги подтипов или подклассов) понятиях, точнее, системе видовых понятий. Содержание видового понятия отличает от содержания родового понятия дополнительное условие (видовое отличие), которое истинно для объектов объема этого видового понятия. Таким образом, спуск по «IS-A»-иерархии сопровождается усложнением содержания понятий. Что касается объемов понятий, то чем выше уровень понятия в «IS-A»-иерархии, тем его объем шире, и, наоборот, опускаясь вниз, мы переходим от объемов родовых понятий к их подмножествам – объемам видовых понятий.

В моделировании данных, исходя из практических потребностей, в одну специализацию суперкласса могут собирать подклассы, образованные по разным основаниям. Это нарушает логическую стройность схемы данных и часто приводит к проблемам при реализации этой схемы. Предпочтительнее вслед за логиками и Смитам рассматривать только специализации с одним основанием деления суперкласса на подклассы.

А вот в чем следует отступить от канонов логики, так это второе и третье условия правильного деления – требования попарной несовместимости видовых понятий и покрытия их объемами объема родительского понятия. Необходимость этого в моделировании данных доказала практика использования специализаций. В некоторых моделях по этому поводу введены даже специальные типы специализаций: полные непересекающиеся (аналог правильного деления в логике), полные пересекающиеся, частичные непересекающиеся, частичные пересекающиеся.

Таким образом, приходим к необходимости расширенной, по сравнению с логиками, трактовки операции деления понятий, учитывающей все особенности специализаций в моделировании данных.

3. Понятие «IS-THE»-отношения

Подход логиков в вопросе «IS-A»-отношений кажется более основательным, и его нельзя обойти стороной при синтезе понятий, описывающих эти отношения, но ориентированных на моделирование данных. Вторым краеугольным камнем этой системы понятий являются отображения и теория семантически значимых отображений [15]. Анализ отображений, сопутствующих «IS-A»-отношениям, позволяет глубже рассмотреть этот объект исследований.

Прежде чем говорить о представлении «IS-A»-отношений в моделях данных, введем понятие «IS-THE»-отношения.

Два элемента объемов разных понятий связаны «IS-THE»-отношением (отношением «ЕСТЬ-ЭТОТ») тогда и только тогда, когда они представляют один и тот же объект предметной области (ПрО) (отсюда использование определенного артикля «THE»). Например, фраза «Пациент Иванов есть этот врач Иванов» описывает ситуацию с заболевшим врачом. «IS-THE»-отношение – это бинарное 1 : 1-отношение, определяющее пару взаимобратных функциональных «IS-THE»-отображений. В этом отношении и отображениях участвуют непосредственно объекты ПрО (когда речь идет о законе соответствия объектов) или их абстракции – данные (когда речь идет об аналогичном законе соответствия данных, представляющих объекты).

Структуры данных (множества сущностей, отношения) предназначены для представления данных об объектах ПрО, и каждая из них соответствует определенному понятию (о предметах, n -ках предметов), экстенционал которого и представлен данными, принадлежащими этой структуре. Каждый объект ПрО определен в ней своей абстракцией (сущностью, кортежем) в контексте этого понятия. Таким об-

разом, один и тот же объект может быть представлен в нескольких различных структурах данных. Именно между этими элементами-абстракциями одного и того же объекта и устанавливается «IS-THE»-отношение в данных. Если известные трактовки «IS-A»-отношения связывают класс с объектом или класс с классом, то в «IS-THE»-отношения вступают объекты.

В максимально обобщенном понимании «IS-THE»-отношение позволяет указывать в каком-то смысле «близость» понятий: понятия близки, если найдется хотя бы один объект, подпадающий под оба эти понятия. «IS-THE»-отношение в таком варианте определяет два частичных функциональных отображения (действительно, соответствия между элементами объемов произвольных понятий может и не быть).

«IS-THE»-отношения, определенные на объемах понятий, представляются в схеме данных специальными универсальными взаимосвязями между соответствующими структурными компонентами – классами. Если представить классы в виде вершин, а взаимосвязи – в виде ребер, то получится в общем случае граф. Назовем его «IS-THE»-графом.

4. Разновидности «IS-THE»-отношений и дополнительные понятия

Если рассматривать два понятия, одно из которых является обобщением другого (а то, в свою очередь, является ограничением первого), то в этом случае «IS-THE»-отношение становится родовидовым отношением. «IS-THE»-отображение из объема родового понятия (суперкласса) в объем видового (подкласса) по-прежнему частичное функциональное (объект подпадает под видовое понятие только в случае удовлетворения содержания этого понятия). А вот «IS-THE»-отображение из подкласса в суперкласс стало полным функциональным (объект видового понятия всегда подпадает под родовое понятие). Это становится очевидным, если представить, как выглядит содержание любого видового понятия. В нем к родовому признаку конъюнктивно добавлено условие видового отличия. Понятно, что каждый элемент каждого подкласса удовлетворяет обоим признакам – и родовому и видовому.

«IS-THE»-кластер (или, по аналогии со Смитами, просто кластер) – это система классов, построенная для ограничений (возможно, единственного ограничения) понятия одного и того же суперкласса, такая, что все понятия подклассов используют в своих основаниях выделения одно и то же отображение (предметную функцию). Основание выделения подкласса – это видовое отличие объектов этого подкласса, которое представляет собой дополнительное условие содержания понятия подкласса. Единство отображения в основаниях выделения подклассов означает для дихотомического деления разбиение на подклассы в зависимости от наличия и отсутствия образов у элементов суперкласса при одном и том же отображении, для деления по видоизменению основания – объединение в подклассы элементов суперкласса, имеющих совпадающие образы при одном и том же отображении.

Такие отображения можно назвать дискриминирующими, ведь именно эти признаки объектов суперкласса определяют их попадание в подклассы. В этом качестве могут использоваться любые признаки – свойства, характеристики объектов, отношения с другими объектами. Роль прообразов таких отображений играют объекты суперкласса или их абстракции, а роль образов – истинностные значения, произвольные значения и другие объекты (абстракции) соответственно.

Дуальными к экземплярам этих отображений выступают экземпляры отображений между суперклассом и объединением подклассов кластера. С понятийной точки зрения это объединение представляет объем понятия, содержанием которого является дизъюнкция условий всех видовых понятий. По сути эти и упомянутые ранее отображения являются двумя альтернативными формами представления одной и той же информации о ПрО.

Для определения кластеров в схеме данных могут использоваться ориентированные гиперребра – ребра, соединяющие в общем случае несколько вершин: начальные вершины, представляющие подклассы, и конечную вершину, представляющую суперкласс.

Понятие «IS-THE»-кластера совпадает по смыслу с понятием двухуровневой специализации, подклассы которой образованы в соответствии с одним и тем же основанием. Например, рассмотрим специализацию людей на мужчин и женщин (рис. 1). Видовое отличие мужчин – «Пол = 'мужской'», женщин – «Пол = 'женский'». И в том и в другом случае используется одно и то же отображение (Пол). От-

личаются объекты подклассов только своими образами при этом отображении. С точки зрения логики имеем деление по видоизменению основания.

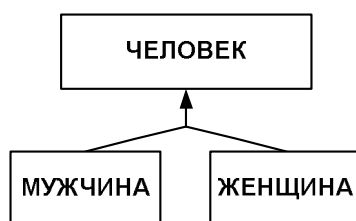


Рис. 1. «IS-THE»-кластер

Последовательно применяя операцию деления к полученным на предыдущем шаге видовым понятиям, можно построить иерархию кластеров (когда нет классов, входящих в качестве подклассов в разные кластеры) или граф кластеров (в противном случае).

Многие исследователи семантических моделей и авторы, повествующие о них, используют термин «специализация» и для обозначения процесса порождения подклассов, и для обозначения его результата – двухуровневой иерархии. Так, Т. Халпин определяет специализацию и как «процедуру, в ходе которой вводятся подклассы более общих типов сущностей» [12. Р. 260], и как «граф подтипов, который может возникнуть нисходящим способом, специализируя тип объекта в подтипы» [Там же. Р. 244]. На наш взгляд, корректнее использовать этот термин только для процесса (и тут все авторы солидарны), а результат именовать «IS-THE»-кластером (в случае двух уровней) или графом кластеров (в случае нескольких уровней). Такой подход позволит к тому же непротиворечиво ввести типизацию структур, некоторые элементы которой не имеют ничего общего со специализациями.

Подавляющее большинство «IS-THE»-отношений, представляющих интерес при построении БД, определяется «IS-THE»-кластерами и подразумевает полное наследование родительских признаков (атрибутов и связей) объектами дочерних понятий. То есть эти признаки на видовом уровне не декларируются, а предполагаются.

В ERM-модели значимые объекты ПрО представляются, так же как и в ER-модели, сущностями. Каждая сущность является абстракцией объекта, построенной в соответствии с некоторым понятием о предметах. Один и тот же объект ПрО, таким образом, может быть представлен несколькими сущностями разных типов. После введения таких представлений можно утверждать, что каждый «IS-THE»-кластер определяет полное функциональное «IS-THE»-отображение между объединением множеств сущностей видовых понятий и множеством сущностей родового понятия. Это является отличительной особенностью «IS-THE»-отношений с полным наследованием. Также действует полное функциональное «IS-THE»-отображение между каждым из множеств сущностей видовых понятий и множеством сущностей родового понятия.

В том случае, когда множество сущностей является подклассом только в одном «IS-THE»-кластере (соответствующее видовое понятие связано только с одним родовым понятием), говорят, что осуществляется полное единичное наследование. Когда множество сущностей входит подклассом в несколько «IS-THE»-кластеров (соответствующее видовое понятие связано с несколькими родовыми понятиями), говорят, что осуществляется полное множественное наследование. Но и в этом случае сущности подкласса наследуют признаки всех своих родителей. В обоих случаях полного наследования «IS-THE»-отношение представляет собой широко известное отношение наследования. Направление гиперссылки «IS-THE»-кластера указывает на родовое понятие, определяющее наследуемые признаки.

В моделировании данных введен особый подвид двухуровневых «IS-A»-иерархий – категоризации, определяющие особые «IS-THE»-отношения между объединением множеств сущностей-суперклассов и множеством сущностей-подклассов, в этом случае называемым категорией [4. С. 209]. Чтобы лучше всего понять суть категоризации, надо представить содержание понятия, объем которого представлен множеством сущностей-подклассов. Первая часть его логического выражения – дизъюнкция условий содержаний множеств сущностей-суперклассов, а вторая часть – условие, которому долж-

ны удовлетворять сущности этих множеств, чтобы быть представленными в категории. Такие двух-уровневые «IS-A»-иерархии будем называть «IS-THE»-категоризациями (или просто категоризациями).

В определении категоризации подчеркивается, что суперклассы не содержат абстракции-сущности одних и тех же объектов и настолько различны, что их не связывают общие классы-предки. Другими словами, понятия суперклассов попарно несовместимы. В таком случае элемент категории всегда наследует признаки только одного суперкласса, того, где содержится связанный с ним родительский элемент. Поэтому говорят, что такие двухуровневые «IS-A»-иерархии предполагают выборочное наследование.

По аналогии с «IS-THE»-кластерами можно утверждать, что с каждой «IS-THE»-категоризацией связано полное функциональное отображение между множеством сущностей-категорий и объединением множеств сущностей-суперклассов. А вот между множеством сущностей-категорий и каждым из множеств сущностей-суперклассов имеется частичное функциональное отображение. Что касается обратных отображений, то в зависимости от семантики ПрО они могут быть как полностью определенными, так и нет.

Для представления «IS-THE»-категоризаций также можно использовать ориентированные гиперребра, только в отличие от «IS-THE»-кластеров они соединяют начальную вершину, представляющую подкласс (категорию), и в общем случае несколько конечных вершин, представляющих суперклассы.

Типичным примером «IS-THE»-категоризации являются «IS-THE»-отношения между понятием «субъект права» и понятиями, объекты которых могут выступать в роли владельцев прав, – физическое лицо, юридическое лицо и административно-территориальное образование (рис. 2). Действительно, каждый субъект права – это или гражданин, или организация, или территория, и в зависимости от этого наследуются признаки соответствующего суперкласса.

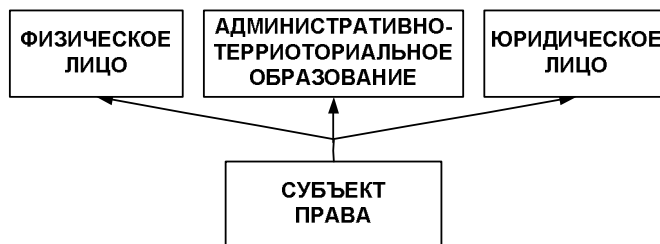


Рис. 2. «IS-THE»-категоризация

Теперь для «IS-THE»-кластеров и «IS-THE»-категоризаций (и только для них) можно ввести обобщенные понятия «суперкласс» и «подкласс». Суперкласс определяет наследуемые признаки, а элементы подклассов их наследуют. В сторону суперклассов ориентированы гиперребра, представляющие эти структуры.

В зависимости от применяемого при образовании «IS-THE»-кластера принципа и основания деления суперкласса на подклассы выделяют кластеры с динамичным и статичным набором подклассов. Если в ходе эволюции данных возможно появление нового подкласса – кластер динамичен, если нет – статичен. При использовании метода дихотомии всегда образуются два статичных подкласса: для объектов одного признак деления истинен, для объектов другого – ложен. Чаще всего второй подкласс не создается, а его экстенционал определяется как разность экстенционала суперкласса и экстенционала другого подкласса. В случае деления по видоизменению основания статичность и динамичность кластера определяет то, под все возможные образы дискриминирующего отображения созданы подклассы или нет.

Предыдущие типы «IS-THE»-кластеров характеризовали интенционал данных, т.е. определялись на основании статичности или динамичности схемы данных. Статичность или динамичность самих данных (экстенционала) приводит к кластерам со стабильными или мигрирующими экземплярами подклассов. Т. Халпин называет эти разновидности кластеров строгими (rigid) и ролевыми (role) соответственно [12. Р. 251].

Во многих моделях данных (в частности, в реляционной модели, модели «Объект–Роль» Халпина, ERMM) в схеме данных выделяются так называемые хранимые и получаемые структуры. Т. Халпин говорит даже о трех состояниях данных – хранимые, получаемые и частично получаемые.

«Получаемый (англ. derived) факт – это факт, который выводится из других фактов математическими вычислениями или логическим выводом. Факт, который нельзя вывести из других фактов, называется хранимым или утверждаемым пользователем (англ. asserted) фактом. Для каждого получаемого факта в схеме данных задается правило его получения» [12. Р. 33]. «Частично получаемый (англ. semi-derived) тип фактов определяется в том случае, когда ряд фактов этого типа можно вывести, а другие факты будут заданы пользователем» [12. Р. 99].

Что касается «IS-THE»-кластеров, то, как уже отмечалось, альтернативной формой представления информации о принадлежности объекта подклассу является наличие соответствующего экземпляра дискриминирующего отображения. Именно между этими формами данных и осуществляет выбор проектировщик схемы, в каждом конкретном случае предпочитая либо хранение экстенсиналов подклассов (хранимый кластер), либо хранение экземпляров отображения (получаемый кластер), либо их смесь (частично получаемый кластер).

Кстати, именно характер дискриминирующего отображения определяет еще одну типизацию «IS-THE»-кластеров, которую упоминают многие авторы [4, 12]. Это деление специализаций на полные и частичные, пересекающиеся и непересекающиеся. Некоторые говорят об этих видах специализаций как об особом типе ограничений целостности на специализации [4. С. 208].

Вот как определяются эти типы «IS-THE»-кластеров в ERM-модели в том случае, когда при делении по видоизменению основания создаются подклассы для всех образов. Если дискриминирующее отображение – полное функциональное, кластер является полным непересекающимся. Если дискриминирующее отображение – частичное функциональное, кластер является частичным непересекающимся. Если дискриминирующее отображение – полное нефункциональное, кластер является полным пересекающимся. Если дискриминирующее отображение – неограниченное, кластер является частичным пересекающимся.

Аналогичные высказывания можно сделать и по поводу отображений между суперклассом и объединением подклассов, дуальных дискриминирующим отображениям. Причем указанное отображение в любом случае определяет характер кластера, чего не скажешь о дискриминирующем отображении. Так, если для некоторых образов дискриминирующего отображения подклассы не определяются, кластер будет частичным независимо от свойств отображения. Функциональность этого отображения гарантирует непересечение подклассов. Но даже нефункциональное дискриминирующее отображение может в этом случае привести к непересекающемуся кластеру.

Если кластер образуется методом дихотомии и включает два подкласса, он всегда будет полным непересекающимся. Если же в таком кластере создается только один подкласс, кластер будет частичным непересекающимся.

Для «IS-THE»-категоризаций существует похожее деление на полные и частичные [4. С. 210]. Если каждый объект всех суперклассов присутствует в подклассе-категории, категоризация считается полной, если есть объекты суперклассов, не отображенные в категории, категоризация частичная. Или на языке отображений: если отображение из объединения суперклассов в категорию – полное функциональное, категоризация полная, если оно – частичное функциональное, категоризация частичная.

И, наконец, в ERM-модели можно строить «IS-THE»-графы на классах одного, но любого типа – множествах сущностей, множествах связей или множествах значений. Это имеет практическое значение и хорошо соотносится с логикой: там наряду с понятиями о предметах (сущностях и значениях) рассматривают и понятия об *n*-ках предметов (связях). Обобщение множеств сущностей и множеств значений (типов данных) стало уже стандартом во многих областях информатики. Потребность в построении «IS-A»-иерархий множеств связей встречалась до сих пор только в системах представления знаний. Теперь такое возможно и в системах баз данных.

Заключение

Значение «IS-A»-иерархий в моделировании данных многими недооценивается. А ведь с их помощью, помимо прочего, еще и восстанавливается единство объекта, рассыпанного в базе данных в виде своих абстракций-сущностей по несвязанным между собой множествам. В семантической модели

ERM предпринята попытка непротиворечивого включения «IS-A»-иерархий в правила построения структур и ограничений целостности. Основным принципом является следование безупречным логическим канонам при сохранении той функциональности этих структур, которая сложилась в моделировании данных. Настоящая статья закладывает основы правил использования «IS-THE»-отношений в виде определений базовых понятий и рассмотрения системы типизации структур, предназначенных для представления этих отношений в данных. В последующих статьях авторы постараются полнее раскрыть феномен «IS-THE»-отношений и их использование в моделировании данных.

ЛИТЕРАТУРА

1. Wang P. From Inheritance Relation to Nonaxiomatic Logic // International Journal of Approximate Reasoning. 1994. No. 11. P. 281–319.
2. Palomäki J., Kangassalo H. That IS-IN Isn't IS-A: A Further Analysis of Taxonomic Links in Conceptual Modelling // Advances in Knowledge Representation. 2012. P. 3–18.
3. Brachman R.J. What IS-A Is and Isn't: An Analysis of Taxonomic Links in Semantic Networks // IEEE Computer. 1983. V. 16, No. 10. P. 30–36.
4. Коннолли Т., Бегг К. Базы данных. Проектирование, реализация и сопровождение. Теория и практика. 2-е изд. М. : Вильямс, 2000. 1120 с.
5. Teorey T., Yang D., Fry J. A Logical Design Methodology for Relational Databases Using the Extended Entity-Relationship Model // Computing Surveys. 1986. V. 18, No. 2. P. 197–222.
6. Бабанов А.М. Семантическая модель «Сущность – Связь – Отображение» // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2007. № 1. С. 77–91.
7. Войшвилло Е.К., Дегтярев М.Г. Логика как часть теории познания и научной методологии (фундаментальный курс) : в 2 кн. М. : Наука, 1994.
8. Бачаров В.А., Маркин В.И. Основы логики. М. : Инфра-М, 1998. 297 с.
9. Coronel C., Morris S., Rob P. Database Systems: Design, Implementation, and Management. 10th ed. Course Technology, 2013. 1054 p.
10. Connolly T., Begg C. Database Systems: A Practical Approach to Design, Implementation, and Management. 4th ed. Addison-Wesley, 2010. 1400 p.
11. Elmasri R., Navathe S. Fundamentals of Database Systems. 6th ed. Addison-Wesley, 2010. 1200 p.
12. Halpin T., Morgan T. Information Modeling and Relational Databases. 2nd ed. Morgan Kaufman, 2008. 943 p.
13. Ferragine V., Doorn J., Rivero L. Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends. Information Science Reference, 2009. 1124 p.
14. Smith J., Smith D. Database Abstractions: Aggregation and Generalization // ACM Transactions on Database Systems. 1977. V. 2, No. 2. P. 105–133.
15. Бабанов А.М. Теория семантически значимых отображений // Вестник Томского государственного университета. Сер. Математика. Кибернетика. Информатика. 2003. № 280. С. 239–248.

Бабанов Алексей Михайлович, канд. техн. наук, доцент. E-mail: babanov2000@mail.ru
Томский государственный университет

Квач Елена Сергеевна. E-mail: kvachelena93@gmail.com
Томский государственный университет

Поступила в редакцию 1 декабря 2015 г.

Babanov Alexey M., Kvach Elena S. (Tomsk State University, Russian Federation).

«IS-THE»-relations in semantic data models: basic notions and variants.

Keywords: «IS-THE»-relation; «IS-THE»-cluster; «IS-THE»-categorization; semantic data model; ERM-model.

DOI: 10.17223/19988605/34/8

«IS-A»-relations are important intellectual tools allowing to pass from the individual phenomena to their abstractions of the different levels and on the contrary. Importance of these relations in data modeling is underestimated. In fact, these relations in addition recover the unity of an object scattered in a database as unconnected abstract entities. In semantic model ERM (Entity-Relationship-Mapping) attempt of consistent «IS-A»-relation inclusion to rules of construction of structures and integrity constraints is undertaken. Main principle is maintenance of faultless logic canons with preservation of that functionality of these structures which has developed in data modeling.

There this natural for people form of the information representation has been offered not at once. In semantic models it has appeared together with the Enhanced Entity-Relationship Model in 80th years of the last century. In this model such structural concepts as specialization and categorization are entered for representation of «IS-A»-relations. But even in data modeling there is no common opinion about «IS-A»-structures so it is no wonder that these representations are different from logic bases.

The logician approach to the «IS-A»-relation problem looks more thorough, and it cannot be bypassed at the definition of the concepts describing these relations, but focused on data modeling. The second corner stone of this system of concepts are mappings and the

semantically significant mapping theory. The analysis of the mappings accompanying «IS-A»-relations allows to consider this subject of inquiry more deeply.

Two elements of scopes of the different concepts are connected by «IS-THE»-relation if and only if they represent the same application domain (AD) object. «IS-THE»-relation is the binary 1:1-relation determining pair of inverse functional «IS-THE»-mappings. AD objects or their abstractions (data) participate in these relation and mappings. If the known treatments of «IS-A»-relations connect a class with an object or a class with a class, «IS-THE»-relations connect objects.

«IS-THE»-relation in maximum generalized understanding allows specifying «nearness» of concepts in such sense: concepts are near if there is at least one object falling under both these concepts. «IS-THE»-relation in such variant determines two partial functional mappings.

If two concepts, one of which is generalization of another, are considered, then «IS-THE»-relation becomes a subsumption relation. «IS-THE»-mapping from scope of generic concept (superclass) to scope of specific concept (subclass) is still partial functional. And here, «IS-THE»-mapping from a subclass to a superclass becomes full functional.

«IS-THE»-cluster (or, simply, cluster) is a system of classes constructed for restrictions of the same superclass concept; such, that all subclass concepts use the same mapping in the bases of them subtyping. Such mapping can be named discriminating; in fact it is this characteristic of superclass objects that determines their hits in subclasses. When an entity set is a subclass only in one «IS-THE»-cluster, it is said that full single inheritance takes place. When an entity set enters as a subclass in more than one «IS-THE»-cluster, it is said that full multiple inheritance is carried out. In both cases entities of subclass inherit characteristics of all their parent classes.

In data modeling the special subspecies of two-level «IS-A»-hierarchies are entered. This is a categorization determining special «IS-THE»-relations between the union of superclass entity sets and subclass entity set (in this case named a category). To understand core of categorization it is necessary to present the content of the concept which scope is represented by subclass entity set. The first part of its logic expression is a disjunction of content conditions of superclass entity sets, and the second part is a condition to which should satisfy entities of these sets to be presented in a category. Such two-level «IS-A»-hierarchy we name «IS-THE»- categorization (or, simply, categorization).

In categorization definition it is emphasized that superclasses do not contain entities been abstractions of the same objects and are so various that they are not connected with the common ancestor. In other words, concepts of superclasses are incompatible in pairs. In that case the element of a category always inherits characteristics only one superclass, where the parental element connected to it contains. Therefore it is said that such two-level «IS-A»-hierarchies assume selective inheritance.

The definitions submitted in this article are based ERMM rules of data structurization and integrity constraint definition concerning representation of «IS-THE»- relations.

REFERENCES

1. Wang, P. (1994) From Inheritance Relation to Nonaxiomatic Logic. *International Journal of Approximate Reasoning*. 11. pp. 281-319. DOI: 10.1016/0888-613X(94)90021-3
2. Palomäki, J. & Kangassalo, H. (2012) That IS-IN Isn't IS-A: A Further Analysis of Taxonomic Links in Conceptual Modelling. In: Gutiérrez, C.R. (ed.) *Advances in Knowledge Representation*. Publisher: InTech. pp. 3-18. DOI: 10.5772/36484
3. Brachman, R.J. (1983) What IS-A Is and Isn't: An Analysis of Taxonomic Links in Semantic Networks. *IEEE Computer*. 16(10). pp. 30-36. DOI: 10.1109/MC.1983.1654194
4. Connolly, T. & Begg, C. (2000) *Базы данных. Проектирование, реализация и сопровождение. Теория и практика* [Database Systems. A Practical Approach to Design, Implementation, and Management]. Translated from English by R. Imamutdinova, K. Ptitsyn. 2nd ed. Moscow: Williams.
5. Teorey, T., Yang, D. & Fry, J. (1986) A Logical Design Methodology for Relational Databases Using the Extended Entity-Relationship Model. *Computing Surveys*. 18(2). pp. 197-222. DOI: 10.1145/7474.7475
6. Babanov, A.M. (2007) Semantic model "Entity – Relationship – Mapping". *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 1(1). pp. 77-91. (In Russian).
7. Voyshvillo, E.K. & Degtyarev, M.G. (1994) *Logika kak chast' teorii poznaniya i nauchnoy metodologii (fundamental'nyy kurs): v 2 kn.* [Logic as a part of the theory of knowledge and scientific methodology]. Moscow: Nauka.
8. Bacharov, V.A. & Markin, V.I. (1998) *Osnovy logiki* [Foundation of logic]. Moscow: Infra-M.
9. Coronel C., Morris S. & Rob, P. (2013) *Database Systems: Design, Implementation, and Management*. 10th ed. Course Technology.
10. Connolly, T. & Begg, C. (2010) *Database Systems: A Practical Approach to Design, Implementation, and Management*. 4th ed. Addison-Wesley.
11. Elmasri, R. & Navathe, S. (2010) *Fundamentals of Database Systems*. 6th ed. Addison-Wesley.
12. Halpin, T. & Morgan, T. (2008) *Information Modeling and Relational Databases*. 2nd ed. Morgan Kaufman.
13. Ferragine, V., Doorn, J. & Rivero, L. (2009) *Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends*. Information Science Reference.
14. Smith, J. & Smith, D. (1977) Database Abstractions: Aggregation and Generalization. *ACM Transactions on Database Systems*. 2(2). pp. 105-133. DOI: 10.1145/320544.320546
15. Babanov, A.M. (2003) Theory of semantically significant mapping. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 280. pp. 239-248. (In Russian).

Е.В. Царева

**РАЗРЕШЕНИЕ КОНФЛИКТНЫХ СИТУАЦИЙ ПРИ СИНХРОНИЗАЦИИ
МНОГОПОЛЬЗОВАТЕЛЬСКИХ ONLINE-ПРИЛОЖЕНИЙ**

Приведено сравнение конфликтных ситуаций для задачи синхронизации многопользовательских online-приложений и других задач синхронизации. Предложены алгоритмы разрешения конфликтов на стороне клиента. Предлагается методика синхронизации приложений на основе указанных алгоритмов, в том числе способ программной реализации. Рассмотрена синхронизация flash-приложения «Мозговой шторм» в системе вебинаров Adobe Connect Meeting 9.

Ключевые слова: синхронизация; многопользовательское приложение; конфликт доступа; блокировка; совместное редактирование.

Термин «синхронизация» сегодня применяется чрезвычайно широко и означает приведение некоторых процессов к одному виду, совместное использование ресурсов, обмен сообщениями между процессами. Термин применяется при обеспечении многозадачности в операционных системах, в многопоточном и параллельном (на многих процессорах) выполнении задач, в базах данных для приведения двух копий базы данных к одной версии или при выполнении конфликтующих транзакций, при приведении зеркальных копий какого-либо сервера к одной версии.

При организации сетей ЭВМ термин применяется на уровне передачи данных (сверка правильности доставленных данных), при синхронизации времени на узлах в сети ЭВМ. Синхронизацией называется приведение к одному виду большого объема данных, например в мобильных сетях, когда при подключении мобильной платформы в сеть происходит синхронизация измененных данных с удаленным сервером (отправка изображений на сервис хранения личной информации, обновление комментариев к новости и др.) [1]. Синхронизацией называется передача больших объемов данных по технологии BitTorrent [2].

Термин сетевой синхронизации применяется также в случае работы многопользовательского online-приложения, когда внесенные несколькими пользователями изменения в рабочую область сразу же (в режиме реального времени) отображаются на экране других пользователей. Примером могут служить совместное редактирование диаграмм и документов, проведение виртуальных совещаний (вебинаров), многопользовательские игры, виртуальные миры.

Перечисленные примеры могут быть отнесены к синхронизации данных (приведение к одному виду двух наборов данных) или к синхронизации процессов (организация непротиворечивого выполнения нескольких процессов). Синхронизируемые процессы могут происходить в системе с разделяемой или распределенной памятью, а синхронизация данных может выполняться однократно по запуску [3] или в режиме реального времени. Далее сравним особенности синхронизации многопользовательских приложений с перечисленными подвидами задачи синхронизации относительно основной проблемы синхронизации – разрешения конфликтных ситуаций.

Если при синхронизации процессов используется общая память, то разработчик должен обеспечить бесконфликтное использование общих переменных, блокируя их на время использования, или, если выражаться точнее, блокируя остальные процессы-претенденты на время выполнения операции. В базах данных для этого предусмотрены блокировки на чтение и запись, а при многопоточном программировании используются семафоры и основанные на них мьютексы, рандеву, барьеры и другие конструкции [4]. Так как процессы могут друг друга блокировать, то возможна ситуация взаимной блокировки, например, когда два процесса обращаются к двум ресурсам в разном порядке.

Для синхронизации многопользовательских приложений проблема блокировок не стоит так остро. Если клиентские и серверные приложения однопоточные, то одновременная работа с общей памятью для них невозможна и, значит, невозможны конфликты записи или чтения. В случае многопоточной архитектуры приложений проблемы общего использования памяти решаются на стороне клиента или сервера и не относятся к сетевой синхронизации приложения в целом.

Конечно, пользователи могут вносить изменения в приложение одновременно, создавая конфликт, однако заявка на блокировку объекта создаст дополнительную временную задержку, а блокировать приложение пользователя на время выполнения операции и вовсе недопустимо. Поэтому при возникновении конфликта все процессы, кроме выбранного, не блокируются, а отменяются. Правило выбора при этом становится концептуально важным и может зависеть только от предметной области синхронизируемого приложения и различаться для разных объектов.

В связи с этим возникают две особенности синхронизации приложения. Во-первых, это дополнительный тип конфликта, когда для двух связанных событий предусмотрены разные правила выбора и при их появлении выбираются одновременно два пользователя. Например, если движению объекта неразрывно предшествует его выделение, но при одновременном выделении объекта выбирается последний пользователь, а при одновременном движении объекта выбирается первый пользователь. Во-вторых, применить правило выбора пользователя в полной мере может только сервер, решение которого, впрочем, может быть неверным, так как какое-либо сообщение серверу еще не поступило. Поэтому клиентские приложения (имеются в виду авторы событий) должны уметь отменять свои решения (которые они выполнили для удобства пользователя, не дожидаясь ответа сервера) или отвергать поступающие сообщения до получения соответствующей инструкции сервера.

Пользователь может заблокировать один из объектов приложения, например для его перемещения, однако ни отсутствие этой блокировки, ни взаимная блокировка не становятся критически важными, поскольку пользователи в таких достаточно редких ситуациях смогут идентифицировать проблему и договориться об использовании объектов. Программное решение этих проблем является только вопросом комфорта (что, разумеется, тоже важно), но не вопросом безопасности используемой информации, как, например, в базах данных.

При синхронизации процессов с распределенной памятью (многопроцессорные кластеры) процессы обмениваются сообщениями, используя функции «отправить» и «получить» [5]. На время выполнения этих функций блокируется либо процесс, либо только буфер передачи данных (так называемый асинхронный вызов, при котором возможно продолжение работы программы). Взаимная блокировка возможна, если два процесса бесконечно ожидают получения или отправки данных.

Подобные функции также используются при передаче данных по сети на транспортном уровне (используются, например, протоколы TCP, UDP) [6]. Для исключения взаимной блокировки для чтения и отправки данных в системе выделяются отдельные потоки, которые при наступлении событий вызывают соответствующие функции-обработчики (слушатели) [7].

Сходство синхронизации процессов с распределенной памятью и синхронизации многопользовательского приложения состоит в использовании сообщений. Действительно, только посредством сообщений серверное приложение может обмениваться данными с клиентами. Различие такое же, как и было описано выше, состоит в разном понимании конфликтных ситуаций и использования блокировок.

Синхронизируемое многопользовательское приложение является приложением прикладного уровня, поэтому конфликты, возможные на транспортном уровне, решаются за счет протоколов более низкого уровня. Отметим, что протокол TCP обеспечивает единый порядок поступления сообщений, так что при его использовании все клиенты получают сообщения сервера если и не в одно и то же время, то как минимум в одном порядке, что сможет заменить временные метки, которые потребуются при использовании менее надежных протоколов.

Рассмотрим подвид задачи синхронизации – синхронизацию данных. Различие синхронизации данных с однократным выполнением и синхронизации многопользовательских приложений очевидно: в первом случае выполняется один процесс над двумя копиями одной модели предметной области по заранее заданным правилам, во втором случае происходит множество процессов, которые к тому же не детерминированы, так как выполняются пользователями. Сходство заключается в потенциальном мно-

гообразии правил, применяемых при возникновении конфликтных ситуаций, и в подходе, что только один вариант какого-либо объекта должен быть в итоге принят как верный.

Синхронизацию данных в режиме реального времени кратко рассмотрим на примере распределенных баз данных [8]. Распределенные базы данных состоят из нескольких узлов, каждый из которых обслуживает некоторое количество пользователей и которые периодически обмениваются обновленной информацией. Обмен может производиться журнальным методом, при котором на каждом узле в специальном журнале сохраняются все внесенные изменения, содержимое журнала периодически отправляется другим узлам системы, где при возникновении конфликтов применяются единые для всей системы правила [9]. Также в узлах системы может использоваться кэш, в котором хранится ограниченный набор используемых данных. Кэш может обновляться с каждой транзакцией (активный кэш) либо ждать появления несоответствующих данных для выполнения запроса у других узлов системы (пассивный кэш) [10].

Несмотря на сходство задачи синхронизации распределенных баз данных с задачей синхронизации многопользовательских приложений, различия между ними все же имеются. Многопользовательские приложения не обязательно подразумевают распределенное хранение данных, сервер в системе может быть один. Данные системы в конечном итоге хранятся не в виде записей, а визуализируются в клиентском приложении, функциональные зависимости между ними обеспечиваются конечным приложением и не создают дополнительных проблем, как при синхронизации распределенных баз данных. В силу трудоемкости разработки каждого события объем хранимой информации в многопользовательских приложениях будет существенно меньше, чем в среднестатистических базах данных.

Распределенные базы данных, в которых есть сильные функциональные связи, чтобы не блокировать на время редактирования большое число объектов, выполняют транзакции на некотором срезе данных; при возникновении конфликта в таких базах данных транзакции откатываются целиком. И хотя откат транзакции отдаленно схож с необходимостью отменять свои действия в клиентских приложениях, он предполагает повтор этой операции, возможно, в автоматическом режиме, что увеличивает временные задержки. Кроме того, распределенные базы данных не требуют ввода каких-либо условий, согласно которым узлы системы отвергают сообщения сервера до определенного момента. То есть научные публикации по вопросам синхронизации распределенных баз данных освещают несколько иные проблемы, чем те, которые возникают при синхронизации многопользовательских приложений.

Таким образом, синхронизация многопользовательских приложений стоит обособленно в ряду других подобных задач. Для определения подвида рассматриваемой задачи примем во внимание, что многопоточное программирование обычно используется для вычисления наукоемких задач или при организации автоматизированных систем управления, т.е. в случаях, когда заранее детерминированные процессы должны работать согласованно. Поскольку поведение пользователей заранее непредсказуемо, а задача состоит в поддержании одного состояния предметной области на всех клиентских приложениях, то сделаем вывод, что синхронизация многопользовательских онлайн-приложений является синхронизацией данных в режиме реального времени.

Синхронизируемые многопользовательские приложения можно подразделить на задачи совместного редактирования и онлайн-игры. Различия заключаются в том, что при совместном редактировании менее важны временные задержки, нет соперничества (т.е. необходимости справедливо отдать предпочтение одному из игроков в конфликтных ситуациях), пользователи редактируют одни и те же объекты, в то время как в играх обрабатываются их взаимодействия (столкновения, выстрелы и др.). Конфликтной ситуацией в онлайн-играх становится не одновременное использование ресурсов, а разное (в силу временных задержек) развитие сюжета в разных клиентских приложениях. Для разрешения подобных конфликтных ситуаций используется предсказывание движения игроков (например, на основе кубических сплайнов [11]), обман (когда приложение корректирует действия других игроков так, чтобы проигрыш пользователя выглядел реалистично), использование балластных действий на время ожидания отклика из сети, а также общее замедление игры [12]. Правила выбора наиболее правильного варианта могут быть такими же, как и при совместном использовании ресурсов: прав первый, прав последний, прав владелец объекта и, возможно, прав тот, в чьем приложении сложнее подделать реалистичную картину проигрыша.

В рассмотренной литературе концепция синхронизации многопользовательских приложений наиболее подробно рассматривается в публикациях, посвященных онлайн-играм, однако и в них не приводятся конкретные алгоритмические решения. Целью настоящей работы является предложение алгоритмов синхронизации событий многопользовательского приложения, предложение конкретных шагов по синхронизации произвольного онлайн-приложения, т.е. предложение методики синхронизации многопользовательских онлайн-приложений (в том числе формулируется алгоритм обработки поступающих событий в приложении клиента), а также применение методики на примере синхронизации flash-приложения «Мозговой штурм» в системе вебинаров Adobe Connect Meeting 9.

1. Алгоритмы разрешения конфликтов

Далее будем предполагать, что синхронизируемая система имеет клиент-серверную архитектуру, в которой клиент отправляет сообщение о произошедшем изменении серверу, а сервер рассылает полученную информацию другим клиентам, в том числе автору события, в неизменном виде (т.е. возвращает эхо). В качестве протокола связи используются надежные протоколы (TCP), которые обеспечивают порядок доставки сообщений. Обработка событий (в том числе разрешение конфликтов) происходит на стороне клиента, а сервер осуществляет хранение информации о текущем состоянии приложения, которое отправляет каждому вновь подключившемуся пользователю. Для разрешения конфликтов будут использоваться два способа: определение очередности поступления событий (важно первое или последнее событие) и блокировка используемого объекта.

Если важно первое событие, то все последующие аналогичные действия над объектами будут отклоняться (пока событие не будет отменено, например, выделение не будет сброшено). Если важно последнее, то, напротив, они будут приниматься к исполнению. Выполнение этого алгоритма очевидно для всех сторон работы многопользовательского приложения, кроме авторов событий, которые должны ориентироваться на ответ сервера (эхо своего запроса), чтобы определить, было ли их событие первым или последним в случае возникновения конфликта. Возвращаемое эхо собственного события при этом не несет какой-либо полезной нагрузки для приложения, вся информация о произошедшем событии хранится в памяти клиента, эхо используется только для определения очередности происходящих событий.

Схема обработки сообщений в приложении приведена на рис. 1.

Указанные алгоритмы позволяют отменить собственное синхронизируемое действие, если для него важно первое событие и ранее в сети произошло аналогичное событие, и отвергнуть сообщения о событиях, если для них важно последнее событие и они произошли раньше собственного.

Вторым способом предотвращения конфликтов является блокировка ресурсов на время редактирования одним из пользователей. События блокировки, изменения (редактирования) объекта и разблокировки, а также сброс выделения доступны только тому пользователю, который ранее объект выбрал (с помощью щелчка мыши или иным способом). Для этого при выделении объекта сохраняется идентификатор пользователя, который затем сравнивается с идентификаторами авторов поступающих событий. Такой порядок позволит избежать конфликт разных правил выбора для связанных событий, когда одно из них является, по сути, выделением, а второе – редактированием.

Если для события выделения объекта важно первое событие, то события блокировки и разблокировки можно не реализовывать, так как функции блокировки будут выполняться событиями выделения и сброса выделения. Если для события выделения важно последнее событие, то, поскольку выделить объект беспрепятственно может другой пользователь, события блокировки и разблокировки объекта на время редактирования объекта (например, движения) обязательны для реализации.

События редактирования объекта (во время его блокировки) и сброса выделения не требуют возвращения эха, так как идентификатор автора события однозначно определяет, допустимо ли событие к исполнению. События блокировки и разблокировки объекта должны происходить синхронно в приложении, поэтому они требуют возвращения эха.

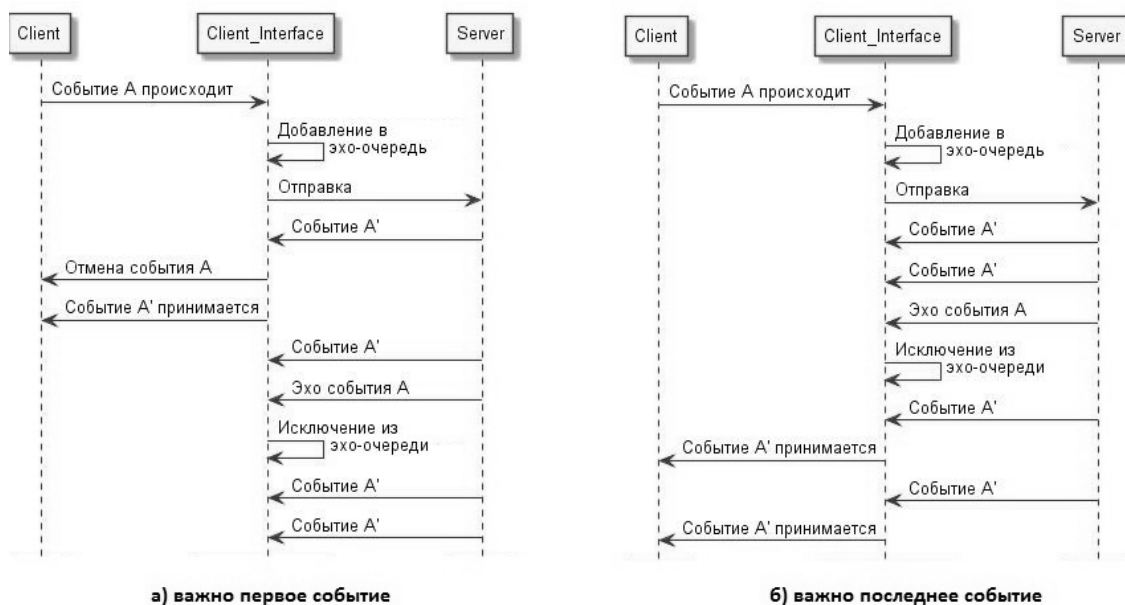


Рис. 1. Работа с событиями одного вида в приложении с учетом их очередности: *а* – важно первое событие; *б* – важно последнее событие. *A* – собственное событие клиента; *A'* – аналогичное сообщение другого пользователя

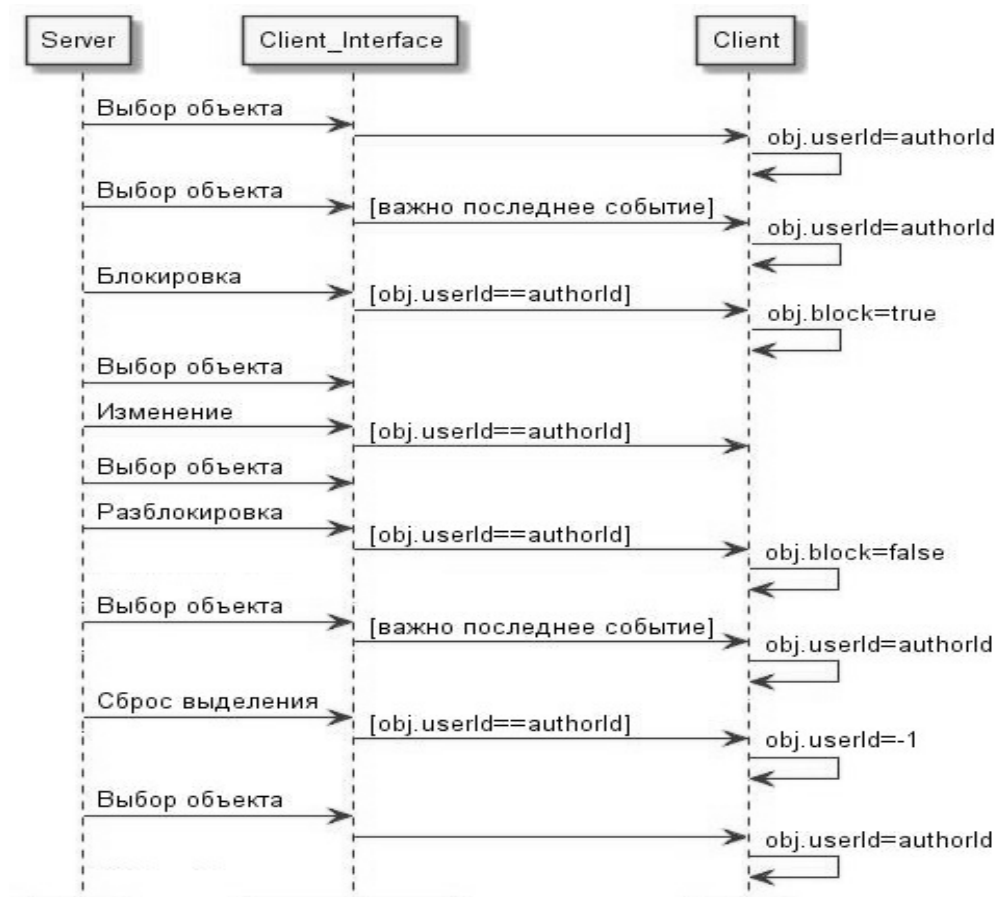


Рис. 2. Обработка событий блокирования и разблокирования ресурса клиентом – сторонним наблюдателем: *obj* – объект, над которым производится действие; *obj.userId* – идентификатор пользователя, который выбрал объект; *obj.block* – индикатор блокировки объекта; *authorId* – идентификатор автора события

Если объект заблокирован (событие блокировки было принято), запрет на его использование также действует внутри приложения.

На рис. 2 приведена последовательность обработки событий блокировки клиентом, который наблюдает происходящие события, на рис. 3 – клиентом, который является автором этих событий. На этих рисунках события блокировки и разблокировки не удалены для случая, когда для события выделения объекта важно первое событие. Наличие стрелки от Client_Interface к Client означает, что событие принимается, т.е. в приложении производятся необходимые изменения, в квадратных скобках указывается условие принятия события к исполнению, отсутствие стрелки означает отказ от выполнения события.



Рис. 3. Обработка событий блокирования и разблокирования ресурса клиентом – автором событий:
ownUserId – идентификатор пользователя клиентского приложения. На рисунке опущено возвращение эха разблокировки объекта и добавление собственных событий в эхо-очередь

2. Сообщения синхронизации

Рассмотрим способы обмена сообщениями в синхронизируемой системе и их возможный состав. Клиент может отправлять сообщения по факту наступления события либо формировать очередь сообщений о произошедших событиях и отправлять их каждый заданный временной интервал. Для непрерывных действий (например, движения объектов), которые генерируют слишком много событий, рекомендуется использовать второй способ, поскольку большое число сообщений может затруднять связь.

Сервер может отправлять полученные сообщения по факту получения либо по запросу клиента. В последнем случае клиент может запросить версии последних изменений, чтобы не обновлять не изменившуюся с последнего запроса информацию.

Формат сообщения включает в себя название команды, идентификатор автора события, идентификатор объекта, в отношении которого произошло событие, информацию о состоянии объекта. Сервер может хранить очередь полученных сообщений, если в них указываются изменения состояния, или одно последнее сообщение, если в нем хранится состояние объекта в целом.

Для максимального исключения избыточности информации в пересылаемых сообщениях разработчик может рассматривать следующие варианты формата сообщений:

- 1) хранение всего состояния (всех параметров) или его части (подмножества параметров) приложения в последнем варианте;
- 2) хранение очереди всех поступивших изменений;
- 3) хранение последнего состояния для каждого вида сообщения и для каждого объекта (имя команды получается соединением названия команды и идентификатора объекта);
- 4) хранение очереди поступивших сообщений с регулярным сохранением состояния приложения в целом или его части.

Все указанные варианты имеют очевидную избыточность хранения информации. В первом случае предполагается сохранять состояние объектов, с которыми, возможно, никаких изменений не произошло. Во втором случае сохраняется информация, которая со временем неизбежно устаревает, например, при движении объекта сервер должен будет сохранять все нюансы этого движения. В третьем варианте предлагается генерация слишком большого числа имен сообщений. Четвертый вариант является компромиссным, он позволяет регулярно удалять большие объемы информации об изменениях, но при этом теряется возможность использовать удаленную информацию для отмены действий пользователей.

Для примера приведем порядок синхронизации содержимого виртуальных досок в виртуальном мире vAcademia [13]. Все действия (типовые команды) пользователей сохраняются на сервере в виде бинарного массива. Бинарному массиву ставится в соответствие его последняя версия. Клиенты системы регулярно запрашивают номер последней версии, и если она изменилась, то запрашивают массив всех произведенных действий и перерисовывают содержимое виртуальной доски заново.

Далее будет предполагаться, что клиент накапливает очередь сообщений и отправляет их серверу каждый интервал времени, сервер рассылает сообщения по факту поступления, данные проекта будут храниться на сервере в виде очереди поступающих сообщений с периодическим сохранением полной версии проекта.

3. Методика синхронизации online-приложений

Предлагаемая методика представляет последовательность шагов по синхронизации онлайн-приложения и предполагает наличие следующих позиций:

- отлаженный исходный код приложения;
- инструменты разработки приложения;
- инструменты разработки клиент-серверной архитектуры;
- работающее приложение сервера.

В функции серверного приложения должны входить следующие возможности:

- получение сообщений;
- отправление их участникам, включая автора сообщения;
- сохранение сообщений под указанным именем;
- отправление сохраненных сообщений вновь подключившимся участникам;
- обеспечение режима сохранения сообщения: очередь изменений либо последнее состояние.

Порядок синхронизации приложения включает следующие пункты:

- определение перечня событий, подлежащих синхронизации;
- для каждого события определяется режим синхронизации: важно первое или последнее сообщение либо блокировка (определяется событие выбора объекта);

- для каждого события разрабатывается состав полезного содержимого сообщения (не считая названия команды, идентификаторов автора и объекта);
- для каждого события определяются режим хранения на сервере (очередь изменений или последнее состояние), потребность в эхо-ответе;
- программная реализация;
- отладка и тестирование.

Синхронизируемые события могут подразделяться на следующие группы:

- создание или загрузка проекта;
- редактирование объектов;
- навигация (например, переход на стартовую страницу);
- организация (например, установка временного интервала синхронизации).

При создании или загрузке проекта до возвращения эха следует отменять все поступающие сообщения относительно предметной области приложения, чтобы не принять к исполнению команды, предназначенные для предыдущего проекта. Такие события, как сохранение проекта на стороннем сервере хранения, открытие страницы справки и другие, синхронизировать не обязательно.

4. Программная реализация

В рамках программной реализации предлагается добавить в код приложения следующие пункты (рис. 4):

- класс, отвечающий за связь с сервером (NetInterface);
- класс-журнал, отвечающий за накопление событий для отправки, обработку конфликтных ситуаций, применение полученных сообщений (Journal);
- главный класс (MainClass) приложения для каждого синхронизируемого события, предоставляющий функцию, которая вызывается из внутренних частей кода приложения и отправляет событие в журнал для регистрации;
- классы синхронизируемых объектов (SyncObject), включающие в себя параметры userId и blocked, вызовы функций регистрации в журнале событий.

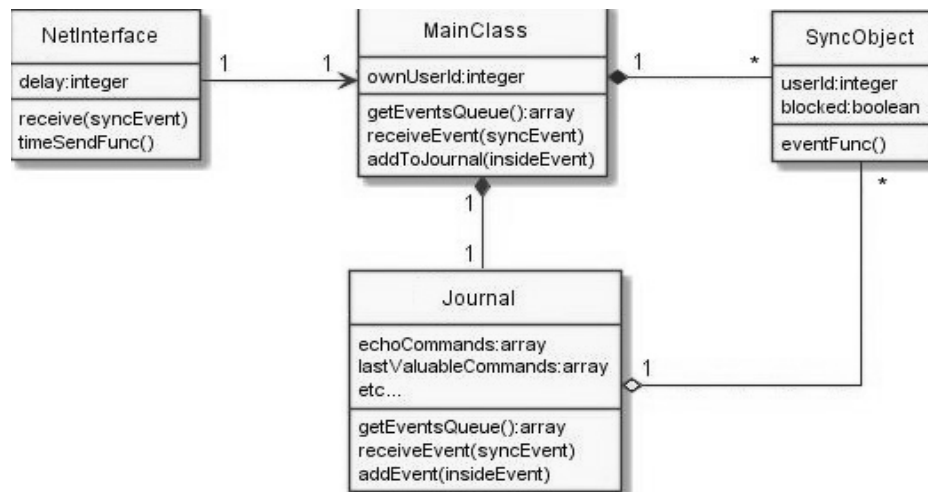


Рис. 4. Внедрение разработанных алгоритмов в код приложения

Самым сложным для реализации является класс-журнал. Журнал предоставляет функцию добавления внутреннего события, получение очереди накопленных событий (сообщений) для отправки, функцию обработки поступающих из сети событий. Непрерывные события (вроде движения) помещаются в очередь накопленных событий непосредственно перед отправлением. При добавлении события оно сохраняется в формате, удобном для отправки: сохраняется имя команды, идентификатор автора и объекта, полезная нагрузка, необходимость возвращения эха, идентификатор эха (уникальный в преде-

лах клиентского приложения), параметр сохранения информации на сервере в виде очереди изменений или последнего состояния объекта. В журнале должен быть создан массив событий (эхо-очередь), для которых ожидается возвращение эха.

Для определения вида события в журнале используются списки названий команд:

- для которых необходимо эхо;
- для которых необходимо сохранение в виде очереди на сервере;
- события создания проекта;
- события, требующие идентичности автора (блокировка, разблокировка, редактирование, сброс выделения объекта);
- события, не связанные с предметной областью (установка интервала синхронизации, переход на стартовую страницу);
- события, для которых важно последнее событие.

Функция обработки полученного из сети события должна реализовывать следующий алгоритм:

1. Если ожидается эхо создания проекта:

1.1) если вернулось эхо любого события, удалить его из эхо-очереди;

1.2) иначе если пришло событие, не связанное с предметной областью, принять согласно пунктам 4 и 5.

2. Иначе если вернулось эхо (определить по идентификатору автора):

2.1) если речь идет о блокировке либо разблокировке и `userId` объекта равен идентификатору автора (т.е. собственному `ownUserId`), принять;

2.2) удалить из эхо-очереди.

3. Иначе если пришло событие, требующее идентичности автора:

3.1) если совпал идентификатор, принять.

4. Иначе если пришло событие, для которого важно последнее событие:

4.1) если в эхо-очереди нет подобного события, принять;

4.1.1) если это событие выбора и в эхо-очереди для выбранного объекта ожидается эхо блокировки, отменить редактирование объекта.

5. Иначе если пришло событие, для которого важно первое событие:

5.1) если в эхо-очереди есть подобное событие, отменить его;

5.2) если событие не произошло (было отменено), принять событие.

Пункты следует выполнять в указанном порядке. На первом шаге важно отклонить и не обработать сообщения, которые относятся к предыдущему проекту. На втором шаге важно не обработать собственные сообщения так же, как сообщения других пользователей. Если событие требует идентичности автора, оно не заносится в список событий, для которых важно последнее событие, т.е. на шагах 4 и 5 будет распознано программой как событие, для которого важно первое событие, и должно быть обработано только на 3-м шаге. Последние два пункта можно поменять местами. Другая организация проверки типов событий может привести к росту числа операторов сравнения и усложнить процесс разработки приложения.

В журнале также должны быть определены функции, которые могут менять состояние приложения, т.е. могут «принять» событие, и другие вспомогательные параметры и функции.

5. Синхронизация flash-приложения в Adobe Connect Meeting 9

Система вебинаров Adobe Connect Meeting 9 предназначена для проведения виртуальных совещаний и, в частности, позволяет загрузку синхронизированных flash-приложений для совместного использования. Для синхронизации приложения предоставляется инструмент разработки Collaboration Builder Toolkit SDK, доступный в среде разработки Flash Builder (панель Flex) при использовании инструмента разработки SDK 3.6. Инструмент Collaboration Builder Toolkit SDK предоставляет разные возможности, в том числе функцию отправки сообщений серверу с параметрами: название команды; содержимое сообщения; параметры, названные «дельта» (следует ли сохранять очередь поступающих сообщений или

сохранять последнее сообщение) и «эхо» (следует ли вернуть эхо сообщения). Роль сервера здесь играет сама система Adobe Connect Meeting.

Приложение также может быть разработано в среде разработки Flash Professional и экспортировано в Flash Builder при помощи инструмента Flex Component Kit, так что в Flash Builder разрабатывается код (флекс-оболочка), ответственный за связь с сервером, в то время как основное приложение (далее назовем его flash-частью) должно предоставлять необходимый интерфейс для обмена информацией и включает в себя основной функционал приложения (рис. 5).

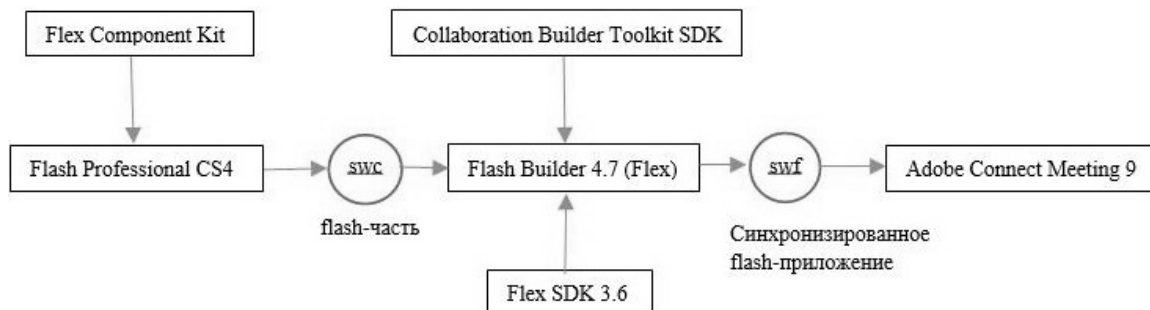


Рис. 5. Схема разработки flash-приложения, синхронизированного в системе вебинаров Adobe Connect Meeting, “swc” и “swf” – расширения выходных продуктов (файлов)

Указанная схема была применена для разработки flash-приложения «Мозговой штурм» [14]. Приложение позволяет совместно построить дерево суждений, состоящее из вершин, соединенных прямыми линиями. Каждой вершине можно назначить текст, размер, цвет и толщину обводки вершины, цвет и толщину линии к родительской вершине. Вершины можно произвольно перемещать. Вершина выделяется желтым цветом, если она выбрана самим пользователем, и красным, если она выбрана другим участником виртуального собрания (рис. 6).

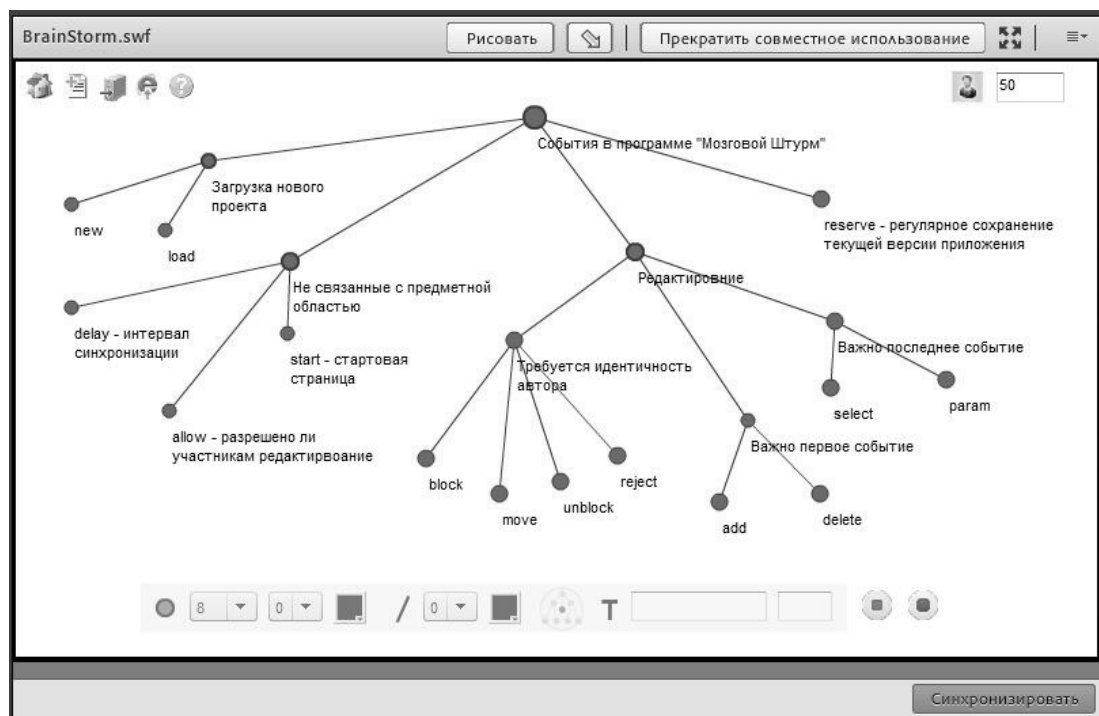


Рис. 6. Программа «Мозговой Штурм» в окне вебинара Adobe Connect Meeting 9; на рабочем поле отображены 14 используемых для синхронизации приложения событий

Очередь сообщений накапливается во flash-части приложения в ходе работы пользователя. Флекс-оболочка извлекает эту очередь и отправляет серверу каждый интервал синхронизации. Событие сохра-

нения полной версии проекта (reserve) создается пользователем с наименьшим идентификатором каждые 100 интервалов синхронизации и отправляется серверу вместе с сообщениями об обнулении всех остальных событий.

Во flash-части приложения синхронизируются следующие 13 видов событий: start (переход на стартовую страницу), new (создать проект), load (загрузить проект), delay (установить интервал синхронизации), allow (разрешить участникам редактировать проект), select (выбрать вершину), param (установить параметры вершины), move (движение вершины), add (добавление вершины), delete (удаление вершины), reject (сброс выделения вершины), block (заблокировать вершину), unblock (разблокировать вершину).

События по своим характеристикам распределены следующим образом:

- не ожидающие эха: move, delete, reject;
- не хранящиеся на сервере в виде очереди: start, new, load, delay, allow;
- не относящиеся к предметной области: start, delay, allow;
- события создания проекта: new, load;
- важно последнее событие: select, param, start, delay, allow, new, load;
- важна идентичность автора: block, move, unblock, reject.

Приложение «Мозговой штурм» опубликовано в сети Интернет, с ним можно ознакомиться по адресу [15].

Заключение

В статье рассматривается синхронизация многопользовательских онлайн-приложений относительно конфликтных ситуаций, когда одно и то же событие происходит в разных клиентских приложениях одновременно. При возникновении конфликта выбирается только один пользователь, чье событие принимается как произошедшее, события остальных пользователей отменяются. В работе используются два правила разрешения конфликтов, основанные на временной последовательности поступления запросов (выбирается первый или последний пользователь), а также механизм блокировки объекта на время редактирования. При этом блокировке (и редактированию) обязательно предшествует событие выделения объекта, для которого, в свою очередь, определяется правило выбора пользователя (первый или последний). Блокировка и редактирование доступны только тому пользователю, который объект выбрал.

Указанные правила реализуются приведенными в статье алгоритмами на основе возвращаемого сервером эха собственных событий (предполагается, что используемые протоколы передачи данных обеспечивают единый порядок поступления сообщений). Для системы с клиент-серверной архитектурой, в которой события обрабатываются на стороне клиента, очередь сообщений отправляется серверу каждый интервал времени, сервер обеспечивает хранение и рассылку сообщений, приложение реализует события редактирования объектов, разработана методика синхронизации многопользовательских онлайн-приложений. В рамках этой методики предлагается вариант программной реализации алгоритмов, в частности приводится результирующий алгоритм по обработке поступающих в приложение клиента сообщений.

Разработанная методика в итоге была применена для синхронизации flash-приложения «Мозговой штурм» в системе вебинаров (т.е. виртуальных совещаний) Adobe Connect Meeting 9, предназначенного для совместного редактирования логического дерева.

ЛИТЕРАТУРА

1. Демиш В.О., Пищик Б.Н. Синхронизация данных на мобильных платформах // Вестник Новосибирского государственного университета. Сер. Информационные технологии. 2013. Т. 11, № 4. С. 46–58.
2. Макаревич А.С. Механизм защищенной синхронизации данных на базе протокола BitTorrent в распределенной вычислительной среде // Электронные средства и системы управления. 2007. № 2. С. 63–65.
3. Баранов Д.Е., Давыденко В.А., Цупранков А.В., Янишевская А.Г., Василевский В.Б., Соседко В.В. Разработка программного приложения для синхронизации базы данных справочника материалов и сортиментов // Инженерный вестник Дона. 2013. Т. 25, № 2 (25). С. 55.

4. Allen B. Downey The little book of semaphores. URL: <http://greenteapress.com/semaphores/downey08semaphores.pdf> (access date: 22.10.15).
5. Gropp W. Tutorial on MPI: The message-passing interface. URL: <http://www.mcs.anl.gov/research/projects/mpi/tutorial/gropp/talk.html> (access date: 05.11.15).
6. Сетевое программирование в ОС UNIX. URL: <http://www.opennet.ru/docs/RUS/tcpip/netprg0.html> (access date: 05.11.15).
7. Семёнов Д.Е. Использование протоколов и их синхронизация для обмена данными в клиент-серверных приложениях // Современные техника и технологии : 13-я Междунар. науч.-практическая конференция студентов, аспирантов и молодых ученых, г. Томск, 26–30 марта 2007 г. : в 3 т. / Томский политехнический университет. Томск, 2007. Т. 2. С. 428–430.
8. Демидов А.А. Топологические методы в проектировании системы синхронизации конкурирующих транзакций распределенной базы данных // Программные системы: теория и приложения. 2011. Т. 2, № 4. С. 139–152.
9. Плутенко А.Д., Ситников А.А. Проблемы имитационного моделирования процесса журнальной репликации данных в распределенных СУБД // Информатика и системы управления. 2005. № 1. С. 3–15.
10. Демидов А.А. Об особенностях организации СУБД в MPP-системе. URL: http://psta.psiras.ru/read/psta2014_4_195-205.pdf (дата обращения: 22.10.15).
11. Калдвел Н. Устранение эффекта запаздывания с использованием кубических сплайнов. URL: <http://www.gamedev.ru/code/articles/?id=4257> (дата обращения: 22.10.15).
12. Карпов А. Как создать онлайн-игру. URL: <http://www.ant-karlov.ru/kak-sozdat-mnogopolzovatel'skuu-igru-anons.html> (дата обращения: 22.10.15).
13. Сморгалов А.Ю. Реализация инструмента графического комментирования на виртуальной доске на графических потоковых процессорах // Образовательные технологии и общество. 2012. Т. 15, № 4. С. 444–456.
14. Царева Е.В. Использование flash-технологий для расширения функциональности LMS Moodle и системы вебинаров Adobe Connect Meeting // Высшее образование сегодня. 2014. № 8. С. 30–33.
15. Мозговой шторм. URL: http://flash_services.lcg.tpu.ru/brain_storm/index.html (дата обращения: 22.10.15).

Царева Елена Владимировна. E-mail: tsareva@tpu.ru
Томский политехнический университет

Поступила в редакцию 21 августа 2015 г.

Tsareva Elena V. (Tomsk Polytechnic University, Russian Federation).

Resolution of conflicts during synchronization multiplayer online-application.

Keywords: synchronization; multiplayer application; access conflict; locking; collaborative use.

DOI: 10.17223/19988605/34/9

In information technology synchronization problem is quite urgent: the synchronization required when providing multi-tasking operating systems, in parallel and multi-threaded programming, with multi-user database. Besides synchronization process there is also the problem of synchronization of data, i.e. bringing copies of large amounts of data to one version, for example, use of cloud services, update mirroring servers or remote databases. Data synchronization can occur singly or in real time, for example, in the distributed databases and the multi-user online applications.

Task of synchronization of multiplayer online applications is different from other similar tasks. Its features are a critical time delays, reject all but one of the conflicting requests, the execution of their own actions (or cancellation of incoming events) in the application before return of a special response from the server, optionally (but preferably) lock objects on the editing, the possibility of conflict of different user selection rules for related events.

Multi-user applications can be divided into online games and applications of collaborative editing. Multiplayer games differ from collaborative editing systems by the presence of competition, conflicts arise there not only due to simultaneous access, but also different (due to time delays) developments in various client applications of system.

For conflict resolution two selection rules can be used, they based on the time sequence of receipt requests: select the first or the last user. For the authors of events the sequence is determined by the returned from the server echo events: before the return of the echo received events are considered as events occurred before, and after returning of the echo – as events occurred later. Similar conclusions are qualified because used data transfer protocols provides a unified order for receipt of messages to all customers.

For conflict resolution is also used the lock of the editing object. Selecting an object certainly precedes block event (and then edit event), for which there is determined the choice rule (first or last user). Lock, unlock, edit and deselection are only available to the user that has selected an object. Locked element is also not available for the new events of editing.

These rules are implemented in the proposed algorithms. For a system with a client-server architecture, in which events are processed on the client side, messages are sent to the server each time interval, the server provides the storage of messages, and the application implements the events editing objects, there is developed methodology of synchronizing of multi-user online applications. As part of this methodology there are proposed to list all the synchronized events, to define a policy for them for resolving the conflicts, to develop content and order of storage on the server.

For a software implementation in the application code there are proposed to add a class of network interface responsible for connection to the server, the class journal in which application events are recorded, and where network events are

processed. The main class must provide the registration function of the events in the journal. In classes of used objects there are offered to add a variable that stores the user ID, flag of the object lock, function calls of event logging, which is determined in the main class. The paper also presents the algorithm for processing of incoming messages to the client application designed for journal class.

The developed methodology was eventually used to synchronize the flash-application "Brainstorm" in the webinars (i.e., virtual meetings) Adobe Connect Meeting 9. The application allows you to build logic tree together, to determine the color, stroke width, and other parameters of vertices and connecting them lines, specify the text labels to the tops. There are 14 kinds of events that synchronized in application, including removing and adding vertices, changing settings and vertices moving, loading the project, the interval timing, and other events.

REFERENCES

1. Demish, V.O. & Pishchik, B.N. (2013) Data synchronization on mobile platforms. *Vestnik Novosibirskogo gosudarstvennogo universiteta. Ser. Informatsionnye tekhnologii – Novosibirsk State University Journal of Information Technologies*. 11(4). pp. 46-58. (In Russian).
2. Makarevich, A.S. (2007) Mekhanizm zashchishchennoy sinkhronizatsii dannykh na baze protokola BitTorrent v raspredelennoy vychislitel'noy srede [A mechanism for protected data synchronization based on BITTORRENT protocol in the distributed computer network]. *Elektronnye sredstva i sistemy upravleniya*. 2. pp. 63-65.
3. Baranov, D.E., Davydenko, V.A., Tsuprankov, A.V., Yanishevskaya, A.G., Vasilevskiy, V.B. & Sosedko, V.V. (2013) Development of software applications for database synchronization reference of materials and assortments. *Inzhenernyy vestnik Dona – Engineering Journal of Don*. 2(25). pp. 55.
4. Downey, A.B. (n.d.) *The little book of semaphores*. [Online] Available from: <http://greenteapress.com/semaphores/downey08semaphores.pdf>. (Accessed: 22nd October 2015).
5. Gropp, W. (n.d.) *Tutorial on MPI: The message-passing interface*. [Online] Available from: <http://www.mcs.anl.gov/research/projects/mpl/tutorial/gropp/talk.html>. (Accessed: 5th November 2015).
6. Fedoruk, V.G. (n.d.) *Setevoye programirovanie v OS UNIX* [Network Programming in the OS UNIX]. [Online] Available from: <http://www.opennet.ru/docs/RUS/tcpip/netprg0.html>. (Accessed: 5th November 2015).
7. Semenov, D.E. (2007) [Using the protocols and timing for exchanging data in a client-server applications]. *Sovremennye tekhnika i tekhnologii* [Modern technique and technologies]. The 13th International Scientific Conference of Students, Postgraduates and Young Scientists. Tomsk. 26 to 30 March, 2007. Vol. 2. Tomsk: Tomsk Polytechnic University. pp. 428-430.
8. Demidov, A.A. (2011) Topologicheskie metody v proektirovanii sistemy sinkhronizatsii konkuriruyushchikh tranzaktsiy raspredelennoy bazy dannykh [Topological methods in the synchronization system design for concurrent transactions in distributed databases]. *Programmye sistemy: teoriya i prilozheniya*. 2(4). pp. 139-152.
9. Plutenko, A.D. & Sitnikov, A.A. (2005) Problemy imitatsionnogo modelirovaniya protsessa zhurnal'noy replikatsii dannykh v raspredelennykh SUBD [Problems of simulation of magazine data replication in distributed database]. *Informatika i sistemy upravleniya*. 1. pp. 3-15.
10. Demidov, A.A. (n.d.) *Ob osobennostyakh organizatsii SUBD v MPP-sisteme* [About the peculiarities of the organization database in the MRP-system]. [Online] Available from: http://psta.psiras.ru/read/psta2014_4_195-205.pdf. (Accessed: 22nd October 2015).
11. Caldwell, N. (2002) *Ustranenie effekta zapazdyvaniya s ispol'zovaniem kubicheskikh splaynov* [Defeating lag with cubic splines]. [Online] Available from: <http://www.gamedev.ru/code/articles/?id=4257>. (Accessed: 22nd October 2015).
12. Karpov, A. (n.d.) *Kak sozdat' onlayn-igru* [How to design an online game]. [Online] Available from: <http://www.ant-karlov.ru/kak-sozdat-mnogopolzovatelskuu-igru-anons.html>. (Accessed: 22nd October 2015).
13. Smorkalov, A.Yu. (2012) Realizatsiya instrumenta graficheskogo kommentirovaniya na virtual'noy doske na graficheskikh potokovykh protsessorakh [Tool implementation for graphic commenting on the virtual board on graphic stream processor]. *Obrazovatel'nye tekhnologii i obshchestvo*. 15(4). pp. 444-456.
14. Tsareva, E.V. (2014) The use of flash-technology for extension of LMS Moodle and webinar system Adobe Connect Meeting. *Vysshee obrazovanie segodnya – The Higher Education Today*. 8. pp. 30-33. (In Russian).
15. Tomsk Polytechnic University. (n.d.) *Mozgovoy shturm* [Brain Storm]. [Online] Available from: http://flash_services.lcg.tpu.ru/brain_storm/index.html. (Accessed: 22nd October 2015).

СВЕДЕНИЯ ОБ АВТОРАХ

БАБАНОВ Алексей Михайлович – доцент, кандидат технических наук, доцент кафедры программной инженерии факультета информатики Томского государственного университета. E-mail: babanov2000@mail.ru

БАХОЛДИНА Мария Алексеевна – аспирантка факультета прикладной математики и кибернетики Национального исследовательского Томского государственного университета. E-mail: maria.bakholdina@gmail.com

БРОНЕР Валентина Игоревна – аспирантка кафедры теории вероятностей и математической статистики факультета прикладной математики и кибернетики Томского государственного университета. E-mail: valsubbotina@mail.ru

ВОЖОВ Станислав Сергеевич – аспирант кафедры теоретической и прикладной информатики Новосибирского государственного технического университета. E-mail: vss920414@gmail.com

ГОРЦЕВ Александр Михайлович – профессор, доктор технических наук, декан факультета прикладной математики и кибернетики Томского государственного университета. E-mail: gam@mail.fpmk.tsu.ru

ДОМБРОВСКИЙ Владимир Валентинович – профессор, доктор технических наук, заведующий кафедрой математических методов и информационных технологий в экономике экономического факультета Томского государственного университета. E-mail: dombrovs@ef.tsu.ru

КВАЧ Елена Сергеевна – магистрант факультета информатики Томского государственного университета. E-mail: kvachelena93@gmail.com

КИМ Константин Станиславович – магистрант кафедры исследования операций Томского государственного университета. E-mail: kks93@rambler.ru

НАЗАРОВ Анатолий Андреевич – профессор, доктор технических наук, заведующий кафедрой теории вероятностей и математической статистики факультета прикладной математики и кибернетики Томского государственного университета. E-mail: nazarov.tsu@gmail.com

САМОРОДОВА Мария Владимировна – аспирант кафедры прикладной математики факультета прикладной математики и кибернетики Томского государственного университета. E-mail: samorodova21@gmail.com

СИРОТИНА Мария Николаевна – аспирантка факультета прикладной математики и кибернетики Томского государственного университета. E-mail: mashuliagol@mail.ru

СМАГИН Валерий Иванович – профессор, доктор технических наук, профессор кафедры исследования операций Томского государственного университета. E-mail: vsm@mail.tsu.ru

ЦАРЕВА Елена Владимировна – программист центра программно-технологического обеспечения электронного обучения (ЦПТОЭО) Института электронного обучения (ИнЭО) Томского политехнического университета. E-mail: tsareva@tpu.ru

ЦИЦИАШВИЛИ Гурам Шалвович – профессор, доктор физико-математических наук, профессор кафедры алгебры, геометрии и анализа Дальневосточного федерального университета, заведующий лабораторией вероятностных методов и системного анализа Института прикладной математики ДВО РАН. E-mail: gu-ram@iam.dvo.ru

ЧИМИТОВА Екатерина Владимировна – доцент, кандидат технических наук, доцент кафедры теоретической и прикладной информатики Новосибирского государственного технического университета. E-mail: chimitova@corp.nstu.ru

Научный журнал

**ВЕСТНИК
ТОМСКОГО
ГОСУДАРСТВЕННОГО
УНИВЕРСИТЕТА**

**УПРАВЛЕНИЕ,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И ИНФОРМАТИКА**

**TOMSK STATE UNIVERSITY
JOURNAL OF CONTROL AND COMPUTER SCIENCE**

2016. № 1 (34)

Редактор Н.А. Афанасьева
Корректор Е.Г. Шумская
Оригинал-макет А.И. Лелоюр
Редакторы-переводчики: Г.М. Кошкин; В.Н. Горенинцева
Дизайн обложки Л.Д. Кривцова

Подписано к печати 15.03.2016 г. Формат 60x84¹/₈.

Гарнитура Times.

Печ. л. 11,75; усл. печ. л. 10,9.

Тираж 250 экз. Заказы № 1671, 1671/1.

Журнал отпечатан на полиграфическом оборудовании
Издательского Дома Томского государственного университета

634050, г. Томск, Ленина, 36

Тел. 8+(382-2)-53-15-28

Сайт: <http://publish.tsu.ru>; E-mail: rio.tsu@mail.ru