

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ

**Научный журнал Российской академии наук
(издается под руководством Отделения нанотехнологий
и информационных технологий РАН)**

Издается с 1989 года
Журнал выходит ежеквартально

Учредители:
Российская академия наук
Институт проблем информатики Российской академии наук

РЕДАКЦИОННЫЙ СОВЕТ

академик РАН И. А. Соколов — председатель Редакционного совета
академик РАН Г. И. Савин
академик РАН А. Л. Стемпковский
член-корреспондент РАН Ю. Б. Зубарев
профессор Ш. Долев (S. Dolev, Beer-Sheva, Israel)
профессор Ю. Кабанов (Yu. Kabanov, Besancon, France)
профессор М. Никулин (M. Nikulin, Bordeaux, France)
профессор В. Ротарь (V. Rotar, San-Diego, USA)
профессор И. Ушаков (I. Ushakov, San-Diego, USA)
профессор М. Финкельштейн (M. Finkelstein, Rostok, Germany)

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

академик РАН И. А. Соколов — главный редактор
профессор, д.ф.-м.н. С. Я. Шоргин — заместитель главного редактора
д.т.н. В. Н. Захаров
проф., д.т.н. В. Д. Ильин
проф., д.ф.-м.н. Л. А. Калининченко
д.т.н. В. А. Козмидиади
проф., д.т.н. К. К. Колин
проф., д.ф.-м.н. В. Ю. Королев
проф., д.ф.-м.н. А. В. Печинкин
проф., д.г.-м.н. Р. Б. Сейфуль-Муллюков
проф., д.т.н. И. Н. Сеницын
к.т.н. А. В. Филин
к.ф.-м.н. С. А. Христочевский

Редакция

профессор, д.г.-м.н. Р. Б. Сейфуль-Муллюков
к.ф.-м.н. Е. Н. Арутюнов
С. Н. Стригина (ответственный секретарь)

© Институт проблем информатики Российской академии наук, 2014

Журнал входит в систему Российского индекса научного цитирования (РИНЦ):

http://elibrary.ru/title_about.asp?id=28980

Журнал включен в базу данных CrossRef (систему DOI — Digital Object Identifier),
в базу данных Ulrich's periodicals directory
и в информационную систему «Общероссийский математический портал Math-Net.Ru»

Журнал реферируется в «Реферативном журнале» ВИНТИ
и в системе Google Scholar

Журнал «Системы и средства информатики»
включен в «Перечень российских рецензируемых журналов,
в которых должны быть опубликованы основные научные результаты диссертаций
на соискание ученых степеней доктора и кандидата наук», утвержденный ВАК

<http://www.ipiran.ru/journal/collected>

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ

Том 24 № 1 Год 2014

СОДЕРЖАНИЕ

Математическое обеспечение параметрического моделирования распределений в интегродифференциальных стохастических системах И. Н. Сеницын, И. В. Сергеев, В. И. Сеницын, Э. Р. Корепанов, В. В. Белоусов	4
Параллелизм в микропроцессорах А. К. Горшенин, С. В. Замковец, В. Н. Захаров	46
О теоретически стойких шифрах С. М. Рацеев	61
Методология мутационного тестирования для наземных испытаний бортовой аппаратуры космических аппаратов Д. А. Недорезов, А. И. Легалов, О. В. Непомнящий, С. С. Красненко, А. В. Анкудинов	73
Динамическое создание карт уровня WiFi-сигналов для систем локального позиционирования Р. В. Воронов, С. В. Малодушев	80
Автоматизированная система разработки и контроля оперативного графика движения поездов в период предоставления «окон» (АС ОГДПО) С. К. Дулин, И. Н. Розенберг, В. И. Уманский	93
Информационная технология исследования тонкой структуры хаотических процессов в плазме с помощью анализа спектров А. К. Горшенин	116
Алгоритм и программный комплекс для анализа характеристик акустических волн плазменного актуатора, встроенного в модельное сопло самолета А. Д. Черноусов, Д. В. Малахов, Н. Н. Скворцова	128

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ

Том 24 № 1 Год 2014

СОДЕРЖАНИЕ

Метод формирования системы показателей живучести информационно-телекоммуникационных сетей А. А. Зацаринный, Н. Г. Буроменский, А. И. Гаранин	138
Функциональная гибридная интеллектуальная система для поддержки принятия решений при диагностике артериальной гипертензии И. А. Кириков, А. В. Колесников, С. Б. Румовская	153
Вопросы выбора медицинской информационной системы и облачные технологии Г. Я. Илюшин, В. И. Лиманский	180
Внедрение и использование в университете технологии виртуализации рабочих мест М. А. Вышиванов, Д. В. Гмарь, В. В. Крюков, К. С. Нагорняк, К. И. Шахгельдян	193
Методика построения веб-сервисов распределенных платформенно-независимых систем автоматизированного проектирования В. Н. Гридин, Г. Д. Дмитриевич, Д. А. Анисимов	213
Задачи идентификации информационных объектов в распределенных массивах данных М. М. Гершкович, Т. К. Бирюкова	224
Библиографическая информационно-аналитическая система ИПИ РАН М. Ю. Заикин, О. Л. Обухова, И. В. Соловьев	244

МАТЕМАТИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ПАРАМЕТРИЧЕСКОГО МОДЕЛИРОВАНИЯ РАСПРЕДЕЛЕНИЙ В ИНТЕГРОДИФФЕРЕНЦИАЛЬНЫХ СТОХАСТИЧЕСКИХ СИСТЕМАХ*

*И. Н. Синицын¹, И. В. Сергеев², В. И. Синицын³, Э. Р. Корепанов⁴,
В. В. Белоусов⁵*

Аннотация: Рассматриваются методы и алгоритмы аналитического и статистического моделирования одно- и многомерных распределений в эрдитарных интегродифференциальных стохастических системах (ИДСтС). Приводятся нелинейные стохастические интегродифференциальные уравнения процессов. Для затухающих физически возможных ядер рассматриваются два способа их аппроксимации: на основе линейных операторных уравнений и вырожденных ядер. Устанавливаются алгоритмы приведения ИДСтС к дифференциальным стохастическим системам (ДСтС). Дается подробный анализ подходов к аналитическому и статистическому моделированию распределений в ИДСтС, приводимых к ДСтС. В основу подходов положены как методы прямого численного интегрирования уравнений ДСтС, так и численного интегрирования для параметров ортогонального разложения плотностей (моментов, коэффициентов ортогонального разложения и др.). Рассматриваются алгоритмы аналитического и статистического моделирования, основанные на методе статистической линеаризации (МСЛ) и методе нормальной аппроксимации (МНА). Приведены условия устойчивости алгоритмов на основе МСЛ и МНА. Для задач МСЛ рассматриваются прямые одношаговые сильные методы и алгоритмы численного интегрирования (различной точности) для гладких и разрывных правых частей уравнений ИДСтС. Приводятся комплекс тестовых примеров для инструментального программного обеспечения IDStS в среде MATLAB.

Ключевые слова: аналитическое моделирование; дифференциальная стохастическая система; инструментальное программное обеспечение; интегродифференциальная стохастическая система; обобщенная формула Ито; сингулярные ядра; система с запаздыванием; статистическое моделирование; стохастическое дифференциальное уравнение в смысле Ито; эрдитарная система

DOI: 10.14357/08696527140101

* Работа выполнена при финансовой поддержке Программы ОНИТ РАН «Интеллектуальные информационные технологии, системный анализ и автоматизация» (проект 1.7).

¹Институт проблем информатики Российской академии наук, sinitsin@dol.ru

²Институт проблем информатики Российской академии наук, ISergeev@ipiran.ru

³Институт проблем информатики Российской академии наук, VSinitsin@ipiran.ru

⁴Институт проблем информатики Российской академии наук, EKorepanov@ipiran.ru

⁵Институт проблем информатики Российской академии наук, VBelousov@ipiran.ru

1 Введение

Интегродифференциальные стохастические системы являются подходящими математическими моделями так называемых эредитарных систем. Такие системы, также известные как системы с последствием, часто называют просто системами с памятью [1–4]. Процессы в ИДСтС, как правило, являются немарковскими.

В прикладных задачах для затухающей памяти путем аппроксимации ядер, определяющих память линейными обыкновенными дифференциальными уравнениями или вырожденными ядрами, удается привести интегродифференциальные уравнения к дифференциальным. В результате процессы в таких ДСтС становятся марковскими. Поэтому, как показано в [2–4], оказывается возможным использовать богатый опыт аналитического и статистического моделирования.

Вопросам параметрического аналитического и статистического моделирования распределений в ДСтС (см. библиографические замечания в [5–8]) посвящена обширная литература.

Статья посвящена математическому обеспечению аналитического и статистического моделирования в негауссовских ИДСтС, описываемых интегродифференциальными стохастическими уравнениями Ито с винеровскими и пуассоновскими шумами в конечномерных пространствах.

В разд. 2 рассматриваются уравнения ИДСтС. Алгоритмы приведения ИДСтС к ДСтС описаны в разд. 3 и 4. В разд. 5 приводится анализ подходов к моделированию распределений в ИДСтС. В разд. 6 и 7 обсуждаются алгоритмы статистического моделирования, а также совместного аналитического и статистического моделирования. Раздел 8 посвящен алгоритмам нелинейного корреляционного статистического и аналитического моделирования. В приложении 1 рассматриваются алгоритмы численного интегрирования приведенных ИДСтС. Приложения 2 и 3 содержат тестовые примеры.

2 Уравнения интегродифференциальных стохастических систем

Обобщая на случай [4] уравнений, содержащих параметр, рассмотрим ИДСтС, описываемую интегродифференциальным уравнением Ито следующего вида:

$$dX = \left[a(X, \Theta, t) + \int_{t_0}^t a_1(X(\tau), \Theta, \tau, t) d\tau \right] dt + \\ + \left[b(X, \Theta, t) + \int_{t_0}^t b_1(X(\tau), \Theta, \tau, t) d\tau \right] dW_0 +$$

$$+ \int_{R_0^q} \left[c(X, \Theta, t, v) + \int_{t_0}^t c_1(X(\tau), \Theta, \tau, t, v) d\tau \right] dP^0(t, dv) \quad (1)$$

с начальным условием $X(t_0) = X_0$.

В (1) приняты следующие обозначения и допущения:

- ▷ $X = X(t)$ — p -мерный вектор состояния;
- ▷ Θ — p^Θ -мерный вектор случайных параметров, в общем случае зависящих от времени t ;
- ▷ W_0 — r -мерный винеровский процесс интенсивности $\nu_0 = \nu_0(t)$;
- ▷ $\int_{\Delta_t} dP^0(t, A)$ — центрированная пуассоновская мера, удовлетворяющая условию:

$$\int_{\Delta_t} dP^0(t, A) = \int_{\Delta_t} dP(t, A) - \int_{\Delta_t} \nu_P(t, A) dt,$$

где $\int_{\Delta_t} dP(t, A)$ — число скачков пуассоновского процесса в интервале Δ_t ;

- ▷ $\nu_P(t, A)$ — интенсивность пуассоновского процесса $P(t, A)$;
- ▷ v — q -мерный вспомогательный параметр;
- ▷ A — некоторое борелевское множество пространства R_0^q с выколотым началом.

Функции $a = a(X, \Theta, t)$, $a_1 = a_1(X(\tau), \Theta, \tau, t)$, $b = b(X, \Theta, t)$, $b_1 = b_1(X(\tau), \Theta, \tau, t)$, $c = c(X, \Theta, t, v)$, $c_1 = c_1(X(\tau), \Theta, \tau, t, v)$ имеют размерности $p \times 1$, $p \times 1$, $p \times r$, $p \times r$, $p \times 1$, $p \times 1$ и допускают представления вида:

$$\left. \begin{aligned} a_1 &= A(t, \tau)\varphi(X(\tau), \Theta, \tau); \\ b_1 &= B(t, \tau)\psi(X(\tau), \Theta, \tau); \\ c_1 &= C(t, \tau)\chi(X(\tau), \Theta, \tau, v). \end{aligned} \right\} \quad (2)$$

Здесь ядра $A(t, \tau) = [A_{ij}(t, \tau)]$ ($i, j = \overline{1, p}$); $B(t, \tau) = [B_{il}(t, \tau)]$ ($i = \overline{1, p}$, $l = \overline{1, r}$); $C(t, \tau) = [C_{ij}(t, \tau)]$ ($i, j = \overline{1, p}$) имеют размерности $p \times p$, $p \times r$, $p \times p$, удовлетворяют следующим условиям физической реализуемости и асимптотического затухания:

$$A_{ij}(t, \tau) = 0; \quad B_{il}(t, \tau) = 0; \quad C_{ij}(t, \tau) = 0 \quad \forall \tau > t; \quad (3)$$

$$\int_{-\infty}^{\infty} |A_{ij}(t, \tau)| d\tau < \infty; \quad \int_{-\infty}^{\infty} |B_{il}(t, \tau)| d\tau < \infty; \quad \int_{-\infty}^{\infty} |C_{ij}(t, \tau)| d\tau < \infty. \quad (4)$$

Нелинейные в общем случае функции $\varphi = \varphi(X(\tau), \Theta, \tau)$, $\psi = \psi(X(\tau), \Theta, \tau)$, $\chi = \chi(X(\tau), \Theta, \tau, v)$ отражают нелинейные и параметрические свойства ИДСтС, зависят от τ и имеют размерности $p \times 1$, $p \times p$, $p \times 1$ соответственно.

В случае, если ядра удовлетворяют условиям

$$A_{ij}(t, \tau) = \tilde{A}_{ij}(u); \quad B_{il}(t, \tau) = \tilde{B}_{il}(u); \quad C_{ij}(t, \tau) = \tilde{C}_{ij}(u) \quad (u = t - \tau), \quad (5)$$

говорят об ИДСтС со стационарным затуханием.

Важный класс ядер представляют собой сингулярные (вырожденные) ядра, когда имеют место представления:

$$\left. \begin{aligned} A_{ij}(t, \tau) &= A_{ij}^+(t)A_{ij}^-(\tau); \\ B_{il}(t, \tau) &= B_{il}^+(t)B_{il}^-(\tau); \\ C_{ij}(t, \tau) &= C_{ij}^+(t)C_{ij}^-(\tau), \end{aligned} \right\} \quad (6)$$

где $i, l = \overline{1, p}$, $j = \overline{1, r}$.

Замечание 2.1. В случае, когда подынтегральные функции $c(X, \Theta, t, v)$ и $c_1(X(\tau), \Theta, \tau, v)$ в (1) допускают представления:

$$\begin{aligned} c(X, \Theta, t, v) &= b(X, \Theta, t)c'(v); \\ c_1(X(\tau), \Theta, \tau, v) &= b(X(\tau), \Theta, \tau)c'(v), \end{aligned}$$

ИДСтС (1) приводится к виду:

$$\begin{aligned} \dot{X} &= a(X, \Theta, t) + \int_{t_0}^t a_1(X(\tau), \Theta, \tau, t) d\tau + \\ &+ \left[b(X, \Theta, t) + \int_{t_0}^t b_1(X(\tau), \Theta, \tau, t) d\tau \right] V, \quad (7) \end{aligned}$$

если принять

$$V = \dot{W}, \quad W(t) = W_0(t) + \int_{R_0^g} c'(v)P^0(t, dv).$$

Замечание 2.2. Путем введения блочных матриц рассматриваются случаи, обобщающие соответственно представления (2) и (6):

$$\begin{aligned} a_1 &= \sum_{k=1}^N w'_k(t, \tau)\varphi'_k(X(\tau), \Theta, \tau); & b_1 &= \sum_{k=1}^N w''_k(t, \tau)\psi'_k(X(\tau), \Theta, \tau); \\ c_1 &= \sum_{k=1}^N w'''_k(t, \tau)\chi'_k(X(\tau), \Theta, \tau, v); \end{aligned}$$

$$\left. \begin{aligned} a_1 &= \sum_{k=1}^N \xi_k'(t) \varphi_k'(X(\tau), \Theta, \tau); \\ b_1 &= \sum_{k=1}^N \xi_k''(t) \psi_k'(X(\tau), \Theta, \tau); \\ c_1 &= \sum_{k=1}^N \xi_k'''(t, \tau) \chi_k'(X(\tau), \Theta, \tau, v). \end{aligned} \right\} \quad (8)$$

Замечание 2.3. Очевидно, что случай ИДСтС, описываемых по одной части переменных состояния системы стохастическим интегральным уравнением, а второй части — стохастическим дифференциальным уравнением является частным случаем (1) и (7).

Замечание 2.4. В некоторых случаях вводят расширенный вектор состояния $\bar{X} = [X^T \Theta^T]^T$ размерности $\bar{p} = p + p^\theta$, тогда уравнения (1) совместно с уравнением формирующего фильтра для Θ приводятся к уравнениям вида (1) или (7).

Решим сначала вспомогательную задачу приведения ИДСтС (1) к ДСтС, предполагая выполненными условия (2)–(4) и (2), (3) и (5) соответственно. Для ИДСтС (7) задача решена в [2–6].

3 Аппроксимация ядер линейными обыкновенными дифференциальными уравнениями

Рассмотрим ИДСтС (1) при условиях (2)–(4). Так же как в [4], будем считать, что ядра $A(t, \tau)$, $B(t, \tau)$ и $C(t, \tau)$ удовлетворяют следующим линейным операторным уравнениям относительно $D = d/dt$:

$$F^{At} A(t, \tau) = H^{At} \delta(t - \tau);$$

$$F^{Bt} B(t, \tau) = H^{Bt} \delta(t - \tau);$$

$$F^{Ct} C(t, \tau) = H^{Ct} \delta(t - \tau);$$

$$A(t, \tau) = A'(t, \tau)^T (H^{A*\tau})^T; \quad A'(t, \tau)^T (F^{A*\tau})^T = I_h^A \delta(t - \tau);$$

$$B(t, \tau) = B'(t, \tau)^T (H^{B*\tau})^T; \quad B'(t, \tau)^T (F^{B*\tau})^T = I_h^B \delta(t - \tau);$$

$$C(t, \tau) = C'(t, \tau)^T (H^{C*\tau})^T; \quad C'(t, \tau)^T (F^{C*\tau})^T = I_h^C \delta(t - \tau).$$

Здесь

$$\left. \begin{aligned} F^A &= F^A(t, D) = \sum_{l=0}^{n_A} \alpha_l^A(t) D^l; & H^A &= H^A(t, D) = \sum_{l=0}^{m_A} \beta_l^A(t) D^l; \\ F^B &= F^B(t, D) = \sum_{l=0}^{n_B} \alpha_l^B(t) D^l; & H^B &= H^B(t, D) = \sum_{l=0}^{m_B} \beta_l^B(t) D^l; \\ F^C &= F^C(t, D) = \sum_{l=0}^{n_C} \alpha_l^C(t) D^l; & H^C &= H^C(t, D) = \sum_{l=0}^{m_C} \beta_l^C(t) D^l \end{aligned} \right\} \quad (9)$$

есть известные матричные дифференциальные операторы размерности $h_A \times h_A$, $h_B \times h_B$, $h_C \times h_C$ порядков $n_A, m_A, n_B, m_B, n_C, m_C$, ($n_A > m_A, n_B > m_B, n_C > m_C$) соответственно; индекс t у операторов означает, что оператор действует на функцию от t при фиксированном τ ; звездочкой обозначен символ сопряжения оператора; I_h^A, I_h^B и I_h^C — единичные $h \times h$ матрицы.

Введем h^A -, h^B - и h^C -мерные векторы инструментальных переменных посредством соотношений:

$$\begin{aligned} U' &= \int_{t_0}^t A(t, \tau) \varphi(X(\tau), \Theta, \tau) d\tau; & U'' &= \int_{t_0}^t B(t, \tau) \psi(X(\tau), \Theta, \tau) d\tau; \\ U''' &= \int_{t_0}^t C(t, \tau) \chi(X(\tau), \Theta, \tau, v) d\tau. \end{aligned}$$

Тогда, как известно из теории линейных дифференциальных систем [4, 5], переменные U', U'', U''' будут удовлетворять следующим линейным дифференциальным уравнениям:

$$\left. \begin{aligned} F^A(t, D)U' &= H^A(t, D)\varphi(X, \Theta, t); \\ F^B(t, D)U'' &= H^B(t, D)\psi(X, \Theta, t); \\ F^C(t, D)U''' &= H^C(t, D)\chi(X, \Theta, t, v). \end{aligned} \right\} \quad (10)$$

Применяя стандартную технику приведения уравнений (10) к форме Коши [5, 6], придем к искомой ДСтС для вектора $Z = [X^T Z_1'^T Z_1''^T Z_1'''^T]^T$ ($Z_1' = U', Z_1'' = U'', Z_1''' = U'''$):

$$dZ = a_1^z(Z, \Theta, t) dt + b_1^z(Z, \Theta, t) dW_0 + \int_{R_0^g} c_1^z(Z, \Theta, t, v) dP^0(t, dv). \quad (11)$$

В подробной записи для случая $h_A = h_B = h_C = h$, $n_A = n_B = n_C = n$, $m_A = m_B = m_C = m$ функции $a_1^z(Z, \Theta, t)$, $b_1^z(Z, \Theta, t)$, $c_1^z(Z, \Theta, t, v)$ имеют следующий вид:

$$\left. \begin{aligned} a_1^z(Z, \Theta, t) &= \begin{bmatrix} a(X, \Theta, t) + Z_1' \\ a'(t)Z_1' \\ a''(t)Z_1'' \\ a'''(t)Z_1''' \end{bmatrix}; & b_1^z(Z, \Theta, t) &= \begin{bmatrix} b(X, \Theta, t) + Z_1'' \\ b''(t)Z_1'' \\ 0 \\ 0 \end{bmatrix}; \\ c_1^z(Z, \Theta, t, v) &= \begin{bmatrix} c(X, \Theta, t, v) + Z_1''' \\ c'''(t)Z_1''' \\ 0 \\ 0 \end{bmatrix}. \end{aligned} \right\} \quad (12)$$

При условии существования обратных матриц $(\alpha_n^A)^{-1}$, $(\alpha_n^B)^{-1}$, $(\alpha_n^C)^{-1}$ входящие в (12) переменные и коэффициенты допускают следующую запись:

$$\left. \begin{aligned} Z_{j+1}' &= \dot{Z}_j' - q_j' \varphi(X, \Theta, t); & Z_{j+1}'' &= \dot{Z}_j'' - q_j'' \psi(X, \Theta, t); \\ Z_{j+1}''' &= \dot{Z}_j''' - q_j''' \chi(X, \Theta, t, v) \quad (j = \overline{1, (n-1)}); \end{aligned} \right\} \quad (13)$$

$$\left. \begin{aligned} a'(t) &= \begin{bmatrix} I_h & 0 & \cdots & 0 \\ 0 & I_h & \ddots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & I_h \\ -(\alpha_n^A)^{-1} \alpha_0^A - (\alpha_n^A)^{-1} \alpha_1^A & \cdots & \cdots & -(\alpha_n^A)^{-1} \alpha_{n-1}^A \end{bmatrix}; \\ a''(t) &= \begin{bmatrix} I_h & 0 & \cdots & 0 \\ 0 & I_h & \ddots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & I_h \\ -(\alpha_n^B)^{-1} \alpha_0^B - (\alpha_n^B)^{-1} \alpha_1^B & \cdots & \cdots & -(\alpha_n^B)^{-1} \alpha_{n-1}^B \end{bmatrix}; \\ a'''(t) &= \begin{bmatrix} I_h & 0 & \cdots & 0 \\ 0 & I_h & \ddots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & I_h \\ -(\alpha_n^C)^{-1} \alpha_0^C - (\alpha_n^C)^{-1} \alpha_1^C & \cdots & \cdots & -(\alpha_n^C)^{-1} \alpha_{n-1}^C \end{bmatrix}, \end{aligned} \right\} \quad (14)$$

$$\left. \begin{aligned} q'_j &= (\alpha_n^A)^{-1} \left[\beta_{n-j}^A - \sum_{i=0}^{j-1} \sum_{l=0}^{j-i} C_{n-j-l}^{n-j} \alpha_{n-j+i+l}^A q_i'^{(l)} \right], \\ q'_n &= (\alpha_n^A)^{-1} \left[\beta_0^A - \sum_{i=0}^{n-1} \sum_{l=0}^{n-i} \alpha_{i+l}^A q_i'^{(l)} \right], \\ q''_j &= (\alpha_n^B)^{-1} \left[\beta_{n-j}^B - \sum_{i=0}^{j-1} \sum_{l=0}^{j-i} C_{n-j-l}^{n-j} \alpha_{n-j+i+l}^A q_i''^{(l)} \right], \\ q''_n &= (\alpha_n^B)^{-1} \left[\beta_0^B - \sum_{i=0}^{n-1} \sum_{l=0}^{n-i} \alpha_{i+l}^A q_i''^{(l)} \right], \\ q'''_j &= (\alpha_n^C)^{-1} \left[\beta_{n-j}^C - \sum_{i=0}^{j-1} \sum_{l=0}^{j-i} C_{n-j-l}^{n-j} \alpha_{n-j+i+l}^C q_i'''^{(l)} \right], \\ q'''_n &= (\alpha_n^C)^{-1} \left[\beta_0^C - \sum_{i=0}^{n-1} \sum_{l=0}^{n-i} \alpha_{i+l}^C q_i'''^{(l)} \right]. \end{aligned} \right\} \quad (15)$$

Здесь $C_m^n = n!/(m!(n-m)!)$, индекс l означает, что суммирование проводится по всем индексам, исключая l .

Теперь рассмотрим ИДСтС со стационарным затуханием, когда ядра удовлетворяют условиям (5). Тогда вместо выполнения условия (4) достаточно потребовать, чтобы преобразования Лапласа ядер были бы рациональными функциями скалярной переменной s :

$$\begin{aligned} \int_0^{\infty} \tilde{A}(u) e^{-su} du &= (F^A(s))^{-1} H^A(s); \\ \int_0^{\infty} \tilde{B}(u) e^{-su} du &= (F^B(s))^{-1} H^B(s); \\ \int_0^{\infty} \tilde{C}(u) e^{-su} du &= (F^C(s))^{-1} H^C(s), \end{aligned}$$

Здесь порядок матричных полиномов $H^i(s)$ равен m^i , а порядок полиномов $F^i(s)$ равен n^i , причем $n^i \geq m^i$ ($i = A, B, C$).

Таким образом, справедливо следующее утверждение.

Теорема 3.1. Пусть эрдитарные ядра $A(t, \tau)$, $B(t, \tau)$, $C(t, \tau)$ в ИДСтС (1) удовлетворяют условиям (2)–(4) или (10), причем матрицы α_n^A , α_n^B , α_n^C в (9) обратимы, а функции φ, ψ, χ дифференцируемы достаточное число раз. Тогда ИДСтС (1) приводится к ДСтС (11) на основе (12)–(15).

Замечание 3.1. В задачах практики элементами эредитарных ядер $A(t, \tau)$, $B(t, \tau)$, $C(t, \tau)$ являются следующие типовые функции: $e^{-\alpha|u|}$ и $e^{-\alpha|u|}(\cos \omega u + \gamma \sin \omega |u|)$, удовлетворяющие обыкновенным дифференциальным уравнениям первого и второго порядка (см. приложение 2).

4 Аппроксимация вырожденными ядрами

В том случае, когда выполнены условия (2)–(4), а функции φ , ψ , χ не дифференцируемы, целесообразна аппроксимация вырожденными ядрами [4]. В этом случае имеют место следующие соотношения:

$$\left. \begin{aligned} \int_{t_0}^t A(t, \tau) \varphi(X(\tau), \Theta, \tau) d\tau &= A^+ Y'; \\ \int_{t_0}^t B(t, \tau) \psi(X(\tau), \Theta, \tau) d\tau &= B^+ Y''; \\ \int_{t_0}^t C(t, \tau) \chi(X(\tau), \Theta, \tau, v) d\tau &= C^+ Y''' \end{aligned} \right\} \quad (16)$$

$$\left. \begin{aligned} \dot{Y}' &= A^- \varphi; & Y'(t_0) &= 0; \\ \dot{Y}'' &= B^- \psi; & Y''(t_0) &= 0; \\ \dot{Y}''' &= C^- \chi; & Y'''(t_0) &= 0; \end{aligned} \right\} \quad (17)$$

$$Z = [X^T Y'^T Y''^T Y'''^T]^T;$$

$$dZ = a_2^z(Z, \Theta, t) dt + b_2^z(Z, t, \Theta, t) dW_0 + \int_{R_0^d} c_2^z(Z, \Theta, t, v) dP^0(t, dv); \quad (18)$$

$$a_2^z(Z, \Theta, t) = \begin{bmatrix} a(X, \Theta, t) + A^+ Y' \\ A^- \varphi \\ B^- \psi \\ C^- \chi \end{bmatrix}; \quad b_2^z(Z, \Theta, t) = \begin{bmatrix} b(X, \Theta, t) + B^+ Y'' \\ 0 \\ 0 \\ 0 \end{bmatrix};$$

$$c_2^z(Z, \Theta, t, v) = \begin{bmatrix} c(X, \Theta, t, v) + C^+ Y''' \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Таким образом, имеем следующий результат.

Теорема 4.1. Пусть эредитарные ядра $A(t, \tau)$, $B(t, \tau)$, $C(t, \tau)$ в ИДСтС (1) удовлетворяют условиям (3), (4) и (6), а функции φ , ψ , χ не дифференцируемы по переменным состояния. Тогда ИДСтС (1) приводится к ДСтС (18) на основе (16) и (17).

Замечание 4.1. Аналогичная теорема устанавливается для ИДСтС (7).

Замечание 4.2. Аналогично рассматривается случай, когда вырожденные ядра определяются формулами (8).

В приложении 2 рассмотрены примеры приведения линейных и нелинейных ИДСтС к ДСтС методом аппроксимации вырожденными ядрами.

5 Анализ подходов к моделированию распределений в интегродифференциальных стохастических системах

В стохастической информатике для вычисления вероятностей событий, связанных со случайными функциями, достаточно знания многомерных распределений. Различают три подхода к моделированию распределений в ИДСтС.

Первый, общий, основан на прямом численном решении уравнений (1) или (7), а для ИДСтС, приводимых к ДСтС интегрированием, — (11) или (18) с последующей статистической обработкой результатов.

Второй подход для ИДСтС, приводимых к ДСтС, основан на теории марковских процессов и предполагает аналитическое моделирование, т. е. решение детерминированных уравнений в функциональных пространствах (Фоккера–Планка–Колмогорова, Пугачева и др.) для одно- и многомерных распределений.

В практических задачах часто можно рекомендовать комбинированный метод статистического и аналитического моделирования. При этом будем предполагать, что, во-первых, существуют одно- и многомерные плотности процессов, во-вторых, плотности одно- и многомерных распределений можно параметризовать с помощью параметров Ξ_n : вероятностных моментов, квазимоментов, семиинвариантов, коэффициентов ортогональных разложений плотностей, канонических разложений и др. — и, в-третьих, по обобщенной формуле Ито составить для Ξ_n стохастические дифференциальные уравнения. При использовании такого подхода под расширенным вектором состояния следует рассматривать вектор $\bar{Z} = [\bar{Z}^T \Xi_n^T]^T = [Z^T \Theta^T \Xi_n^T]^T$ [5, 6, 8–11].

Наконец, отметим, что всегда приведенные нелинейные уравнения ИДСтС линейны относительно интегральных инструментальных переменных Z' , Z'' , Z''' или Y' , Y'' , Y''' .

6 Алгоритмы параметрического статистического моделирования

Как известно [5, 6, 8–11], численное интегрирование дифференциальных стохастических уравнений имеет некоторые особенности. Дело в том, что все

численные методы интегрирования обыкновенных дифференциальных уравнений, кроме простейшего метода Эйлера, основаны на вычислении приращений искомых функций на каждом шаге путем применения интегральной теоремы о среднем значении. В соответствии с этим правые части уравнений (производные искомых функций) берутся в средних точках интервалов. Различные методы численного интегрирования отличаются один от другого, по существу, только способом приближенного нахождения средних значений правых частей уравнений.

Предположим, что требуется заменить уравнение (11) или (18) разностным уравнением для значений процесса $Z(t)$ в заданном дискретном ряду равноотстоящих точек $\{t_k\}$, $t_k = kh$, где h — интервал между соседними точками t_k , $h = t_{k+1} - t_k$. Принципиально задача решается точно, так как значения марковского процесса $Z(t)$ в точках t_k образуют марковскую случайную последовательность $\{\bar{Z}_k\}$, $\bar{Z}_k = Z(kh)$, а всякая марковская последовательность определяется некоторым стохастическим разностным уравнением. Однако составить это точное разностное уравнение по данному дифференциальному уравнению практически невозможно. Для его составления необходимо прежде всего найти переходное распределение марковского процесса $Z(t)$, определяемого уравнением (11) или (18), а потом по найденному переходному распределению можно составить разностное уравнение для последовательности $\{\bar{Z}_n\}$. Но точное определение переходного распределения процесса $Z(t)$ возможно только в некоторых частных случаях. В общем же случае придется довольствоваться приближенным определением переходного распределения процесса $Z(t)$. В результате по этому переходному распределению можно будет получить только приближенное разностное уравнение. Последнее делает нецелесообразным применение чрезвычайно сложного алгоритма вывода точного разностного уравнения для последовательности $\{\bar{Z}_n\}$. Раз уж все равно приходится довольствоваться лишь приближенным разностным уравнением даже при использовании данного алгоритма, то целесообразно сделать это более простыми способами.

Прежде всего заменим интеграл по переменной v в уравнении (11) соответствующей интегральной суммой. В результате (11) заменится уравнением

$$dZ = a(Z, t) dt + b(Z, t) dW_0 + \sum_{i=1}^N c_i(Z, t) dP_i^0, \quad (19)$$

где $c_i(z, t)$ — p -мерные векторные функции, представляющие собой значения функции $c(z, t, v)$ в некоторых средних точках v_i соответствующих элементов A_i разбиения q -мерного шара достаточно большого радиуса, $v_i \in A_i$ ($i = 1, \dots, N$), а $P_i^0(t)$ — центрированные простые пуассоновские процессы:

$$P_i^0(t) = P^0(\Delta_t, A_i) - \mu(\Delta_t, A_i), \quad i = 1, 2, \dots, N.$$

Интенсивности этих процессов определяются через математическое ожидание $\mu(\Delta_t, A)$ пуассоновской меры $P(\Delta_t, A)$ по формуле:

$$\nu_i(t) = \frac{d\mu(\Delta_t, t, A_i)}{dt}. \quad (20)$$

Простейший способ замены уравнения (19) разностным уравнением состоит в замене всех дифференциалов элементами интегральных сумм:

$$\begin{aligned} Z((n+1)h) - Z(nh) = & a(Z(nh), nh)h + b(Z(nh), nh) [W_0((n+1)h) - \\ & - W_0(nh)] + \sum_{i=1}^N c_i(Z(nh), nh) [P_i^0((n+1)h) - P_i^0(nh)]. \end{aligned} \quad (21)$$

Положим

$$\left. \begin{aligned} \bar{Z}_n &= Z(nh); \quad \varphi_n(\bar{Z}_n) = Z(nh) + a(Z(nh), nh); \\ \psi_{1n}(\bar{Z}_n) &= b(Z(nh), nh); \quad \psi_{in}(\bar{Z}_n) = c_{i-1}(Z(nh), nh); \\ V_{1n} &= W_0((n+1)h) - W_0(nh); \\ V_{in} &= P_{i-1}^0((n+1)h) - P_{i-1}^0(nh), \quad i = 2, \dots, N+1. \end{aligned} \right\} \quad (22)$$

Учитывая (20)–(22), получим разностное стохастическое уравнение

$$\bar{Z}_{n+1} = \varphi_n(\bar{Z}_n) + \sum_{i=1}^{N+1} \psi_{in}(\bar{Z}_n) V_{in}.$$

Вводя блочную матрицу $p \times (r + N)$

$$\psi_n(\bar{Z}_n) = [\psi_{1n}(\bar{Z}_n) \cdots \psi_{N+1,n}(\bar{Z}_n)] \quad (23)$$

и $(r + N)$ -мерный случайный вектор

$$V_n = [V_{1n}^T \ V_{2n} \ \cdots \ V_{N+1,n}^T]^T, \quad (24)$$

можем записать полученное разностное уравнение коротко в виде:

$$\bar{Z}_{n+1} = \omega_n(\bar{Z}_n, V_n) = \varphi_n(\bar{Z}_n) + \psi_n(\bar{Z}_n) V_n, \quad (25)$$

где φ_n и ψ_n определены (22)–(24).

Так как винеровский и пуассоновский процессы являются процессами с независимыми приращениями, то случайные векторы V_n образуют последовательность независимых случайных векторов $\{V_n\}$, причем блоки V_{1n} векторов V_n имеют нормальное распределение $\mathcal{N}(0, \bar{G}_n)$, где

$$\bar{G}_n = \int_{nh}^{(n+1)h} \nu_0(\tau) d\tau \cong \nu_0(nh)h;$$

$\nu_0(t)$ — интенсивность винеровского процесса $W_0(t)$; скалярные блоки V_{in} ($i = 2, \dots, N + 1$) имеют пуассоновские распределения с параметрами

$$\mu_{in} = \int_{nh}^{(n+1)h} \nu_i(\tau) d\tau \cong \nu_i(nh)h.$$

Эти распределения полностью определяют распределения случайных векторов V_n . Ковариационная матрица G_n вектора V_n представляет собой блочно-диагональную матрицу

$$G_n = \begin{bmatrix} \bar{G}_n & 0 & 0 & \cdots & 0 \\ 0 & \mu_{2n} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \mu_{N+1,n} \end{bmatrix}.$$

Уравнение (25) определяет \bar{Y}_{n+1} при данном \bar{Z}_n с точностью до h в детерминированном слагаемом $\varphi_n(\bar{Z}_n)$ и с точностью до $h^{1/2}$ в случайном слагаемом $\psi_n(\bar{Z}_n)V_n$.

Изложенный метод замены стохастического дифференциального уравнения разностными, по существу, не отличается от метода Эйлера численного интегрирования обыкновенных дифференциальных уравнений.

Более точные уравнения (до h^2 и h^3 , $h^{3/2}$ и $h^{5/2}$ соответственно в детерминированном и случайном слагаемом), лежащие в основе метода статистического моделирования (МСМ), приведены в приложении 1.

Таким образом, с точностью до h в детерминированном слагаемом φ_n и $h^{1/2}$ в случайном слагаемом ψ_n в основе алгоритмов параметрического МСМ лежат разностные стохастические уравнения (25). Пользуясь уточненными уравнениями (П.1.34) приложения 1, получим соответствующие уравнения с точностью h^2 и $h^{3/2}$.

7 Алгоритмы параметрического совместного аналитического и статистического моделирования

Как известно [5–8], в задачах корреляционного аналитического моделирования процессов в СтС с аддитивными шумами широкое распространение получил МСЛ. Для СтС с мультипликативными шумами развит МНА.

Обобщением МНА распределений являются различные приближенные методы, основанные на параметризации распределений. Аппроксимация одномерной характеристической функции $g_1(\lambda; t)$ и соответствующей плотности $f_1(z, t)$ известными функциями $g_1^*(\lambda; \Xi_1)$, $f_1^*(z; \Xi_1)$, зависящими от конечномерного

векторного параметра Ξ_1 , сводит задачу приближенного определения одномерного распределения к выводу из уравнения для характеристических функций обыкновенных дифференциальных уравнений, определяющих Ξ как функцию времени. Это относится и к остальным многомерным распределениям. При аппроксимации многомерных распределений целесообразно выбирать последовательности функций $\{g_n^*(\lambda_1, \dots, \lambda_n; \Xi_n)\}$ и $\{f_n^*(z_1, \dots, z_n; \Xi_n)\}$, каждая пара которых находилась бы в такой зависимости от векторного параметра Ξ_n , чтобы при любом n множество параметров, образующих вектор Ξ_n , включало в качестве подмножества множество параметров, образующих вектор Ξ_{n-1} . Тогда при аппроксимации n -мерного распределения придется определять только те координаты вектора Ξ_n , которые не были определены ранее при аппроксимации функций $g_1, f_1, \dots, g_{n-1}, f_{n-1}$.

В зависимости от того, что представляют собой параметры, от которых зависят функции $f_n^*(z_1, \dots, z_n; \Xi_n)$ и $g_n^*(\lambda_1, \dots, \lambda_n; \Xi_n)$, аппроксимирующие неизвестные многомерные плотности $f_n(z_1, \dots, z_n; t_1, \dots, t_n)$ и характеристические функции $g_n(\lambda_1, \dots, \lambda_n; t_1, \dots, t_n)$, используются различные приближенные методы решения уравнений, определяющих многомерные распределения вектора состояния системы Y_t , в частности методы моментов, семиинвариантов, ортогональных разложений и др.

Уравнения метода аналитического моделирования (МAM) для (25), если использовать метод ортогональных разложений (МОР) [5, 6], имеют следующий вид:

$$f_n^*(z_1, \dots, z_n; \Xi_n) = w_n(z_1, \dots, z_n; \bar{m}_n, \bar{K}_n) \times \left\{ 1 + \sum_{\rho=3}^N \sum_{|\nu_1|+\dots+|\nu_n|=\rho} c_{\nu_1, \dots, \nu_n}^{l_1, \dots, l_n} p_{\nu_1, \dots, \nu_n}(z_1, \dots, z_n) \right\} \quad (n = 1, 2, \dots); \quad (26)$$

$$c_{\nu}^{l+1} = M q_{\nu}(\omega_l(Z_l, V_l)) = \left[q_{\nu} \left(\frac{\partial}{\partial \lambda} \right) M \exp \left\{ i \lambda^T \omega_l(Z_l, V_l) \right\} \right]_{\lambda=0}; \quad (27)$$

$$c_{\nu_1, \dots, \nu_n}^{l_1, \dots, l_n+1} = M q_{\nu_1, \dots, \nu_n}(Z_{l_1}, \dots, Z_{l_{n-1}}, \omega_{l_n}(Y_{l_n}, V_{l_n})). \quad (28)$$

Здесь введены обозначения:

$$\bar{m}_n = [m_{l_1}^T \dots m_{l_n}^T]^T; \quad \bar{K}_n = \begin{bmatrix} K_{l_1} & K_{l_1 l_2} & \dots & K_{l_1 l_n} \\ K_{l_1 l_2}^T & K_{l_2} & \dots & K_{l_2 l_n} \\ \vdots & \vdots & \ddots & \vdots \\ K_{l_1 l_n}^T & K_{l_2 l_n}^T & \dots & K_{l_n} \end{bmatrix}, \quad (29)$$

$$c_{\nu_1, \dots, \nu_n}^{l_1, \dots, l_n} = \left[q_{\nu_1, \dots, \nu_n} \left(\frac{\partial}{i \partial \lambda_1} \dots \frac{\partial}{i \partial \lambda_n} \right) g_{l_1, \dots, l_n}(\lambda_1, \dots, \lambda_n) \right]_{\lambda_1 = \dots = \lambda_n = 0},$$

$\{p_{\nu_1, \dots, \nu_n}, q_{\nu_1, \dots, \nu_n}\}$ — биортогональные полиномы, причем

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} w(z_1, \dots, z_n; \bar{m}, \bar{K}) p_{\nu_1, \dots, \nu_n}(z_1, \dots, z_n) \times \\ \times q_{\mu_1, \dots, \mu_n}(z_1, \dots, z_n) dz_1 \cdots dz_n = \delta_{\nu_1 \mu_1} \cdots \delta_{\nu_n \mu_n},$$

где $\nu_l = [\nu_{l_1} \cdots \nu_{l_p}]^T$, $l = 1, 2, \dots$; $\nu_{l_1}, \dots, \nu_{l_n} = 0, 1, \dots, N$; $|\nu_1|, \dots, |\nu_n| = 1, \dots, N - n + 1$; $|\nu_l| = |\nu_{l_1}| + \cdots + |\nu_{l_p}| = n, \dots, N$; $w_n(z_1, \dots, z_n; \bar{m}_n, \bar{K}_n)$ — эталонная плотность, как правило, нормальная:

$$w_n(z_1, \dots, z_n; \bar{m}_n, \bar{K}_n) = \\ = [(2\pi)^n |\bar{K}_n|]^{-1/2} \exp \left\{ -\frac{1}{2} (u_n - \bar{m}_n)^T \bar{K}_n^{-1} (u_n - \bar{m}_n) \right\},$$

где в дополнение к обозначениям (29) принято $u_n = [z_1^T z_2^T \cdots z_n^T]^T$. Отсюда как частный случай вытекают уравнения методов моментов, квазимоментов, семиинвариантов и др., а также параметров структурной параметризации [5–8].

Таким образом, с точностью до h и $h^{1/2}$ соответственно при детерминированном и случайном слагаемом алгоритмы совместного параметрического МСМ и МАМ служат уравнения (25), (26)–(28).

8 Алгоритмы нелинейного корреляционного статистического и аналитического моделирования

Основываясь на результатах разд. 6 и 7, ограничимся уравнениями с точностью до h в детерминированном слагаемом φ_n и с точностью до $h^{1/2}$ в случайном слагаемом ψ_n и заменим нелинейные функции $\varphi_n, \psi_{1n}, \psi_{in}, \psi_n$ в (25) статистически линеаризованными зависимостями:

$$\varphi_n \approx \hat{\varphi}_n = \bar{Z}_n + k_0^a(m_n, K_n)m_n + k_1^a(m_n, K_n)(\bar{Z}_n - m_n) = \\ = \alpha_n(m_n, K_n)\bar{Z}_n + \alpha_{0n}(m_n, K_n); \\ \psi_{1n} \approx \hat{\psi}_{1n} = k_0^b(m_n, K_n)m_n + k_1^b(m_n, K_n)(\bar{Z}_n - m_n); \\ \psi_{in} \approx \hat{\psi}_{in} = k_0^{c_i-1}(m_n, K_n)m_n + k_1^{c_i-1}(m_n, K_n)(\bar{Z}_n - m_n); \\ \psi_n \approx \hat{\psi}_n = k_0^{\psi_n}(m_n, K_n)m_n + k_1^{\psi_n}(m_n, K_n)(\bar{Z}_n - m_n),$$

где $m_n = M\bar{Z}_n = MZ(nh)$, $K_n = M\bar{Z}_n\bar{Z}_n^T = K(nh)$, $k_0^\bullet(m_n, K_n)$ и $k_1^\bullet(m_n, K_n)$ — коэффициенты нормальной статистической линеаризации соответствующих нелинейных функций. Тогда, введя обозначение

$$\hat{\psi}_n(m_n, K_n) = \gamma_{0n}(m_n, K_n) + \sum_{\rho=1}^p \gamma_{\rho n} \bar{Y}_{\rho n},$$

представим искомое уравнение алгоритма МСМ в виде:

$$\begin{aligned} \bar{Z}_{n+1} = & \alpha_n(m_n, K_n) \bar{Z}_n + \alpha_{0n}(m_n, K_n) + \\ & + \left[\gamma_{0n}(m_n, K_n) + \sum_{\rho=1}^p \gamma_{\rho n}(m_n, K_n) \bar{Z}_{\rho n} \right] V_n, \end{aligned} \quad (30)$$

где $\bar{Z}_{\rho n}$ — компонента \bar{Z}_n с номером ρ ($\rho = 1, \dots, p$).

Замечание 8.1. Уравнение (30) является разностным стохастическим уравнением Ито с мультипликативными шумами. Оно подробно изучено в [4, 5].

Применяя к (30) теорию дискретных стохастических систем с мультипликативными шумами [4, 5], получим следующие детерминированные уравнения для m_n , K_n и $K(j, l)$ вместе с соответствующими начальными условиями:

$$m_{n+1} = \alpha(m_n, K_n) m_n + \alpha_{0n}(m_n, K_n); \quad m_1 = MZ_1; \quad (31)$$

$$\begin{aligned} K_{n+1} = & \alpha_n(m_n, K_n) K_n \alpha_n(m_n, K_n)^T + \gamma_{0n}(m_n, K_n) G_n \gamma_{0n}(m_n, K_n)^T + \\ & + \sum_{j=1}^p \left[\gamma_{0n}(m_n, K_n) G_n \gamma_{jn}(m_n, K_n)^T + \gamma_{jn}(m_n, K_n) G_n \gamma_{0n}(m_n, K_n)^T \right] + \\ & + \sum_{j,l=1}^p \sum_{j,l=1}^p \gamma_{jn}(m_n, K_n) G_n \gamma_{ln}(m_n, K_n) [m_{nj} m_{nl} + K_{njl}], \quad K_1 = MZ_1 Z_1^T; \end{aligned} \quad (32)$$

$$\left. \begin{aligned} K(j, l+1) &= K(j, l) \alpha_l(m_l, K_l)^T; \\ K(j, j) &= K_j \quad (l > j); \quad K(j, l) = K(l, j)^T \quad (l < j). \end{aligned} \right\} \quad (33)$$

Таким образом, совокупность стохастического разностного уравнения (30) и детерминированных уравнений (31)–(33) представляет собой систему искомых нелинейных корреляционных уравнений МСМ и МАМ.

Точность совместного корреляционного статистического и аналитического моделирования особенно для разрывных нелинейных функций a, b, c в (19) можно повысить, если вместо (31)–(33) воспользоваться дискретной версией непрерывных уравнений МНА [5, 6].

При практическом использовании разработанных алгоритмов следует учесть, что в общем случае приведенные уравнения ИДСтС обладают следующими особенностями:

- (1) уравнения линейны относительно инструментальных переменных Z', Z'', Z''' или Y', Y'', Y''' ;
- (2) диффузионная матрица имеет блочную структуру и всегда вырождена, т. е. уравнения для инструментальных переменных не содержат шумов.

В приложениях 2 и 3 приведены тестовые примеры, иллюстрирующие разработанное математическое обеспечение.

9 Заключение

Для ИДСтС разработаны приближенные аналитические и численные методы, алгоритмы и инструментальное программное обеспечение IDStS-2.0 в среде MATLAB, а также тестовые примеры. Особое внимание уделено ИДСтС, приводимым к ДСтС. Результаты допускают обобщение на случай использования канонических разложений марковских стохастических процессов [7, 12, 13].

Приложение 1

1. Следуя [8], заменим (19) соответствующим интегральным уравнением:

$$\Delta Z_n = \int_{nh}^{(n+1)h} a(Z_\tau, \tau) d\tau + \int_{nh}^{(n+1)h} b(Z_\tau, \tau) dW_0(\tau) + \sum_{i=1}^N \int_{nh}^{(n+1)h} c_i(Z_\tau, \tau) dP_i^0(d\tau). \quad (\text{П1})$$

С целью приближенного вычисления интегралов определим Z_τ путем линейной интерполяции случайной функции $Z(t)$ на интервале $(nh, (n+1)h)$. Тогда, полагая по-прежнему $\Delta Z_n = Z((n+1)h) - Z(nh)$, будем иметь:

$$a(YZ_\tau, \tau) \cong a\left(\bar{Z}_n + \frac{\tau - nh}{h} \Delta Z_n, \tau\right); \quad b(Z_\tau, \tau) \cong b\left(\bar{Z}_n + \frac{\tau - nh}{h} \Delta Z_n, \tau\right);$$

$$c_i(Z_\tau, \tau) \cong c_i\left(\bar{Z}_n + \frac{\tau - nh}{h} \Delta Z_n, \tau\right), \quad i = 1, \dots, N.$$

Для вычисления правых частей в этих формулах применим обобщенную формулу Ито [2, 3]:

$$\begin{aligned} \varphi(Z + dZ, t + dt) = & \varphi(Z, t) + \{\varphi_t(Z, t) + \varphi_z(Z, t)^T a(Z, t) + \frac{1}{2} \varphi_{zz}(Z, t) : \sigma(Z, t) + \\ & + \sum_{i=1}^N [\varphi(Z + c_i(Z, t), t) - \varphi(Z, t) - \varphi_z(Z, t)^T c_i(Z, t)] \nu_i(t)\} dt + \\ & + \varphi_z(Z, t)^T b(X, t) dW_0 + \sum_{i=1}^N [\varphi(Z + c_i(Z, t), t) - \varphi(Z, t)] dP_i^0, \quad (\text{П2}) \end{aligned}$$

где $\varphi_i(z, t)$ — частная производная функции $\varphi(yz, t)$ (возможно, векторной) по времени t ; $\varphi_z(z, t)$ — матрица, строки которой представляют собой частные производные матрицы-строки $\varphi(z, t)^T$ по компонентам вектора z ; $\varphi_{zz}(z, t) : \sigma(z, t)$ — вектор, компонентами которого служат следы произведений матриц вторых производных соответствующих компонент векторной функции $\varphi(z, t)$ по компонентам вектора z на матрицу $\sigma(z, t)$:

$$[\varphi_{zz}(z, t) : \sigma(z, t)]_k = \text{tr}[\varphi_{kzz}(z, t)\sigma(y, t)] ;$$

$$\sigma(z, t) = b(z, t)\nu_0(t)b(z, t)^T. \quad (\text{П3})$$

Формулу (П2) удобно преобразовать так, чтобы в нее входили непосредственно пуассоновские процессы $P_i(t)$, а не центрированные процессы $P_i^0(t)$. Имея в виду, что

$$P_i^0(t) = P_i(t) - \int_0^t \nu_i(\tau) d\tau, \quad (\text{П4})$$

можем переписать формулу (П2) в виде:

$$\begin{aligned} \varphi(Z + dZ, t + dt) = & \varphi(Z, t) + \\ & + \left\{ \varphi_t(Z, t) + \varphi_z(Z, t)^T \left[a(Z, t) - \sum_{i=1}^N c_i(Z, t)\nu_i(t) \right] + \frac{1}{2} \varphi_{zz}(Z, t) : \sigma(Z, t) \right\} dt + \\ & + \varphi_z(Z, t)^T b(Z, t) dW_0 + \sum_{i=1}^N [\varphi(Z + c_i(Z, t), t) - \varphi(Z, t)] dP_i. \quad (\text{П5}) \end{aligned}$$

Для распространения формул (П2) и (П5) на матричные функции φ необходимо видоизменить запись некоторых ее членов, чтобы в рамках алгебры матриц выражения $\varphi_z(z, t)$ имели смысл и для матричной функции φ . Имея в виду, что в случае скалярной или векторной функции φ

$$\varphi_z(z, t)^T u = \sum_{i=1}^n \frac{\partial \varphi(z, t)}{\partial z_i} u_i = u^T \frac{\partial}{\partial z} \varphi(z, t)$$

для любого p -мерного вектора u , можем заменить слагаемые $\varphi_z(Z, t)^T a(Z, t)$, $\varphi_z(Z, t)^T b(Z, t) dW_0$ и $\varphi_z(Z, t)^T c_i(Z, t)$ слагаемыми $a(Z, t)^T (\partial/\partial z)^T \varphi(Z, t)$, $dW_0^T b(Z, t) (\partial/\partial z) \varphi(Z, t)$ и $c_i(Z, t)^T (\partial/\partial z) \varphi(Z, t)$ соответственно. Поэтому формулы (П2) и (П5) будут справедливы и для матричных функций φ , так как $u^T (\partial/\partial y)$ представляет собой скалярный дифференциальный оператор для любого p -мерного вектора u и его применение к векторной или матричной функции φ означает его применение ко всем компонентам вектора φ или ко всем элементам матрицы φ . Величина $\varphi_{zz}(Z, t) : \sigma(Z, t)$ в случае матричной функции φ представляет собой матрицу, элементами которой служат следы произведений на матрицу $\sigma(z, t)$ матриц вторых производных соответствующих элементов матрицы $\varphi(z, t)$ по компонентам вектора y на матрицу $\sigma(z, t)$:

$$[\varphi_{zz}(z, t) : \sigma(z, t)]_{kl} = \text{tr}[\varphi_{klzz}(z, t)\sigma(z, t)].$$

Используя формулу (П5) и ее модификацию, пригодную для матричной функции φ , а также учитывая (П3) и то, что согласно (1) и (П4)

$$\begin{aligned} \frac{\tau - nh}{h} \Delta Z_n &= \frac{\tau - nh}{h} a(\bar{Z}_n, nh)h + \frac{\tau - nh}{h} b(\bar{Z}_n, nh) \Delta W_n + \\ &+ \sum_{i=1}^N \frac{\tau - nh}{h} c_i(\bar{Z}_n, nh) (\Delta P_{in} - \nu_{in}h), \quad (\text{П6}) \end{aligned}$$

где, как и раньше, $\Delta W_n = W_0((n+1)h) - W_0(nh)$, $\Delta P_{in} = P_i((n+1)h) - P_i(nh)$, будем иметь с точностью до малых высшего порядка относительно h следующие равенства:

$$\begin{aligned} a\left(\bar{Z}_n + \frac{\tau - nh}{h} \Delta Z_n, \tau\right) &= \\ &= a + a_t(\tau - nh) + \frac{\tau - nh}{h} a_z^T \left(a - \sum_{i=1}^N c_i \nu_{in}\right) h + \frac{1}{2} \left(\frac{\tau - nh}{h}\right)^2 (a_{zz} : \sigma)h + \\ &+ \frac{\tau - nh}{h} a_z^T b \Delta W_n + \sum_{i=1}^N \left[a\left(\bar{Z}_n + \frac{\tau - nh}{h} c_i, nh\right) - a \right] \Delta P_{in}; \quad (\text{П7}) \end{aligned}$$

$$\begin{aligned} b\left(\bar{Z}_n + \frac{\tau - nh}{h} \Delta Z_n, \tau\right) &= \\ &= b + b_t(\tau - nh) + \frac{\tau - nh}{h} \left(a^T - \sum_{i=1}^N c_i^T \nu_{in}\right) \frac{\partial}{\partial z} b h + \frac{1}{2} \left(\frac{\tau - nh}{h}\right)^2 (b_{zz} : \sigma)h + \\ &+ \frac{\tau - nh}{h} \Delta W_n^T b^T \frac{\partial}{\partial z} b + \sum_{i=1}^N \left[b\left(\bar{Z}_n + \frac{\tau - nh}{h} c_i, nh\right) - b \right] \Delta P_{in}; \quad (\text{П8}) \end{aligned}$$

$$\begin{aligned} c_i\left(\bar{Z}_n + \frac{\tau - nh}{h} \Delta Z_n, \tau\right) &= \\ &= c_i + c_{it}(\tau - nh) + \frac{\tau - nh}{h} c_{iz}^T \left(a - \sum_{j=1}^N c_j \nu_{jn}\right) h + \frac{1}{2} \left(\frac{\tau - nh}{h}\right)^2 (c_{izz} : \sigma)h + \\ &+ \frac{\tau - nh}{h} c_{iz}^T b \Delta W_n + \sum_{j=1}^N \left[c_i\left(\bar{Z}_n + \frac{\tau - nh}{h} c_j, nh\right) - c_i \right] \Delta P_{in}. \quad (\text{П9}) \end{aligned}$$

Здесь аргументы \bar{Z}_n и nh у всех функций, зависящих от них, для краткости опущены.

2. Для повышения точности вычислений, в частности ковариационной матрицы нормально распределенного случайного вектора $\Delta W_n = W_0((n+1)h) - W_0(nh)$ и параметров пуассоновских распределений случайных величин $\Delta P_{in} = P_i((n+1)h) - P_i(nh)$, можно взять значения интенсивностей $\nu_0(t)$ процесса $W_0(t)$ в формуле (П3)

и $\nu_i(t)$ пуассоновских потоков, порождающих процессы $P_i(t)$, в средней точке $nh + h/2$ интервала $(nh, (n+1)h)$:

$$\left. \begin{aligned} \sigma(\bar{Z}_n, nh) &= b(\bar{Z}_n, nh)\nu_0\left(nh + \frac{h}{2}\right)b(\bar{Z}_n, nh)^T; \\ \nu_{in} &= \nu_i\left(nh + \frac{h}{2}\right), \quad i = 1, \dots, N. \end{aligned} \right\} \quad (\text{П10})$$

Для облегчения дальнейших вычислений найдем приращения функций a , b и c_i в суммах уравнения (П6) путем линейной интерполяции на малом интервале $(nh, (n+1)h)$:

$$\left. \begin{aligned} a\left(\bar{Z}_n + \frac{\tau - nh}{h}c_i, nh\right) - a &\cong \frac{\tau - nh}{h}\Delta_i a_n; \\ b\left(\bar{Z}_n + \frac{\tau - nh}{h}c_i, nh\right) - b &\cong \frac{\tau - nh}{h}\Delta_i b_n; \\ c_i\left(\bar{Y}_n + \frac{\tau - nh}{h}c_j, nh\right) - c_i &\cong \frac{\tau - nh}{h}\Delta_j c_{in}; \end{aligned} \right\} \quad (\text{П11})$$

$$\begin{aligned} \Delta_i a_n &= a(\bar{Z}_n + c_i, nh) - a(\bar{Z}_n, nh); \\ \Delta_i b_n &= b(\bar{Z}_n + c_i, nh) - b(\bar{Z}_n, nh); \\ \Delta_j c_{in} &= c_i(\bar{Y}_n + c_j, nh) - c_i(\bar{Z}_n, nh). \end{aligned}$$

Используя формулы (П7)–(П11), найдем приближенные выражения интегралов в (П1):

$$\begin{aligned} \int_{nh}^{(n+1)h} a(Z_\tau, \tau) d\tau &= ah + \frac{1}{2} \left[a_t + a_z^T \left(a - \sum_{i=1}^N c_i \nu_{in} \right) + \frac{1}{3} a_{zz} : \sigma \right] h^2 + \\ &\quad + \frac{1}{2} \left(a_z^T b \Delta W_n + \sum_{i=1}^N \Delta_i a_n \Delta P_{in} \right) h; \\ \int_{nh}^{(n+1)h} b(Z_\tau, \tau) dW_0(\tau) &= b \Delta W_n + \left\{ \left[b_t + \left(a^T - \sum_{i=1}^N c_i^T \nu_{in} \right) \frac{\partial}{\partial z} b \right] h + \Delta W_n^T b^T \frac{\partial}{\partial z} b + \right. \\ &\quad \left. + \sum_{i=1}^N \Delta_i b_n \Delta P_{in} \right\} \int_{nh}^{(n+1)h} \frac{\tau - nh}{h} dW_0(\tau) + \frac{1}{2} (b_{zz} : \sigma) \int_{nh}^{(n+1)h} \left(\frac{\tau - nh}{h} \right)^2 dW_0(\tau); \end{aligned}$$

$$\begin{aligned}
 & \int_{nh}^{(n+1)h} c_i(Z_\tau, \tau) dP_i^0(\tau) = c_i \Delta P_{in} + \left\{ \left[c_{it} + c_{iz}^T \left(a - \sum_{j=1}^N c_j \nu_{jn} \right) \right] h + c_{iz}^T b \Delta W_n + \right. \\
 & \left. + \sum_{j=1}^N \Delta_j c_{in} \Delta P_{jn} \right\} \int_{nh}^{(n+1)h} \frac{\tau - nh}{h} dP_i^0(\tau) + \frac{1}{2} (c_{izz} : \sigma) \int_{nh}^{(n+1)h} \left(\frac{\tau - nh}{h} \right)^2 dP_i^0(\tau); \\
 & \int_{nh}^{(n+1)h} c_i(Z_\tau, \tau) \nu_i(\tau) d\tau = \left\{ c_i h + \frac{1}{2} \left[c_{it} + c_{iz}^T \left(a - \sum_{j=1}^N c_j \nu_{jn} \right) + \frac{1}{3} c_{izz} : \sigma \right] h^2 + \right. \\
 & \left. + \frac{1}{2} \left(c_{iz}^T b \Delta W_n + \sum_{j=1}^N \Delta_j c_{in} \Delta P_{jn} \right) h \right\} \nu_{in}.
 \end{aligned}$$

Далее, подставив полученные приближенные выражения интегралов в (П1) и приняв во внимание, что $\Delta Z_n = Z((n+1)h) - Z_i(nh) = \bar{Z}_{n+1} - \bar{Z}_n$, придем к разностному уравнению

$$\bar{Z}_{n+1} = \varphi_n(\bar{Z}_n) + \sum_{i=1}^{N+1} [\psi_{in}(\bar{Z}_n) V_{in} + \psi'_{in}(\bar{Z}_n, V_{in}^{(1)}) V'_{in} + \psi''_{in}(\bar{Z}_n) V''_{in}]. \quad (\text{П12})$$

Здесь введены следующие обозначения:

$$\begin{aligned}
 \varphi_n(\bar{Z}_n) = & \bar{Z}_n + \left[a(\bar{Z}_n, nh) - \sum_{j=1}^N c_j(\bar{Z}_n, nh) \nu_{jn} \right] h + \\
 & + \frac{1}{2} \left\{ a_t(\bar{Z}_n, nh) - \sum_{j=1}^N c_{jt}(\bar{Z}_n, nh) \nu_{jn} + \right. \\
 & + \left[a_x(\bar{Z}_n, nh)^T - \sum_{j=1}^N c_{jz}(\bar{Z}_n, nh)^T \nu_{jn} \right] \left[a(\bar{Z}_n, nh) - \sum_{j=1}^N c_j(\bar{Z}_n, nh) \nu_{jn} \right] + \\
 & \left. + \frac{1}{3} \left[a_{zz}(\bar{Z}_n, nh) - \sum_{j=1}^N c_{jzz}(\bar{Z}_n, nh) \nu_{jn} \right] : \sigma(\bar{Z}_n, nh) \right\} h^2. \quad (\text{П13})
 \end{aligned}$$

$$\psi_{1n}(\bar{Z}_n) = b(\bar{Z}_n, nh) + \frac{1}{2} \left[a_z(\bar{Z}_n, nh)^T - \sum_{j=1}^N c_j(\bar{Z}_n, nh)^T \nu_{jn} \right] b(\bar{Z}_n, nh); \quad (\text{П14})$$

$$\psi_{in}(\bar{Z}_n) = c_{i-1}(\bar{Z}_n, nh) + \left[\Delta_{i-1} a_n - \sum_{j=1}^N \Delta_{i-1} c_{jn} \nu_{jn} \right] h; \quad (\text{П15})$$

$$\psi'_{1n}(\bar{Z}_n, V_n^{(1)}) = \left\{ b(\bar{Z}_n, nh) + \left[a(\bar{Z}_n, nh)^T - \sum_{j=1}^N c_j(\bar{Z}_n, hn)^T \nu_{jn} \right] \frac{\partial}{\partial z} b(\bar{Z}_n, nh) + \right. \\ \left. + V_{1n}^T b(\bar{Z}_n, nh)^T \frac{\partial}{\partial z} b(\bar{Z}_n, nh) \right\} h + \sum_{j=1}^N \Delta_j b_n V_{j+1,n}; \quad (\text{II16})$$

$$\psi'_{in}(\bar{Z}_n, V_n^{(1)}) = \left\{ c_{i-1,t}(\bar{Z}_n, nh) + c_{i-1,y}(\bar{Z}_n, nh)^T \left[a(\bar{Z}_n, nh) - \sum_{j=1}^N c_j(\bar{Z}_n, hn) \nu_{jn} \right] \right\} h + \\ + c_{i-1,z}(\bar{Z}_n, nh)^T b(\bar{Z}_n, nh) V_{1,n} + \sum_{j=1}^N \Delta_j c_{j-1,n} V_{j+1,n}; \quad (\text{II17})$$

$$\psi''_{1n}(\bar{Z}_n) = \frac{1}{2} [b_{zz}(\bar{Z}_n, nh) : \sigma(\bar{Z}_n, nh)]; \quad (\text{II18})$$

$$\psi''_{in}(\bar{Z}_n) = \frac{1}{2} [c_{i-1,zz}(\bar{Z}_n, nh) : \sigma(\bar{Z}_n, nh)]; \quad (\text{II19})$$

$$V_{1n} = \Delta W_n = W_0((n+1)h) - W_0(nh); \quad (\text{II20})$$

$$V_{in} = \Delta P_{i-1,n} = P_{i-1}((n+1)h) - P_{i-1}(nh), \quad i = 2, \dots, N; \quad (\text{II21})$$

$$V'_{1n} = \int_{nh}^{(n+1)h} \frac{\tau - nh}{h} dW_0(\tau); \quad V''_{1n} = \int_{nh}^{(n+1)h} \left(\frac{\tau - nh}{h} \right)^2 dW_0(\tau); \quad (\text{II22})$$

$$V'_{in} = \int_{nh}^{(n+1)h} \frac{\tau - nh}{h} dP_{i-1}^0(\tau); \quad (\text{II23})$$

$$V''_{in} = \int_{nh}^{(n+1)h} \left(\frac{\tau - nh}{h} \right)^2 dP_{i-1}^0(\tau), \quad i = 2, \dots, N+1. \quad (\text{II24})$$

Обозначим $V_n^{(1)} = [V_{1n}^T \ V_{2n} \ \dots \ V_{n+1}]^T$ и введем блочную матрицу

$$\psi_n(\bar{Z}_n, V_n^{(1)}) = [\psi_{1n}(\bar{Z}_n) \ \psi'_{1n}(\bar{Z}_n, V_n^{(1)}) \ \psi''_{1n}(\bar{Z}_n) \ \dots \\ \dots \ \psi_{N+1,n}(\bar{Z}_n) \ \psi'_{N+1,n}(\bar{Z}_n, V_n^{(1)}) \ \psi''_{N+1,n}(\bar{Z}_n)]$$

и блочный случайный вектор

$$V_n = [V_{1n}^T \ V_{1n}'^T \ V_{1n}''^T \ V_{2n}^T \ V_{2n}'^T \ V_{2n}''^T \ \dots \ V_{N+1,n} \ V_{N+1,n}' \ V_{N+1,n}'']^T. \quad (\text{II25})$$

Тогда можем коротко записать уравнение (II12) в виде:

$$\bar{Z}_{n+1} = \varphi_n(\bar{Z}_n) + \psi_n(\bar{Z}_n, V_n^{(1)}) V_n. \quad (\text{II26})$$

Входящие в уравнения (II26) величины определены формулами (II13)–(II25).

3. Найдем распределение случайного вектора V_n . Ясно, что математические ожидания случайных величин V'_{1n}, V''_{1n} равны нулю и что случайный вектор $[V_{1n}^T V'_{1n} V''_{1n}^T]$ имеет нормальное распределение, а скалярные случайные величины $V_{2n}, \dots, V_{N+1,n}$ — пуассоновские распределения с параметрами:

$$\mu_{in} = \int_{nh}^{(n+1)h} \nu_{i-1}(\tau) d\tau = \nu_{i-1} \left(nh + \frac{1}{2} h \right) h, \quad i = 2, \dots, N + 1.$$

Очевидно также, что тройки случайных величин $V_{1n}, V'_{1n}, V''_{1n}, \dots, V_{N+1,n}, V'_{N+1,n}, V''_{N+1,n}$ независимы в силу независимости процессов $W_0(t), P_1(t), \dots, P_N(t)$ и что при различных n величины V_n независимы. Однако при любых данных i, n величины $V_{in}, V'_{in}, V''_{in}$ зависимы.

Для полного определения распределения случайного вектора V_n в (П26) достаточно найти ковариационную матрицу нормально распределенного случайного вектора $[V_{in}^T V'_{in} V''_{in}^T]$. Пользуясь известными формулами ковариационных и взаимных ковариационных матриц стохастических интегралов, находим блоки ковариационной матрицы K_{1n} случайного вектора $[V_{in}^T V'_{in} V''_{in}^T]$:

$$\begin{aligned} K_{1n,11} &= MV_{1n}V_{1n}^T = \int_{nh}^{(n+1)h} \nu_0(\tau) d\tau \cong \nu_0 \left(nh + \frac{h}{2} \right) h; \\ K_{1n,12} &= MV_{1n}V'_{1n}{}^T = \int_{nh}^{(n+1)h} \frac{\tau - nh}{h} \nu_0(\tau) d\tau \cong \frac{1}{2} \nu_0 \left(nh + \frac{h}{2} \right) h; \\ K_{1n,13} &= MV_{1n}V''_{1n}{}^T = \int_{nh}^{(n+1)h} \left(\frac{\tau - nh}{h} \right)^2 \nu_0(\tau) d\tau \cong \frac{1}{3} \nu_0 \left(nh + \frac{h}{2} \right) h; \\ K_{1n,21} &= K_{1n,12}; \quad K_{1n,31} = K_{1n,13}; \\ K_{1n,23} &= MV'_{1n}V''_{1n}{}^T = \int_{nh}^{(n+1)h} \left(\frac{\tau - nh}{h} \right)^3 \nu_0(\tau) d\tau \cong \frac{1}{4} \nu_0 \left(nh + \frac{h}{2} \right) h; \\ K_{1n,31} &= K_{1n,13}; \quad K_{1n,32} = K_{1n,23}; \\ K_{1n,33} &= MV''_{1n}V''_{1n}{}^T = \int_{nh}^{(n+1)h} \left(\frac{\tau - nh}{h} \right)^4 \nu_0(\tau) d\tau \cong \frac{1}{5} \nu_0 \left(nh + \frac{h}{2} \right) h. \end{aligned}$$

Практически целесообразно аппроксимировать стохастические интегралы от неслучайных функций в (П23), определяющие величины V'_{in} ($i = 2, \dots, N + 1$), с помощью аналога интегральной теоремы о среднем для стохастических интегралов:

$$\begin{aligned}
 V'_{in} &= \int_{nh}^{(n+1)h} \frac{\tau - nh}{h} dP_{i-1}^0(\tau) \cong \\
 &\cong \Delta P_{i-1,n} \int_{nh}^{(n+1)h} \frac{\tau - nh}{h} \nu_{i-1}(\tau) d\tau \bigg/ \int_{nh}^{(n+1)h} \nu_{i-1}(\tau) d\tau \cong \frac{1}{2} \Delta P_{i-1,n} = \frac{1}{2} V_{in}; \\
 V''_{in} &= \int_{nh}^{(n+1)h} \left(\frac{\tau - nh}{h} \right)^2 dP_{i-1}^0(\tau) \cong \\
 &\cong \Delta P_{i-1,n} \int_{nh}^{(n+1)h} \left(\frac{\tau - nh}{h} \right)^2 \nu_{i-1}(\tau) d\tau \bigg/ \int_{nh}^{(n+1)h} \nu_{i-1}(\tau) d\tau \cong \\
 &\cong \frac{1}{3} \Delta P_{i-1,n} = \frac{1}{3} V_{in}, \quad i = 2, \dots, N + 1.
 \end{aligned}$$

4. При статистическом моделировании системы с помощью уравнения (П26) также не представляет трудностей моделировать случайные величины, распределенные по нормальному и пуассоновским законам. Легко видеть, что правая часть разностного уравнения (П26) определена с точностью до h^2 в детерминированном (при данном \bar{Z}_n) слагаемом $\varphi_n(\bar{Z}_n)$ и с точностью до $h^{3/2}$ в случайном слагаемом $\psi_n(\bar{Z}_n, V_n^{(1)})V_n$.

При выводе уравнения (П26) были допущены две небольшие неточности. Во-первых, при замене \bar{Z}_τ величиной $\bar{Z}_n + (\tau - nh)\Delta Y_n/h$ случайные функции $b(Z_\tau, \tau)$ и $c_i(Z_\tau, \tau)$, независимые от $dW_0(\tau)$ и $dP_i^0(\tau)$ в силу конструкции интеграла Ито, были заменены неслучайными функциями, зависящими от случайного параметра ΔZ_n , который в свою очередь зависит от значений $dW_0(\tau)$ и $dP_i^0(\tau)$ в интервале $(nh, (n + 1)h)$. Во-вторых, если $c_i(z, t) \neq 0$ хотя бы при одном i , реализации случайного процесса $Z(t)$ имеют разрывы первого рода в случайных точках, несмотря на его среднеквадратическую непрерывность. Поэтому линейную интерполяцию данного процесса, строго говоря, проводить нельзя. Первую из этих неточностей можно устранить двумя способами. Первый состоит в замене интерполяции процесса $Z(t)$ экстраполяцией, что равноценно замене ΔZ_n в получаемом выражении для Y_τ величиной $\Delta \bar{Z}_{n-1}$. Однако это приведет к появлению в правой части разностного уравнения величин \bar{Z}_{n-1} и $V_{n-1}^{(1)}$, т. е. к замене уравнения первого порядка разностным уравнением второго порядка. Второй способ состоит в отказе от интерполяции процесса $Z(t)$ на интервале $(nh, (n + 1)h)$ и непосредственном выражении приращений функций $a(Z_\tau, \tau)$, $b(Z_\tau, \tau)$, $c_i(Y_\tau, \tau)$ на малом интервале $(nh, (n + 1)h)$ по обобщенной формуле Ито с заменой в ней дифференциалов приращениями. При этом способе устраняется и вторая допущенная неточность. Но полученное таким путем разностное уравнение будет более сложным. В него войдут случайные величины, представляющие собой двойные интегралы по компонентам винеровского процесса $W(t) = W_0(t)$ и по пуассоновским процессам:

$$\begin{aligned} & \int_{nh}^{(n+1)h} \int_{nh}^{\tau} dW_j(\sigma) dW_j(\tau); & \int_{nh}^{(n+1)h} \int_{nh}^{\tau} dP_i(\sigma) dP_j^0(\tau); \\ & \int_{nh}^{(n+1)h} \int_{nh}^{\tau} dP_i^0(\sigma) dW_j(\tau); & \int_{nh}^{(n+1)h} \int_{nh}^{\tau} dP_j^0(\sigma) dW_i(\tau). \end{aligned}$$

Распределения этих случайных величин найти чрезвычайно сложно, и только первые два из них легко вычисляются при $j = i$:

$$\begin{aligned} & \int_{nh}^{(n+1)h} \int_{nh}^{\tau} dW_i(\sigma) dW_i(\tau) = \frac{[\Delta W_{in}]^2 - \nu_{ii}(nh + h/2)h}{2}; \\ & \int_{nh}^{(n+1)h} \int_{nh}^{\tau} dP_i^0(\sigma) dP_i^0(\tau) = \frac{[\Delta P_{in}]^2 - \Delta P_{in}}{2}. \end{aligned}$$

Что касается второй неточности, то она не может существенно повлиять на результат, так как вероятность появления скачка пуассоновского процесса в достаточно малом интервале $(nh, (n+1)h)$ ничтожно мала.

Точность аппроксимации дифференциального стохастического уравнения разностным можно повышать и дальше. В частности, в одном из способов сначала достаточно выразить $a(Z_\tau, \tau)$, $b(Z_\tau, \tau)$, $c_i(Z_\tau, \tau)$ на интервале $(nh, (n+1)h)$ интегральной формулой Ито, соответствующей дифференциальной формуле:

$$\begin{aligned} a(Z_\tau, \tau) = & a(\bar{Z}_n, nh) + \int_{nh}^{\tau} \left\{ a_t(Z_s, s) + \right. \\ & \left. + a_x(Z_s, s)^T \left[a(Z_s, s) - \sum_{i=1}^N c_i(Z_s, \nu_i(s)) + \frac{1}{2} a_{zz}(Z_s, s) : \sigma(Z_s, s) \right] \right\} ds + \\ & + \int_{nh}^{\tau} a_z(Z_s, s)^T b(Z_s, s) dW_0(s) + \sum_{i=1}^N \int_{nh}^{\tau} [a(Z_s + c_i(Z_s, s), s) - a(Z_s, s)] dP_i^0(s). \end{aligned}$$

При этом формулы для $b(Z_\tau, \tau)$, $c_i(Z_\tau, \tau)$ аналогичны. Затем применить к полученным интегралам тот же прием, который использовался для вычисления интегралов в (П1). В результате получим правую часть разностного уравнения с точностью до h^3 в детерминированном (при данном \bar{Z}_n) слагаемом и $h^{5/2}$ в случайном слагаемом. Процесс уточнения разностного уравнения, соответствующего данному стохастическому дифференциальному уравнению, можно продолжать неограниченно. Однако каждое новое уточнение требует существования производных функций a , b , c_i все более высоких порядков.

5. Полученные разностные уравнения можно использовать как при теоретических исследованиях, так и для численного интегрирования дифференциальных стохастических уравнений. При этом нужно знать распределение всех случайных величин, входящих в разностные уравнения. В данном случае разностные уравнения будут представлять собой так называемую сильную аппроксимацию стохастических дифференциальных уравнений [9–11]. При численном интегрировании такая аппроксимация нужна, когда требуется получать реализации процесса $Z(t)$. Однако часто нет нужды в получении реализаций процесса, а достаточно иметь лишь оценки моментов или математических ожиданий каких-либо функций от значения случайного процесса $Z(t)$ в определенный момент. В таких случаях можно отказаться от использования точных распределений, входящих в разностные уравнения случайных величин, а заменить их какими-нибудь более простыми распределениями с теми же моментными характеристиками. Например, нормально распределенную скалярную величину с нулевым математическим ожиданием и дисперсией D можно заменить дискретной случайной величиной, принимающей два значения $\pm\sqrt{D}$ с вероятностями $1/2$. При замене случайных величин на величины с более простыми распределениями разностное уравнение будет представлять собой слабую аппроксимацию дифференциального стохастического уравнения [9–11].

Приложение 2

Тестовые примеры приведения интегродифференциальных стохастических систем к дифференциальным

1. Рассмотрим скалярную нелинейную ИДСтС

$$\dot{X} = a_0 + a_1 X + \int_0^t e^{-\alpha|t-\tau|} \varphi(X(\tau)) d\tau + b\dot{W}, \quad X_0 = 0,$$

где a_0 , a_1 , α и b — константы. Полагая

$$\int_0^t e^{-\alpha|t-\tau|} \varphi(X(\tau)) d\tau = Y,$$

приведем ИДСтС к ДСтС:

$$\begin{aligned} \dot{X} &= a_0 + a_1 X + \frac{1}{\alpha} Y + b\dot{W}; & \dot{Y} &= -\alpha Y + \varphi(X); \\ X(0) &= Y(0) = 0. \end{aligned}$$

2. Рассмотрим скалярную линейную ИДСтС

$$a\ddot{X} + b \int_0^t \dot{X}(\tau) e^{-\beta|t-\tau|} d\tau + c \int_0^t X(\tau) e^{-\gamma|t-\tau|} d\tau = Q(t), \quad X_0 = 0,$$

где a , b , c , β и γ — константы. Расширим вектор состояния системы следующим образом: $X_1 = X$; $X_2 = \dot{X}$; $\dot{X}_1 = X_2$. Полагая

$$\int_0^t e^{-\beta|t-\tau|} X_2(\tau) d\tau = Y_1; \quad \int_0^t e^{-\gamma|t-\tau|} X_1(\tau) d\tau = Y_2,$$

приведем ИДСтС к ДСтС:

$$\begin{aligned} \dot{X}_1 &= X_2; & \dot{X}_2 &= -\frac{b}{a} Y_1 - \frac{c}{a} Y_2 + \frac{1}{a} Q(t); & \dot{Y}_1 &= -\beta Y_1 + X_2; & \dot{Y}_2 &= -\gamma Y_2 + X_1; \\ X_1(0) &= X_2(0) = Y_1(0) = Y_2(0) = 0. \end{aligned}$$

3. Рассмотрим скалярную нелинейную ИДСтС

$$\ddot{X} + \int_0^t \varphi(X(\tau), \dot{X}(\tau)) e^{-\alpha|t-\tau|} d\tau = Q(t), \quad X_0 = 0,$$

где α — константа. Расширим вектор состояния системы следующим образом: $X_1 = X$; $X_2 = \dot{X}$; $\dot{X}_1 = X_2$. Полагая

$$\int_0^t \varphi(X_1(\tau), X_2(\tau)) e^{-\alpha|t-\tau|} d\tau = Y,$$

придем к следующей ДСтС:

$$\begin{aligned} \dot{X}_1 &= X_2; & \dot{X}_2 &= -\frac{1}{\alpha} Y + Q(t); & \dot{Y} &= -\alpha Y + \varphi(X_1, X_2); \\ X_1(0) &= X_2(0) = Y(0) = 0. \end{aligned}$$

4. Рассмотрим скалярную нелинейную стохастическую ИДСтС

$$\ddot{X} + \int_0^t \varphi(X(\tau), \dot{X}(\tau)) e^{-\alpha|t-\tau|} \cos(\omega_0(t-\tau)) d\tau = Q(t), \quad X_0 = 0,$$

где α и ω_0 — константы. Расширим вектор состояния системы: $X_1 = X$; $X_2 = \dot{X}$; $\dot{X}_1 = X_2$. Положим

$$\int_0^t \varphi(X_1(\tau), X_2(\tau)) e^{-\alpha|t-\tau|} \cos(\omega_0(t-\tau)) d\tau = Y$$

и найдем производную \dot{Y} :

$$\begin{aligned} \dot{Y} &= \int_0^t \left(-\alpha e^{-\alpha|t-\tau|} \cos(\omega_0(t-\tau)) \varphi(X_1(\tau), X_2(\tau)) - \right. \\ &\quad \left. - \omega_0 e^{-\alpha|t-\tau|} \sin(\omega_0(t-\tau)) \varphi(X_1(\tau), X_2(\tau)) \right) d\tau + \varphi(X_1, X_2). \end{aligned}$$

Представляя интеграл суммы суммой интегралов, получим

$$\dot{Y} = -\alpha Y - \omega_0 \int_0^t e^{-\alpha|t-\tau|} \cos(\omega_0(t-\tau)) \varphi(X_1(\tau), X_2(\tau)) d\tau + \varphi(X_1, X_2).$$

Введем новую переменную:

$$I = \int_0^t \varphi(X_1(\tau), X_2(\tau)) e^{-\alpha|t-\tau|} \sin(\omega_0(t-\tau)) d\tau.$$

Тогда полученное уравнение запишется в виде:

$$\dot{Y} = -\alpha Y - \omega_0 I + \varphi(X_1, X_2).$$

Выразим из последнего уравнения I :

$$I = \frac{1}{\omega_0} \left(-\dot{Y} - \alpha Y + \varphi(X_1, X_2) \right).$$

Найдем производную \dot{I} :

$$\begin{aligned} \dot{I} = \int_0^t \left(-\alpha e^{-\alpha|t-\tau|} \sin(\omega_0(t-\tau)) \varphi(X_1(\tau), X_2(\tau)) + \right. \\ \left. + \omega_0 e^{-\alpha|t-\tau|} \cos(\omega_0(t-\tau)) \varphi(X_1(\tau), X_2(\tau)) \right) d\tau = -\alpha I + \omega_0 Y. \end{aligned}$$

Найдем вторую производную \ddot{Y} :

$$\begin{aligned} \ddot{Y} = -\alpha \dot{Y} - \omega_0 \dot{I} + \frac{\partial \varphi}{\partial X_1} \dot{X}_1 + \frac{\partial \varphi}{\partial X_2} \dot{X}_2 = \\ = \alpha^2 Y + 2\alpha \omega_0 I - \omega_0^2 Y - \alpha \varphi(X_1, X_2) + \frac{\partial \varphi}{\partial X_1} \dot{X}_1 + \frac{\partial \varphi}{\partial X_2} \dot{X}_2 = \\ = -2\alpha \dot{Y} - (\alpha^2 + \omega_0^2) Y + \alpha \varphi + \frac{\partial \varphi}{\partial X_1} \dot{X}_1 + \frac{\partial \varphi}{\partial X_2} \dot{X}_2. \end{aligned}$$

Расширим вектор состояния $Y_1 = Y$; $Y_2 = \dot{Y}$; $\dot{Y}_1 = Y_2$. В итоге придем к ДСтС вида:

$$\begin{aligned} \dot{X}_1 = X_2; \quad \dot{X}_2 = -Y_1 + Q(t); \quad \dot{Y}_1 = Y_2; \\ \dot{Y}_2 = -2\alpha Y_2 - (\alpha^2 + \omega_0^2) Y_1 + \alpha \varphi + \frac{\partial \varphi}{\partial X_1} X_2 + \frac{\partial \varphi}{\partial X_2} [Q(t) - Y]; \\ X_1(0) = X_2(0) = Y_1(0) = Y_2(0) = 0. \end{aligned}$$

5. Рассмотрим скалярную нелинейную стохастическую ИДСтС

$$\ddot{X} + \int_0^t \varphi(X(\tau), \dot{X}(\tau)) e^{-\alpha|t-\tau|} (\cos(\omega_0(t-\tau)) + \gamma \sin(\omega_0(t-\tau))) d\tau = Q(t), \quad X_0 = 0,$$

где α , ω_0 и γ — константы. Расширим вектор состояния системы $X_1 = X$; $X_2 = \dot{X}$; $\dot{X}_1 = X_2$. Примем

$$\int_0^t \varphi(X_1(\tau), X_2(\tau)) e^{-\alpha|t-\tau|} (\cos(\omega_0(t-\tau)) + \gamma \sin(\omega_0(t-\tau))) d\tau = Y.$$

Найдем производную \dot{Y} :

$$\begin{aligned} \dot{Y} = & \int_0^t \left(-\alpha e^{-\alpha|t-\tau|} (\cos(\omega_0(t-\tau)) + \gamma \sin(\omega_0(t-\tau))) \varphi(X_1(\tau), X_2(\tau)) + \right. \\ & \left. + \omega_0 e^{-\alpha|t-\tau|} (-\sin(\omega_0(t-\tau)) + \gamma \cos(\omega_0(t-\tau))) \varphi(X_1(\tau), X_2(\tau)) \right) d\tau + \\ & + \varphi(X_1(t), X_2(t)). \end{aligned}$$

Отсюда имеем

$$\begin{aligned} \dot{Y} = & -\alpha Y + \\ & + \omega_0 \int_0^t e^{-\alpha|t-\tau|} (-\sin(\omega_0(t-\tau)) + \gamma \cos(\omega_0(t-\tau))) \varphi(X_1(\tau), X_2(\tau)) d\tau + \\ & + \varphi(X_1(t), X_2(t)). \end{aligned}$$

Введем новую переменную:

$$I = \int_0^t \varphi(X(\tau), \dot{X}(\tau)) e^{-\alpha|t-\tau|} (-\sin(\omega_0(t-\tau)) + \gamma \cos(\omega_0(t-\tau))) d\tau.$$

Учитывая, что $\dot{Y} = -\alpha Y + \omega_0 I + \varphi(X_1, X_2)$, выразим из последнего уравнения I :

$$I = \frac{1}{\omega_0} \left(\dot{Y} + \alpha Y - \varphi(X_1, X_2) \right).$$

Найдем производную \dot{I} :

$$\begin{aligned} \dot{I} = & \int_0^t \left(-\alpha e^{-\alpha|t-\tau|} (-\sin(\omega_0(t-\tau)) + \gamma \cos(\omega_0(t-\tau))) \varphi(X_1(\tau), X_2(\tau)) + \right. \\ & \left. + \omega_0 e^{-\alpha|t-\tau|} (-\cos(\omega_0(t-\tau)) - \gamma \sin(\omega_0(t-\tau))) \varphi(X_1(\tau), X_2(\tau)) \right) d\tau + \\ & + \gamma \varphi(X_1, X_2). \end{aligned}$$

Отсюда получим $\dot{I} = -\alpha I - \omega_0 Y + \gamma \varphi(X_1, X_2)$. Найдем вторую производную \ddot{Y} :

$$\begin{aligned} \ddot{Y} = & -\alpha \dot{Y} + \omega_0 \dot{I} + \frac{\partial \varphi}{\partial X_1} \dot{X}_1 + \frac{\partial \varphi}{\partial X_2} \dot{X}_2 = \\ = & (\alpha^2 - \omega_0^2) Y - 2\alpha \omega_0 I + (-\alpha + \omega_0 \gamma) \varphi(X_1, X_2) + \frac{\partial \varphi}{\partial X_1} \dot{X}_1 + \frac{\partial \varphi}{\partial X_2} \dot{X}_2 = \\ = & -2\alpha \dot{Y} - (\omega_0^2 + \alpha^2) Y + (\alpha + \omega_0 \gamma) \varphi(X_1, X_2) + \frac{\partial \varphi}{\partial X_1} \dot{X}_1 + \frac{\partial \varphi}{\partial X_2} \dot{X}_2. \end{aligned}$$

Расширим вектор состояния: $Y_1 = Y$; $Y_2 = \dot{Y}$; $\dot{Y}_1 = Y_2$. В итоге ИДСтС приводится к ДСтС

$$\begin{aligned} \dot{X}_1 = X_2; \quad \dot{X}_2 = -Y_1 + Q(t); \quad \dot{Y}_1 = Y_2; \\ \dot{Y}_2 = -2\alpha Y_2 - (\omega_0^2 + \alpha^2) Y_1 + (\alpha + \omega_0 \gamma) \varphi(X_1, X_2) + \frac{\partial \varphi}{\partial X_1} X_2 + \frac{\partial \varphi}{\partial X_2} [Q(t) - Y_1]; \\ X_1(0) = X_2(0) = Y_1(0) = Y_2(0) = 0. \end{aligned}$$

Приложение 3

Тестовые примеры анализа и моделирования интегродифференциальных стохастических систем при ударных воздействиях

Пример 1. Рассмотрим ИДСтС из примера 2 приложения 2. Переобозначим параметры полученной систем следующим образом:

$$\frac{b}{2} = 2\epsilon_0 \omega_0; \quad \frac{c}{a} = \omega_0^2; \quad \frac{Q(t)}{a} = \sigma(t).$$

Тогда приходим к ДСтС вида:

$$\begin{aligned} \dot{X}_1 = X_2; \quad \dot{X}_2 = -2\epsilon_0 \omega_0 Y_1 - \omega_0^2 Y_2 + \sigma(t); \quad \dot{Y}_1 = -\beta Y_1 + X_2; \quad \dot{Y}_2 = -\gamma Y_2 + X_1; \\ X_1(0) = X_2(0) = Y_1(0) = Y_2(0) = 0. \end{aligned}$$

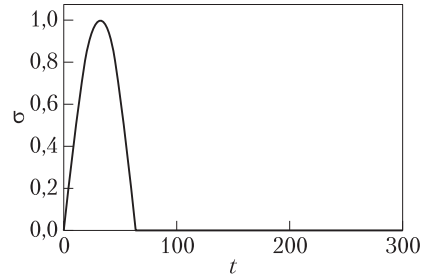


Рис. 1 Ударное воздействие $\sigma(t)$ для $\tau = 20\pi$

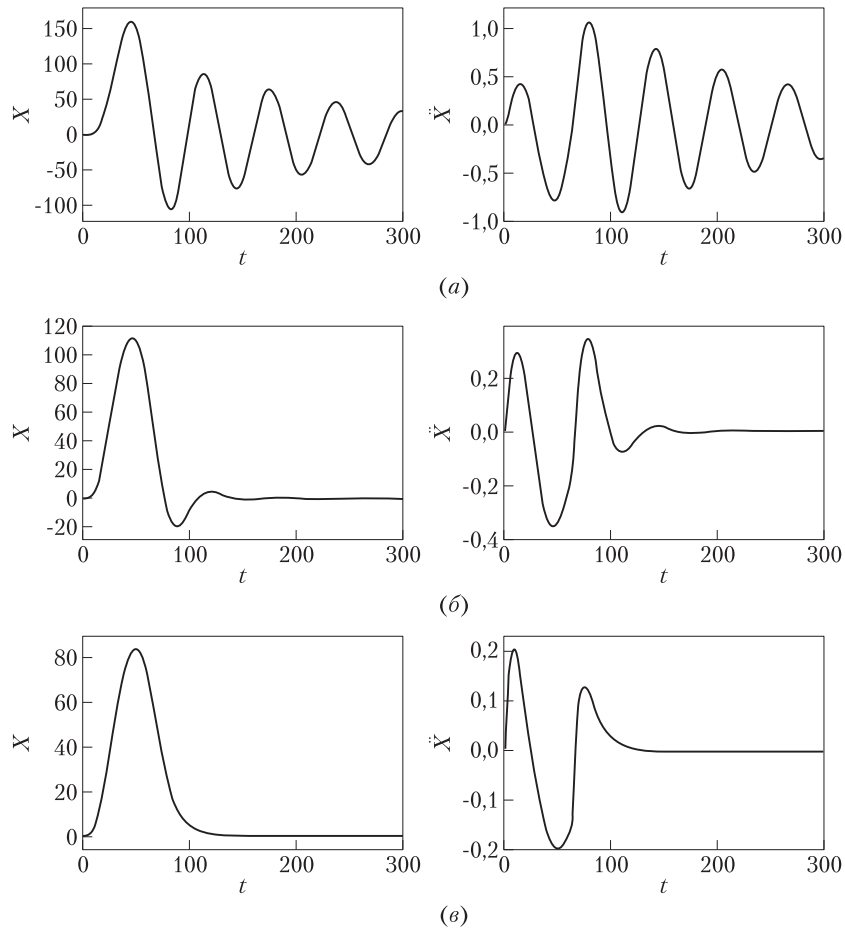


Рис. 2 Значение координаты X и ускорения \ddot{X} : (а) $\epsilon_0 = 0,1$; (б) $0,5$; (е) $\epsilon_0 = 1,0$

Проведем вычислительный эксперимент для детерминированного полусинусоидального входного воздействия

$$\sigma(t) = \begin{cases} \sigma_0 \sin\left(\frac{\pi}{\tau} t\right), & t < \tau; \\ 0, & t > \tau, \end{cases}$$

и следующих значений параметров $\omega_0 = 0,1$, $\epsilon_0 = 0,1$, $\sigma_0 = 1$ на промежутке времени $t \in [0, 300]$. Параметр τ выбирается из следующих соображений. Собственная частота колебаний системы $\omega_0 = 0,1$. Выберем параметр τ таким образом, чтобы частота ударного воздействия была в 2 раза меньше собственной частоты. Искомый параметр будет принимать значение $\tau = 20\pi$.

Входное воздействие изображено на рис. 1.

Результаты вычислительного эксперимента представлены на рис. 2, а.

Изменим величину демпфирования системы. Результаты вычислительного эксперимента для $\epsilon_0 = 0,5$ представлены на рис. 2, б, результаты вычислительного эксперимента для $\epsilon_0 = 1$ представлены на рис. 2, в.

Из результатов численного эксперимента можно сделать вывод о том, что колебания координаты X и ускорения \ddot{X} системы с течением времени затухают. При этом значение величины демпфирования ϵ_0 системы влияет на скорость затухания колебаний:

- при значении $\epsilon_0 = 0,1$ и координата X , и ускорение \ddot{X} колеблются на всем рассматриваемом промежутке времени;
- при значении $\epsilon_0 = 0,5$ колебания координаты X и ускорения \ddot{X} затухают к моменту времени $t = 200$;
- при значении $\epsilon_0 = 1$ колебания координаты X и ускорения \ddot{X} затухают к моменту времени $t = 150$.

Пример 2. Рассмотрим полученную в примере 1 систему на промежутке времени $t \in [0, 300]$. Проведем вычислительный эксперимент для системы дифференциальных уравнений. При этом в качестве входного воздействия возьмем экспоненциальный ударный импульс

$$\sigma(t) = \begin{cases} \sigma_0 e^{-at}, & t < \tau; \\ 0, & t > \tau. \end{cases}$$

Значения параметров возьмем аналогичными значениям параметров примера 1. Параметр a ударного воздействия выберем равным 1.

Входное воздействие и результаты моделирования изображены на рис. 3 и 4.

Вывод: колебания координаты X и ускорения \ddot{X} системы с течением времени затухают. При этом значение величины демпфирования ϵ_0 системы влияет на скорость затухания колебаний:

- при значении $\epsilon_0 = 0,1$ и координата X , и ускорение \ddot{X} колеблются на всем рассматриваемом промежутке времени;
- при значении $\epsilon_0 = 0,5$ колебания координаты X и ускорения \ddot{X} затухают к моменту времени $t = 150$;

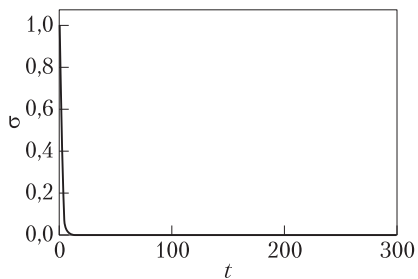


Рис. 3 Ударное воздействие $\sigma(t)$ для $\tau = 20\pi$

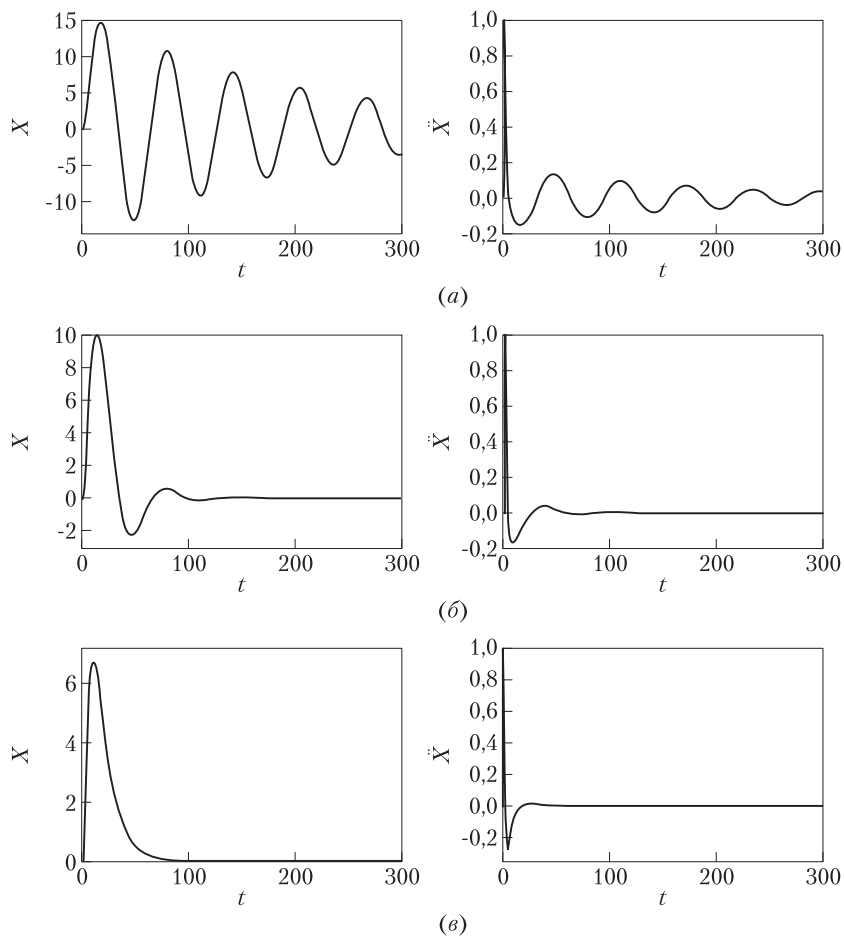


Рис. 4 Значение координаты X и ускорения \ddot{X} : (а) $\epsilon_0 = 0,1$; (б) $0,5$; (в) $\epsilon_0 = 1,0$

- при значении $\epsilon_0 = 1$ колебания координаты X и ускорения \ddot{X} затухают к моменту времени $t = 100$.

Таким образом, для полусинусоидального и экспоненциального ударных импульсов результаты получились схожими.

Пример 3. Рассмотрим полученную в примере 1 систему на промежутке времени $t \in [0, 300]$. При этом в качестве входного воздействия возьмем полусинусоидальный ударный импульс

$$\sigma(t) = \begin{cases} \sigma_0(1 + \xi(t)) \sin\left(\frac{\pi}{\tau} t\right) + \eta(t), & t < \tau; \\ \eta(t), & t > \tau, \end{cases}$$

где $\xi(t)$ и $\eta(t)$ — гауссовские белые шумы с нулевыми математическими ожиданиями и дисперсиями, равными 0,1. Значения параметров возьмем аналогичными значениям параметров в примере 1. На рис. 5 изображена реализация входного воздействия $\sigma(t)$ и стандартное квадратическое отклонение $\sqrt{D_\sigma}$.

Результаты вычислительного эксперимента представлены на рис. 6 и 7.

Таким образом, колебания координаты X и ускорения \ddot{X} системы с течением времени затухают. При этом значение величины демпфирования ϵ_0 системы влияет на скорость затухания колебаний:

- при значении $\epsilon_0 = 0,1$ и координата X , и ускорение \ddot{X} колеблются на всем рассматриваемом промежутке времени;
- при значении $\epsilon_0 = 0,5$ колебания координаты X и ускорения \ddot{X} затухают к моменту времени $t = 150$;
- при значении $\epsilon_0 = 1$ колебания координаты X и ускорения \ddot{X} затухают к моменту времени $t = 100$.

Кроме того, предел изменения стандартного отклонения координаты $\sqrt{D_X}$ уменьшается с ростом величины демпфирования.

Пример 4. Рассмотрим полученную в примере 1 систему на промежутке времени $t \in [0, 300]$. При этом в качестве входного воздействия возьмем экспоненциальный ударный импульс:

$$\sigma(t) = \begin{cases} \sigma_0(1 + \xi(t))e^{-at} + \eta(t), & t < \tau; \\ \zeta(t), & t > \tau, \end{cases}$$

где $\xi(t)$ и $\eta(t)$ — гауссовские белые шумы с нулевыми математическими ожиданиями и дисперсиями, равными 0,1, $a = 1$. Значения параметров возьмем аналогичными значениям параметров в примере 1. На рис. 8 изображена реализация входного воздействия $\sigma(t)$ и стандартное квадратическое отклонение $\sqrt{D_\sigma}$. Результаты вычислительного эксперимента представлены на рис. 9 и 10.

Таким образом, заключаем, что колебания координаты X и ускорения \ddot{X} системы с течением времени затухают. При этом значение величины демпфирования ϵ_0 системы влияет на скорость затухания колебаний:

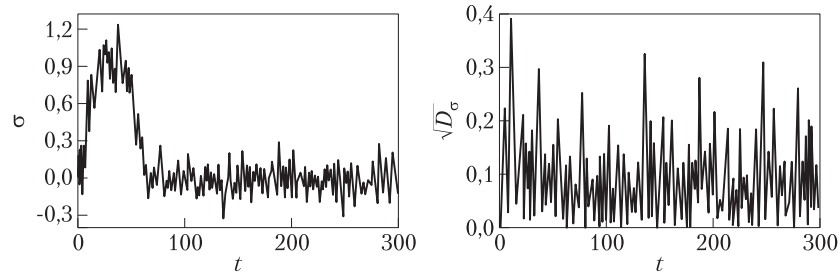


Рис. 5 Входное воздействие $\sigma(t)$ и стандартное квадратическое отклонение $\sqrt{D_\sigma}$

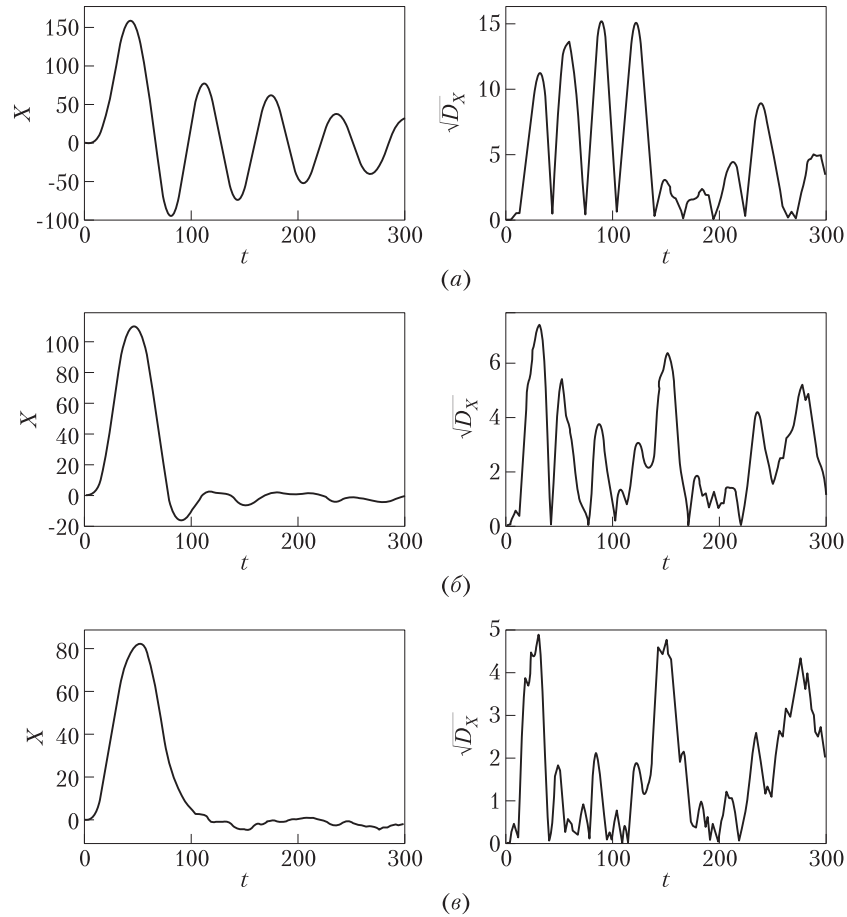


Рис. 6 Координата X и стандартное отклонение координаты $\sqrt{D_X}$ системы: (а) $\epsilon_0 = 0,1$; (б) $0,5$; (е) $\epsilon_0 = 1,0$

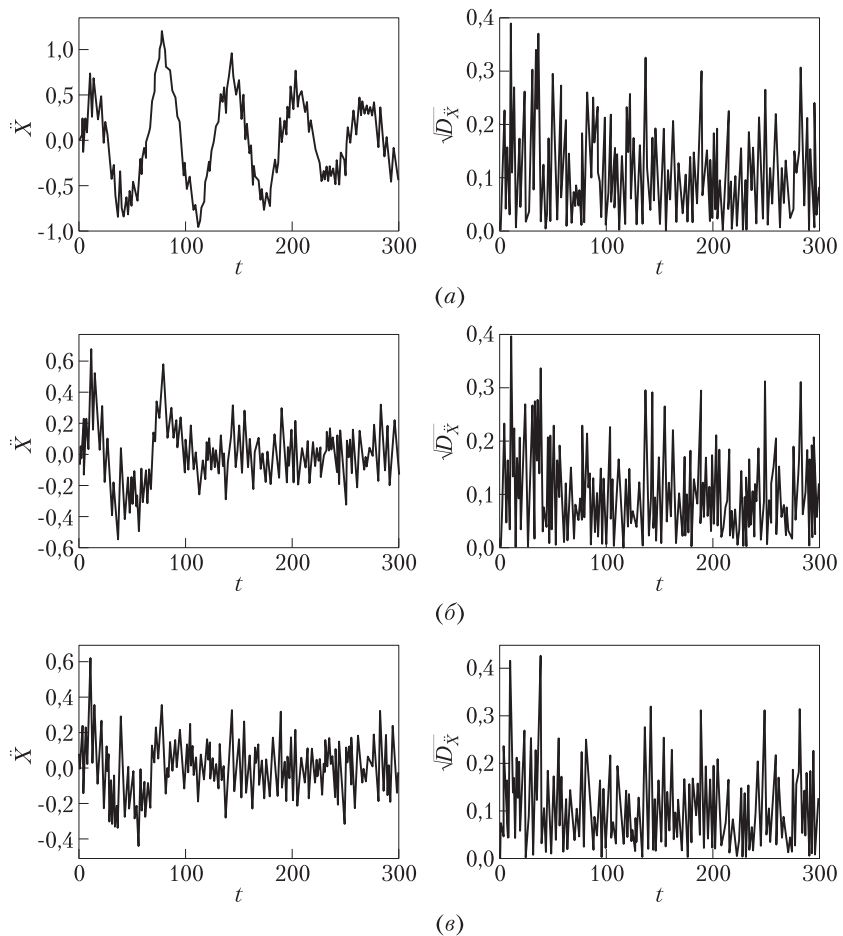


Рис. 7 Ускорение \ddot{X} и стандартное отклонение ускорения $\sqrt{D\ddot{X}}$ системы: (а) $\epsilon_0 = 0,1$; (б) $0,5$; (в) $\epsilon_0 = 1,0$

- при значении $\epsilon_0 = 0,1$ и координата X , и ускорение \ddot{X} колеблются на всем рассматриваемом промежутке времени;
- при значении $\epsilon_0 = 0,5$ колебания координаты X и ускорения \ddot{X} затухают к моменту времени $t = 150$;
- при значении $\epsilon_0 = 1$ колебания координаты X и ускорения \ddot{X} затухают к моменту времени $t = 100$.

Для полусинусоидального и экспоненциального ударных импульсов с аддитивными и мультипликативными помехами результаты получились схожи.

В качестве других тестовых примеров можно выбрать задачи из [6, 7, 14, 15].

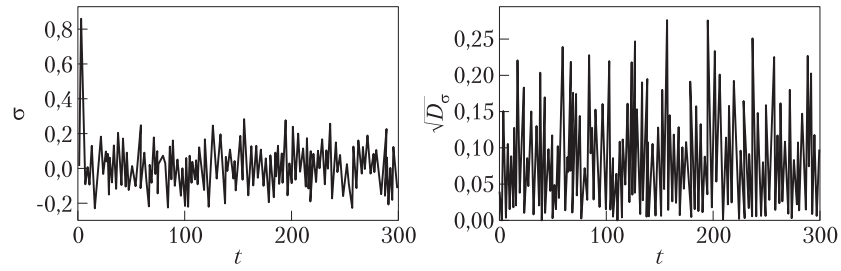


Рис. 8 Входное воздействие $\sigma(t)$ и стандартное квадратическое отклонение $\sqrt{D_\sigma}$

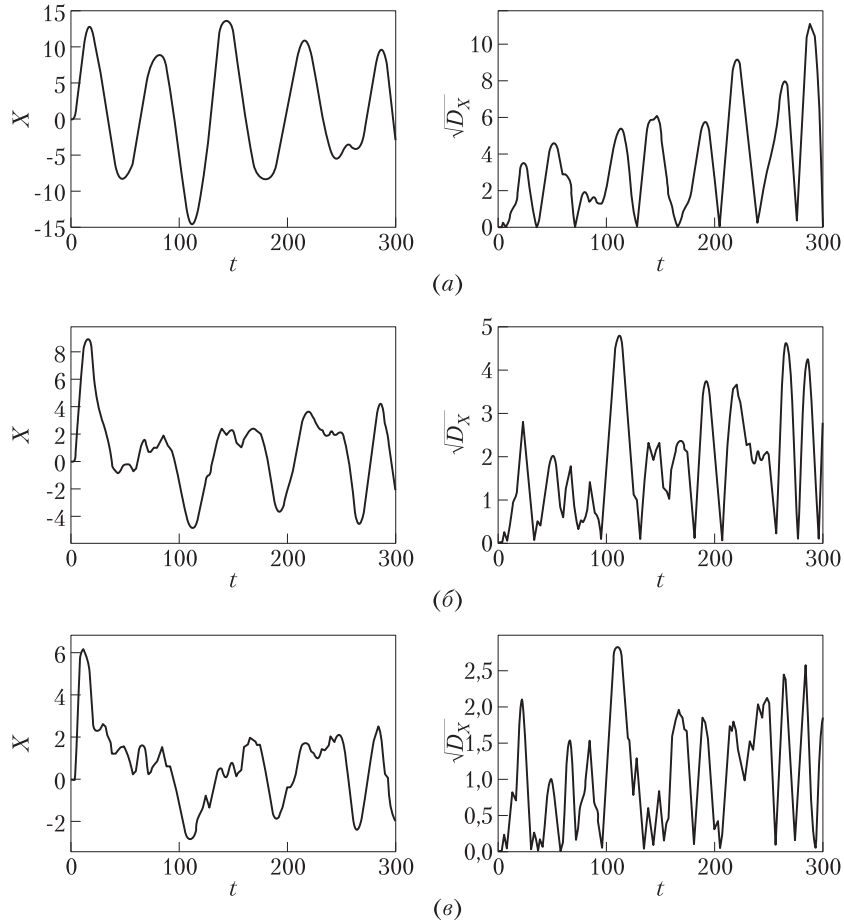


Рис. 9 Координата X и стандартное отклонение координаты $\sqrt{D_X}$ системы: (а) $\epsilon_0 = 0,1$; (б) $0,5$; (в) $\epsilon_0 = 1,0$

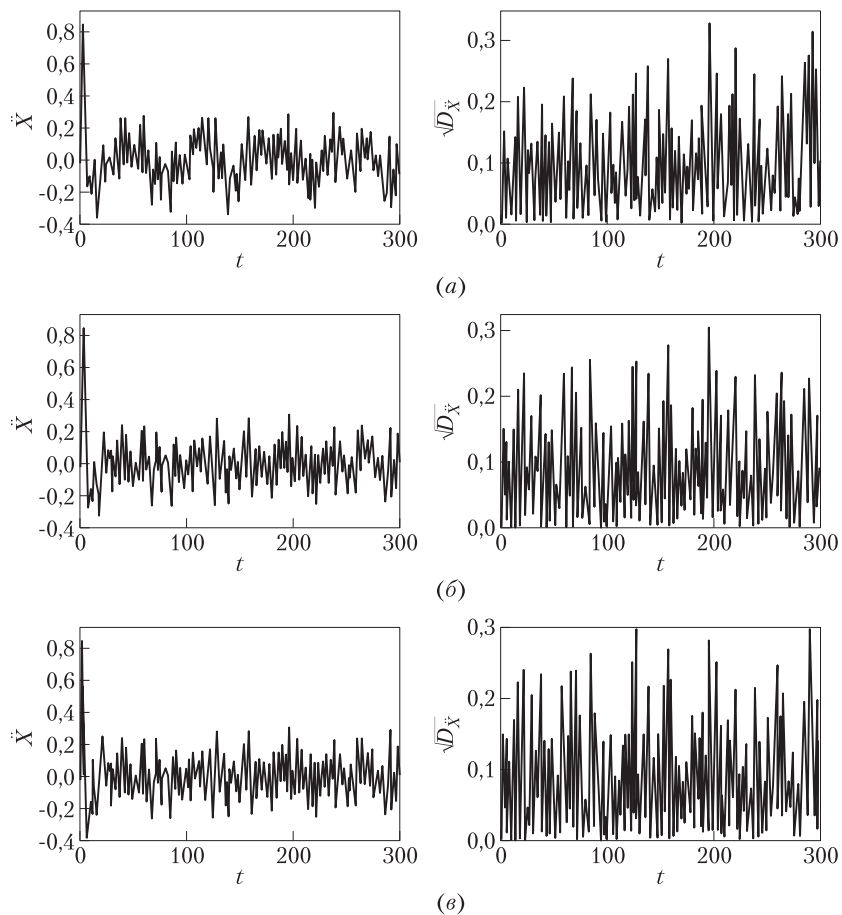


Рис. 10 Ускорение \ddot{X} и стандартное отклонение ускорения $\sqrt{D\ddot{X}}$ системы: (а) $\epsilon_0 = 0,1$; (б) $0,5$; (в) $\epsilon_0 = 1,0$

Литература

1. Колмановский В. Б., Новов В. Р. Устойчивость и периодические режимы регулируемых систем с последействием. — М.: Наука, 1981. 386 с.
2. Сеницын И. Н. Stochastic hereditary control systems // Проблемы управления и теории информатики, 1986. Т. 15. № 4. С. 287–298.
3. Сеницын И. Н. Конечномерные распределения процессов в стохастических интегральных и интегродифференциальных системах // 2nd Symposium (International) IFAC on Stochastic Control — Pergamon Press, 1987. Pt. 1. P. 144–153.
4. Сеницын И. Н. Анализ и моделирование распределений в эрдитарных стохастических системах // Информатика и её применения, 2014. Т. 8. Вып. 1. С. 2–11.

5. Пугачев В. С., Сеницын И. Н. Стохастические дифференциальные системы. Анализ и фильтрация. — М.: Наука, 1990. 632 с. [Stochastic differential systems. Analysis and filtering. — Chichester, New York: John Wiley, 1987. 549 p.]
6. Пугачев В. С., Сеницын И. Н. Теория стохастических систем. — М.: Логос, 2000; 2004. 1000 с. [Stochastic systems. Theory and applications. — Singapore: World Scientific, 2001. 968 p.]
7. Сеницын И. Н. Канонические представления случайных функций и их применение в задачах компьютерной поддержки научных исследований. — М.: ТОРУС ПРЕСС, 2009. 768 с.
8. Сеницын И. Н. Параметрическое статистическое и аналитическое моделирование распределений в нелинейных стохастических системах на многообразиях // Информатика и её применения, 2013. Т. 7. Вып. 2. С. 4–16.
9. Kloeden P., Platen E. Numerical solution of stochastic differential equations. — Berlin–Heidelberg–New York: Springer, 1992. 636 p.
10. Артемьев С. С. Численные методы решения задачи Коши для систем обыкновенных и стохастических дифференциальных уравнений. — Новосибирск: ВЦ СО РАН, 1993. 156 с.
11. Кузнецов Д. Ф. Численное интегрирование стохастических дифференциальных уравнений. — СПб.: СПбГУ, 2001. 712 с.
12. Сеницын И. Н., Сеницын В. И. Лекции по нормальной и эллипсоидальной аппроксимации распределений в стохастических системах. — М.: ТОРУС ПРЕСС, 2013. 488 с.
13. Сеницын И. Н., Сеницын В. И., Корепанов Э. Р., Белоусов В. В., Сергеев И. В., Басилашвили Д. А. Опыт моделирования эргодических стохастических систем // Кинетика и высокие технологии XXI века (С&Т): Тр. XIII Междунар. науч.-технич. конф. — Воронеж: Саквое, 2012. С. 346–357.
14. Артемьев С. С., Михайличенко И. Г., Сеницын И. Н. Статистическое моделирование сложных финансовых операций. — Новосибирск: ВЦ СО РАН, 1996. Кн. 1 и 2. 232 с.
15. Сеницын И. Н., Шаламов А. С., Сергеев И. В., Сеницын В. И., Корепанов Э. Р., Белоусов В. В., Агафонов Е. С., Шоргин В. С. Методы и средства анализа и моделирования стохастических систем интегрированной логистической поддержки // Системы и средства информатики, 2012. Вып. 22. № 2. С. 3–28.

Поступила в редакцию 17.10.13

DISTRIBUTIONS PARAMETRICAL MODELING SOFTWARE FOR INTEGRODIFFERENTIAL STOCHASTIC SYSTEMS

I. N. Sinitsyn, I. V. Sergeev, V. I. Sinitsyn, E. R. Korepanov, and V. V. Belousov

Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str.,
Moscow 119333, Russian Federation

Abstract: Methods and algorithms for analytical and statistical modeling of one- and multidimensional distributions in integrodifferential stochastic system (IDStS) are considered. Nonlinear Ito equations for IDStS with Wiener and Poisson noises are presented. For dying physically realizable integral kernels, two methods of IDStS reduction to differential stochastic system (DStS) are given for dying physical realizable integral kernels (based on linear operator equations and singular kernels). Different approaches for numerical analytical and statistical modeling for IDStS reducible to DStS are discussed. Two numerical methods for practical applications are presented: normal approximation method (for nonadditive noises) and statistical linearization method (for additive noises). Special attention is paid to pure direct numerical methods for continuous and discontinuous right hands of equations. Test examples and results of computer experiments carried out by the “IDStS” software tool in MATLAB are given.

Keywords: analytical modeling; differential stochastic system; generalized Ito formula; hereditary system; integrodifferential stochastic system; Ito stochastic differential equation; singular kernels; software; statistical modeling

DOI: 10.14357/08696527140101

Acknowledgments

The work was financially supported by the Program of the RAS Department for Nanotechnologies and Information Technologies “Intelligent information technology, system analysis, and automation” (project 1.7).

References

1. Kolmanovskij, V. B., and V. R. Nosov. 1981. *Ustoychivost' i periodicheskie rezhimy reguliruemyykh sistem s posledeystviem* [Stability of hereditary systems]. Moscow: Nauka. 386 p.
2. Sinitsyn, I. N. 1986. Stochastic hereditary control systems. *Problemy Upravleniya i Teorii Informatiki* [Problems Control Information Theory] 15 (4):287–298.
3. Sinitsyn, I. N. 1986. Konkretnye raspredeleniya protsessov v stokhasticheskikh integral'nykh i integrodifferentsial'nykh sistemakh [Finite-dimensional distributions in stochastic integral and integrodifferential systems]. *2nd Symposium (International) IFAC on Stochastic Control Proceedings: Preprints*. Pergamon Press. 1:144–153.

4. Sinitsyn, I. N. 2014. Analiz i modelirovanie raspredeleniy v ereditarnykh stokhasticheskikh sistemakh [Analysis and modeling of distributions in hereditary stochastic systems]. *Informatika i ee Primeneniya — Inform. Appl.* 8(1):2–11.
5. Pugachev, V. S., and I. N. Sinitsyn. 1987. *Stochastic differential systems. Analysis and filtering*. Chichester, New York: John Wiley. 549 p.
6. Pugachev, V. S., and I. N. Sinitsyn. 2001. *Stochastic systems. Theory and applications*. Singapore: World Scientific. 968 p.
7. Sinitsyn, I. N. 2009. Kanonicheskie predstavleniya sluchaynykh funktsiy i ikh primeneniye v zadachakh komp'yuternoy podderzhki nauchnykh issledovaniy [Canonical expansions of random functions and their application to scientific computer aided support]. Moscow: TORUS PRESS. 768 p.
8. Sinitsyn, I. N. 2013. Parametricheskoe statisticheskoe i analiticheskoe modelirovanie raspredeleniy v nelineynykh stokhasticheskikh sistemakh na mnogoobraznykh [Parametrical statistical and analytical modeling of distributions in stochastic systems on manifolds]. *Informatika i ee Primeneniya — Inform. Appl.* 7(2):4–16.
9. Kloeden, P., and E. Platen. 1992. *Numerical solution of stochastic differential equations*. Berlin – Heidelberg – New York: Springer. 636 p.
10. Artemiev, S. S. 1993. *Chislennyye metody resheniya zadachi Koshi dlya sistem obyknovennykh i stokhasticheskikh differentsial'nykh uravneniy* [Cauchy problem numerical methods for systems of ordinary and stochastic equations]. Novosibirsk: VC SO RAS, 1993. 156 p.
11. Kuznetsov, D. F. 2001. *Chislennoe integririrovanie stokhasticheskikh differentsial'nykh uravneniy* [Numerical methods for stochastic differential equations]. St. Petersburg: SPU. 712 p.
12. Sinitsyn, I. N., and V. I. Sinitsyn. 2013. *Lektsii po normal'noy i ellipsoidal'noy aproksimatsii raspredeleniy v stokhasticheskikh sistemakh* [Lectures on normal and ellipsoidal distributions approximations]. Moscow: TORUS PRESS. 488 p.
13. Sinitsyn, I. N., V. I. Sinitsyn, E. R. Korepanov, V. V. Belousov, I. V. Sergeev, and D. A. Basilashvili. 2012. Opyt modelirovaniya ereditarnykh stokhasticheskikh sistem [Experience in modeling of hereditary stochastic systems]. *Kinetika i Vysokie Tekhnologii XXI Veka (C&T): Trudy XIII Mezhdunarodnoy nauchno-tekhnicheskoy konferentsii* [Kinetics and High Technologies in XXI Century (C&T): 13th Conference (International) Proceedings]. Voronezh: Sakvoe. 346–357.
14. Artem'ev, S. S., I. G. Mikhaylichenko, and I. N. Sinitsyn. 1996. Statisticheskoe modelirovanie slozhnykh finansovykh operatsiy. [Statistical modeling of complex financial operations]. Novosibirsk: VC RAS. Books 1, 2. 232 p.
15. Sinitsyn, I. N., A. S. Shalamov, I. V. Sergeev, V. I. Sinitsyn, E. R. Korepanov, V. V. Belousov, E. S. Agafonov, and V. S. Shorgin. 2012. *Metody i sredstva analiza i modelirovaniya stokhasticheskikh sistem integral'noy logisticheskoy podderzhki* [Methods and means of analysis and modeling of complex stochastic systems of integral logistic support]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 22(2):3–28.

Received October 17, 2013

Contributors

Sinitsyn Igor N. (b. 1940)— Doctor of Science in technology, Professor, Honored scientist of Russian Federation, Head of Department, Institute of Information Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; sinitsin@dol.ru

Sergeev Igor V. (b. 1965) — Candidate of Science (PhD) in technology, Deputy Director, Institute of Information Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; ISergeev@ipiran.ru

Sinitsyn Vladimir I. (b. 1968) — Doctor of Science in physics and mathematics, associate professor, Head of Department, Institute of Information Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; VSinitsin@ipiran.ru

Korepanov Eduard R. (b. 1966)— Candidate of Science (PhD) in technology, Head of Laboratory, Institute of Information Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; EKorepanov@ipiran.ru

Belousov Vasilij V. (b. 1977) — Candidate of Science (PhD) in technology, Head of Laboratory, Institute of Information Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; VBelousov@ipiran.ru

ПАРАЛЛЕЛИЗМ В МИКРОПРОЦЕССОРАХ*

А. К. Горшенин¹, С. В. Замковец², В. Н. Захаров³

Аннотация: Производительность работы микропроцессора в значительной мере определяется степенью организации параллельной работы различных блоков. Рассмотрены основные возможности распараллеливания работы микропроцессора. Для параллельной обработки команд используется конвейерный метод, для параллельной обработки данных применяется архитектура SIMD (Single Instruction – Many Data, одна команда – много данных). Реализованный метод распараллеливания на уровне потоков послужил основой для создания многоядерных микропроцессоров. Основу многоядерного микропроцессора составляет один или несколько мощных процессоров, окруженных множеством вспомогательных ядер, которые предназначены для более эффективной обработки сложных мультимедийных приложений в многопоточном режиме. Архитектуры с поддержкой многопроцессорной обработки на уровне кристалла (chip-level multiprocessing, CMP) представляют будущее микропроцессоров, потому что такие архитектуры позволяют достичь огромных уровней производительности при более приемлемых тактовых частотах благодаря параллельному выполнению множества операций.

Ключевые слова: архитектура микропроцессора; распараллеливание; конвейер (pipeline); суперскалярный микропроцессор; архитектура MMX и SSE; Hyper-Threading; многоядерные процессоры

DOI: 10.14357/08696527140102

1 Введение

Производительность микропроцессоров зависит от двух составляющих. Первая составляющая — это тактовая частота, и на протяжении длительного времени производительность микропроцессоров в основном связывалась именно с этой составляющей. В начале нашего века в производственных планах основных производителей микропроцессоров на конец десятилетия предполагалось достичь тактовой частоты 10 ГГц [1]. Однако в связи с рядом технологических трудностей

*Работа выполнена при поддержке грантов РФФИ № 13-07-00579-а и № 12-07-00109-а, а также Гранта Президента РФ для государственной поддержки молодых российских ученых МК-4103.2014.9.

¹Институт проблем информатики Российской академии наук; Московский государственный технический университет радиотехники, электроники и автоматики; agorshenin@ipiran.ru

²Институт проблем информатики Российской академии наук; Московский государственный технический университет радиотехники, электроники и автоматики; zamkovetc@ipiran.ru

³Институт проблем информатики Российской академии наук, vzakharov@ipiran.ru

(основные проблемы связаны с увеличением токов утечки и вызванным этим неприемлемым ростом тепловыделения, но рассмотрение этих вопросов не входит в круг интересов данной статьи) этим планам не суждено было сбыться. По опубликованным данным, наибольшая тактовая частота 5,5 ГГц была достигнута в 2012 г. в микропроцессоре zE12 фирмы IBM [2].

Вторая составляющая, связанная с производительностью, определяет число инструкций, выполняемых микропроцессором за один такт. Эта составляющая связана с реализацией в микропроцессоре параллельной работы различных функциональных блоков, и необходимо отметить, что для обеспечения такой параллельности разработчиками аппаратной части микропроцессоров была проделана гигантская работа. Можно сказать, что параллелизм или параллельные вычисления стали доминирующей парадигмой в архитектуре микропроцессоров.

В качестве характерного примера начала использования принципа распараллеливания в микропроцессорах можно рассмотреть особенности микроархитектуры (способа реализации архитектуры) микропроцессора i80386 фирмы INTEL, разработанного в 1985 г. [3]. В первую очередь можно заметить, что по своей структуре этот микропроцессор был разделен на две части, и когда первая исполнительная часть этого микропроцессора выполняла работу по исполнению инструкции, вторая интерфейсная часть осуществляла подкачку информации из памяти в микропроцессор, если шина передачи данных была в это время свободна. Кроме того, для поддержания некоторого количества получаемой из памяти информации в исполнительной части микропроцессора использовался буфер для хранения следующих для выполнения инструкций. Этот буфер получил название буфера предвыборки (prefetch unit).

Для распараллеливания работы микропроцессора имеется несколько возможностей, и среди этих возможностей можно выделить следующие основные:

- распараллеливание на уровне инструкций;
- распараллеливание на уровне обработки данных;
- распараллеливание на уровне потоков (TLP — thread-level parallelism).

2 Распараллеливание на уровне инструкций

На начальном этапе усилия разработчиков были направлены на такую реализацию архитектуры, при которой обеспечивалось параллельное выполнение команд микропроцессора. И здесь значительным шагом внедрения параллелизма в работу микропроцессора стало использование конвейерной техники выполнения команд.

Характерным примером является тот факт, что в технической документации на микропроцессор i80386 [3] указывалось время выполнения (в тактах) каждой инструкции. Например, для команды ADD r8, r8 (сложение двух операндов, находящихся в 8-разрядных регистрах) время выполнения равнялось двум тактам, а для команды ADD r8, m8 (сложение двух операндов, один из которых нахо-

дится в регистре, а второй — в памяти) время выполнения равнялось 7 тактам. Для современных микропроцессоров подобные данные отсутствуют, поскольку в этих микропроцессорах используются конвейеры. При конвейерной организации время выполнения всех команд равно времени нахождения их на конвейере.

Использование конвейера позволяет микропроцессору одновременно обрабатывать столько команд, сколько ступеней имеется в конвейере. Кроме того, в случае использования большого количества ступеней (а это означает, что на каждой ступени выполняется меньший объем работы) можно увеличивать тактовую частоту работы микропроцессора. Однако большое число ступеней также может отрицательно влиять на производительность микропроцессора, поскольку в случае необходимости сброса конвейера его новое заполнение может потребовать значительных затрат времени. Поиск оптимальной длины конвейеров можно объяснить различным, а в некоторых случаях заметно отличающееся число ступеней в различных моделях. Например, в микропроцессоре Pentium длина конвейера составляла 5 ступеней, а в Pentium 4 — уже 20 ступеней [4].

Преимущество конвейерной обработки перед последовательной наблюдается в идеальном конвейере, в котором отсутствуют конфликты. Конфликты — это такие ситуации при работе конвейера, которые препятствуют последовательному выполнению очередной команды в предназначенном для нее такте. Конфликты, возникающие в конвейере, можно разделить на три группы:

- (1) структурные конфликты;
- (2) конфликты по управлению;
- (3) конфликты по данным.

Структурные конфликты возникают в тех случаях, когда аппаратные средства конвейера микропроцессора не могут поддерживать для всех команд одинаковое время нахождения на какой-либо ступени конвейера. Чаще всего такая ситуация возникает на ступени выполнения инструкции. В этом случае работа конвейера приостанавливается до тех пор, пока команда не будет обработана на этой ступени.

Как правило, обычный способ преодоления структурного конфликта предполагает применение таких аппаратных решений, которые позволяют снизить затраты времени на выполнение данного этапа (например, использование специальных схем умножения). Однако такой способ может привести к значительному усложнению схемы микропроцессора. Поскольку такие конфликты в работе конвейера возникают при выполнении команд, относительно редко встречающихся в программах, то обычно при разработке ищут компромисс между увеличением длительности работы конвейера и усложнением схемы микропроцессора.

Конфликты по управлению возникают в тех случаях, когда в конвейере находятся команды переходов или другие команды, изменяющие значение счетчика команд. При выполнении команды перехода возникает необходимость сброса конвейера, поскольку должна выполняться не та команда, которая находится в конвейере, а некоторая другая команда. В соответствии со статистикой, в типичной программе на каждые 6–8 команд приходится 1 команда перехода и по-

следствия выполнения перехода очевидны: при наличии конвейера через каждые 6–8 команд его необходимо очищать и заполнять заново в соответствии с новым адресом. Таким образом, все преимущества конвейеризации теряются. Ярким примером конфликта по управлению может служить команда безусловного перехода с косвенным адресом, при встрече с которой возникает необходимость сбрасывать конвейер, так как в счетчике команд появляется абсолютно непредсказуемый адрес следующей команды.

В случае конфликта по управлению, возникающего с командами условного перехода, для снижения потерь работы конвейера в процессорах используется метод предсказания переходов. Основная идея этого метода заключается в том, что в процессоре используется специальный блок, который определяет наиболее вероятное направление перехода, не дожидаясь формирования признаков, на основании анализа которых этот переход реализуется. После этого процессор выполняет команды по предсказанной ветви программы (так называемое спекулятивное выполнение). Чтобы обеспечить корректное выполнение программы, процессор записывает получаемые результаты не в память или регистры, а в специальный буфер. Если направление перехода предсказано правильно, то процессор продолжает выполнять программу дальше, а полученные результаты переписываются из буфера по местам назначения. В случае неправильного предсказания направления перехода результаты выполнения данного участка программы уничтожаются.

Конфликты по данным возникают в случаях, когда выполнение одной команды зависит от результата выполнения предыдущей команды. Например, команда i предшествует команде j , а команда j должна использовать операнд, который формирует команда i . Если эти команды находятся достаточно близко в конвейере, то может возникнуть ситуация, при которой предшествующая команда не успеет записать значение этого операнда. Вопросы, связанные с конвейеризацией, подробнее рассмотрены в [5].

Для решения проблемы с конфликтами по данным используется один из методов параллелизма, а именно способ внеочередного выполнения команд (Out-of-Order — OoO). При использовании этого метода декодированная команда попадает не в конвейер, а в специальный буфер-диспетчер, который накапливает подготавливаемые к исполнению инструкции, отслеживает выполнение ранее запущенных команд и по мере освобождения исполнительных устройств отправляет на них подготовленные к выполнению инструкции. Основной парадигмой OoO-обработки является возможность разрешить процессору избежать остановки конвейера, которая должна произойти, когда данные, необходимые для выполнения команды, являются недоступными. Процессор заполняет эти «промежутки времени» другими командами, которые готовы для выполнения, затем переупорядочивает результаты, чтобы на выходе они появлялись в нормальном порядке. Для преобразования из программного порядка в порядок, в котором команды обрабатываются на самом деле, требуется достаточно сложная схемотехника.

Следующим шагом в развитии параллелизма в микропроцессорах стало использование нескольких конвейеров в одном микропроцессоре. Такие микропроцессоры получили название суперскалярных. Необходимо отметить, что при выполнении программы подбором инструкций для конвейеров занимаются аппаратные средства самих микропроцессоров, поскольку различные конвейеры обладают специфическими требованиями и работа с ними должна учитывать их специфику. Кроме того, аппаратура должна следить и за синхронностью работы конвейеров, так как приостановка одного из них может нарушить правильное выполнение всей программы.

Один из методов организации параллельной работы отдельных блоков микропроцессора связан с переименованием регистров. Идея этого метода возникла в связи с тем, что при выполнении некоторых команд передачи управления появляется необходимость сохранять содержимое регистров общего назначения и, как правило, такое сохранение осуществляется при помощи стека. Такая операция сохранения содержимого регистров в стеке (а затем и обратная запись из стека в регистры) требует определенного времени на ее выполнение. Для того чтобы сократить это время, было предложено использовать дополнительные физические регистры, т. е. в регистровом файле микропроцессора размещать большое количество физических регистров. Эти физические регистры группируются в регистровые окна, и непосредственно в каждый момент работы доступно только одно окно. Необходимо иметь в виду, что непосредственно сами команды при помощи соответствующих полей могут обращаться только к так называемым архитектурным (логическим) регистрам, и понятно, что количество физических и логических регистров в микропроцессорах может отличаться в значительной степени. Например, количество архитектурных регистров в микропроцессорах Pentium равно 8, а в то же время число физических регистров равно 40. В типичной SPARC-архитектуре (scalable processor architecture) используется регистровый файл из 128 регистров, причем пользователю из них одновременно доступны только расположенные подряд 24 регистра. Процессор всегда содержит указатель текущего окна с возможностью его модификации в рамках реализованной глубины вложения процедур. Номера логических регистров динамически отображаются на номера физических регистров посредством таблиц отображения. Таким образом, при выполнении команд передачи управления микропроцессор просто переключает регистровое окно, сохраняя содержимое тех регистров, которые должны быть сохранены, и освобождая новые регистры для дальнейшего использования.

Другой метод использования регистрового окна связан с применением переменных. В современных языках программирования, например в языке С, переменные принято разделять на локальные и глобальные. Локальные переменные используются только одной конкретной функцией и существуют только то время, пока эта функция работает. Глобальные переменные существуют все время, пока выполняется программа, и доступны всем функциям программы. Когда вызывается определенная функция, то необходимые для ее работы ис-

ходные данные записываются в конец окна, а процессор при вызове функции смещает окно по регистровому файлу таким образом, чтобы записанные данные оказались в начале нового, пока пустого окна. Требовавшие сохранения временные данные вызывавшего функцию кода оказываются за пределами окна, вследствие чего становятся недоступными. После завершения работы функции полученные результаты записываются в те же самые регистры в начале окна, после чего процессор смещает окно обратно.

В определенном смысле к реализации парадигмы параллелизма можно отнести и аппаратную поддержку виртуализации, т. е. возможность реализовывать на одном процессоре несколько виртуальных машин (ВМ). Изначально в микропроцессорах наиболее популярных семейств Intel и AMD не предусматривались режимы работы, при которых возможна эффективная виртуализация. Поэтому для построения монитора ВМ (МВМ, управляющего блока для виртуализации) приходилось пользоваться пошаговой эмуляцией инструкций ВМ или же выполнять бинарную трансляцию исполняемого кода, что резко снижало эффективность работы всей системы.

Но уже в начале 2005 г. появились общедоступные спецификации расширения для архитектуры IA-32 под первоначальным названием Vanderpool Technology (другое известное название расширения — VT-x), а также для архитектуры Itanium, называемые VT-I (Vanderpool Technology for the Intel Itanium architecture), позволяющие эффективно реализовывать МВМ и тем самым обеспечивать возможность одновременной работы нескольких операционных систем и сред на одной аппаратной платформе. В мае 2005 г. и компания AMD опубликовала полную спецификацию технологии расширения архитектуры AMD64 (под названием Pacifica, иначе называемая AMD-V), которая позволяет реализовать МВМ. В настоящее время обе фирмы поддерживают технологию виртуализации на всех вновь выпускаемых семействах микропроцессоров.

3 Параллельная обработка данных

Рассмотренные методы реализации параллелизма относятся к микропроцессорам с принципом обработки операндов SISD (Single Instruction – Single Data, одна команда – одни данные). В этих микропроцессорах одна команда осуществляет обработку одних данных. Впоследствии в микропроцессорах был реализован принцип обработки операндов SIMD. В соответствии с таким подходом в микропроцессорах одна команда может одновременно обрабатывать много данных, поэтому можно сказать, что таким образом был реализован параллелизм в обработке данных.

Принцип обработки SIMD впервые был реализован в микропроцессоре Pentium III для ускорения выполнения мультимедиа и коммуникационных программ и получил название архитектуры MMX (MultiMedia eXtension) [6]. Эта архитектура включает в себя новые команды и типы данных, что позволяет создавать

приложения нового уровня. Новые MMX-команды увеличивают производительность прикладных программ в таких областях, как видео, комбинирование графики и видео, обработка изображений, звуковой синтез, распознавание и синтез речи, телефония, видеоконференции, двух- и трехмерная графика.

В MMX-архитектуре вводятся новые типы данных:

- упакованные байты — 8 байт, упакованные в одно 64-разрядное число;
- упакованные слова — 4 слова, упакованные в одно 64-разрядное число;
- упакованные двойные слова — 2 двойных слова, упакованные в одно 64-разрядное число;
- учетверенное слово — одно 64-разрядное число.

Команды MMX могут обрабатывать группы из 8 байт, 4 слов, двух двойных слов и одно учетверенное слово. Такие группы из 64 разрядов рассматриваются как упакованные данные и могут трактоваться как один из вышеперечисленных типов данных.

Основной выигрыш в производительности обработки данных при использовании SIMD-метода осуществляется за счет одновременной обработки нескольких операндов при помощи одной команды. Набор MMX-команд состоит из команд пересылки данных, упаковки/распаковки, сложения/вычитания, умножения, сдвига, сравнения и поразрядных логических операций.

Дальнейшее развитие MMX-архитектура получила в технологии, получившей название SSE (Streaming SIMD Extensions — потоковое расширение SIMD) [4]. Технология SSE была направлена на реализацию принципа SIMD для вещественных чисел. Для реализации этой технологии в микропроцессор были добавлены восемь новых 128-разрядных регистров, получивших название XMM. В каждом из регистров XMM помещаются четыре вещественных 32-разрядных числа. Каждое 32-разрядное число с плавающей запятой имеет 1 знаковый бит, 8 битов порядка и 23 бита мантиссы, что соответствует стандарту IEEE-754 на формат представления чисел одинарной точности с плавающей запятой.

Технология SSE включает 70 новых команд и позволяет выполнять векторные (пакетные) и скалярные инструкции, при этом векторные инструкции выполняют операции одновременно над четырьмя наборами операндов. В именах команд, выполняющих векторные операции, присутствует суффикс PS. Скалярные инструкции работают только с одним комплектом операндов — младшим 32-разрядным набором. В именах команд, выполняющих скалярные операции, присутствует суффикс SS.

В микропроцессорах Pentium 4 была использована технология SSE2, которая улучшает возможности архитектуры MMX и технологии SSE. В технологии SSE2 используются те же самые регистры, которые были введены в технологии SSE, т. е. 128-разрядные регистры XMM. Однако в отличие от SSE, где эти регистры использовались только для хранения и обработки вещественных

чисел (32-разрядные операнды с плавающей запятой и одинарной точности), в технологии SSE2 эти регистры могут хранить и обрабатывать целочисленные операнды, представляющие: 16 байт; 8 слов; 4 двойных слова; 2 четверных слова; одно 128-разрядное слово. Кроме того, в этих регистрах может храниться и обрабатываться новый формат вещественных чисел — два 64-разрядных числа с плавающей запятой двойной точности.

Технология SSE2 включает в себя набор инструкций, который производит операции со скалярными и упакованными типами данных. Эта технология предназначена для разработки приложений трехмерной графики, обработки сигналов, кодирования и декодирования, распознавания речи и т. п. К набору команд SSE в SSE2 добавлены 144 новые команды, которые позволяют:

- разрабатывать алгоритмы, в которых одновременно можно обрабатывать смешанные типы данных: упакованные числа с плавающей запятой с одинарной точностью, упакованные числа с плавающей запятой с двойной точностью, а также целые 64- и 128-разрядные числа;
- работать с данными различной размерности: байтом, словом, двойным словом, учетверенным словом и двойным учетверенным словом.

Следующим шагом совершенствования SSE-технологии стало SSE3-расширение, которое появилось в микропроцессоре Pentium 4. Расширение SSE3 добавило еще 13 новых команд. Все они, за исключением трех, используют XMM-регистры и предназначены для повышения производительности при выполнении следующих операций:

- быстрое преобразование вещественного числа в целое (соответствующая команда `fisttp` заменяет семь «обычных» команд);
- сложные арифметические вычисления (команды `addsubps`, `addsubpd`, `movsldup`, `movshdup`, `movddup`);
- кодирование видео (команда `1ddqu`);
- обработка графики (команды `haddps`, `hsubps`, `haddpd`, `hsubpd`);
- синхронизация потоков (команды `monitor`, `mwait`).

И, наконец, в микропроцессорах Intel Core была реализована технология SSE4, использующая еще 54 новых инструкций.

4 Распараллеливание на уровне потоков

Как уже отмечалось выше, распараллеливание работы микропроцессоров осуществляется также на уровне потоков, и такой подход получил название Hyper-Threading. Основная идея данной технологии заключается в том, чтобы обеспечить одновременную работу тех функциональных блоков, которые имеются в микропроцессорах. Это достигается за счет одновременного запуска нескольких вычислительных потоков.

В настоящее время рынок полностью завоевали многозадачные операционные системы, идеология которых построена на одновременном существовании нескольких вычислительных потоков (threads), относящихся к одному или разным активным приложениям, либо к самой операционной системе. Потоки часто называют также потоками выполнения (thread of execution), иногда называют «нитеями» (буквальный перевод thread). Если на многопроцессорных системах эти потоки могут выполняться одновременно (по одному на процессор), то в однопроцессорных системах микропроцессоры вынуждены непрерывно переключаться между потоками, квантуя процессорное время между исполнением их различных частей.

Таким образом, если разрешить процессору одновременное выполнение более чем одного потока, то его работа может осуществляться более эффективно. Именно в этом и состоит основная идея Hyper-Threading. Благодаря этой технологии один физический процессор воспринимается операционной системой и приложениями как два логических процессора. Соответственно, операционная система предполагает, что процессор с технологией Hyper-Threading может одновременно выполнять два потока и загружает такой процессор работой гораздо полнее. Поскольку физический процессор с технологией Hyper-Threading представляет собой два логических процессора, некоторые блоки в таких процессорах должны быть продублированы. Причем продублированы должны быть лишь отдельные управляющие элементы, основные же исполнительные ресурсы остаются теми же — теперь они попросту загружаются более плотно и эффективно. В итоге в процессорах с Hyper-Threading продублированы регистры, включая регистры общего назначения и регистры Instruction Pointer — указатели на очередную инструкцию для исполнения, некоторые управляющие регистры, усовершенствованный программируемый контроллер прерываний (Advanced Programmable Interrupt Controller — APIC). Остальные же ресурсы, включая кэши, исполнительные блоки, логику предсказания переходов, контроллер шины и т. п., логические процессоры разделяют друг с другом. Именно поэтому реализация технологии Hyper-Threading обошлась разработчикам достаточно дешево — площадь процессорного ядра увеличилась всего лишь примерно на 5%, но позволила увеличить производительность на 15%–30% [7].

Имеется несколько концепций реализации многонитевой архитектуры (Multithreading Architecture — MTA). Наиболее распространенной является архитектура с одновременным выполнением нитей (SMT — Simultaneous Multithreading). Само ее название указывает на допустимость одновременного выполнения нескольких нитей, когда на каждом новом такте на выполнение в какое-либо функциональное устройство может направляться команда любой нити. В SMT, если программные коды нити обеспечивают выполнение большого количества команд за один такт, такая нить будет использовать большинство ресурсов процессора. «Плохие» же, в смысле числа выполняемых команд, нити будут разделять ресурсы с другими подобными себе. Для реализации SMT необходимы следующие аппаратные средства:

- несколько счетчиков команд (по одному на нить) с возможностью выбора любого из них на каждом такте;
- средства, ассоциирующие команды с нитью, которой они принадлежат (необходимые, в частности, для работы механизмов предсказания переходов и переименования регистров);
- несколько стеков адресов возврата (по одному на нить) для предсказания адресов возврата из подпрограмм;
- специальная дополнительная память в процессоре (в расчете на каждую нить) для осуществления процедуры удаления из буфера выполненных внеочередным образом команд.

Можно сделать вывод о следующих преимуществах Hyper-Threading: улучшенная поддержка многопоточного кода, позволяющая запускать потоки одновременно; улучшенная реакция и время отклика; увеличенное число пользователей, которое может поддерживаться сервером. Впервые технология Hyper-Threading была реализована в микропроцессорах Pentium 4 [7].

5 Многоядерные микропроцессоры

Следующим логическим и наиболее значимым шагом по внедрению параллелизма, поддержанным технологическими возможностями по наращиванию числа транзисторов (закон Мура продолжает действовать), стал шаг по размещению на одном чипе нескольких ядер. Можно сказать, что многоядерные процессоры реализуют параллелизм работы в квадрате, поскольку не только непосредственно в самих ядрах могут быть обеспечены те виды параллелизма, которые были представлены выше, но и сами ядра могут работать параллельно.

Первый многоядерный микропроцессор появился в 2001 г. — фирма IBM выпустила процессор Power4, который мог быть использован только для решения некоторых узкопрофильных задач. Пользователям персональных компьютеров пришлось ждать появления двухъядерного процессора целых 4 года — когда в 2005 г. вслед за двухъядерным процессором Opteron для серверов компании AMD появился двухъядерный микропроцессор Intel Pentium D для домашних персональных компьютеров. Первый четырехъядерный микропроцессор Intel Core 2 Quad появился в конце 2007 г. В 2009 г. фирма Intel выпустила микропроцессоры Core i3, i5 и i7, которые производились по 32-нанометровому техпроцессу. А 14 октября 2011 г. увидел свет шестиядерный процессор Intel Core i7-3960X на базе архитектуры Sandy Bridge-E.

В корпорации Intel уже говорят не о «мультиядерности» (multicore) процессоров, как это делается в отношении 2-, 4-, 8-, 16- или даже 32-ядерных решений, а о «многоядерности» (many-core), подразумевая совершенно новую архитектурную макроструктуру чипа [8]. Многоядерные микропроцессоры могут иметь гомогенную или гетерогенную архитектуру. В гомогенной архитектуре все

ядра одинаковые и выполняют одни и те же функции. В гетерогенной архитектуре ядра процессора выполняют разные задачи. Например, в процессоре Cell альянса IBM, Sony и Toshiba одно ядро является ядром общего назначения, а восемь — специализированными микропроцессорами, оптимизированными для векторных операций.

На основании данных, представленных в [8], структуру многоядерного микропроцессора можно описать следующим образом, при этом учитывая, что каждое ядро должно быть независимым — с независимым энергопотреблением и управляемой потребляемой мощностью:

Основу структуры составляют одно или несколько мощных центральных ядер, число таких ядер зависит от класса решаемых данным процессором задач. Эти ядра окружены множеством вспомогательных ядер, которые предназначены для более эффективной обработки сложных мультимедийных приложений в многопоточном режиме. Кроме этих ядер, которые можно назвать ядрами «общего назначения», процессоры должны также обладать специализированными ядрами для выполнения различных классов задач, таких как графическая обработка, алгоритмы распознавания речи, обработка коммуникационных протоколов.

Учитывая возрастающую в среде исследователей и пользователей популярность проведения вычислений на специализированных графических видеокартах (технологии CUDA от NVIDIA, ATI Stream Technology от AMD), очень перспективной представляется идея интеграции сложных графических процессоров вместе с центральными многозадачными процессорами на одном кристалле. Первые шаги к такой интеграции как раз и представлены в технологии APU (Accelerated Processing Unit). Очевидно, что большую часть вычислительной нагрузки можно будет снять с центрального процессора, оставив ему задачи управления, синхронизации и т. д., а вычисления реализовывать с помощью графических ядер.

По мнению сотрудников Intel, архитектуры с поддержкой многопроцессорной обработки на уровне кристалла (chip-level multiprocessing, CMP) представляют будущее микропроцессоров, потому что такие архитектуры позволяют достичь огромных уровней производительности при более приемлемых тактовых частотах благодаря параллельному выполнению множества операций, и в то же время обеспечить эффективное управление питанием и эффективный режим охлаждения.

Архитектуры CMP фирмы Intel — это не только путь к огромному росту производительности, но и возможность свести к минимуму потребляемую мощность и теплоотдачу. В отличие от ориентации на большие, энергоемкие вычислительные ядра с большой теплоотдачей, кристаллы CMP фирмы Intel будут активизировать только те ядра, которые необходимы для выполнения текущей задачи, тогда как остальные ядра будут отключены. Такое мелко модульное управление вычислительными ресурсами позволяет кристаллу потреблять ровно столько электроэнергии, сколько нужно в данный момент времени.

Однако, несмотря на видимый подъем интереса к многоядерным и многопоточным процессорам, следует признать, что ключевой момент грядущей «параллельной революции» заключается не в собственно процессорах, а в грядущих

радикальных изменениях в программном обеспечении. Оно должно стать таким, чтобы сохранилась возможность использовать преимущества экспоненциального роста производительности процессоров. Оптимальная производительность приложений в многоядерных системах может быть достигнута за счет рационального использования программных потоков для «правильного» распределения подзадач, создаваемых программным обеспечением. Выполнение потоков можно оптимизировать для выполнения на различных между собой физических ядрах, что повышает производительность системы.

При написании подобных программ необходимо придерживаться принципа декомпозиции [9], т. е. стараться выделять в рамках решения глобальной задачи связанные между собой подзадачи, которые могут выполняться параллельно. При этом возможно несколько видов декомпозиции:

- (1) декомпозиция по заданиям, т. е. присвоение различных функций разным потокам. Это наиболее простой способ создания параллельных программ, подразумевающий одновременное выполнение для задач, которые изначально предполагались независимыми между собой. Например, программа может рассчитывать параметры модели, однако пользователь может в это время просматривать и отображать (таблично, графически) первоначальные данные, для которых она строится;
- (2) декомпозиция по данным предполагает обработку информации разными процессами, однако каждый процесс обрабатывает свой непересекающийся блок данных. Например, при обработке результатов физического эксперимента, в котором система может находиться в относительно независимых между собой состояниях, возможно строить модель для каждого из режимов с помощью разных процессов, хотя изначально это были данные одного эксперимента;
- (3) декомпозиция по информационным потокам, наиболее сложная с точки зрения реализации параллелизма. Дело в том, что во многих реальных задачах встречаются ситуации, когда выходные данные одного подзадания являются входными данными для другого. Очевидно, что пока входные данные не будут получены, второй процесс не может быть полноценно запущен. Например, пока не будет сформирована тестовая выборка экспериментальных данных достаточного размера, процесс моделирования простаивает.

Таким образом, выбирая определенный тип декомпозиции для реализации программного кода (или используя некие гибридные варианты), можно добиться наиболее эффективного использования аппаратных ресурсов, предоставляемых в современных и перспективных многопроцессорных архитектурах. При этом для решения конкретных задач, возможно, потребуется индивидуальный подход, так как технологии параллельного программирования не только предоставляют возможности по увеличению производительности, но и предъявляют повышенные требования к разработчикам программного обеспечения в силу необходимости разрешения ряда возникающих специфических трудностей.

6 Заключение

Анализируя процесс дальнейшего развития микропроцессоров, можно предположить, что многоядерные процессоры будут выделяться не только высокой производительностью самой по себе, но также богатыми и разнообразными вычислительными и коммуникационными возможностями, управлением питанием, повышенной надежностью, безопасностью и управляемостью, а также полной интеграцией со всеми остальными компонентами платформы.

Литература

1. Черняк Л. Многоядерные процессоры и грядущая параллельная революция // Открытые системы, 2007. № 4. <http://www.osp.ru/os/2007/04/4219910/>.
2. IBM zEnterprise EC12 Technical Introduction. — IBM, 2012. <http://www.redbooks.ibm.com/redbooks/pdfs/sg248050.pdf>.
3. Брамм П., Брамм Д. Микропроцессор 80386 и его программирование / Пер. с англ. — М.: Мир, 1990. 448 с. (*Brumm P., Brumm D. 80386: A programming and design handbook. Blue Ridge Summit, PA: TAB Professional and Reference Books, 1987. 452 p.*)
4. Замковец С. В. Развитие микропроцессорных архитектур. — М.: ИПИ РАН, 2012. 148 с.
5. Замковец С. В., Захаров В. Н., Красовский В. Е. Эволюция архитектур современных микропроцессоров // Системы и средства информатики, 2011. Т. 21. № 1. С. 34–46.
6. Бердышев Е. Технология MMX (новые возможности процессоров P5 и P6). — М.: Диалог-МИФИ, 1997. 214 с.
7. Новый процессор Prescott // Компьютер Пресс, 2003. № 4. <http://compress.ru/article.aspx?id=10396>.
8. Borkar Sh. Y., Dubey P., Kahn K. C., Kuck D. J., Mulder H., Pawlowski S. S., Rattner J. R. Platform 2015: Intel processor and platform evolution for the next decade. — Intel, 2006. http://heim.ifi.uio.no/~inf3410/docs/Intel_Corp_Platform_2015.pdf.
9. Эхтер Ш., Робертс Дж. Многоядерное программирование / Пер. с англ. — СПб.: Питер, 2010. 316 с. (*Akhter S., Roberts J. Multi-core programming. Increasing performance through software multi-threading. Intel Press, 2006. 358 p.*)

Поступила в редакцию 28.02.14

PARALLELISM IN MICROPROCESSORS

A. K. Gorshenin^{1,2}, *S. V. Zamkovets*^{1,2}, and *V. N. Zakharov*¹

¹Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

²Moscow Institute of Radio, Electronics, and Automation (MIREA), 78 Vernadskogo Ave., Moscow 119454, Russian Federation

Abstract: Microprocessor performance is largely determined by the degree of organization of parallel work of various units. Different ways of microprocessor parallelization are considered. For parallel processing of commands, the pipeline method is used; for parallel data processing, the SIMD (Single Instruction–Many Data) architecture is applied. The implemented method of thread-level parallelism was the basis for creation of multicore microprocessors. A multicore microprocessor is one of more powerful processors that are surrounded by a multitude of auxiliary engines, which are designed for more efficient processing of complex multimedia applications in the multithreaded mode. Architectures with support of chip-level multiprocessing represent the future of microprocessors, because such architecture can achieve huge productivity levels with more acceptable frequencies through parallel execution of many operations.

Keywords: microprocessor architecture; parallelization; pipeline; superscalar microprocessor; architecture of MMX and SSE; Hyper-Threading; multicore processors

DOI: 10.14357/08696527140102

Acknowledgments

The research was supported by the Russian Foundation for Basic Research (grants Nos. 13-07-00579-a and 12-07-00109-a) and by the President of Russian Federation Grant for Government Support of Young Russian Scientists No. MK-4103.2014.9.

References

1. Chernyak, L. 2007. Mnogoyadernye protsessory i gryadushchaya parallel'naya revolyutsiya [Multicore processors and forthcoming parallel revolution]. *Otkrytye Sistemy [Open Systems]* 4. Available at: <http://www.osp.ru/os/2007/04/4219910/> (accessed April 3, 2014).
2. IBM zEnterprise EC12 Technical Introduction. 2012. Available at: <http://www.redbooks.ibm.com/redbooks/pdfs/sg248050.pdf> (accessed April 3, 2014).
3. Brumm, P., and D. Brumm. 1987. *80386: A programming and design handbook*. Blue Ridge Summit, PA: TAB Professional and Reference Books. 452 p.

4. Zamkovets, S. V. 2012. *Razvitie mikroprotsessornykh arkhitektur* [Evolution of micro-processors architectures]. Moscow: IPI RAN. 148 p.
5. Zamkovets, S. V., V. N. Zakharov, and V. E. Krasovsky. 2011. Evolyutsiya arkhitektur sovremennykh mikroprotssessorov [Evolution of modern microprocessors architecture]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 21(1):34–46.
6. Berdyshev, E. 1997. *Tekhnologiya MMX (novye vozmozhnosti protssessorov P5 i P6)* [MMX technology (new capabilities of P5 and P6 processors)]. Moscow: Dialog-MIFI. 214 p.
7. Novyy protssessor Prescott [New Prescott processor]. 2003. *Komp'yuter Press* [Computer Press] 4. Available at: <http://compress.ru/article.aspx?id=10396> (accessed April 3, 2014).
8. Borkar, Sh. Y., P. Dubey, K. C. Kahn, D. J. Kuck, H. Mulder, S. S. Pawłowski, and J. R. Rattner. 2006. Platform 2015: Intel processor and platform evolution for the next decade. — Intel. Available at: http://heim.ifi.uio.no/~inf3410/docs/Intel_Corp_Platform_2015.pdf (accessed April 9, 2014).
9. Akhter, S., and J. Roberts. 2006. *Multi-core programming. Increasing performance through software multi-threading*. Intel Press. 358 p.

Received February 28, 2014

Contributors

Gorshenin Andrey K. (b. 1986) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; associate professor, Moscow Institute of Radio, Electronics, and Automation (MIREA), 78 Vernadskogo Ave., Moscow 119454, Russian Federation; agorshenin@ipiran.ru

Zamkovets Sergey V. (b. 1944) — Candidate of Science (PhD) in technology, Head of Laboratory, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; associate professor, Moscow Institute of Radio, Electronics, and Automation (MIREA), 78 Vernadskogo Ave., Moscow 119454, Russian Federation; szamkovetc@ipiran.ru

Zakharov Victor N. (b. 1948) — Doctor of Science in technology, associate professor; Scientific Secretary, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; vzakharov@ipiran.ru

О ТЕОРЕТИЧЕСКИ СТОЙКИХ ШИФРАХ

С. М. Рацеев¹

Аннотация: К. Шеннон в 1940-х гг. ввел понятие совершенного шифра, обеспечивающего наилучшую защиту открытых текстов. Такой шифр не дает криптоаналитику никакой дополнительной информации об открытом тексте на основе перехваченной криптограммы. При этом хорошо известный шифр гаммирования с равновероятной гаммой является совершенным, но максимально уязвимым к попыткам имитации и подмены. Это происходит потому, что в шифре гаммирования алфавиты для записи открытых и шифрованных текстов равнозначны. В данной обзорной работе рассматриваются совершенные шифры, стойкие к имитации и подмене шифрованных сообщений, построенные на основе математической модели шифра замены с неограниченным ключом, предложенной А. Ю. Зубовым. При этом опорный шифр данной модели является совершенным и достигает нижних границ для вероятностей успеха имитации и подмены сообщений. Также приведены некоторые обобщения классической теории К. Шеннона о совершенных шифрах.

Ключевые слова: шифр; совершенный шифр; имитация сообщения

DOI: 10.14357/08696527140103

Введение

Следуя работе [1], введем алгебраическую модель шифра. Пусть X , K и Y — конечные множества возможных открытых текстов, ключей и шифрованных текстов соответственно. Пусть также $E_k : X \rightarrow Y$ — правило зашифрования на ключе $k \in K$. Обозначим через E множество всех правил зашифрования $\{E_k \mid k \in K\}$. Пусть $D_k : E_k(X) \rightarrow X$ — правило расшифрования на ключе $k \in K$. Обозначим через D множество правил расшифрования $\{D_k \mid k \in K\}$.

Алгебраической моделью шифра назовем совокупность

$$\Sigma_A = (X, K, Y, E, D),$$

для которой выполнены следующие свойства:

- (1) для любых $x \in X$ и $k \in K$ выполнено равенство $D_k(E_k(x)) = x$;
- (2) $Y = \bigcup_{k \in K} E_k(X)$.

Пусть P_X и P_K — независимые априорные распределения вероятностей соответственно на конечных множествах X и K , не содержащие нулевых вероятностей. Тогда под вероятностной моделью Σ_B будем понимать совокупность

$$\Sigma_B = (X, K, Y, E, D, P_X, P_K),$$

где множества X , K и Y связаны условиями 1 и 2.

¹Ульяновский государственный университет, RatseevSM@mail.ru

Заметим, что распределения P_X и P_K естественным образом индуцируют распределение вероятностей $P_Y = \{P_Y(y)|y \in Y\}$ на множестве шифрованных текстов Y следующим образом:

$$P_Y(y) = \sum_{(x,k) \in X \times K E_k(x)=y} P_X(x)P_K(k). \quad (1)$$

Заметим также, что для любого $y \in Y$ выполнено неравенство $P_Y(y) > 0$. Это следует из условия 1, равенства (1) и того, что для любых $x \in X$ и $k \in K$ выполнены неравенства $P_X(x) > 0$ и $P_K(k) > 0$.

Обозначим $K(x, y) = \{k \in K | E_k(x) = y\}$. Условная вероятность $P_{Y|X}(y|x)$ определяется естественным образом:

$$P_{Y|X}(y|x) = \begin{cases} \sum_{k \in K(x,y)} P_K(k), & K(x, y) \neq \emptyset; \\ 0, & K(x, y) = \emptyset. \end{cases}$$

С учетом теоремы умножения вероятностей определим условную вероятность $P_{X|Y}(x|y)$:

$$P_{X|Y}(x|y) = \frac{P_X(x)P_{Y|X}(y|x)}{P_Y(y)}.$$

1 Теоретическая стойкость шифров

Напомним, что шифр Σ_B называется совершенным по Шеннону, если для любых $x \in X$, $y \in Y$ выполнено равенство $P_{X|Y}(x|y) = P_X(x)$. Приведем эквивалентные условия совершенного по Шеннону шифра.

Предложение 1. Для произвольного шифра Σ_B следующие условия эквивалентны:

- (i) для любых $x \in X$ и $y \in Y$ выполнено равенство $P_{X|Y}(x|y) = P_X(x)$;
- (ii) для любых $x \in X$ и $y \in Y$ выполнено равенство $P_{Y|X}(y|x) = P_Y(y)$;
- (iii) для любых $x_1, x_2 \in X$ и $y \in Y$ выполнено равенство $P_{Y|X}(y|x_1) = P_{Y|X}(y|x_2)$.

Приведем также некоторые свойства совершенного по Шеннону шифра.

Предложение 2 [1]. Пусть Σ_B — совершенный по Шеннону шифр. Тогда шифру Σ_B присущи следующие свойства:

- (i) для любых $x \in X$, $y \in Y$ найдется такой ключ $k \in K$, что $E_k(x) = y$;
- (ii) для множеств X , Y и K справедливо двойное неравенство $|X| \leq |Y| \leq |K|$.

Предложение 3. Пусть Σ_B — совершенный по Шеннону шифр. Тогда равенство $|K| = |Y|$ выполнено в том и только том случае, когда $|K(x, y)| = 1$ для любых $x \in X$, $y \in Y$.

Теорема 1 (достаточные условия совершенного по Шеннону шифра [2]). Пусть для шифра Σ_B выполнены следующие условия:

- (i) $|K(x, y)| = 1$ для любых $x \in X, y \in Y$;
- (ii) распределение вероятностей P_K является равномерным.

Тогда шифр Σ_B является совершенным по Шеннону, причем распределение вероятностей P_Y будет равномерным и $|K| = |Y|$.

Следствие 1 (теорема К. Шеннона). Пусть Σ_B — некоторый шифр, для которого выполнено равенство $|X| = |K| = |Y|$. Шифр Σ_B является совершенным по Шеннону тогда и только тогда, когда выполнены следующие условия:

- (i) $|K(x, y)| = 1$ для любых $x \in X, y \in Y$;
- (ii) распределение вероятностей P_K является равномерным.

Наряду с теоремой Шеннона приведем еще один критерий совершенных шифров в классе шифров с равномерным распределением вероятностей на множестве ключей K .

Теорема 2 [3]. Пусть для шифра Σ_B выполнено двойное неравенство $|X| \leq |Y| \leq |K|$ и распределение вероятностей P_K является равномерным. Шифр Σ_B является совершенным по Шеннону тогда и только тогда, когда выполнены следующие условия:

- (i) для любых $x \in X, y \in Y$ найдется такой ключ $k \in K$, что $E_k(x) = y$;
- (ii) для любых $x_1, x_2 \in X, y \in Y$ выполнено равенство $|K(x_1, y)| = |K(x_2, y)|$.

Следствие 2. Пусть для шифра Σ_B выполнено соотношение $|X| \leq |Y| = |K|$ и распределение вероятностей P_K является равномерным. Шифр Σ_B является совершенным тогда и только тогда, когда $|K(x, y)| = 1$ для любых $x \in X, y \in Y$.

Шифр Σ_B называется $(k|y)$ -совершенным, если для любых $k \in K, y \in Y$ выполнено равенство $P_{K|Y}(k|y) = P_K(k)$. Естественным образом вводятся следующие вероятности:

$$P_{Y|K}(y|k) = \begin{cases} \sum_{x \in X(k,y)} P_X(x), & X(k, y) \neq \emptyset; \\ 0, & X(k, y) = \emptyset, \end{cases} \quad (2)$$

где $X(k, y) = \{x \in X \mid E_k(x) = y\}$;

$$P_K(k)P_{Y|K}(y|k) = P_Y(y)P_{K|Y}(k|y).$$

Заметим, что для любых $k \in K, y \in Y$ множество $X(k, y)$ либо пусто, либо одноэлементно. Как и в предложениях 1, 2 и 3 приведем эквивалентные условия $(k|y)$ -совершенных шифров и некоторые их свойства.

Предложение 4. Для произвольного шифра Σ_B следующие условия эквивалентны:

- (i) для любых $k \in K$ и $y \in Y$ выполнено равенство $P_{K|Y}(k|y) = P_K(k)$;
- (ii) для любых $k \in K$ и $y \in Y$ выполнено равенство $P_{Y|K}(y|k) = P_Y(y)$;
- (iii) для любых $k_1, k_2 \in K$ и $y \in Y$ выполнено равенство $P_{Y|K}(y|k_1) = P_{Y|K}(y|k_2)$.

Предложение 5. Для произвольного шифра Σ_B следующие условия эквивалентны:

- (i) $|X| = |Y|$;
- (ii) для любой пары $(k, y) \in K \times Y$ найдется $x \in X$ такое, что $E_k(x) = y$.

Предложение 6. Пусть шифр Σ_B является $(k|y)$ -совершенным. Тогда $|X| = |Y|$.

Теорема 3 (достаточные условия $(k|y)$ -совершенного шифра). Пусть для шифра Σ_B выполнены следующие условия:

- (i) $|X| = |Y|$;
- (ii) распределение вероятностей P_X является равномерным.

Тогда шифр Σ_B является $(k|y)$ -совершенным, причем распределение вероятностей P_Y также будет равномерным.

Доказательство. Пусть выполнены условия теоремы. Тогда из условия (i), предложения 5 и определения понятия шифра следует, что для любой пары $(k, y) \in K \times Y$ найдется, и притом единственное, $x \in X$, такое что $E_k(x) = y$. Поэтому, учитывая условие (ii) и равенство (2), будет выполнено такое равенство:

$$P_{Y|K}(y|k) = P_X(x) = \frac{1}{|X|} = \frac{1}{|Y|}.$$

Следовательно, для любых $y_1, y_2 \in Y$, $k_1, k_2 \in K$ справедливо следующее равенство:

$$P_{Y|K}(y_1|k_1) = \frac{1}{|Y|} = P_{Y|K}(y_2|k_2).$$

Таким образом, выполнено условие (iii) предложения 4, поэтому шифр Σ_B является $(k|y)$ -совершенным.

Из того же предложения 4 следует, что для любых $y \in Y$, $k \in K$ будет выполнено равенство $P_{Y|K}(y|k) = P_Y(y)$. Поэтому

$$P_Y(y) = P_{Y|K}(y|k) = \frac{1}{|Y|}. \quad \square$$

Теорема 4. Пусть для шифра Σ_B выполнено равенство $|X| = |Y| = |K|$. Шифр Σ_B является совершенным по Шеннону и $(k|y)$ -совершенным тогда и только тогда, когда выполнены следующие условия:

- (i) $|K(x, y)| = 1$ для любых $x \in X$ и $y \in Y$;
- (ii) распределение вероятностей P_K является равномерным;
- (iii) распределение вероятностей P_X является равномерным.

Доказательство. Пусть выполнены условия (i)–(iii). Тогда совершенность по Шеннону и $(k|y)$ -совершенность шифра Σ_B следует из теорем 1 и 3.

Покажем в обратную сторону. Пусть шифр Σ_B является совершенным по Шеннону и $(k|y)$ -совершенным. Тогда условия (i) и (ii) следуют из теоремы Шеннона. Покажем справедливость условия (iii).

Предварительно заметим, что распределение вероятностей P_Y является равномерным (теорема 1).

Зафиксируем произвольное значение $k_0 \in K$. Из предложений 5 и 6 следует, что для любого $y \in Y$ найдется, и притом единственное, $x \in X$ такое, что $E_{k_0}(x) = y$. Пронумеруем элементы множеств $X = \{x_1, \dots, x_N\}$ и $Y = \{y_1, \dots, y_N\}$ таким образом, чтобы выполнялись равенства

$$E_{k_0}(x_i) = y_i, \quad i = 1, \dots, N.$$

Тогда

$$P_X(x_i) = P_{Y|K}(y_i|k_0) = P_Y(y_i) = \frac{1}{|Y|}, \quad i = 1, \dots, N.$$

Поэтому распределение вероятностей P_X является равномерным. \square

2 Совершенные шифры замены с неограниченным ключом

Определенная вероятностная модель шифра Σ_B позволяет рассматривать в качестве множества открытых текстов X лишь последовательности в некотором конечном алфавите A , длины которых ограничены некоторой заранее определенной константой. В работе [4] приводятся модели шифров замены с ограниченным и неограниченным ключом, для которых, в частности, на множество X такое ограничение не накладывается. Поскольку, в общем случае, шифр замены с ограниченным ключом совершенным не является (см. [4]), будем рассматривать шифр замены с неограниченным ключом. Такая математическая модель имеет ряд полезных свойств, например, она позволяет строить модели совершенных шифров и оптимальных кодов аутентификации, стойких к имитации и подмене [2, 3, 5].

Приведем модель данного шифра.

Пусть U — конечное множество возможных «шифрвеличин», а V — конечное множество возможных «шифробозначений». Пусть также имеются r ($r > 1$) инъективных отображений из U в V . Пронумеруем данные отображения:

E_1, E_2, \dots, E_r . Данные отображения называются простыми заменами. Обозначим $\mathbb{N}_r = \{1, 2, \dots, r\}$. Опорным шифром замены назовем совокупность $\Sigma = (U, \mathbb{N}_r, V, E, D)$, обладающую следующими свойствами:

- (1) для любых $u \in U$ и $j \in \mathbb{N}_r$ выполнено равенство $D_j(E_j(u)) = u$;
- (2) $V = \bigcup_{j \in \mathbb{N}_r} E_j(U)$.

При этом $E = \{E_1, \dots, E_r\}$, $D = \{D_1, \dots, D_r\}$, $D_j : E_j(U) \rightarrow U$, $j \in \mathbb{N}_r$. l -й степенью опорного шифра Σ назовем совокупность

$$\Sigma^l = \left(U^l, \mathbb{N}_r^l, V^l, E^{(l)}, D^{(l)} \right),$$

где U^l, \mathbb{N}_r^l, V^l — декартовы степени соответствующих множеств U, \mathbb{N}_r, V . Множество $E^{(l)}$ состоит из отображений $E_{\bar{j}} : U^l \rightarrow V^l$, $\bar{j} \in \mathbb{N}_r^l$, таких, что для любых $\bar{u} = u_1 \dots u_l \in U^l$, $\bar{j} = j_1 \dots j_l \in \mathbb{N}_r^l$ выполнено равенство

$$E_{\bar{j}}(\bar{u}) = E_{j_1}(u_1) \dots E_{j_l}(u_l) = v_1 \dots v_l \in V^l,$$

а множество $D^{(l)}$ состоит из отображений $D_{\bar{j}} : E_{\bar{j}}(U^l) \rightarrow U^l$, $\bar{j} \in \mathbb{N}_r^l$, таких, что для любых $\bar{v} = v_1 \dots v_l \in V^l$, $\bar{j} = j_1 \dots j_l \in \mathbb{N}_r^l$ выполнено равенство

$$D_{\bar{j}}(\bar{v}) = D_{j_1}(v_1) \dots D_{j_l}(v_l) = u_1 \dots u_l \in U^l.$$

Отметим такой важный момент. В ряде случаев не всякое слово длиной l в алфавите U может появиться в открытом тексте. Поэтому обозначим через $U^{(l)}$ подмножество всех таких слов во множестве U^l , появление которых в открытом тексте имеет ненулевую вероятность:

$$U^{(l)} = \left\{ \bar{u} \in U^l \mid P_{U^l}(\bar{u}) > 0 \right\}.$$

Тогда

$$V^{(l)} = \bigcup_{\bar{j} \in \mathbb{N}_r^l} E_{\bar{j}}(U^{(l)}).$$

Пусть ψ_c — случайный генератор ключевого потока, который для любого натурального числа l вырабатывает случайный ключевой поток j_1, \dots, j_l , где все $j_i \in \mathbb{N}_r$.

Обозначим через Σ_H^l следующую совокупность величин:

$$\Sigma_H^l = \left(U^{(l)}, \mathbb{N}_r^l, V^{(l)}, E^{(l)}, D^{(l)}, P_{U^{(l)}}, P_{\mathbb{N}_r^l} \right).$$

Шифром замены с неограниченным ключом назовем семейство

$$\Sigma_H = \left(\Sigma_H^l, l \in \mathbb{N}; \psi_c \right).$$

При этом независимые и не содержащие нулевых вероятностей распределения $P_{U^{(l)}}$ и $P_{\mathbb{N}_r^l}$ индуцируют распределения вероятностей на множестве $V^{(l)}$:

$$P_{V^{(l)}}(\bar{v}) = \sum_{(\bar{u}, \bar{j}) \in U^{(l)} \times \mathbb{N}_r^l, E_{\bar{j}}(\bar{u}) = \bar{v}} P_{U^{(l)}}(\bar{u}) P_{\mathbb{N}_r^l}(\bar{j}).$$

Также определим условные вероятности $P_{U^{(l)}|V^{(l)}}(\bar{u}|\bar{v})$ и $P_{V^{(l)}|U^{(l)}}(\bar{v}|\bar{u})$:

$$P_{V^{(l)}|U^{(l)}}(\bar{v}|\bar{u}) = \sum_{\bar{j} \in \mathbb{N}_r^l(\bar{u}, \bar{v})} P_{\mathbb{N}_r^l}(\bar{j}); \quad P_{U^{(l)}|V^{(l)}}(\bar{u}|\bar{v}) = \frac{P_{U^{(l)}}(\bar{u}) P_{V^{(l)}|U^{(l)}}(\bar{v}|\bar{u})}{P_{V^{(l)}}(\bar{v})},$$

где $\mathbb{N}_r^l(\bar{u}, \bar{v}) = \{\bar{j} \in \mathbb{N}_r^l \mid E_{\bar{j}}(\bar{u}) = \bar{v}\}$.

Говорят, что шифр Σ_H является совершенным тогда и только тогда, когда для любого натурального l шифр Σ_H^l является совершенным по Шеннону.

Предложение 7. Для шифра Σ_H следующие условия эквивалентны:

- (i) для любого $l \in \mathbb{N}$ и любых $\bar{u} \in U^{(l)}$, $\bar{v} \in V^{(l)}$ выполнено равенство $P_{U^{(l)}|V^{(l)}}(\bar{u}|\bar{v}) = P_{U^{(l)}}(\bar{u})$;
- (ii) для любого $l \in \mathbb{N}$ и любых $\bar{u} \in U^{(l)}$, $\bar{v} \in V^{(l)}$ выполнено равенство $P_{V^{(l)}|U^{(l)}}(\bar{v}|\bar{u}) = P_{V^{(l)}}(\bar{v})$;
- (iii) для любого $l \in \mathbb{N}$ и любых $\bar{u}_1, \bar{u}_2 \in U^{(l)}$, $\bar{v} \in V^{(l)}$ выполнено равенство $P_{V^{(l)}|U^{(l)}}(\bar{v}|\bar{u}_1) = P_{V^{(l)}|U^{(l)}}(\bar{v}|\bar{u}_2)$.

Предложение 8. Пусть шифр замены с неограниченным ключом Σ_H является совершенным. Тогда данный шифр обладает следующими свойствами:

- (i) для любого натурального числа l и любых $\bar{u} \in U^{(l)}$, $\bar{v} \in V^{(l)}$ найдется такой ключевой поток $\bar{j} \in \mathbb{N}_r^l$, что $E_{\bar{j}}(\bar{u}) = \bar{v}$;
- (ii) для любого натурального числа l справедливо двойное неравенство $|U^{(l)}| \leq |V^{(l)}| \leq |\mathbb{N}_r^l|$.

Теорема 5 (достаточные условия совершенности шифра Σ_H [2]). Пусть шифр замены Σ_H удовлетворяет следующим условиям:

- (i) правила зашифрования E_1, E_2, \dots, E_r шифра Σ_H обладают тем свойством, что для любых $u \in U$, $v \in V$ найдется, и притом единственный, элемент $j = j(u, v) \in \mathbb{N}_r$, такой что $E_j(u) = v$;
- (ii) распределение вероятностей $P_{\mathbb{N}_r}$ является равномерным.

Тогда шифр Σ_H является совершенным, причем для любого $l \in \mathbb{N}$ выполнено равенство $|V^{(l)}| = r^l$ и распределение вероятностей $P_{V^{(l)}}$ является равномерным.

Теорема 6. Пусть для шифра Σ_H выполнено равенство $|U| = |\mathbb{N}_r| = |V|$. Шифр Σ_H является совершенным тогда и только тогда, когда выполнены следующие условия:

- (i) правила зашифрования E_1, E_2, \dots, E_r шифра Σ_H обладают тем свойством, что для любых $u \in U$, $v \in V$ найдется, и притом единственный, элемент $j = j(u, v) \in \mathbb{N}_r$, такой что $E_j(u) = v$;
- (ii) распределение вероятностей $P_{\mathbb{N}_r}$ является равномерным.

Доказательство следует из теоремы Шеннона и теоремы 5. \square

Пусть для шифра Σ_H выполнено равенство $|U| = |\mathbb{N}_r| = |V|$ и распределение вероятностей $P_{\mathbb{N}_r}$ является равномерным. Тогда из теоремы 6 следует, что шифр Σ_H является совершенным тогда и только тогда, когда матрица зашифрования опорного шифра для Σ_H

$\mathbb{N}_r \setminus U$	u_1	\dots	u_r
1	$E_1(u_1)$	\dots	$E_1(u_r)$
\vdots	\vdots	\vdots	\vdots
r	$E_r(u_1)$	\dots	$E_r(u_r)$

является латинским квадратом (подробнее о латинских квадратах см. [6]). Поэтому шифры табличного и модульного гаммирования с равновероятной гаммой являются совершенными. О применении латинских квадратов в криптографии можно посмотреть в [7].

Рассмотрим еще один критерий совершенных шифров замены с неограниченным ключом в классе шифров с равномерным распределением вероятностей на множестве \mathbb{N}_r .

Теорема 7 [3]. Пусть для шифра Σ_H выполнены неравенства $|U| \leq |V| \leq |\mathbb{N}_r|$ и распределение вероятностей $P_{\mathbb{N}_r}$ является равномерным. Шифр Σ_H является совершенным тогда и только тогда, когда выполнены следующие условия:

- (i) для любых $u \in U$ и $v \in V$ найдется такое $j \in \mathbb{N}_r$, что $E_j(u) = v$;
- (ii) для любых $u_1, u_2 \in U$, $v \in V$ выполнено равенство $|\mathbb{N}_r(u_1, v)| = |\mathbb{N}_r(u_2, v)|$.

Следствие 3. Пусть для шифра Σ_H выполнено соотношение $|U| \leq |V| = |\mathbb{N}_r|$ и распределение вероятностей $P_{\mathbb{N}_r}$ является равномерным. Шифр Σ_H является совершенным тогда и только тогда, когда $|\mathbb{N}_r(u, v)| = 1$ для любых $u \in U$ и $v \in V$.

3 Совершенные имитостойкие шифры

Рассмотрим вероятностное пространство $(\Omega = K, F_K, P_K)$. Зафиксируем $y \in Y$. Обозначим через $K(y)$ следующее множество: $K(y) = \{k \in K \mid y \in E_k(X)\}$. Под обозначением $K(y)$ будем также понимать событие $(K(y) \in F_K)$, заключающееся в том, что при случайном выборе ключа $k \in K$ шифрованный текст y можно расшифровать на ключе k , т.е. $y \in E_k(X)$. Тогда событию $K(y)$ будут благоприятствовать все элементы из множества $K(y)$, и только они. Поэтому

$$P(K(y)) = \sum_{k \in K(y)} P_K(k).$$

Если канал связи готов к работе и на приеме установлены действующие ключи, но в данный момент времени никакого сообщения не передается, то в этом случае противником может быть предпринята попытка имитации сообщения. Тогда вероятность успеха имитации определяется следующим образом:

$$P_{im} = \max_{y \in Y} P(K(y)).$$

Если же в данный момент передается некоторое сообщение $y \in Y$ (которое получено из открытого текста $x \in X$ на ключе $k \in K$), то противник может заменить его на $\tilde{y} \in Y$, отличное от y . При этом он будет рассчитывать на то, что на действующем ключе k криптограмма \tilde{y} будет воспринята как некий осмысленный открытый текст \tilde{x} , отличный от x . Пусть $K(\tilde{y}) \mid K(y)$ — событие, заключающееся в попытке подмены сообщения y сообщением \tilde{y} . Применяя теорему о произведении вероятностей, получаем, что

$$P(K(\tilde{y}) \mid K(y)) = \frac{P(K(y) \cap K(\tilde{y}))}{P(K(y))} = \frac{\sum_{k \in K(y, \tilde{y})} P_K(k)}{\sum_{k \in K(y)} P_K(k)},$$

где $K(y, \tilde{y}) = K(y) \cap K(\tilde{y})$. Тогда вероятность успеха подмены сообщения будет вычисляться по следующей формуле:

$$P_{podm} = \max_{y, \tilde{y} \in Y, y \neq \tilde{y}} P(K(\tilde{y}) \mid K(y)).$$

Теорема 8 [4]. Для любого шифра Σ_B справедливы неравенства

$$P_{im} \geq \frac{|X|}{|Y|}; \quad P_{podm} \geq \frac{|X| - 1}{|Y| - 1}.$$

При этом $P_{im} = |X|/|Y|$ тогда и только тогда, когда для любого $y \in Y$ выполнено равенство $P(K(y)) = |X|/|Y|$. Также $P_{podm} = (|X| - 1)/(|Y| - 1)$

тогда и только тогда, когда для любых $y, \tilde{y} \in Y$, $y \neq \tilde{y}$, выполнено равенство $P(K(\tilde{y}) | K(y)) = (|X| - 1) / (|Y| - 1)$.

Далее везде предполагается, что для любого натурального l выполнены равенства $U^{(l)} = U^l$, $V^{(l)} = V^l$. Для шифра замены с неограниченным ключом Σ_H обозначим через P_{im}^l вероятность имитации сообщения для шифра Σ_H^l , а через $P_{\text{podm}}^l(s)$ — вероятность подмены в сообщении длиной l ровно s символов для шифра Σ_H^l , где $s \leq l$. Из теоремы 8 следует, что если для некоторого шифра Σ_H выполнено равенство $|U| = |V|$, то $P_{\text{im}}^l = P_{\text{podm}}^l(s) = 1$ для любого натурального l и любого $s \leq l$, т. е. такие шифры максимально уязвимы к угрозам имитации и подмены сообщения. Приведем некоторые конструкции шифра замены с неограниченным ключом, в котором одновременно обеспечивается стойкость и имитостойкость. При этом матрица зашифрования строится только для опорного шифра, поэтому ее размерность не зависит от длины открытого текста.

Будем говорить, что матрица $A = A(m, n)$, $m \geq n$, над некоторым m -элементным множеством S является латинским прямоугольником, если каждый столбец матрицы A является перестановкой элементов множества S , причем в строках каждый элемент встречается не более одного раза.

Предложение 9 [2]. Пусть $A = A(n+1, n)$ — некоторая матрица над множеством «шифробозначений» $V = \{v_1, \dots, v_{n+1}\}$, которая является латинским прямоугольником, и пусть матрица A является матрицей зашифрования для опорного шифра замены с неограниченным ключом Σ_H . Пусть также случайный генератор ключевых последовательностей ψ_c из конструкции шифра Σ_H имеет равномерное распределение. Тогда для любого натурального l шифр Σ_H^l является совершенным по Шеннону и выполнены следующие равенства:

$$P_{\text{im}}^l = \left(\frac{n}{n+1} \right)^l ; \quad P_{\text{podm}}^l(s) = \left(\frac{n-1}{n} \right)^s .$$

Пусть S_n — симметрическая группа степени n , $T^j \in S_n$ — циклическая перестановка на j позиций влево. Обозначим через $A_j = A_j(n, 2)$ матрицу размера $n \times 2$ над множеством \mathbb{N}_n , имеющую такой вид:

$$A_j = \begin{pmatrix} 1 & 2 & \dots & n \\ T^j(1) & T^j(2) & \dots & T^j(n) \end{pmatrix}^T, \quad j = 1, \dots, n-1.$$

Из матриц A_j , $j = 1, \dots, n-1$, составим матрицу $M = M(n^2 - n, 2)$ размера $(n^2 - n) \times 2$ путем последовательной графической записи матриц A_1, \dots, A_{n-1} одной под другой.

Предложение 10 [3]. Пусть $M = M(n^2 - n, 2)$ — матрица над множеством $V = \{v_1, \dots, v_n\}$, построенная выше, $r = n^2 - n$, $|U| = 2$ и пусть матрица M является матрицей зашифрования для опорного шифра замены

с неограниченным ключом Σ_H . Пусть также случайный генератор ключевых последовательностей ψ_c из конструкции шифра Σ_H имеет равномерное распределение. Тогда для любого натурального l шифр Σ_H^l является совершенным по Шеннону и выполнены следующие равенства:

$$P_{\text{im}}^l = \left(\frac{2}{n}\right)^l; \quad P_{\text{podm}}^l(s) = \left(\frac{1}{n-1}\right)^s.$$

Заметим, что в предложениях 9 и 10 $P_{\text{im}}^l \rightarrow 0$ при $l \rightarrow \infty$, $P_{\text{podm}}^l(s) \rightarrow 0$ при $s \rightarrow \infty$.

Литература

1. Алферов А. П., Зубов А. Ю., Кузьмин А. С., Черемушкин А. В. Основы криптографии. — М.: Гелиос АРВ, 2005. 480 с.
2. Рацев С. М. О совершенных имитостойких шифрах // Прикладная дискретная математика, 2012. № 3(17). С. 41–46.
3. Рацев С. М. О совершенных имитостойких шифрах замены с неограниченным ключом // Вестник Самарского гос. ун-та. Естественно-научная сер., 2013. № 9/1(110). С. 42–48.
4. Зубов А. Ю. Криптографические методы защиты информации. Совершенные шифры. — М.: Гелиос АРВ, 2005. 192 с.
5. Рацев С. М. Об оптимальных кодах аутентификации // Системы и средства информатики, 2013. Т. 23. № 1. С. 53–57.
6. Холл М. Комбинаторика / Пер. с англ. — М.: Мир, 1970. 424 с. (*Hall M. Combinatorial theory.* — Waltham, MA: Blaisdell Publ., 1967. 310 p.)
7. Тужилин М. Э. Латинские квадраты и их применение в криптографии // Прикладная дискретная математика, 2012. № 3(17). С. 47–52.

Поступила в редакцию 03.07.13

ON THEORETICALLY PERFECT CIPHERS

S. M. Ratseev

Department of Information Security and Control Theory, Faculty of Mathematics and Information Technologies, Ulyanovsk State University, 42 Lev Tolstoy Str., Ulyanovsk 432017, Russian Federation

Abstract: K. Shannon in the 1940s introduced the concept of a perfect cipher, which provides the best protection of plaintexts. Perfect secrecy means that

a cryptanalyst can obtain no information about the plaintext by observing the ciphertext. It is well known that the Vernam cipher with equiprobable gamma is a perfect cipher but it is not imitation resistant because it uses equipotent alphabets for plaintexts and ciphertexts. On the basis of A. Yu. Zubov's mathematical model of the substitution cipher with the unbounded key, the models of perfect and imitation resistant ciphers were constructed. In this case, the bearing cipher of the model is perfect and reaches the lower bounds for the probability of success of imitation and substitution of ciphertexts. The paper also gives some generalizations of the classical theory of K. Shannon on perfect ciphers.

Keywords: cipher; perfect cipher; imitation resistant ciphers

DOI: 10.14357/08696527140103

References

1. Alferov, A. P., A. Yu. Zubov, A. S. Kuz'min, and A. V. Cheremushkin. 2005. *Osnovy kriptografii* [Foundations of cryptography]. Moscow: Gelios ARV. 480 p.
2. Ratseev, S. M. 2012. O sovershennykh imitostoykikh shifrakh [On perfect imitation resistant ciphers]. *Prikladnaya Diskretnaya Matematika* [Applied Discrete Mathematics] 17(3):41–47.
3. Ratseev, S. M. 2013. O sovershennykh imitostoykikh shifrakh zameny s neogranichenym klyuchom [On perfect imitation resistant ciphers with unbounded key]. *Vestnik Samarskogo Gosudarstvennogo Universiteta* [Vestnik Samara State University] 110(9/1):42–48.
4. Zubov, A. Yu. 2005. *Kriptograficheskie metody zashchity informatsii. Sovershennyye shifry* [Cryptographic methods of information security. Perfect ciphers]. Moscow: Gelios ARV. 192 p.
5. Ratseev, S. M. 2013. Ob optimal'nykh kodakh autentifikatsii [On optimal authentication code]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 23(1):53–57.
6. Holl, M. 1967. *Combinatorics*. Waltham, MA: Blaisdell Publ., 1967. 310 p.
7. Tuzhilin, M. E. 2012. Latinskie kvadraty i ikh primeneniye v kriptografii [Latin squares and their applications in cryptography]. *Prikladnaya Diskretnaya Matematika* [Applied Discrete Mathematics] 17(3):47–52.

Received July 03, 2013

Contributor

Ratseev Sergey M. (b. 1979) — Candidate of Science (PhD) in physics and mathematics, associate professor, Department of Information Security and Control Theory, Faculty of Mathematics and Information Technologies, Ulyanovsk State University, 42 Lev Tolstoy Str., Ulyanovsk 432017, Russian Federation; RatseevSM@mail.ru

МЕТОДОЛОГИЯ МУТАЦИОННОГО ТЕСТИРОВАНИЯ ДЛЯ НАЗЕМНЫХ ИСПЫТАНИЙ БОРТОВОЙ АППАРАТУРЫ КОСМИЧЕСКИХ АППАРАТОВ

*Д. А. Недорезов¹, А. И. Легалов², О. В. Непомнящий³, С. С. Красненко⁴,
А. В. Анкудинов⁵*

Аннотация: Рассматриваются методы повышения эффективности формирования тестов, используемых при проведении наземных испытаний бортовой аппаратуры космических аппаратов. Предложен подход, позволяющий применять мутационное тестирование для аппаратно-программных средств, разрабатываемых на основе динамически реконфигурируемых интегральных схем. Рассмотрены два способа моделирования аппаратных мутаций для тестирования. Первый подход основан на классическом использовании средств автоматизации моделирования электронных приборов (EDA — electronic design automation) с автоматическим формированием пакетов тестов. Второй способ использования мутационного тестирования может опираться на применение динамически реконфигурируемых архитектур в аппаратной части проектируемого устройства в сочетании с программными средствами, имитирующими логику работы реального объекта испытаний. Описаны особенности создания аппаратной и программной составляющих тестовых комплексов, разрабатываемых на основе предлагаемой методики. Определены основные пути развития предлагаемого подхода.

Ключевые слова: мутационное тестирование; языки описания аппаратуры; наземные испытания; аппаратура; комплекс

DOI: 10.14357/08696527140104

1 Введение

Постоянное повышение сложности бортовой радиоэлектронной и вычислительной аппаратуры космических аппаратов требует проведения сложных и тщательных наземных испытаний для обнаружения и устранения возможных ошибок и недоработок. Такие испытания требуют больших временных затрат, что при растущих объемах производства критически влияет на экономическую эффективность проектов. Для решения этих задач создаются системы наземных

¹Сибирский федеральный университет, Красноярск, nedorezovd@mail.ru

²Сибирский федеральный университет, Красноярск, legalov@mail.ru

³Сибирский федеральный университет, Красноярск, olegn_68@mail.ru

⁴Сибирский федеральный университет, Красноярск, t_150@list.ru

⁵ОАО «Информационные спутниковые системы» им. академика М. Ф. Решетнева, Железногорск, nd@iss-reshetnev.ru

испытаний, в которых используются принципы и технологии модульного построения аппаратного и программного обеспечения для систем тестирования, гибко настраиваемых на класс решаемых задач.

Как правило, программно-аппаратные комплексы для наземных испытаний реализуются на базе «готовых решений», например на основе платформ ADLINK или National Instruments [1, 2]. При этом для имитации отдельных частей тестируемой аппаратуры применяют реконфигурируемые однокристалльные модули, реализуемые с использованием программируемых логических интегральных схем (FPGA). Поскольку логика функционирования FPGA может быть представлена на языке описания аппаратуры, то это позволяет применять для отладки оборудования методы тестирования, аналогичные тем, что используются и при программировании с применением универсальных языков.

В этом случае на первый план выступает качество создаваемых тестов, их способность покрыть множество некорректных ситуаций, которые при создании аппаратуры могут возникать не только вследствие ошибок разработчиков, но и в ходе эксплуатации, например из-за воздействия радиации и ионизирующего излучения. Последнее особенно важно для бортовой аппаратуры, эксплуатируемой в сложных условиях, так как ошибки в топологии возникают не по вине разработчиков, а вследствие внешних факторов во время эксплуатации. Поэтому весьма важно, чтобы тестирование аппаратуры могло выявить возникновение некорректных модификаций схемных решений.

2 Мутационное тестирование

Для проверки качества тестов и анализа покрытия ими всевозможных некорректных ситуаций используются различные подходы, среди которых можно выделить метод мутационного тестирования [3–5], основанный на включении в код программы небольших изменений. Результаты этих изменений называются мутациями, которые происходят при использовании мутационных операторов, имитирующих типичные ошибки программистов (например, использование неправильной операции или имени переменной). В разрабатываемую программу P вносят изменения (мутации), т. е. искусственно создают программы-мутанты P_1 , P_2 , P_3 и т. д. Затем программа P и ее мутанты тестируются на одном и том же наборе тестов. Если на этом наборе тестов подтверждается правильность программы P и выявляются все ошибки в программах-мутантах, то набор тестов соответствует мутационному критерию, а программа P объявляется правильной. Если в некоторых мутантах не были выявлены все мутации, то набор тестов рассматривается как недостаточный и требует расширения.

3 Применение метода мутационного тестирования на аппаратном уровне

До недавнего времени мутационное тестирование не имело широкого распространения, что во многом обуславливалось высокими требованиями к вы-

числительным ресурсам, используемым для анализа покрытия тестов. Однако применение многоядерных процессоров и графических ускорителей способствовало появлению и массовому использованию высокопроизводительных вычислительных систем в различных областях, что позволило интенсивно внедрять в практику и ресурсоемкие методы тестирования программ.

Дополнительным аргументом, указывающим на перспективность использования мутационного тестирования на аппаратном уровне, является различие в уровне базовых операций между универсальными языками программирования и языками описания аппаратуры. При традиционном программировании не вполне очевидно, какие операции имеет смысл подвергать мутации, так как их возможная замена на некорректные программистом зачастую видится маловероятной. Но при разработке топологии интегральных схем, когда осуществляются манипуляции на уровне вентиляционных операций, ошибки, вносимые как разработчиками, так и внешними факторами во время эксплуатации, становятся более очевидными. Более того, во многих случаях мутации могут быть сгенерированы не вручную, а автоматически. Следует также отметить, что непосредственная реализация тестовых мутаций на аппаратном уровне практически невозможна из-за необходимости изменения схемных решений, которые практически являются ненужными в реальном изделии. Поэтому макетирование мутаций является нецелесообразным. Вместо этого предлагаются два подхода к моделированию аппаратных мутаций для тестирования.

Первый подход основан на классическом использовании средств EDA. В этом случае в проектируемую систему можно вносить мутации на различных уровнях ее описания, получая тем самым схемы-мутанты, и в автоматическом режиме формировать пакеты тестов для каждого конкретного случая на требуемом уровне тестирования. Такой подход обладает большой гибкостью и обеспечивает максимальное покрытие с минимальными аппаратными затратами, поскольку все мутации обрабатываются в режиме симуляции функционирования реального прибора. Проблемы использования данного подхода могут быть связаны с отсутствием систем моделирования электронных схем с поддержкой мутационного тестирования.

Другой подход к использованию мутационного тестирования может опираться на применение динамически реконфигурируемых архитектур в аппаратной части проектируемого устройства в сочетании с программными средствами, имитирующими логику работы реального объекта испытаний. Используемое на этапе тестирования программное обеспечение (ПО) является более гибким носителем логических структур, нежели логика, реализованная аппаратным способом. Мутация аппаратной части может выполняться в базе программируемой логической интегральной схемы с реконфигурируемой архитектурой, например FPGA, топология которой представляется на языках описания аппаратуры (HDL-языках), которые, по сути, являются языками программирования. К таким языкам относятся VHDL, Verilog, SystemC и др. HDL-языки получили широкое распространение и на сегодняшний день являются стандартом

де-факто при проектировании однокристальных систем с перестраиваемой архитектурой [6].

Современные FPGA функционируют на скоростях, зачастую превосходящих скорости функционирования бортовой аппаратуры. При этом логические емкости этих микросхем позволяют реализовывать сложнейшие устройства комбинационной логики, микропроцессорные системы на кристалле, счетчики, таймеры, генераторы импульсов, измерители частотно-временных параметров и импульсных последовательностей, аппаратные вычислительные и коммуникационные устройства и др. Таким образом, при использовании FPGA в качестве базы для испытательного комплекса остается только формализовать логику поведения прибора и включить в нее аппаратные мутации.

4 Программирование тестовых комплексов

При организации системы мутационного тестирования ПО комплекса тестирования делится на две составляющие: Host и Target. Host — управляющая программа, служащая интерфейсом между пользователем и FPGA, используемой для имитации необходимого оборудования. Этот программный модуль располагается в памяти управляющей вычислительной машины. Через Host осуществляется общее управление содержащейся в FPGA логикой, имитирующей работу реального объекта испытаний. Target — программный модуль, реализуемый в FPGA и управляемый из Host. В Host реализуется часть ПО, которая не требует реального времени, работает под управлением обычной операционной системы и реализует основной алгоритм программы. В Target же реализованы алгоритмы, к которым предъявляются требования реальных промежутков времени.

Предполагается, что во время испытаний логика FPGA мутирует в динамическом режиме согласно требуемому заданию, тем самым порождая нештатные ситуации в алгоритме функционирования прибора. Разумеется, что такой подход возможен только в режиме динамической перестройки FPGA, а реконфигурация должна осуществляться после прохождения очередной порции тестов, определяющих работоспособность текущей мутации. В исключительных ситуациях, например для медленно протекающих процессов, перестройка FPGA может осуществляться непосредственно в процессе функционирования (в промежутках между повторами тестовых комплексов). Помимо этого могут происходить мутации и встроенного ПО, что позволяет осуществлять комплексный анализ покрытия тестами всего аппаратно-программного комплекса. Например, при наличии на кристалле управляющего процессора можно вносить требуемые мутации в его ПО.

Для формирования пакета тестов и разработки процедур тестирования мутации можно разделить по определенной системе, т. е. выделить типы мутаций. При обнаружении ошибки определенного типа в процессе тестирования можно локализовать неотработанный уровень в системной иерархии прибора или модуль

объекта испытаний и, таким образом, выявить потенциально недоработанный узел.

5 Заключение

Использование мутационного тестирования для анализа тестов аппаратно-программных средств позволяет повысить достоверность тестов, что является одной из важных проблем при разработке аппаратуры ответственного применения, функционирующей в агрессивной среде. Применение однокристалльных систем с динамически перестраиваемой архитектурой позволяет гибко реализовать аппаратные мутации. Кроме того, применение предлагаемых подходов для автоматического формирования тестовых воздействий позволит не только расширить номенклатуру тестов, но и перейти на качественно новый уровень проведения наземных испытаний бортовой радиоэлектронной аппаратуры.

Литература

1. *Трэвис Дж., Кринг Дж.* LabVIEW для всех / Пер. с англ. — СПб: ДМК Пресс, 2011. 904 с. (*Travis J., Kring J.* LabView for everyone: Graphical programming made easy and fun. — 3rd ed. — Prentice Hall, 2006. 1032 p.)
2. *Гуров В.* Аппаратные средства Compact PCI производства ADLINK Technology // Современные технологии автоматизации, 2011. № 2. С. 30–36.
3. *DeMillo R. A., Lipton R. J., Sayward F. G.* Hints on test data selection: Help for the practicing programmer // IEEE Computer, 1978. Vol. 11. No. 4. P. 34–41.
4. *Budd T. A.* Mutation analysis of program test data: PhD thesis. — New Haven, CT: Yale University, 1980.
5. *Offutt A. J., Lee S. D.* An empirical evaluation of weak mutation // IEEE Trans. Software Eng., 1994. Vol. 20. No. 5. P. 337–344.
6. *Непомнящий О. В., Алекминский С. Ю.* Проблемы верификации проекта при сквозном проектировании вычислительных систем на кристалле // Нано- и микросистемная техника, 2010. № 9. С. 4–7.

Поступила в редакцию 22.04.13

MUTATION TEST METHODOLOGY FOR ONBOARD SPACECRAFTS RADIOELECTRONIC EQUIPMENT GROUND TEST

*D. A. Nedorezov¹, A. I. Legalov¹, O. V. Nepomnjashhij¹, S. S. Krasnenko¹,
and A. V. Ankudinov²*

¹Siberian Federal University, 79/10 Svobodny Pr., Krasnoyarsk 660041, Russian Federation

²JSC “Information Satellite Systems” named after academician M. F. Reshetnev, 52 Lenin Str., Zheleznogorsk, Krasnoyarsk Region 662972, Russian Federation

Abstract: Improving methods of tests developing effectiveness that are applied in onboard spacecrafts radioelectronic equipment ground test are considered. The approach that allows using mutation testing for hardware and software tools, based on Field-Programmable Gate Arrays is suggested. Two ways of hardware mutations modeling for testing are considered. The first approach is based on classical use of electronic design automation systems with automatic test suites generation. The second approach of mutation testing application can use dynamically reconfigurable architectures (field-programmable gate arrays) in hardware of developed device in combination with software that imitates logic of real test object. The features of hardware and software components of developed test complexes based on the proposed method are described. General ways of introduced approach advance are specified.

Keywords: mutation testing; hardware description languages; ground test; equipment; complexes

DOI: 10.14357/08696527140104

References

1. Travis, J., and J. Kring. 2006. *LabVIEW for everyone: Graphical programming made easy and fun*. 3rd ed. Prentice Hall. 1032 p.
2. Gurov, I. 2011. Apparatus sredstva Compact PCI proizvodstva ADLINK Technology [ADLINK Technology Compact PCI hardware]. *Sovremennye Tekhnologii Avtomatizatsii* [Modern Automation Technology] 2:30–36.
3. Budd, T. A. 1980. Mutation analysis of program test data. PhD Thesis. New Haven CT, Yale University, 1980.
4. DeMillo, R. A., R. J. Lipton, and F. G. Sayward. 1978. Hints on test data selection: Help for the practicing programmer. *IEEE Computer* 11(4):34–41.
5. Offutt, A. J., and S. D. Lee. 1994. An empirical evaluation of weak mutation. *IEEE Trans. Software Eng.* 20:337–344.
6. Nepomnjashhij, O. V., and S. Ju. Alekminskij. 2010. Problemy verifikatsii proekta pri skvoznom proektirovanii vychislitel'nykh sistem na kristalle [Problems of project verification in through design of on-chip calculating systems]. *Nano i Mikrosistemnaya Tekhnika* [Nano and Microsystem Technics] 9:4–7.

Received April 22, 2013

Contributors

Nedorezov Dmitrij A. (b. 1984) — PhD student, Siberian Federal University, 79/10 Svobodny Pr., Krasnoyarsk 660041, Russian Federation; nedorezovd@mail.ru

Legalov Aleksandr I. (b. 1956) — Doctor of Science in technology, Head of Computer Facilities Department, Institute of Space and Information Technologies,

Siberian Federal University, 79/10 Svobodny Pr., Krasnoyarsk 660041, Russian Federation; legalov@mail.ru

Nepomnjashhij Oleg V. (b. 1968) — Candidate of Science (PhD) in technology, Professor, Head of Laboratory, Institute of Space and Information Technologies, Siberian Federal University, 79/10 Svobodny Pr., Krasnoyarsk 660041, Russian Federation; olegn_68@mail.ru

Krasnenko Sergej S. (b. 1988) — PhD student, Siberian Federal University, 79/10 Svobodny Pr., Krasnoyarsk 660041, Russian Federation; t_150@list.ru

Ankudinov Aleksandr V. (b. 1964) — Candidate of Science (PhD) in technology, Deputy Head of Department, JSC “Information Satellite Systems” named after academician M. F. Reshetnev, 52 Lenin Str., Zheleznogorsk, Krasnoyarsk Region 662972, Russia; nd@iss-reshetnev.ru

ДИНАМИЧЕСКОЕ СОЗДАНИЕ КАРТ УРОВНЯ WiFi-СИГНАЛОВ ДЛЯ СИСТЕМ ЛОКАЛЬНОГО ПОЗИЦИОНИРОВАНИЯ*

Р. В. Воронов¹, С. В. Малодушев²

Аннотация: Карты уровня сигналов используются для локации беспроводных мобильных устройств. Увеличение беспроводного сегмента сети требует создания новых карт уровня сигналов для осуществления локации в новых областях. Традиционно подобные карты формируются из большого количества измерений в определенных точках, подобный процесс занимает много времени. В статье исследована задача динамического создания карт уровня сигналов. Описан эвристический алгоритм динамического создания карт уровня сигналов во время эксплуатации системы локации. При оценивании средних значений уровня сигналов использована функция, являющаяся аналогом функции правдоподобия. Разработанный алгоритм позволяет автоматизировать процесс создания подобных карт. В качестве беспроводных сетей использовались WiFi-сети, работающие в частотном диапазоне 2,4 ГГц. Результаты тестов показали, что предложенный алгоритм может быть использован для решения данной задачи.

Ключевые слова: WiFi; базовые станции; опорная точка; уровень сигнала; локация; карта уровня сигналов

DOI: 10.14357/08696527140105

1 Введение

Как известно, спутниковые системы навигации невозможно использовать в закрытых помещениях — в зданиях, шахтах, подземных парковках, переходах метро и т. д. Для таких случаев разрабатываются и применяются специальные системы локации, основанные, например, на обработке уровня принимаемого сигнала от стационарных точек доступа (базовых станций) беспроводной сети.

*Исследования проводились в рамках деятельности МИП ООО «Наносети» и ЗАО «РТЛ-Сервис», а также поддерживались Петрозаводским государственным университетом (Программа стратегического развития ПетрГУ на 2012–2016 гг.), Министерством образования и науки РФ, Фондом СР МФП в НТС и Министерством экономического развития Республики Карелия (Программа «Разработка системы слежения за мобильными абонентами распределенной WiFi-сети» на 2013 г.), а также в рамках программы УМНИК «Анализ и проектирование WiFi-сетей в помещениях» на 2012–2013 гг.

¹Петрозаводский государственный университет, rvoronov@sampo.ru

²Петрозаводский государственный университет, svm@petrsu.ru

Предварительно помещение разбивается на множество опорных точек, для каждой из которых определяются карты уровня принимаемых сигналов от базовых станций [1–3].

Поскольку уровень принимаемого сигнала сильно зависит от числа отражений, любые изменения в помещении или смена ориентации антенн базовых станций приводят к необходимости заново настраивать систему локации — пересчитывать карты уровня сигналов для нескольких или всех опорных точек помещения. Это трудоемкая операция, так как она требует непосредственного обхода оператором системы опорных точек и определения для каждой из них среднего значения уровня сигнала. В рассматриваемой работе предполагается автоматическое формирование карт уровня принимаемых сигналов в ходе эксплуатации системы, что является существенной особенностью рассматриваемого подхода и отличает его от других подходов к решению задачи локации внутри помещений.

Для систем локации, основанных на WiFi-сетях, существуют ограничения: с одной стороны, в каждой опорной точке помещения нужно слышать достаточное для локации число базовых станций (желательно больше трех), с другой стороны, используемые базовыми станциями диапазоны частот по возможности не должны пересекаться, иначе количество интерферирующих сигналов будет высоко.

Введем необходимые обозначения. Пусть N — число опорных точек помещения, в которых может находиться мобильный объект, D — число базовых станций. На приемной стороне уровни сигналов от базовых станций являются дискретными величинами, поскольку радиопередача ведется в цифровой форме. Тогда в ходе настройки системы необходимо для каждой базовой станции d оценить распределение $p_d(x|j)$ условных вероятностей уровней сигнала x в опорных точках $j = 1, \dots, N$.

В ходе эксплуатации системы местоположение мобильного объекта определяется следующим образом. Пусть p_j — априорная вероятность нахождения объекта в опорной точке j ($j = 1, \dots, N$). Предположим, что зарегистрирован вектор уровней сигнала $\vec{x} = (x_1, \dots, x_d)$, где x_d — уровень сигнала от базовой станции d . Предполагая независимость уровней сигнала от различных базовых станций, апостериорную вероятность q_j нахождения объекта в опорной точке j можно найти по формуле:

$$q_j = \frac{p_j \prod_{d=1}^D p_d(x_d|j)}{\sum_{k=1}^N p_k \prod_{d=1}^D p_d(x_d|k)}. \quad (1)$$

Как правило, в качестве искомой опорной точки выбирается та, для которой апостериорная вероятность принимает наибольшее значение.

В данной статье предлагается эвристический метод, позволяющий без перенастройки оператором системы локации в ходе ее эксплуатации автоматически оценивать вероятности $p_d(x|j)$ для добавляемых в систему новых базовых станций, используя значения q_j , полученные на основе имеющихся («старых») базовых станций. Предложенный метод опробован в системе локации, основанной на WiFi-сетях.

2 Построение областей локации

Рассмотрим существующие частотные ограничения, накладываемые на количество используемых базовых станций. Сети WiFi работают в частотном диапазоне 2400–2495 МГц. Данный частотный диапазон делится на 14 каналов, несущая частота каждого из которых отличается на 5 МГц, при этом каждый радиоканал имеет фиксированный частотный диапазон шириной 22 МГц. Для того чтобы сигналы от базовых станций не интерферировали друг с другом, при построении WiFi-сетей их несущие частоты должны отличаться не менее чем на 25 МГц, т. е. в доступном для WiFi частотном диапазоне 2,4 ГГц возможна работа без взаимных помех лишь на трех радиоканалах [4, 5]. В данной статье не рассматривается случай использования диапазона 5 ГГц.

На рис. 1 представлен типичный пример деления WiFi-сети на радиозоны (а) и пример тех же радиозон, используемых при локации (б), — каждой базовой станции соответствует свой номер канала, определяющий частотный диапазон, и радиозона. Расчет позиции мобильного объекта возможен только в областях пересечения радиозон, являющихся областями локации. В данном случае областям локации соответствуют радиозоны с номерами каналов 6 и 11 (см. рис. 1, б), несущие частоты данных радиозон не интерферируют друг с другом. Из рис. 1, б видно, что для систем локации плотность размещения базовых станций выше в сравнении с традиционными WiFi-сетями, ввиду необходимости организации областей локации.

По мере удаления от источника излучения базовой станции радиосигнал затухает, ввиду чего границы радиозон определяются чувствительностью радиоприемника на стороне клиента.

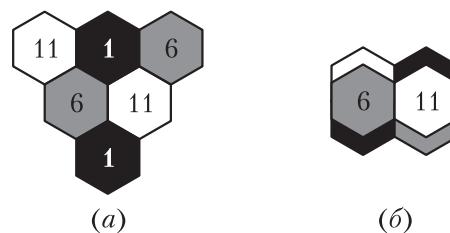


Рис. 1 Использование WiFi-каналов для организации радиозон

3 Применение карт уровня сигналов

Карты уровня сигналов используются в расчетах при локации, и их формирование является необходимой и ресурсоемкой задачей. Рассмотрим, как осуществляется создание и применение данных карт при расчетах в локации.

В момент времени t каждой опорной точке j соответствует уникальный вектор уровней сигналов $\bar{x} = (x_1, \dots, x_d)$ для радиозон соответствующих базовых станций. Для каждой базовой станции формируется карта, содержащая усредненные показания принятого от нее сигнала в заданном наборе опорных точек $j = 1, \dots, N$.

При локации осуществляется сравнение полученного в конкретной точке вектора уровней сигналов RSSI (Received Signal Strength Indication) со значениями, хранимыми во всех картах. По результатам данного сравнения определяется наиболее вероятная позиция мобильного объекта [6].

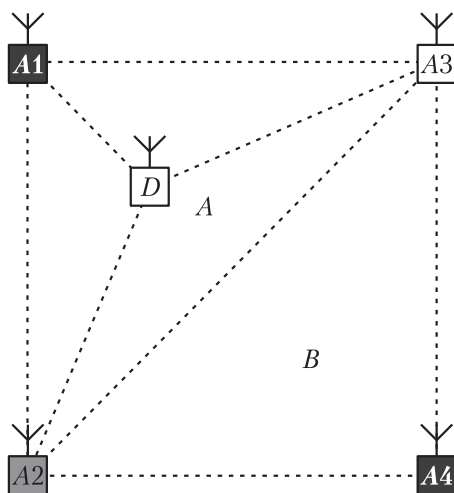


Рис. 2 Положение мобильного объекта относительно базовых станций

Для корректного расчета позиции мобильного объекта необходимо, чтобы данный объект принимал показания RSSI не менее чем от трех базовых станций, для которых сформированы карты уровня сигналов, при этом данный мобильный объект должен находиться внутри области, по периметру которой располагаются эти базовые станции (рис. 2).

На рис. 2 представлены две опорные точки — A и B . Локация мобильного объекта D в опорной точке A осуществляется на основе информации об уровнях сигналов от базовых станций $A1$, $A2$ и $A3$.

При локации в опорной точке A также используются показания уровней сигналов от базовой станции $A4$ в случаях, когда мобильный объект D принимает радиосигнал от данной базовой станции, при этом, ввиду частотных ограничений, базовые станции $A1$ и $A4$ по возможности не должны находиться в зоне слышимости друг друга, так как их сигналы будут интерферировать друг с другом.

Рассматриваемая модель позволяет формировать карты уровня сигналов в ходе эксплуатации системы с целью их дальнейшего использования при локации, что делает возможным автоматизировать процесс создания подобных карт и, как следствие, увеличить количество областей локации.

4 Математическая модель

Пусть имеется некоторое число базовых станций, для которых известны распределения вероятностей уровней сигнала в каждой опорной точке помещения. Тогда, обрабатывая по формуле (1) для мобильного объекта уровни сигналов от базовых станций, можно определять вероятности расположения объекта в опорных точках помещения. Пусть добавляется новая базовая станция, для которой не определены распределения вероятностей уровней сигналов в опорных точках.

Предположим для простоты, что уровни сигнала от новой базовой станции являются непрерывными случайными величинами ξ_1, \dots, ξ_M , которые в результате M независимых измерений приняли значения x_1, \dots, x_M .

В модели примем плотности распределений уровней сигналов в опорных точках j равными

$$p(x|j; a_j) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - a_j)^2}{2\sigma^2}\right), \quad j = 1, \dots, N, \quad (2)$$

где параметр σ является известным и одинаковым для всех опорных точек j . Теоретически в этом случае допускается ненулевая вероятность для отрицательных значений уровня сигналов, но она является ничтожно малой. Параметры a_j — математические ожидания значений уровня сигнала в опорных точках $j = 1, \dots, N$ — априори не известны. Задача состоит в поиске их точечных оценок.

Пусть q_{ij} — известные вероятности нахождения объекта в опорной точке j в момент регистрации i -го измерения, определенные по формуле (1).

Плотность распределения уровня сигнала x в момент регистрации i -го измерения вычисляется по формуле полной вероятности:

$$p_i(x) = \sum_{j=1}^N p(x|j; a_j) q_{ij}.$$

Составим функцию, являющуюся аналогом логарифмической функции правдоподобия для измерений x_1, \dots, x_M :

$$L(\bar{x}; \bar{a}) = \sum_{i=1}^M \ln \left(\sum_{j=1}^N p(x_i|j; a_j) q_{ij} \right), \quad (3)$$

где $\bar{x} = (x_1, \dots, x_M)$ и $\bar{a} = (a_1, \dots, a_N)$. В случае, если вероятности q_{ij} меняются от измерения к измерению, случайные величины ξ_1, \dots, ξ_M будут распределены по-разному, и в строгом смысле функция (3) не является функцией правдоподобия. Задача оценивания параметров в такой постановке сама по себе

является нетривиальной. Предлагается оценивать вектор \bar{a} при помощи поиска максимума этой функции.

Для поиска максимума функции (3) найдем ее частные производные по a_k :

$$\frac{\partial L(\bar{x}, \bar{a})}{\partial a_k} = \sum_{i=1}^M \frac{(dp(x_i|k; a_k)/(da_k)) q_{ij}}{\sum_{j=1}^N p(x_i|j; a_j) q_{ij}}, \quad k = 1, \dots, N. \quad (4)$$

Учитывая (2), получаем производные $p(x|k; a_k)$ по a_k :

$$\frac{dp(x|k; a_k)}{da_k} = \frac{1}{\sigma^3 \sqrt{2\pi}} (x - a_k) \exp\left(-\frac{(x - a_k)^2}{2\sigma^2}\right), \quad k = 1, \dots, N.$$

Подставляем их в (4), получаем:

$$\frac{\partial L(\bar{x}, \bar{a})}{\partial a_k} = \sum_{i=1}^M \frac{(1/\sigma^2) (x_i - a_k) \exp(-(x_i - a_k)^2/(2\sigma^2)) q_{ik}}{\sum_{j=1}^N \exp(-(x_i - a_j)^2/(2\sigma^2)) q_{ij}}, \quad k = 1, \dots, N.$$

Обозначим

$$c_{ik}(\bar{a}) = \frac{\exp(-(x_i - a_k)^2/(2\sigma^2)) q_{ik}}{\sum_{j=1}^N \exp(-(x_i - a_k)^2/(2\sigma^2)) q_{ij}}.$$

Тогда

$$\frac{\partial L(\bar{x}, \bar{a})}{\partial a_k} = \sum_{i=1}^M \frac{1}{\sigma^2} (x_i - a_k) c_{ik}(\bar{a}), \quad k = 1, \dots, N.$$

Приравняв к нулю производные, получаем систему

$$\sum_{i=1}^M \frac{1}{\sigma^2} (x_i - a_k) c_{ik}(\bar{a}) = 0, \quad k = 1, \dots, N. \quad (5)$$

Выразим из (5) a_k :

$$a_k = \frac{\sum_{i=1}^M c_{ik}(\bar{a}) x_i}{\sum_{i=1}^M c_{ik}(\bar{a})}, \quad k = 1, \dots, N.$$

Введем обозначения:

$$\gamma_{ik}(\bar{a}) = c_{ik}(\bar{a}) / \sum_{i=1}^M c_{ik}(\bar{a});$$

$$f_k(\bar{a}) = \sum_{i=1}^M \gamma_{ik}(\bar{a})x_i, \quad k = 1, \dots, N;$$

$$f(\bar{a}) = (f_1(\bar{a}), \dots, f_N(\bar{a})).$$

Тогда система уравнений (5) примет наиболее простой вид:

$$\bar{a} = f(\bar{a}). \quad (6)$$

Утверждение. Уравнение (6) имеет решение.

Доказательство. Обозначим

$$Y = \left\{ \bar{y} = (y_1, \dots, y_N) \mid \min_i x_i \leq y_l \leq \max_i x_i, \quad l = 1, \dots, N \right\}.$$

Очевидно, что Y является выпуклым и компактным множеством, а отображение f — непрерывным. Так как $0 \leq \gamma_{ik}(\bar{a}) \leq 1$, то $\min_i \{x_i\} \leq f_k(\bar{a}) \leq \max_i \{x_i\}$; следовательно, $f : Y \rightarrow Y$ и по теореме Брауэра отображение f имеет неподвижную точку. Утверждение доказано.

Опишем итерационный алгоритм решения уравнения $\bar{a} = f(\bar{a})$. Пусть задана точность вычислений ε .

Алгоритм 1

Положить начальные значения $\bar{a}^{(0)} = (0, \dots, 0)$.

Вычислять последовательность $\bar{a}^{(t)}$ по формуле

$$\bar{a}^{(t+1)} = f(\bar{a}^{(t)}), \quad t = 1, 2, \dots,$$

до тех пор, пока не будет выполнено условие $\|\bar{a}^{(t)} - \bar{a}^{(t+1)}\| < \varepsilon$.

Рассмотрим обобщение задачи. Пусть имеется несколько базовых станций, для которых построены карты уровня сигналов. Добавляется несколько новых базовых станций. Необходимо для них построить карты зон.

Пусть d_1, \dots, d_K — базовые станции, для которых имеются карты зон, d_{K+1}, \dots, d_{K+R} — базовые станции, для которых требуется построить карты зон.

Предлагается следующий алгоритм решения поставленной задачи.

Алгоритм 2

Для $t = 1, \dots, R$ выполнить:

Получить M наборов измерений уровня сигнала от базовых станций d_1, \dots, d_{K+t} .

Используя карты уровня сигналов для базовых станций d_1, \dots, d_{K+t-1} , вычислить q_{ij} , $i = 1, \dots, M$, $j = 1, \dots, N$.

Используя алгоритм 1, построить карту уровня сигнала для базовой станции d_{K+t} .

5 Тестовые испытания

Опишем эксперимент, проведенный на реальных данных. В эксперименте использовалось оборудование, описанное в табл. 1.

Таблица 1 Использованное оборудование

Оборудование	Описание
Базовая станция (4 шт.)	Беспроводная точка доступа Netgear WAG 102 Мощность передатчика 100 мВт 1 внешняя антенна 5 dBi (2,4 ГГц)
Ноутбук	MacBook Pro, Core i5 2,4 ГГц, 4ГБ ОЗУ, WiFi 802.11n (Broadcom) Mac OS X 10.7.5, 64-битная архитектура

На территории физико-технического факультета Петрозаводского государственного университета была выбрана тестовая площадка размером 12×20 м, на которой было задействовано 4 базовые станции (d_1, \dots, d_4 , $D = 4$, рис. 3). Данная площадка имеет опорные колонны, что позволяет эмитировать условия не прямой видимости какой-либо из базовых станций мобильным объектом. Для определения оценок математических ожиданий a_{jd} уровней сигналов базовых станций были проведены измерения в 22 опорных точках ($N = 22$), по 50 замеров в четырех направлениях относительно стен здания (итого 200 замеров в каждой опорной точке).

Полученные в ходе проведения измерений карты уровня сигналов для базовых станций d_1, \dots, d_4 представлены на рис. 4. Вертикальная ось соответствует уровню сигнала (дБм), горизонтальная плоскость соответствует размерам тестовой площадки, маркерами отмечены реальные усредненные значения.

Далее были использованы те же измерения для тестирования работы алгоритма 1 ($M = 4400$). Используя данные от трех из четырех базовых станций, по формуле (1) в момент каждого измерения i вычислялись вероятности q_{ij} нахождения объекта в опорных точках $j = 1, \dots, N$. Затем с помощью алгоритма 1 вычислялись оценки математических ожиданий a_j в опорных точках $j = 1, \dots, N$ для четвертой базовой станции. Параметр σ был равен трем.

Таким образом, были смоделированы карты уровня сигналов для каждой базовой станции d_1, \dots, d_4 . Пример смоделированных карт уровня сигналов

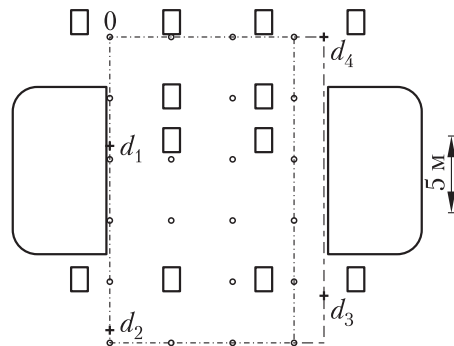


Рис. 3 Опорные точки тестовой площадки

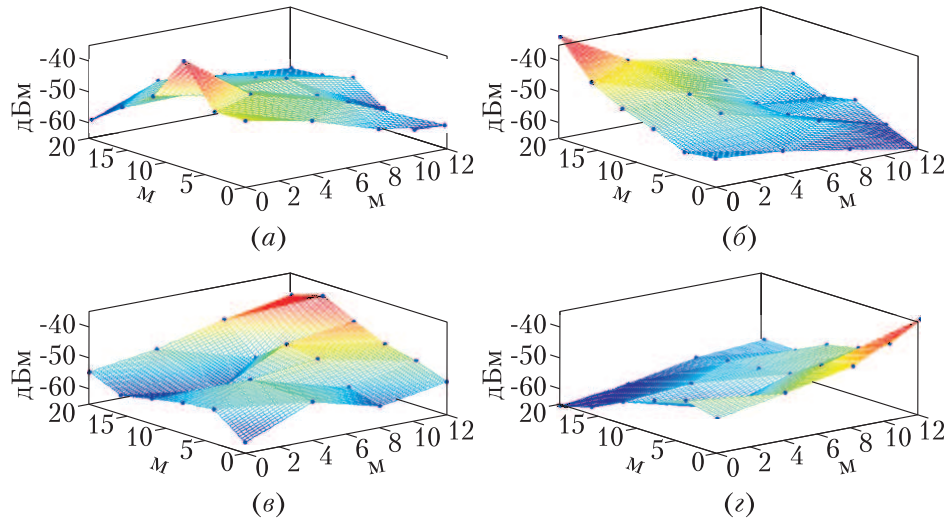


Рис. 4 Эмпирические карты уровня сигналов: (а) d_1 ; (б) d_2 ; (в) d_3 ; (г) d_4

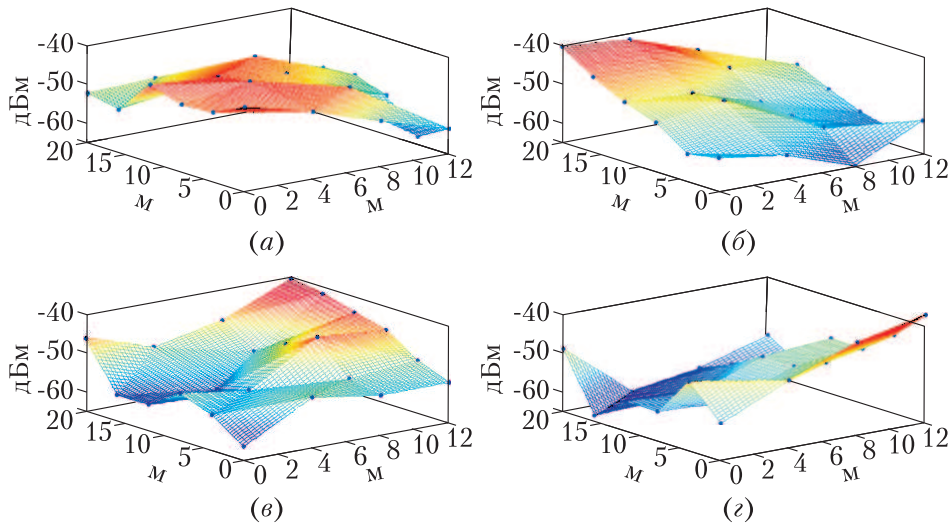


Рис. 5 Смоделированные карты уровня сигналов: (а) d_1 ; (б) d_2 ; (в) d_3 ; (г) d_4

Таблица 2 Разница между реальными и спрогнозированными математическими ожиданиями уровня сигнала

$ a_j - a_j^* $	Число опорных точек	Доля опорных точек, %
[0-1)	35	39,8
[1-2)	24	27,3
[2-3)	11	12,5
[3-4)	11	12,5
4 и больше	7	7,9

для всех базовых станций представлен на рис. 5. Найденные значения a_j сравнивались с реальными средними значениями уровня сигнала a_j^* для данной базовой станции. Полученное в ходе эксперимента число опорных точек с различным значением $|a_j - a_j^*|$ представлено в табл. 2.

В двух случаях был зафиксирован выброс $|a_j - a_j^*| \approx 16$. Однако следует отметить, что такие большие отклонения возможны и при «ручном» способе составления карты уровня сигналов.

Для определения сходимости работы алгоритма 2 был проведен имитационный эксперимент, заключающийся в следующем. Для десяти базовых станций ($D = 10$) были назначены условные математические ожидания уровня сигналов в десяти опорных точках ($N = 10$). Параметр σ был равен трем. В ходе каждого эксперимента выбиралось значение $K < 10$, на основе карт уровня сигналов первых K базовых станций строились карты уровня сигналов R остальных базовых станций ($R = D - K$). В каждом из M измерений ($M = 1000$) случайным образом выбиралась опорная точка и разыгрывались нормально распределенные случайные величины x_{id} — уровни сигнала от базовых станций $d = 1, \dots, D$ в данной опорной точке. Уровни сигналов x_{id} поступали на вход алгоритма 2, при помощи которого на основе карт уровня сигналов первых K базовых станций определялись карты уровня сигналов для базовых станций $d = K + 1, \dots, D$.

Далее полученные результаты использовались для определения позиции мобильного объекта следующим образом. Случайным образом выбиралась опорная точка, генерировались сигналы и определялось местоположение точки при помощи карт уровня сигналов всех базовых станций (начальных и обученных). Для этого по формуле (1) вычислялись апостериорные вероятности опорных точек и в качестве искомой определялась точка с максимальной вероятностью. После этого проверялось, совпала ли найденная таким образом опорная точка с заранее выбранной. Доля неправильно определенных опорных точек в ходе всего имитационного эксперимента представлена в табл. 3.

Из табл. 3 видно, что точность определения опорных точек улучшается при добавлении новых базовых станций, для которых карты уровня принимаемого сигнала строятся при помощи алгоритма 2.

Таблица 3 Доля неправильно определенных опорных точек при различном числе начальных базовых станций

Число добавленных базовых станций R	Начальное число базовых станций K			
	9	7	5	3
1	0,088	0,121	0,204	0,285
2		0,13	0,17	0,227
3		0,114	0,148	0,213
4			0,128	0,163
5			0,077	0,144
6				0,124
7				0,105

6 Заключение

Результаты полученных расчетов показали, что при уменьшении с 9 до 3 числа исходных «обученных» базовых станций число неправильно определенных опорных точек увеличилось менее чем на 2%. Это указывает на практическую применимость описанной модели для формирования карт уровня сигналов в ходе эксплуатации системы и их дальнейшего использования при уточнении позиций мобильных объектов.

Полученные в данной работе результаты также могут быть применены для построения систем локации с резервными базовыми станциями, для которых отсутствуют карты уровня сигналов. Кроме того, описанный алгоритм можно использовать для корректировки уже существующих карт уровня сигналов.

В дальнейшем планируется исследовать задачу построения карт уровня сигналов устройств различных производителей на основе имеющейся карты уровня сигналов одного из этих устройств.

Литература

1. Galov A. S., Moschevikin A. P., Voronov R. Combination of RSS localization and ToF ranging for increasing positioning accuracy indoors // ITST-2011: 11th Conference (International) on ITS Telecommunications Proceedings. — St. Petersburg: IEEE, 2011. P. 299–304.
2. Moschevikin A., Voronov R., Galov A., Soloviev A. Using pressure sensors for floor identification in wireless sensors networks // IDAACS-SWS: 1st Symposium (International) on Wireless Systems Proceedings. — Offenburg: IEEE, 2012. P. 2–6.
3. Galov A., Moschevikin A. Bayesian filters for ToF and RSS measurements for indoor positioning of a mobile object // IPIN-2013: Conference (International) on Indoor Positioning and Indoor Navigation Proceedings. — Montbeliard: IEEE, 2013. P. 310–317.

4. *Under J.* Deploying license-free wireless wide-area networks. — Indianapolis, Indiana: Cisco Press, 2004. 333 p.
5. *Tomsho G.* Guide to networking essentials. — 6th ed. — Stamford, Connecticut: Cengage Learning, 2011. 752 p.
6. *Hassan A.* Advanced location-based technologies and services. — Boca Raton, Florida: CRC Press, 2013. 354 p.

Поступила в редакцию 21.12.13

DYNAMIC CREATION OF WiFi-SIGNAL LEVEL MAPS FOR LOCATION SYSTEMS

R. V. Voronov and S. V. Malodushev

Petrozavodsk State University, 33 Lenin Prosp., Petrozavodsk 185910, Karelia,
Russian Federation

Abstract: Signal level maps are designed to locate wireless mobile devices. To increase a wireless network segment, it is necessary to create new signal level maps for implementation of location of wireless mobile devices in new areas. Traditionally, such maps are based on a great number of measurements taken at certain points, the process of these measurements requires a lot of time. In the article, dynamic creation of signal level maps is investigated. A heuristic algorithm for dynamic creation of signal level maps during operation of a location system is described. A function similar to the likelihood function was used for estimation of the average values of the signal levels. WiFi networks operating in the 2.4-gigahertz frequency band were used as wireless networks. Test results show that the suggested algorithm can be used for solving this issue.

Keywords: WiFi; base station; control point; signal level; location; signal level map

DOI: 10.14357/08696527140105

Acknowledgments

The research was done in the frames of MIP Nanonets Ltd. and RTL-Service Co. activity. The research was also supported by the Petrozavodsk State University (PSU) (Program of strategy development of PSU in 2012–2016), Russian Ministry of Education and Science, SR MFP Foundation and NTS and Ministry of Economic Development of Karelia (Program “Development of sky system for mobile abonents of distributed WiFi network” in 2013) as well as in the frames of program UMNİK (Member of the Youth Research and Innovation Competition) “Analysis and design of WiFi networks inside buildings” in 2012–2013.

References

1. Moschevikin, A., R. Voronov, A. Galov, and A. Soloviev. 2012. Using pressure sensors for floor identification in wireless sensors networks. *1st Symposium (International) on Wireless Systems IDAACS-SWS Proceedings*. Offenburg, Germany: IEEE. 2–6.
2. Galov, A. S., A. P. Moschevikin, and R. Voronov. 2011. Combination of RSS localization and ToF ranging for increasing positioning accuracy indoors. *11th Conference (International) on ITS Telecommunications ITST-2011 Proceedings*. St. Petersburg: IEEE. 299–304.
3. Galov, A., and A. Moschevikin. 2013. Bayesian filters for ToF and RSS measurements for indoor positioning of a mobile object. *Conference (International) on Indoor Positioning and Indoor Navigation (IPIN-2013) Proceedings*. Montbeliard, France: IEEE. 310–317.
4. Under, J. 2004. *Deploying license-free wireless wide-area networks*. Indianapolis, Indiana: Cisco Press. 333 p.
5. Tomsho, G. 2011. *Guide to networking essentials*. 6th ed. Stamford, Connecticut: Cengage Learning. 752 p.
6. Hassan, A. 2013. *Advanced location-based technologies and services*. Boca Raton, Florida: CRC Press. 354 p.

Received December 21, 2013

Contributors

Voronov Roman V. (b. 1976) — Candidate of Science (PhD) in technology, associate professor, Petrozavodsk State University, 33 Lenin Prosp., Petrozavodsk 185910, Karelia, Russian Federation; rvoronov@sampo.ru

Malodushev Sergei V. (b. 1985) — leading engineer, Petrozavodsk State University, 33 Lenin Prosp., Petrozavodsk 185910, Karelia, Russian Federation; svm@petsu.ru

АВТОМАТИЗИРОВАННАЯ СИСТЕМА РАЗРАБОТКИ И КОНТРОЛЯ ОПЕРАТИВНОГО ГРАФИКА ДВИЖЕНИЯ ПОЕЗДОВ В ПЕРИОД ПРЕДОСТАВЛЕНИЯ «ОКОН» (АС ОГДПО)

С. К. Дулин¹, И. Н. Розенберг², В. И. Уманский³

Аннотация: В работе представлены архитектура и технологические основы системы автоматизации управления процессом работы участка в период предоставления «окна» (АС ОГДПО), определен состав, функции и условия взаимодействия с учетом использования спутниковых технологий. Разработана целевая функция системы АС ОГДПО, основанная на выполнении условия обеспечения пропускной способности максимального числа поездов. Определены ограничения и условия, определяющие области допустимых значений расчетных параметров. В качестве инструментальной среды реализована система имитационного моделирования работы участка, учитывающая принципы построения «полигона окна» и движения поездов на его пространственно-временной модели, информацию о подходе поездов с «полигона окна» к «участку окна» и «перегону окна» и обеспечивающая построение прогнозного оперативного графика подхода поездов к «перегону окна» и разработку оперативного пакетного графика движения поездов.

Ключевые слова: оперативный график движения поездов; период предоставления «окна»; имитационное моделирование работы участка

DOI: 10.14357/08696527140106

1 Введение

Проблема автоматизации разработки графика движения возникла практически одновременно с поступлением на железнодорожный транспорт первых электронных вычислительных машин (ЭВМ). Развитие исследований в этой области обеспечило создание теоретических положений, которые стали основой для получения ряда практических решений [1–3].

В опубликованных до настоящего времени работах практически предлагаются методики автоматизированного построения максимальных графиков на период предоставления «окон». При этом за основу принимается исходная информация нормативного графика. По существу, такой график решает вопрос расчета пропускной способности участка в период предоставления «окна» и может служить

¹Научно-исследовательский и проектно-конструкторский институт информатизации, автоматизации и связи на железнодорожном транспорте (ОАО НИИАС), Институт проблем информатики Российской академии наук, s.dulin@ccas.ru

²Научно-исследовательский и проектно-конструкторский институт информатизации, автоматизации и связи на железнодорожном транспорте (ОАО НИИАС), I.Rozenberg@gismps.ru

³Интехгеотранс, umanvi@yandex.ru

ориентиром для руководителей движением поездов о максимальном количестве поездов, которые возможно пропустить в период предоставления «окна» [4, 5].

Однако поездным диспетчерам необходимо учитывать фактический подход поездов, и они вынуждены достаточно часто отступать от предварительно разработанных графиков и регулировать движение в период предоставления «окна», руководствуясь реальным подходом поездов, в том числе по направлениям движения, и своим опытом, т. е. поездные диспетчеры вручную разрабатывают оперативный график в период предоставления «окна». При этом они не в состоянии в полной мере учесть неравномерность подхода поездов по направлениям движения, вес и длину поездов и другие факторы.

В работе [6] расчетные модели прогнозирования графика движения предлагается разрабатывать с учетом переменных значений норм времени для элементов графика (СЭМН — ситуационно-эвристический метод нормирования), получаемых на основе накопления опыта реального движения поездов. Однако накопление такого опыта только в период предоставления «окон» затруднительно и не обеспечивает массовости экспериментальной базы. Неслучайно поэтому, что в работе [6] предоставление «окон» выступает в качестве ограничения при разработке прогнозов графика движения поездов. В то же время выдвинутое положение об использовании оперативных (динамических) исходных данных следует учитывать и при составлении расписаний в период «окон».

Постановка задачи в настоящем исследовании формулируется следующим образом.

Необходимо разработать основы построения автоматизированной системы составления графика (расписания) движения поездов в период предоставления «окна» продолжительностью $T_{\text{ок}}$ за некоторый небольшой период Δt_n до его начала с учетом реального подхода поездов по направлениям движения и возможностей оперативного определения необходимых исходных данных для разработки графика на период «окна». Построенный АС ОГДПО оперативный график должен служить основным руководящим документом для поездных диспетчеров в период предоставления «окна». Если на основе данных подсистемы обратной связи имеет место существенное влияние возмущающих воздействий на движение поездов в период предоставления «окна», то предусматривается итерационная корректировка оперативных графиков, разрабатываемых АС ОГДПО.

2 Архитектура АС ОГДПО

На рис. 1 представлена принципиальная архитектура автоматизированной системы управления с обратной связью технологическим процессом работы участка в период предоставления «окна». Основная функция этой системы — автоматизированная оперативная разработка графика движения и контроль за его выполнением.

Управляемым объектом ($УО_0$) служит совокупность пассажирских и грузовых поездов, следующих по закрываемому для работ перегону в период «окна»,

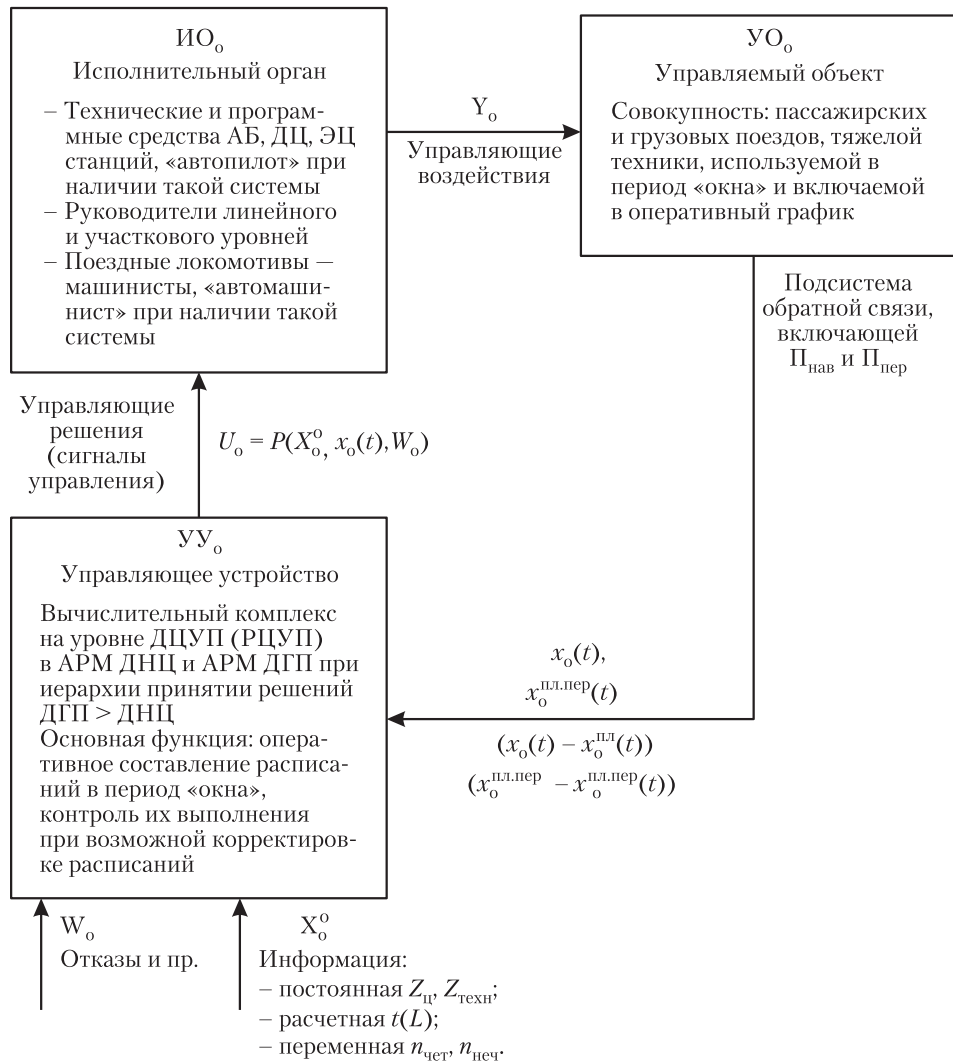


Рис. 1 Принципиальная архитектура АС ОГДПО (УО_о, ИО_о, УУ_о)

а также тяжелая ремонтная техника, обеспечивающая выполнение работ, если есть необходимость включать ее передвижения в оперативный график.

Управляющим устройством (УУ_о) является вычислительный комплекс, который входит в состав технических средств и программного обеспечения (ПО) дорожных (региональных) автоматизированных диспетчерских центров управления перевозочным процессом (ДЦУП или РЦУП) и обеспечивает работу авто-

матризованных рабочих мест поездного диспетчера АРМ ДНЦ и дорожного диспетчера по району управления АРМ ДГП.

В качестве исполнительного органа (ИО_о) выступают технические и программные средства систем автоматики и телемеханики — автоблокировка (АБ), диспетчерская централизация (ДЦ), электрическая централизация на станциях (ЭЦ), а также оперативные руководители линейного (ДСП) и дорожного (ДНЦ и ДГП) уровня. При внедрении системы «Автопилот» (автоматическое приготовление маршрутов по выбранным станциям) данная система также входит в состав ИО_о.

Характерной особенностью системы АС ОГДПО является то, что к ИО_о необходимо, по существу, отнести и комплекс «поездной локомотив – машинист», поскольку для каждого поезда УУ_о выдает задания на выполнение времен хода по отрезкам пути перегонов — $t(L_{\text{ВЫХ}})$, $t(L_{\text{ВЫХ}}^n)$ и перегонам — t_x^n . Машинист должен обеспечивать выполнение этих времен хода. При наличии системы «автомашинист» такая система также входит в состав ИО_о.

Взаимодействие УО_о, УУ_о и ИО_о осуществляется следующим образом. Устройство УУ_о вырабатывает управляющие решения:

$$U_o = P(X_o^o, x_o(t), W_o), \quad (1)$$

где X_o^o — множество параметров графика — постоянных, расчетных и переменных, учитываемых при разработке исходного оперативного графика движения и при его корректировке; $x_o(t)$ — множество данных, поступающих из подсистемы обратной связи о фактическом выполнении разработанного оперативного графика движения; W_o — данные о возмущающих воздействиях, которые следует учитывать при разработке оперативного графика; P — оператор, определяющий действия системы для различных значений X_o^o , $x_o(t)$ и W_o .

Постоянными параметрами $x_o^{\text{пост}}$ в множестве X_o^o являются:

- $Z_{\text{Т-В}}$ — нормативная база технологического или вариантного графика, которая используется и при разработке оперативного графика, — время начала «окна», его продолжительность, расписание пассажирских поездов, порядок подвода тяжелой путевой техники и др.;
- $Z_{\text{Ц}}$ — цифровая модель участка (включая отдельные пункты и перегоны), где выполняются работы в «окно»;
- $Z_{\text{ТЕХН}}$ — технология работы станций и других отдельных пунктов при следовании поездов различных категорий (пассажирских, грузовых, соединенных).

К расчетным параметрам относятся значения скорости движения поездов и времени их хода по отрезкам пути $L_{\text{ВЫХ}}$, $L_{\text{ВЫХ}}^n$ и $L_{\text{П}}$ в зависимости от категории, веса поездов Q_i и ограничений скорости в период предоставления «окна»

$$V_{\text{огр}} : t(L) = f(x_o^{\text{пост}}, Q_i, V_{\text{огр}}).$$

Такую информацию необходимо подготовить заранее, так как иначе придется при составлении оперативного графика для каждого поезда выполнять тяговые расчеты, что необоснованно удлинит время разработки расписаний.

Переменная информация представляет собой данные системы АСОУП-2 и системы навигационного позиционирования локомотивов о фактическом подходе поездов различных направлений на период предоставления «окна». Обозначим эти данные как $n_{\text{чет}}^{\text{поз}}$ и $n_{\text{неч}}^{\text{поз}}$.

К этой категории информации относятся и данные о плане поездообразования по направлениям движения, где учитываются поезда, которые еще не сформированы, но поступят на рассматриваемый участок в период «окна»: $n_{\text{чет}}^{\text{п-о}}$ и $n_{\text{неч}}^{\text{п-о}}$.

Плановый подход поездов по направлениям движения, четному — $n_{\text{чет}}$ и нечетному — $n_{\text{неч}}$, составит:

$$\begin{aligned} n_{\text{чет}} &= n_{\text{чет}}^{\text{поз}} + n_{\text{чет}}^{\text{п-о}}; \\ n_{\text{неч}} &= n_{\text{неч}}^{\text{поз}} + n_{\text{неч}}^{\text{п-о}}. \end{aligned}$$

В целом множество параметров графика X_o^o , являющееся совокупностью постоянных $x_o^{\text{пост}} \in Z_{\text{т-в}}, Z_{\text{ц}}, Z_{\text{техн}}$; расчетных $t(L)$ и переменных $n_{\text{чет}}, n_{\text{неч}}$ параметров определяется выражением:

$$\begin{aligned} X_o^o &\in ((Z_{\text{т-в}}, Z_{\text{ц}}, Z_{\text{техн}}), t(L) = \\ &= f(x_o^{\text{пост}}, Q_i, V_{\text{огр}}), (n_{\text{чет}}^{\text{поз}} + n_{\text{чет}}^{\text{п-о}}), (n_{\text{неч}}^{\text{поз}} + n_{\text{неч}}^{\text{п-о}})). \end{aligned}$$

Значение W_o в зависимости (1) определяет влияние возмущающих воздействий на выполнение оперативного графика. Здесь особенно важно определять те возмущающие воздействия, которые необходимо и возможно учитывать при разработке исходного оперативного графика или при его корректировках. Среди них — фактические отклонения от расписания пассажирских поездов, дополнительные ограничения скоростей движения поездов и другие.

В период действия оперативного графика его выполнение контролируется подсистемой обратной связи, включающей две составляющих:

- (1) подсистему $\Pi_{\text{нав}}$, основанную на использовании данных навигационного позиционирования поездных локомотивов, обеспечивающую возможность сравнения плановых данных о расписаниях движения поездов в оперативном графике $x_o^{\text{пл}}$ с фактическим их выполнением $x_o(t)$ при движении поездов;
- (2) подсистему $\Pi_{\text{пер}}$, обеспечивающую сравнение с фактическим выполнением данных переменной информации $x_o^{\text{пл. пер}}(t)$, полученных без использования системы позиционирования — $x_o^{\text{пл. пер}}$.

Значения $(x_o(t) - x_o^{\text{пл}}(t))$ и $(x_o^{\text{пл. пер}} - x_o^{\text{пл. пер}}(t))$ являются основой для принятия решения о корректировке оперативного графика.

3 Целевая функция системы

Основной целью управления движением поездов в период предоставления «окна» является минимизация потерь поездо-часов. В ряде работ [1, 3, 5] принимается именно такой критерий. Однако построение целевой функции при использовании этого условия носит слишком общий характер и не позволяет ставить четкие задачи при разработке алгоритмов и ПО АС ОГДПО. Это утверждение основывается на следующих положениях.

Определение минимума поездо-часов ($\sum nt_{\min}$) следует относить не только к поезвному участку, на котором выделяется «окно», но и к прилегающим участкам и даже станциям (с расчетом образующихся вагоно-часов). Это следует из того, что участковая скорость (что эквивалентно поездо-часам) зависит от степени внутрисуточной неравномерности поездопотоков. А именно в период «окна» происходят существенные сгущения и разряжения в движении поездов, в том числе и на прилегающих участках. Простой вагонов на станциях также зависит от степени неравномерности подхода поездов, а значит, изменяется при предоставлении «окон», и, следовательно, этот фактор в расчетах также должен учитываться. Поэтому величину $\sum nt_{\min}$ надо искать не только на участке, где предоставляется «окно», но по крайней мере и на прилегающих участках. Кроме того, в целевую функцию надо вводить и поиск минимума вагоно-часов $\sum Nt_{\min}$ на прилегающих станциях.

В такой постановке практическое решение поставленной задачи нереально. Для определения величин $\sum nt_{\min}$ и $\sum Nt_{\min}$ требуется использовать работающую в оперативном режиме модель работы направлений из нескольких участков и станций, по которой даже теоретических решений пока не имеется. При этом следует учитывать наличие многочисленных ограничений при разработке оперативного графика для участка с «окном».

В других работах (см., например, [7]), в качестве критерия предлагается следующая постановка вопроса: «выполнить заданные размеры перевозок и в то же время не ухудшить основные показатели (оборот вагона, производительность труда локомотивных бригад)». Такая постановка требует разработки еще более сложных моделей работы направлений.

В этой ситуации при установлении целевой функции для решаемой задачи необходимо использовать следующее положение: чем больше будет пропущено поездов в период «окна», тем меньше будут задержки поездов, т. е. в качестве целевой функции может быть принято условие построения максимального графика в период действия «окна»: $n_{\text{ок}} \rightarrow \max$, где $n_{\text{ок}}$ — число поездов, следующих в период действия окна $T'_{\text{ок}}$.

Принципиально такой подход принят в работе [8], и с ним следует согласиться. Разработка оперативного графика имеет целый ряд особенностей, так как основывается на текущем состоянии поездного положения, его изменениях и на нормативах графика, часть из которых рассчитывается в оперативном режиме (например, $I_{\text{от}} = f(Q_i)$).

Соответственно, при построении графика требуется обеспечить пропуск максимального числа поездов различного веса из реально следующих в сутки при предоставлении «окна» — $n_{\text{факт}}$, а не количества поездов в нормативном графике. Для оперативного графика учитывают, что

$$T'_{\text{ок}} = T_{\text{ок}}^{\text{до}} + T_{\text{ок}} + T_{\text{ок}}^{\text{посл}},$$

где $T_{\text{ок}}$ — собственно время предоставления «окна»; $T_{\text{ок}}^{\text{до}}$ и $T_{\text{ок}}^{\text{посл}}$ — соответственно время до и после предоставления «окна», на условия организации движения поездов в течение которого оказывает влияние предоставление «окна».

В период $T_{\text{ок}}^{\text{до}}$ важно обеспечить пропуск тяжелых путевых машин и максимально возможного числа пассажирских и грузовых поездов с использованием пропускной способности двухпутного перегона. При этом возможна некоторая небольшая (не оказывающая заметного влияния на условия выполнения работ в «окно») сдвигка (корректировка) времени начала «окна» для снижения потерь пропускной способности временно однопутного перегона. Это особенно актуально при наличии опаздывающих пассажирских поездов. Тогда если обозначить (по технологическому или вариантному графику) текущее время начала «окна» через $t_{\text{н}}$ и его завершения через $t_{\text{зав}}$, то при сдвигке времени начала «окна» на $\Delta t_{\text{ок}}$ текущее время начала и завершения оперативного «окна» составит $t_{\text{н}}^{\text{о}} = t_{\text{н}} + \Delta t_{\text{ок}}$, $t_{\text{зав}}^{\text{о}} = t_{\text{н}} + T_{\text{ок}} + \Delta t_{\text{ок}}$. При $\Delta t_{\text{ок}} = 0$ начало и окончание оперативного «окна» совпадает с соответствующими значениями технологического или вариантного «окна».

Основным требованием к значению $T_{\text{о}}^{\text{посл}}$ является его минимизация, т. е. $T_{\text{о}}^{\text{посл}} \rightarrow \min$, что при выполнении требования максимального пропуска поездов минимизирует потери поездо-часов в период $T_{\text{ок}}^{\text{посл}}$ и в период $T'_{\text{ок}}$.

На основании вышеизложенного целевая функция организации управления поездопотоками в системе АС ОГДПО в период работы участка с «окном» представляется следующим образом:

$$\frac{n_{\text{ок}}}{n_{\text{факт}}} \rightarrow \max .$$

Изложенные положения являются основой создания АС ОГДПО.

4 Технология и инструментальная среда имитационного моделирования работы участка для построения оперативного графика

Ниже рассмотрены вопросы разработки технологических положений формирования и выполнения оперативных графиков движения поездов в период предоставления «окон» для принятой архитектуры АС ОГДПО.

На рис. 2 представлена организационная структура системы.

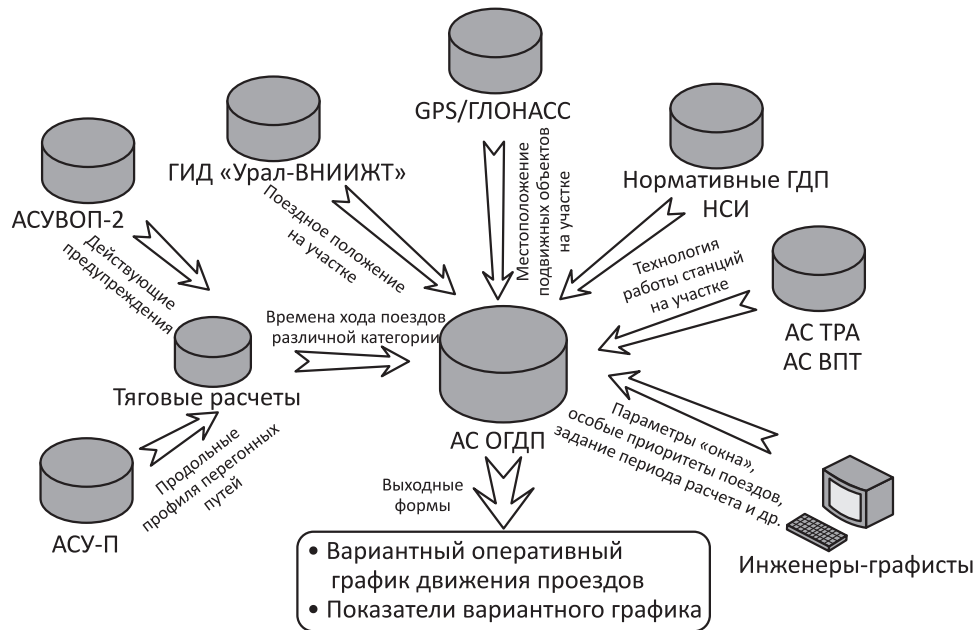


Рис. 2 Организационная структура системы разработки оперативных графиков

Данные о поездном положении считываются из системы ГИД «Урал-ВНИИЖТ» с последующим уточнением по данным источников спутникового позиционирования GPS/ГЛОНАСС. Действующие предупреждения и некоторые такие данные, как будет показано ниже (см. рис. 6), получают дополнительно из АСУОП-2, продольные профили перегонных путей — из АСУ-П. Эти данные совместно с параметрами поездов используются в тяговых расчетах для получения уточненных времен хода поездов в зависимости от веса и ограничений скорости.

Расчетным ядром системы является имитационная модель работы участка, в основу реализации которой положен принцип максимального прямого использования имеющихся в ОАО «РЖД» геоинформационных ресурсов и нормативно-справочной информации. Прежде всего это касается данных цифровых моделей пути и станционного развития, включающих семантическое описание объектов инфраструктуры. В частности, описание станций участка формируется на основе данных автоматизированных систем ведения техническо-распорядительных актов станций (АС ТРА) и ведения технологических процессов работы станций (АС ВТП).

Входной настраиваемой информацией служат параметры «окна», особые приоритеты поездов, вводится информация о дислокации путевых машин и другие данные.

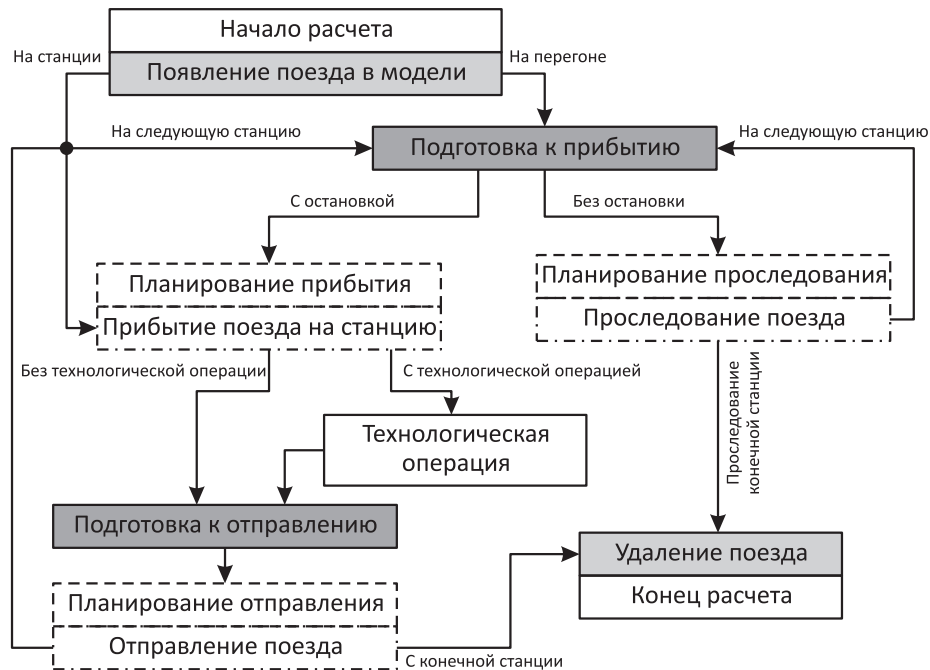


Рис. 3 Структура дискретно-событийного механизма для каждого поезда

Выходной формой системы является оперативный график с указанием порядка пропуска поездов во время «окна». Определяются дополнительные простои по каждому грузовому поезду, отклонения от расписания пассажирских поездов при подходе к участку, где предоставляется «окно», и на выходе с него с указанием времени возможного дальнейшего нагона.

Продвижение времени в модели участка за расчетный период основано на дискретно-событийном механизме, общая схема которого представлена на рис. 3. Событийность механизма заключается в том, что в каждый дискретный момент времени по всем поездам в порядке убывания приоритета анализируется время запланированной операции. Если в текущий момент времени запланирована операция, то запускается ее выполнение, если нет — осуществляется переход к следующему шагу.

Технологическая операция определяется сдвигом времени на период от события «прибытие поезда» до события «подготовка к отправлению».

Необходимо также определить события, связанные с построением оперативного графика в период выделения «окна», а также продолжительности интервалов между ними. Эти события: начало и завершение «окна», текущее время начала и завершения «окна» и др., которые представлены на шкале текущего

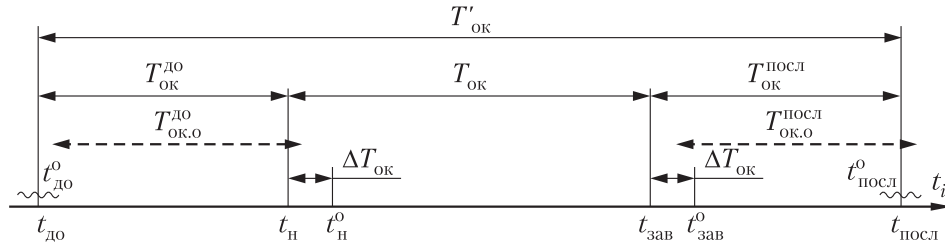


Рис. 4 События в период выделения «окна», представленные на шкале текущего времени

времени t_i (рис. 4). Именно эти события и их временные характеристики должны учитываться при построении оперативного графика для поездного участка, где выделяется «окно» (далее — «участок окна») и перегона, где для выполнения работ закрывается один из главных путей двухпутного участка (далее — «перегон окна»).

На рис. 4 приняты следующие обозначения: t_i — текущее время; $t_н$ и $t_{зав}$ — текущее время начала и завершения «окна» в заблаговременно разрабатываемом графике (технологический — тип 1, вариантный — тип 2); $t_н^o$ и $t_{зав}^o$ — текущее время начала и завершения «окна» в оперативном графике при возможной сдвиге начала «окна» на время $\Delta T_{ок}$; $t_{до}$ — текущее время начала периода влияния «окна» на расписание движения поездов при продолжительности этого периода $T_{ок}^{до}$ в заблаговременно разрабатываемом графике; $t_{посл}$ — текущее время окончания периода влияния «окна» на расписание движения поездов после его завершения при продолжительности этого периода $T_{ок}^{посл}$ в заблаговременно разрабатываемом графике; $t_{до}^o$ и $t_{посл}^o$ — соответствующее текущее время событий в оперативном графике с продолжительностью периодов $T_{ок,о}^{до}$ и $T_{ок,о}^{посл}$.

«Участок окна» располагается на некотором полигоне. В простейшем случае это линия без разветвлений, что характерно, например, для Забайкальской, Дальневосточной и некоторых других железных дорог. Более характерным является расположение «участка окна» на разветвленном полигоне (далее — «полигон окна»). Под «полигоном окна» следует понимать те прилегающие к «участку окна» линии, движение поездов по которым следует учитывать при разработке оперативного графика.

Определим теперь условия реализации дискретно-событийного механизма для каждого из поездов с учетом особенностей их формирования и поступления на участок.

«Полигон окна» может включать в себя помимо «участка окна» другие однопутные и двухпутные (многопутные) участки. На «полигоне окна» необходимо выделить два типа станций:

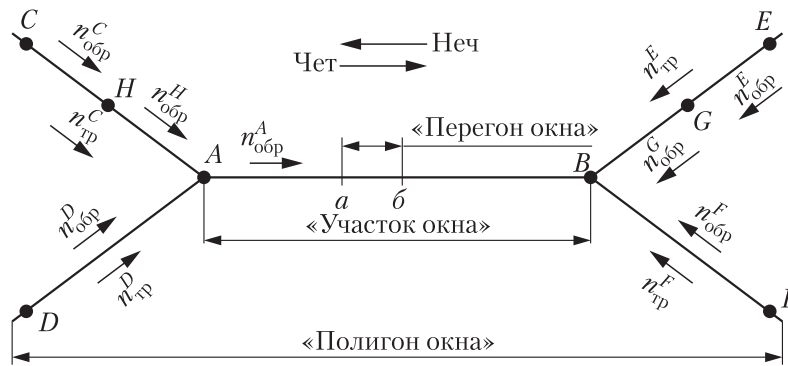


Рис. 5 Схема железнодорожного полигона («полигона окна»), движение поездов на котором должно учитываться в оперативном графике

- (1) станции, пропускающие только транзитный поездопоток;
- (2) станции, пропускающие транзитный поездопоток и осуществляющие переработку вагонопотока, в том числе и с выполнением местной работы (погрузка, выгрузка), в результате чего происходит поездообразование дополнительно к проходящим через такие станции транзитным поездам без переработки.

На рис. 5 представлен разветвленный «полигон окна», включающий поездообразующие станции C, H, D, A, G, E, F , «участок окна» AB и «перегон окна» $a-b$. Учтены только те транзитные $n_{тр}$ и образующиеся $n_{обр}$ поезда, движение которых в четном или нечетном направлении связано с их прохождением через «участок окна» AB .

На рис. 5 приняты следующие обозначения: AB и $a-b$ — соответственно участок и перегон, где предоставляется «окно»; A, C, D, E, F, G, H — технические и грузовые станции, где происходит образование поездопотока (помимо транзитных поездов); $n_{обр}^A, n_{обр}^C, n_{обр}^D, n_{обр}^H$ — поток поездов, образующийся на станциях A, C, D, H в четном направлении движения; $n_{обр}^E, n_{обр}^G, n_{обр}^F, n_{обр}^B$ — то же в нечетном направлении движения; $n_{тр}^C$ и $n_{тр}^D$ — транзитный поездопоток, следующий со станций C и D в четном направлении движения; $n_{тр}^F$ и $n_{тр}^E$ — то же в нечетном направлении движения.

Таким образом, множество поездов, следующих в четном направлении на «участок окна» $N_{чет}^{AB}$, состоит из подмножеств транзитных $N_{тр,чет}^{AB}$ и образующихся $N_{обр,чет}^{AB}$ поездов, т. е.

$$N_{чет}^{AB} = N_{тр,чет}^{AB} \cup N_{обр,чет}^{AB}.$$

В свою очередь,

$$N_{\text{обр.чет}}^{AB} = \sum_{j=1}^l n_{\text{тр.чет.}j};$$

$$N_{\text{обр.чет}}^{AB} = \sum_{\varphi=1}^m n_{\text{обр.чет.}\varphi},$$

где j и φ — станции, с которых отправляется транзитный и образующийся поездопоток, число которых (станций) составляет l и m . При этом

$$N_{\text{тр.чет}}^{AB} \cap N_{\text{обр.чет}}^{AB} \neq 0.$$

Тогда для примера, приведенного на рис. 5, имеем

$$N_{\text{чет}}^{AB} = (n_{\text{тр}}^C + n_{\text{тр}}^D) + (n_{\text{обр}}^C + n_{\text{обр}}^H + n_{\text{обр}}^D + n_{\text{обр}}^A).$$

Для нечетного направления:

$$N_{\text{неч}}^{AB} = (n_{\text{тр}}^E + n_{\text{тр}}^F) + (n_{\text{обр}}^C + n_{\text{обр}}^F + n_{\text{обр}}^G + n_{\text{обр}}^E).$$

Разделение на «полигоне окна» общего поездопотока на транзитный и образующийся связано с различной информационной базой, используемой для этих категорий поездов, и порядком получения информации при построении оперативного графика. При этом справедливо утверждение, что для любого $n_{\text{тр.}j}$ и $n_{\text{обр.}\varphi}$ в рассматриваемой информационной среде АС ОГДПО используются единые действия для получения информации I о поездах этих категорий при построении оперативного графика, т. е.

$$I_{\text{тр}} = \forall n_{\text{тр.}j} p_j (n_{\text{тр.}j});$$

$$I_{\text{обр}} = \forall n_{\text{обр.}\varphi} p_{\varphi} (n_{\text{обр.}\varphi}).$$

Здесь p_j и p_{φ} — операторы, соответствующие действиям в АС ОГДПО для получения информации о транзитных и образующих поездах.

Для решения вопросов, связанных с информационным обеспечением задачи построения оперативного графика, представим «полигон окна» в виде пространственно-временной модели (рис. 6). По существу, это сокращенный график движения поездов на «полигоне окна». На рис. 6 (за исключением «перегона окна») рассмотрено только четное направление движения с подходом поездов к «участку окна» со станций C и H . Подход со станции D учитывается (условно) только по станции A . Для упрощения на рис. 6 не нанесены расписания пассажирских поездов и тяжелых путевых машин, следующих на «перегон окна».

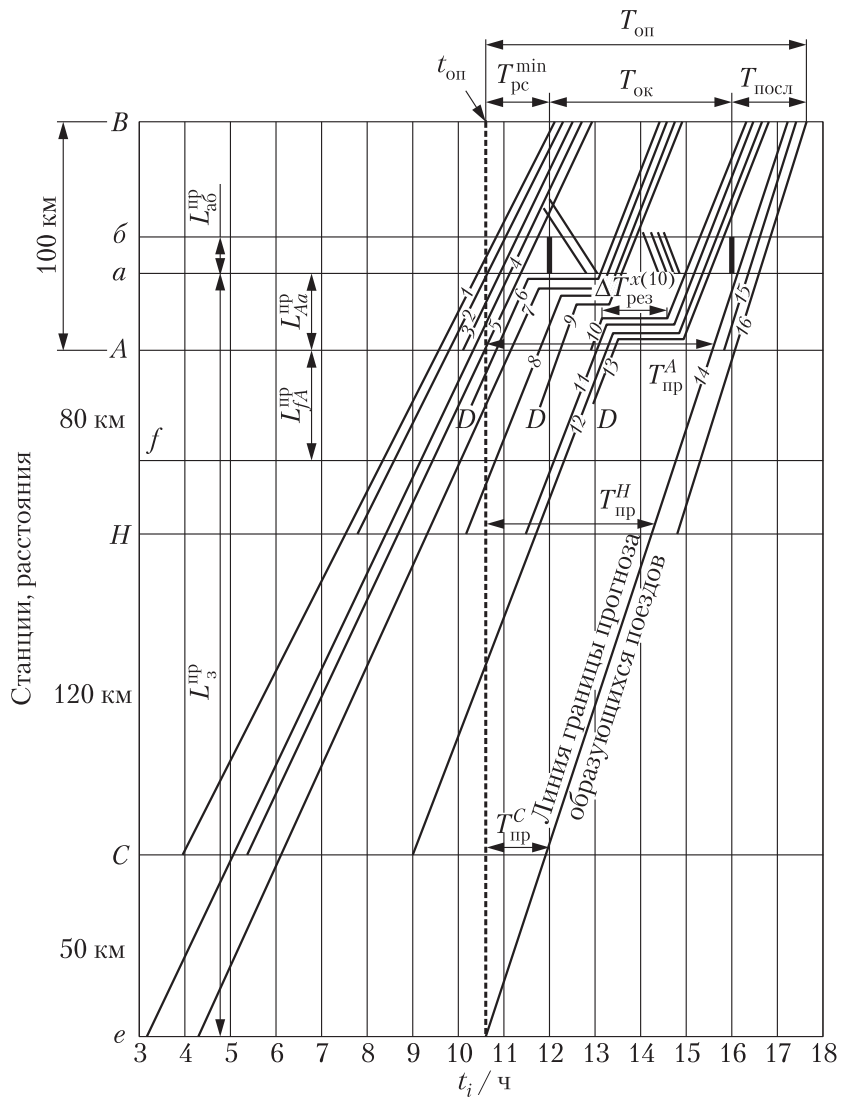


Рис. 6 Пространственно-временная модель «полигона окна» (четное направление движения) для решения вопросов информационного обеспечения разработки оперативного графика по подходу поездов к «перегону окна»

Очевидно, что для повышения точности информации о поездах, следующих в период «окна», момент начала действий по разработке оперативного графика $t_{оп}$ следует максимально возможно приближать ко времени начала «окна», т. е.

$$t_{н}^o - t_{оп} \rightarrow \min . \quad (2)$$

Условие (2) выполняется на основе минимизации времени реакции системы АСОГДПО $T_{рс}$, которое, в свою очередь, равно

$$T_{рс} = T_{инф} + T_{ц} + T_{р.оп} + T_{ДГП} + T_{исп}^{дов} + T_{исп}^{подг} ,$$

где $T_{инф}$ — время на сбор информации о подходе поездов к «участку окна» и к «перегону окна»; $T_{ц}$ — подготовка цифровой модели «участка окна» и других данных к выполнению расчетов с учетом полученных сведений о поездах; $T_{р.оп}$ — время расчета оперативного графика; $T_{ДГП}$ — время согласования рассчитанного оперативного графика с верхним уровнем управления; $T_{исп}^{дов}$ — время доведения результатов расчета до исполнителей (дежурные по станциям — ДСП, машинисты локомотивов); $T_{исп}^{подг}$ — время подготовки исполнителей (ДСП, машинисты локомотивов) к выполнению оперативного графика.

Заметим, что операции, связанные с затратами времени $T_{инф}$ и $T_{ц}$, могут в определенной доле выполняться параллельно.

Горизонтальная ось на рис. 6, соответствующая текущему времени $t_{оп}$ (жирная штриховая линия), определяет перечень поездов, о которых должна поступить информация в АС ОГДПО для разработки оперативного графика. Из рис. 6 следует, что в момент $t_{оп}$ при разработке оперативного графика следует учитывать поезд с № 5 по № 14 (нумерация поездов условная). А длина зоны прогноза $L_{з}^{пр}$ для четных поездов, следующих с примыкающей к «участку окна» некоторой χ -й линии подхода поездов $A-H-C$ (при их общем числе k), составляет расстояние от станции a до некоторой станции e . Зная величину участковой скорости $V_{уч}^{a-e}$ на данной линии, определим длину зоны прогноза (обозначения из рис. 6):

$$L_{з,a-e}^{пр} = T_{оп} V_{уч}^{B-e} - L_{B-a} .$$

Таким образом, в момент $t_{оп}$ при расчете оперативного графика учитываются все поезда, находящиеся от станции a на расстоянии $L_{з,a-e}^{пр}$.

Определим порядок значений величины зоны прогноза $L_{з}^{пр}$ для типичных условий. Приняты следующие значения расчетных величин: $T_{ок} = 3, 4$ и 5 ч; $T_{рс}^{min} \approx 1$ ч; $T_{посл} \approx 2$ ч; $V_{уч} = 30, 40$ и 50 км/ч, «перегон окна» находится в середине «участка окна» длиной 100 км. Результаты расчетов показаны на рис. 7.

Представляет интерес оценка общей протяженности зон прогноза $L_{з,общ}^{пр}$ на k линиях «полигона прогноза». Так, для схемы, представленной на рис. 5, при $V_{уч} = 40$ км/ч и $T_{ок} = 5$ ч $L_{з,общ}^{пр} = \sum_{\chi=1}^k L_{з\chi}^{пр} \approx 1400$ км.

Таким образом, определение подхода поездов к «участку окна» на «полигоне окна» представляет собой достаточно трудоемкую задачу.

Но этим не исчерпывается сбор информации о подходе поездов к «участку окна». Часть поездов может следовать (см. рис. 5 и 6) с поездобразующих станций A , H и C с их отправлением после момента времени $t_{оп}$. При этом линия подхода $D-A$ должна рассматриваться отдельно. Следовательно, на основе программ поездобразования действующих автоматизированных систем (АСОУП-2, АСУСС) в момент $t_{оп}$ необходимо дать прогноз образования поездов по следующим зонам времени $T_{пр}$:

$T_{пр}^C$ — по станции C на период от $t_{оп}$ до $t_{оп} + T_{пр}^C$;

$T_{пр}^H$ — по станции H на период от $t_{оп}$ до $t_{оп} + T_{пр}^H$;

$T_{пр}^A$ — по станции A на период от $t_{оп}$ до $t_{оп} + T_{пр}^A$.

При этом $T_{пр}^C = T_{пр} = T_{оп} - L_{CB}/V_{уч}^{CB}$. Соответственно определяются $T_{пр}^H$ и $T_{пр}^A$.

Из рис. 6 следует, что для рассматриваемого примера при расчете оперативного графика дополнительно к поездам, находящимся в зоне $L_3^{пр}$ в момент $t_{оп}$, необходимо учитывать поезд № 11 по станции H и поезд № 10 по станции A . По станции C образующийся поезд № 5 учитывается в зоне $L_3^{пр}$ в момент $t_{оп}$.

В целом алгоритм действий по сбору информации о «подходе поездов» к «участку окна» и «перегону окна» для их учета при разработке оперативного графика представлен на рис. 8.

Полученная информация определяет сведения о подходе поездов в момент $t_{оп}$. Для расчета текущего времени подхода поездов к станциям, ограничивающим «участок окна», и к станциям, ограничивающим «перегон окна», требуется разработка прогноза времени следования таких поездов до ограничивающих станций. Такой прогноз будет тем точнее, чем меньше величина $T_{рс}^{min}$. Порядок выполнения прогнозных расчетов рассмотрим на примере, представленном на рис. 6 для четного направления движения.

В зону прогноза $L_3^{пр}$ в момент $t_{оп}$ и в периоды прогноза образования поездов $T_{пр}$ по станциям A , H , C попадают поезда №№ 1–5, 7, 8, 10–12, 14. Для линии подхода со станции D строится самостоятельная модель определения перечня и числа подходящих поездов (поезда №№ 6, 9 и 13).

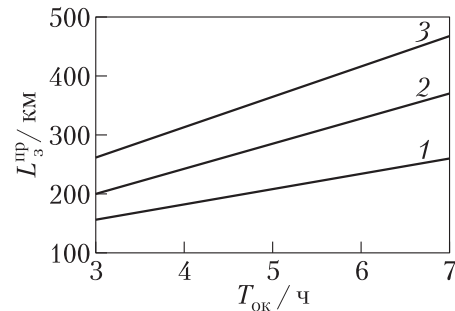


Рис. 7 Расчетная протяженность зоны прогноза $L_3^{пр}$ для различных эксплуатационных условий: 1 — $V_{уч} = 30$ км/ч; 2 — 40; 3 — $V_{уч} = 50$ км/ч

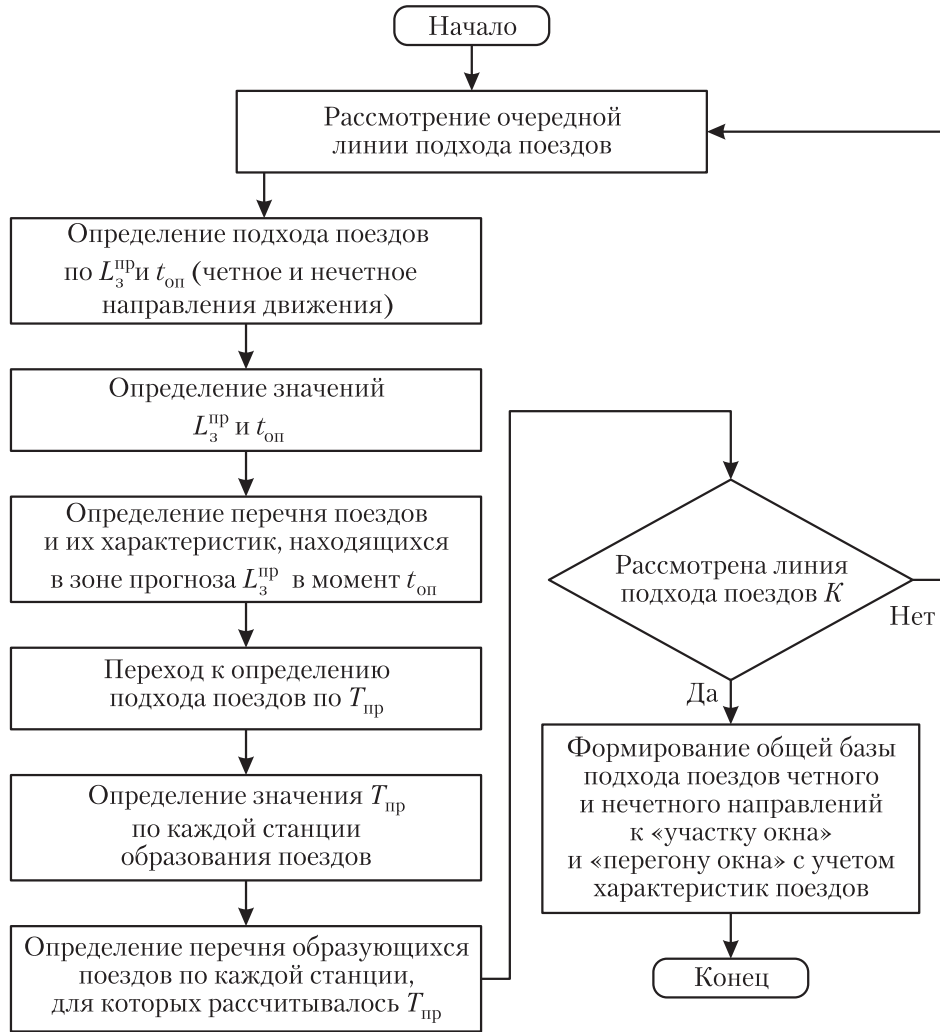


Рис. 8 Алгоритм действий по определению информации о подходе поездов с «полигона окна» к «участку окна» и «перегону окна» для разработки оперативного графика (K — число линий подхода поездов, включая «участок окна»)

Назовем такие поезда прогнозируемыми поездами в период «окна», прибывающими с χ -й линии n_{χ}^{np} , а общее число поездов, подходящих со всех линий:

$$n_k^{np} = \sum_{\chi=1}^k n_{\chi}^{np},$$

где k — общее число подходящих линий.

При определении значений $n_x^{\text{пп}}$ учитываются поезда, находящиеся внутри «линии границы прогноза образующихся поездов» (см. рис. 6). Такая «линия границы» определяется по технологическому или вариантному графику.

Рассмотрим теперь методику определения прогнозируемого времени хода на примере поезда № 8, который образуется на станции H , но фиксируется на длине зоны прогноза $L_3^{\text{пп}}$ в момент $t_{\text{оп}}$. Необходимо сделать прогноз времени хода $t_{\text{пр}}$ поезда № 8 на отрезках зоны прогноза $L_{fA}^{\text{пп}}$ и $L_{Aa}^{\text{пп}}$.

В простейшем случае

$$t_{fA}^{\text{пп}} = \frac{L_{fA}^{\text{пп}}}{V_{\text{уч}}^{fA}}, \quad t_{Aa}^{\text{пп}} = \frac{L_{Aa}^{\text{пп}}}{V_{\text{пр}}^{Aa}},$$

где $V_{\text{уч}}^{fA}$ и $V_{\text{уч}}^{Aa}$ — нормативы участковой скорости на соответствующих отрезках зоны прогноза по технологическому или вариантному графику.

Однако более точным будет использование метода построения прогнозного графика следования всех входящих в зону прогноза поездов $n_k^{\text{пп}}$ отдельно по четному и нечетному направлениям движения. Разработка такого графика должна быть предусмотрена в АС ОГДПО и выполняться с использованием известных методов автоматизированного построения двухпутных графиков [9].

Возможно также использование аппарата автоматизированной системы ГИД «Урал-ВНИИЖТ». При этом управляющие решения \bar{R} по прокладке поездов $x(1), x(2), \dots, x(n_k^{\text{пп}})$ на пространстве $L_3^{\text{пп}}$ в период времени $(t_{\text{оп}}, (t_{\text{оп}} + (T_{\text{рп}}^{\text{мин}} + T_{\text{ок}})))$ принимаются на основе следующего функционала, в котором величина $t_{\text{оп}} + (T_{\text{рп}}^{\text{мин}} + T_{\text{ок}})$ обозначена как $t_{\text{оп}} + \dots$:

$$\begin{aligned} \bar{R}(L_3^{\text{пп}})(x(1), x(2), \dots, x(n_k^{\text{пп}}))(t_{\text{оп}}, (t_{\text{оп}} + \dots)) = \\ = P_{\text{оп}}(Z_{\text{ц}}(L_3^{\text{пп}})(t_{\text{оп}}, (t_{\text{оп}} + \dots)), Z_{\text{техн}}(L_3^{\text{пп}})(t_{\text{оп}}, (t_{\text{оп}} + \dots)), \\ Z_{T-B}(L_3^{\text{пп}})(t_{\text{оп}}, (t_{\text{оп}} + \dots)), W_{\text{o}}(L_3^{\text{пп}})(t_{\text{оп}}, (t_{\text{оп}} + \dots))), \quad (3) \end{aligned}$$

где $P_{\text{оп}}$ — оператор, определяющий порядок действий по выбору расписаний для поездов $x(1), x(2), \dots, x(n_k^{\text{пп}})$ на пространстве $L_3^{\text{пп}}$ в период времени $(t_{\text{оп}}, (t_{\text{оп}} + \dots))$; $Z_{\text{ц}}(L_3^{\text{пп}})(t_{\text{оп}}, (t_{\text{оп}} + \dots))$ — цифровая модель пространства $L_3^{\text{пп}}$ с учетом изменений такой модели, используемой для построения технологического или вариантного графика, характерных для периода $(t_{\text{оп}}, (t_{\text{оп}} + \dots))$; $Z_{\text{техн}}(L_3^{\text{пп}})(t_{\text{оп}}, (t_{\text{оп}} + \dots))$ — то же для описания технологии пропуска поездов (специализации путей, направлений движения поездов и др.); $Z_{T-B}(L_3^{\text{пп}})(t_{\text{оп}}, (t_{\text{оп}} + \dots))$ — данные технологического или вариантного графиков (расписания пассажирских поездов, порядок подвода тяжелой техники и др.), используемые для принятия управляющих решений \bar{R} ; $W_{\text{o}}(L_3^{\text{пп}})(t_{\text{оп}}, (t_{\text{оп}} + \dots))$ — возмущающие воздействия, которые в момент $t_{\text{оп}}$ могут быть учтены при принятии управляющих решений \bar{R} (опоздания пассажирских поездов, фактические ограничения скорости движения на пространстве $L_3^{\text{пп}}$ и др.).

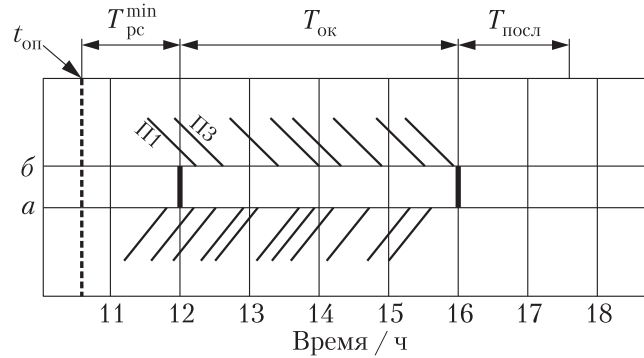


Рис. 9 Прогнозный оперативный график подхода поездов (грузовых и пассажирских) к «перегону окна» $a-b$ с учетом влияния возмущающих воздействий $W_o(L_3^{np})(t_{оп}, (t_{оп} + \dots))$; П1, П3 — пассажирские поезда

Результатом выполнения расчетов на основе зависимости (3) является представленный в качестве примера на рис. 9 прогнозный оперативный график подхода грузовых и пассажирских поездов по четному и нечетному направлениям движения к «перегону окна» $a-b$ с учетом влияния поддающихся учету в момент $t_{оп}$ возмущающих воздействий $W_o(L_3^{np})(t_{оп}, (t_{оп} + \dots))$.

Следующим этапом является разработка оперативного графика для «перегона окна» с переходом от пространства L_3^{np} к пространству $L_{аб}^{np}$ в рассматриваемый период времени $((t_{оп} + T_{рс}^{min}), (t_{оп} + T_{рс}^{min} + T_{посл}))$. Основным условием такой разработки графика является реализация целевой функции, рассмотренной выше.

В результате строится оперативный график движения по «перегону окна» так, как показано в качестве примера на рис. 6. Контроль выполнения оперативного графика с использованием подсистемы обратной связи строится на основе следующих технологических положений.

Разработанный оперативный график для «перегона окна» действует в период его предоставления $(t_{н}^o, t_{зав}^o)$. Основной задачей является точное его выполнение, что отвечает условию целевой функции, предусматривающему максимизацию числа поездов, пропускаемых в период $(t_{н}^o, t_{зав}^o)$.

Подсистема обратной связи обеспечивает контроль следования поездов и поездообразования. Следование поездов в пределах «участка окна» контролируется системой спутниковой навигации, поскольку на подходе к «перегону окна» необходима повышенная точность определения координат поездов. Контроль поездообразования целесообразно осуществлять по данным системы АСОУП-2 (АСУСС).

Различного рода возмущающие воздействия $W_o(L_3^{np})(t_{оп}, (t_{оп} + \dots))$ могут нарушать предусмотренное оперативным графиком расписание движения поез-

дов. Однако не всякое изменение фактических расписаний следования поездов требует пересмотра оперативного графика.

Система толерантна к изменению входных данных по выполнению разработанного оперативного графика до тех пор, пока соблюдаются условия образования предусмотренных пакетов поездов в четном и нечетном направлениях движения. И только в случае невозможности соблюдения этих условий из-за нарушения расписаний подхода поездов к «перегону окна» требуется пересмотр оперативного графика, т. е. его корректировка.

Каждый поезд в оперативном графике, следующий по «перегону окна» в период $(t_{\text{н}}^{\circ}, t_{\text{зав}}^{\circ})$, а именно: $x(1), x(2), \dots, x(\gamma), \dots, x(n_k^{\text{пп}})$, — имеет некоторый резерв времени:

$$\Delta T_{\text{рез}}^{x(\gamma)} = t_{\text{отпр.по}}^{x(\gamma)} - t_{\text{пр.по}}^{x(\gamma)}, \quad (4)$$

где $\Delta T_{\text{рез}}^{x(\gamma)}$ — абсолютное значение резерва времени для поезда $x(\gamma)$; $t_{\text{пр.по}}^{x(\gamma)}$ — то же для прибытия поезда $x(\gamma)$ к «перегону окна» для формирования очередного пакета поездов в соответствии с оперативным графиком.

Если соблюдается условие

$$T_{\text{опозд}}^{x(\gamma)} \leq \Delta T_{\text{рез}}^{x(\gamma)}, \quad (5)$$

где $T_{\text{опозд}}^{x(\gamma)}$ — фактическое или расчетное время отклонения $x(\gamma)$ -го поезда от расписания, предусмотренного в оперативном графике, то корректировка оперативного графика не требуется, т. е. система толерантна к изменению входных данных.

При несоблюдении условия (5) требуется проведение расчетов по корректировке оперативного графика.

Смысл зависимостей (4) и (5) хорошо понятен из рис. 6, на котором представлено значение $\Delta T_{\text{рез}}^{x(\gamma)}$ для поезда № 10 — $\Delta T_{\text{рез}}^{x(10)}$.

Техническое обеспечение системы АС ОГДПО должно предусматривать выполнение условия

$$\Delta T_{\text{тол}} \rightarrow \min,$$

где $\Delta T_{\text{тол}}$ — период проверки системы на толерантность к изменению входных данных по выполнению оперативного графика. При этом очередной период проверки системы на толерантность наступает в текущий момент времени $t_{i+1} \geq t_i + \Delta T_{\text{тол}}$, где t_i и t_{i+1} — начало текущего и последующего периодов проверки системы на толерантность.

Алгоритм работы подсистемы обратной связи представлен на рис. 10.

Рассмотренные технологические обоснования построения оперативных графиков в период предоставления «окон» и организации их выполнения с использованием подсистемы обратной связи обеспечивают минимизацию задержек поездов при снижении пропускной способности временно однопутного «перегона окна».

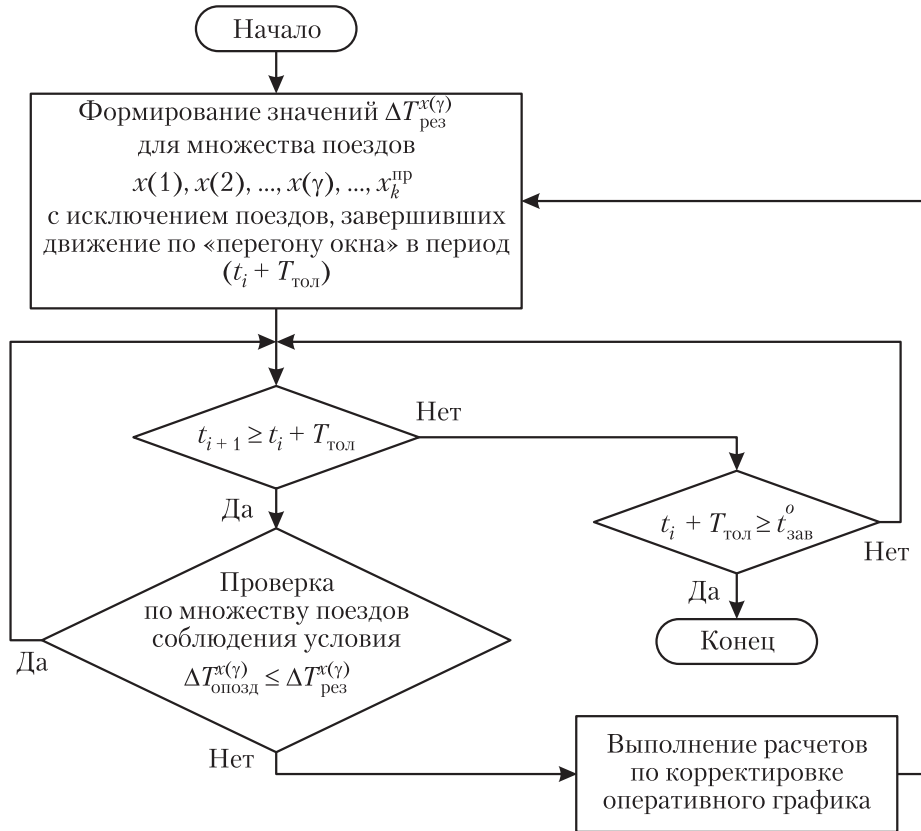


Рис. 10 Алгоритм работы подсистемы обратной связи и корректировки оперативного графика движения в период «окна»

5 Заключение

Разработана архитектура и технологические основы АС ОГДПО автоматизации управления процессом работы участка в период предоставления «окна», определен состав, функции и условия взаимодействия с учетом использования спутниковых технологий. Определена целевая функция АС ОГДПО, основанная на выполнении условия обеспечения пропускания максимального числа поездов различного веса по «перегону окна» из общего числа реально следующих поездов в прогнозный период суток «окна» с учетом максимизации числа поездов в пакетах направления преимущественного движения. Заданы ограничения и условия, определяющие области допустимых значений расчетных параметров.

В качестве инструментальной среды реализована система имитационного моделирования работы участка, учитывающая принципы построения «полигона окна» и движения поездов на его пространственно-временной модели, информацию о подходе поездов с «полигона окна» к «участку окна» и «перегону окна» и обеспечивающая построение прогнозного оперативного графика подхода поездов к «перегону окна» и разработку оперативного пакетного графика движения поездов по нему.

Литература

1. *Сотников Е. А.* Эксплуатационная работа железных дорог. — М.: Транспорт, 1986. 256 с.
2. *Розенберг И. Н., Цветков В. Я., Матвеев С. И., Дулин С. К.* Интегрированная система управления железной дорогой / Под ред. В.И. Якунина. — 2-е изд., перераб. и доп. — М.: Дизайн. Информация. Картография, 2008. 144 с.
3. *Ададуров С. Е., Розенберг Е. Н., Розенберг И. Н.* Оптимизация управления инфраструктурой и безопасностью движения // Железнодорожный транспорт, 2009. № 9. С. 25–30.
4. *Матвеев С. И., Коугия В. А.* Высокоточные цифровые модели пути и спутниковая навигация железнодорожного транспорта. — М.: Маршрут, 2005. 290 с.
5. *Розенберг И. Н., Уманский В. И., Дулин С. К., Калинин С. В.* Автоматизированное построение оперативных графиков движения поездов с учетом данных спутникового позиционирования // Железнодорожный транспорт, 2009. № 11. С. 68–70.
6. *Уманский В. И.* Об организации пропуска поездов в период технологических «окон» // Железнодорожный транспорт, 2010. № 9. С. 21–24.
7. *Дулин С. К., Калинин С. В., Уманский В. И.* Анализ эксплуатационных характеристик станций на базе геоинформационного представления данных. — М.: ВЦ РАН, 2008. 28 с.
8. *Дулин С. К., Селецкий А. С., Уманский В. И.* Система имитационного моделирования движения железнодорожного транспорта на основе интерактивно задаваемых правил организации движения // КИИ-2008: Труды 11-й Национальной конф. по искусственному интеллекту. — М.: Ленанд, 2008. Т. I. С. 77–85.
9. *Лысиков М. Г.* Автоматизация разработки вариантных графиков движения поездов в условиях предоставления окна // Железнодорожный транспорт, 2008. № 11. С. 33–36.

Поступила в редакцию 14.03.13

AUTOMATED SYSTEM DEVELOPMENT AND CONTROL OF OPERATIONAL TRAIN SCHEDULE DURING “FREE ZONES”

S. Dulin^{1,2}, I. Rozenberg¹, and V. Umanskiy³

¹Research and Design Institute for Information Technology, Signalling and Telecommunications on Railway Transport (JSC NIIAS), 27-1 Nizhegorodskaya Str., Moscow 109029, Russian Federation

²Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

³Close Corporation “IntekhGeoTrans,” 5-1 Orlikov per., Moscow 107996, Russian Federation

Abstract: This paper presents the architecture and technological foundations of process control automation of the site during the period of “free zones” to determine the composition, functions and conditions of interaction using the satellite technology. The objective function of the presented system based on the condition ensuring passage of the maximum number of trains was developed. Limits and conditions governing the tolerance range of the design parameters were defined. The authors implemented a system of site simulation, taking into account the principles of the “polygon box” and trains on its space–time model, information on approach of trains with a “polygon box” to “section of the zone” and “stretch of the zone” and allowing to construct a projected operational schedule of trains to approach “the stretch of the zone” and to develop an operational packet traffic schedule.

Keywords: operational train schedule; period of “free zones;” site simulation

DOI: 10.14357/08696527140106

References

1. Sotnikov, E. A. 1986. *Ekspluatatsionnaya rabota zheleznikh dorog* [The operational work of the railways]. Moscow: Transport. 256 p.
2. Rozenberg, I. N., V. Ja. Cvetkov, S. I. Matveev, and S. K. Dulin. 2008. *Integrirovannaya sistema upravleniya zheleznoy dorogoy* [The integrated control system railway]. 2nd ed. Moscow: Dizayn. Informaciya. Kartografiya. 144 p.
3. Adadurov, S. E., E. N. Rozenberg, and I. N. Rozenberg. 2009. Optimizatsiya upravleniya infrastrukturoy i bezopasnost'yu dvizheniya [Optimization of infrastructure management and traffic safety]. *Zheleznodorozhnyy Transport* [Rail Transport] 9:25–30.
4. Matveev, S. I., and V. A. Kougiya. 2005. *Vysokotochnye tsifrovye modeli puti i sputnikovaya navigatsiya zheleznodorozhnogo transporta* [Highly accurate digital models of satellite navigation path and rail transport]. Moscow: Marshrut. 290 p.

5. Rozenberg, I. N., V. I. Umanskij, S. K. Dulin, and S. V. Kalinin. 2009. Avtomatizirovanoe postroenie operativnykh grafikov dvizheniya poezdov s uchetom dannykh sputnikovogo pozitsionirovaniya [Automated construction of operational train schedules based on satellite positioning]. *Zheleznodorozhnyy Transport* [Rail Transport] 11:68–70.
6. Umanskij, V. I. 2010. Ob organizatsii propuska poezdov v period tekhnologicheskikh “okon” [On the organization of trains crossing between technological “windows”]. *Zheleznodorozhnyy Transport* [Rail Transport] 9:21–24.
7. Dulin, S. K., S. V. Kalinin, and V. I. Umanskij. 2008. *Analiz ekspluatatsionnykh kharakteristik stantsiy na baze geoinformatsionnogo predstavleniya dannykh* [Analysis of the performance of the stations on the basis of geodata representation]. Moscow: Dorodnicyn Computing Centre of RAN. 28 p.
8. Dulin, S. K., A. S. Seleckij, and V. I. Umanskij. 2008. Sistema imitatsionnogo modelirovaniya dvizheniya zheleznodorozhnogo transporta na osnove interaktivno zadavaemykh pravil organizatsii dvizheniya [System simulation of the railway transport on the basis of interactively defined rules for the organization of movement]. *Trudy 11-y Natsional’noy Konferentsii po Iskusstvennomu Intellectu “KII 2008”* [11th National Conference on Artificial Intelligence “CAI-2008” Proceedings]. Dubna. 1:77–85.
9. Lysikov, M. G. 2008. Avtomatizatsiya razrabotki variantnykh grafikov dvizheniya poezdov v usloviyakh predostavleniya okna [Automating of the development of variant train schedules in terms of providing a window]. *Zheleznodorozhnyy Transport* [Rail Transport] 3:33–36.

Received March 14, 2013

Contributors

Dulin Sergey K. (b. 1950) — Doctor of Science in technology, professor, principal scientist, Research and Design Institute for Information Technology, Signalling and Telecommunications on Railway Transport (JSC NIIAS), 27-1 Nizhegorodskaya Str., Moscow 109029, Russian Federation; senior scientist, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; s.dulin@ccas.ru

Rozenberg Igor N. (b. 1965) — Doctor of Science in technology, professor, First Deputy Director General, Research and Design Institute for Information Technology, Signalling and Telecommunications on Railway Transport (JSC NIIAS), 27-1 Nizhegorodskaya Str., Moscow 109029, Russian Federation; I.Rozenberg@gismps.ru

Umansky Vladimir I. (b. 1954) — Doctor of Science in technology, Director General, Close Corporation “IntekhGeoTrans,” 5-1 Orlikov per., Moscow 107996, Russian Federation; umanvi@yandex.ru

ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ ИССЛЕДОВАНИЯ ТОНКОЙ СТРУКТУРЫ ХАОТИЧЕСКИХ ПРОЦЕССОВ В ПЛАЗМЕ С ПОМОЩЬЮ АНАЛИЗА СПЕКТРОВ*

*А. К. Горшенин*¹

Аннотация: Рассмотрена информационная технология, созданная для изучения с помощью спектрального анализа специфических структур, формирующих турбулентность в плазме. Ее математическую основу составляет вероятностный подход, использующий оригинальный бутстреп-алгоритм. В качестве базовой модели для описания тонкой структуры хаотических процессов используются конечные смеси различных вероятностных распределений. Для отыскания статистических оценок неизвестных параметров модели в программный комплекс включены специально созданные статистические процедуры, улучшающие свойства классического EM-алгоритма (expectation-maximization algorithm). Значительное внимание в статье уделяется структуре и общей схеме функционирования разработанного программного обеспечения, однако обсуждаются и практические результаты, полученные в важном разделе современной физики (например, определение числа формирующих процессов, значений ряда физических величин и др.). Продемонстрированы наглядные примеры работы созданного программного пакета.

Ключевые слова: спектр; информационная технология; турбулентная плазма; бутстреп; вероятностные модели

DOI: 10.14357/08696527140107

1 Введение

В целях создания эффективных и безопасных научных установок значительного масштаба, основанных на концепции токамаков (например, международного экспериментального термоядерного реактора — ITER), необходимо развитие корректных моделей функционирования турбулентной плазмы. Построение таких моделей невозможно без изучения значительного объема данных, получаемых на установках меньшего масштаба в лабораторных условиях, для чего необходима разработка соответствующих математических методов, алгоритмов и программных средств.

* Работа выполнена при поддержке Гранта Президента РФ для государственной поддержки молодых российских ученых МК-4103.2014.9 и Российского фонда фундаментальных исследований (проекты 12-07-00115а и 14-07-00041а).

¹ Институт проблем информатики Российской академии наук; Московский государственный технический университет радиотехники, электроники и автоматики; agorshenin@ipiran.ru

Задача определения параметров тех или иных процессов по спектрам, полученным в ходе эксперимента при помощи различных способов (например, измерения спектрометром, спектрографом или оценки по временной выборке аналого-цифрового преобразователя), может быть решена с помощью разделения спектра на составляющие: сплошной спектр, полосы, отдельные компоненты. Однако существует проблема, заключающаяся в неоднозначной трактовке одних и тех же результатов экспериментов. Дело в том, что задача выявления компонент спектра является математически некорректной и требует дополнительных предположений, в рамках которых решение может быть найдено. Эта проблема известна, например, в оптической спектроскопии; столь же актуальна она и для физики плазмы, так как плазма представляет собой состояние вещества с большим числом степеней свободы. В данном случае спектр может быть достаточно успешно приближен сразу целым рядом различных между собой моделей (например, моделью дробового шума, моделью Колмогорова–Обухова), причем отдать предпочтение какой-то единственной модели не представляется возможным. Очевидно, что в такой ситуации невозможно получить точную спектральную информацию о механизмах турбулентности. Однако именно исследование спектров позволяет добиться более тонких результатов, поскольку спектр дает возможность определить такие важные показатели, как тип неустойчивости, механизм формирования турбулентности, доли ионно-звуковых солитонов и дрейфовых вихрей и т. д.

Для преодоления указанных сложностей в статье [1] была предложена бутстреп-методология моделирования спектров, полученных в экспериментах с турбулентной плазмой. Исходный спектр рассматривается в качестве плотности неизвестного закона распределения, с помощью вероятностного моделирования строится тестовая выборка нужного (вообще говоря, произвольного) объема. В дальнейшем она используется для построения аппроксимирующей кривой, которая в соответствии с моделями обобщенных процессов Кокса ищется в виде конечной сдвиг-масштабной смеси вероятностных распределений (нормальных, гамма). Важной задачей становится выбор и оптимизация вычислительных процедур, предназначенных для оценивания параметров модели.

В дальнейшем для некоторых экспериментальных режимов функционирования плазмы в работе [2] с применением созданных специализированных программных модулей была продемонстрирована возможность качественного приближения спектров с высокой степенью согласия с ранее полученными результатами. Было получено подтверждение эффективности использования предложенной методологии, так как для ряда режимов функционирования турбулентной плазмы была решена задача приближения спектра и разделения его на компоненты, число и форма которых сохраняются внутри характерных областей в серии. Также было продемонстрировано хорошее соответствие с физической сутью изучаемых процессов для числа значимых компонент в разложении: в реальных системах количество формирующих процессов хоть и велико, но наиболее значимыми оказываются преимущественно 3–4 из них. Таким образом, работы в данной области представляют значительный интерес для продолжения исследова-

дований как с практической, так и с теоретической точки зрения для построения более точных моделей функционирования турбулентной плазмы.

В настоящей работе приводится описание итогового программного комплекса, предназначенного для разностороннего спектрального анализа экспериментальных данных, предлагающего широкие возможности для изучения вида, формы и характерных особенностей спектров. Исходными данными для моделирования послужили экспериментальные выборки со специального стелларатора Л-2М (см., например, работу [3]), полученные в различных режимах функционирования турбулентной плазмы.

2 Математические основы информационной технологии

В этом разделе кратко рассмотрим основные принципы и математические модели, использованные при разработке программного обеспечения.

В ходе исследований низкочастотных пульсаций плазмы было установлено, что плазменная турбулентность успешно описывается моделями типа обобщенного процесса Кокса (см. [4]). Исходный спектр интерпретируется как плотность некоторого неизвестного вероятностного распределения, функцию распределения которого обозначим через $\mathcal{F}(x)$. Затем находим квантиль x_α порядка α для функции распределения $\mathcal{F}(x)$, т. е. решаем уравнение

$$\mathcal{F}(x_\alpha) = \alpha.$$

Поступая так необходимое число раз, можно сформировать выборку \mathcal{X} из неизвестного распределения $\mathcal{F}(x)$ заранее заданного (и при этом сколь угодно большого) объема. Для этой выборки используем базовое предположение о том, что распределение $\mathcal{F}(x)$ представляет собой конечную смесь вероятностных распределений (нормальных, гамма) с неизвестными параметрами, т. е.

$$\mathcal{F}(x) = \sum_{i=1}^k p_i F(x, a_i, b_i), \quad \sum_{i=1}^k p_i = 1, \quad p_i \geq 0,$$

где через $F(\cdot)$ обозначена функция распределения соответствующего типа (нормального закона или гамма-распределения); $k \geq 1$ — известное натуральное число; величины a_i и b_i , $i = 1, \dots, k$, представляют собой параметры соответствующего распределения с необходимыми ограничениями (например, коэффициенты масштаба для нормального распределения строго положительны и т. п.). Величины $F(x, a_i, b_i)$, $i = 1, \dots, k$, называются компонентами смеси, параметры p_1, \dots, p_k — весами соответствующих компонент. Значения p_i , a_i и b_i , $i = 1, \dots, k$, обычно неизвестны, поэтому требуется использовать специальные процедуры оценивания параметров модели по выборке.

Для этого можно использовать различные алгоритмы отыскания оценок максимального правдоподобия. В рамках созданного программного пакета в

качестве основных вычислительных методов использовался классический EM-алгоритм [5] для смесей нормальных и гамма-распределений, а также специализированный сеточный метод [6]. Это позволило реализовать приемлемую точность вычислений за разумное время для различных типов спектров (односторонних, двусторонних и т. п.).

3 Структура программного обеспечения

В данном разделе рассмотрим вопросы функционирования и взаимосвязей программных модулей, составляющих вычислительную часть информационной технологии.

В условиях современных исследований, когда экспериментальные данные имеют весьма значительный объем и требуют существенного времени для обработки, невозможно представить проведение анализа процессов без создания специализированной информационной технологии. На основании упомянутой выше математической модели, основанной на использовании обобщенных процессов Кокса, была разработана соответствующая информационная технология для решения задачи разделения серий спектров на компоненты. При исследовании включенных в состав информационной технологии математических методов были получены новые научные результаты, касающиеся адекватности методов и их быстродействия, которые и предопределили набор алгоритмов, вошедших в конечный пакет. Вычислительный алгоритм был создан с использованием встроенного в пакет MATLAB языка программирования четвертого поколения и оптимизирован для проведения исследований в физике плазмы.

Разработанная информационная технология включает в себя математический аппарат проведения расчетов, расчетные модули, инструменты визуализации полученных результатов с различными возможностями их сохранения, а также пользовательский интерфейс для обеспечения удобства работы с пакетом для исследователей. Рассмотрим подробнее структуру разработанного программного обеспечения (рис. 1).

Экспериментальные данные обычно предоставляются в виде внешних файлов, доступ к которым осуществляется через интерфейсную часть пакета. Здесь существуют различные варианты, так как предусмотрены варианты работы и через стандартную для MATLAB консоль (существует «головная» функция, выполняющая всю обработку), и через использование специализированного интерфейса. Первый способ подходит для использования в «режиме разработчика», в то время как большинству пользователей, безусловно, удобнее работать с интерфейсом, скрывающим детали реализации и упрощающим работу с программой.

Логически функционал программы делится на три основных блока:

- (1) блок моделирования, отвечающий за создание тестовой выборки;
- (2) блок оценивания параметров неизвестного распределения по построенной тестовой выборке, отвечающий за отыскание аппроксимирующей кривой;

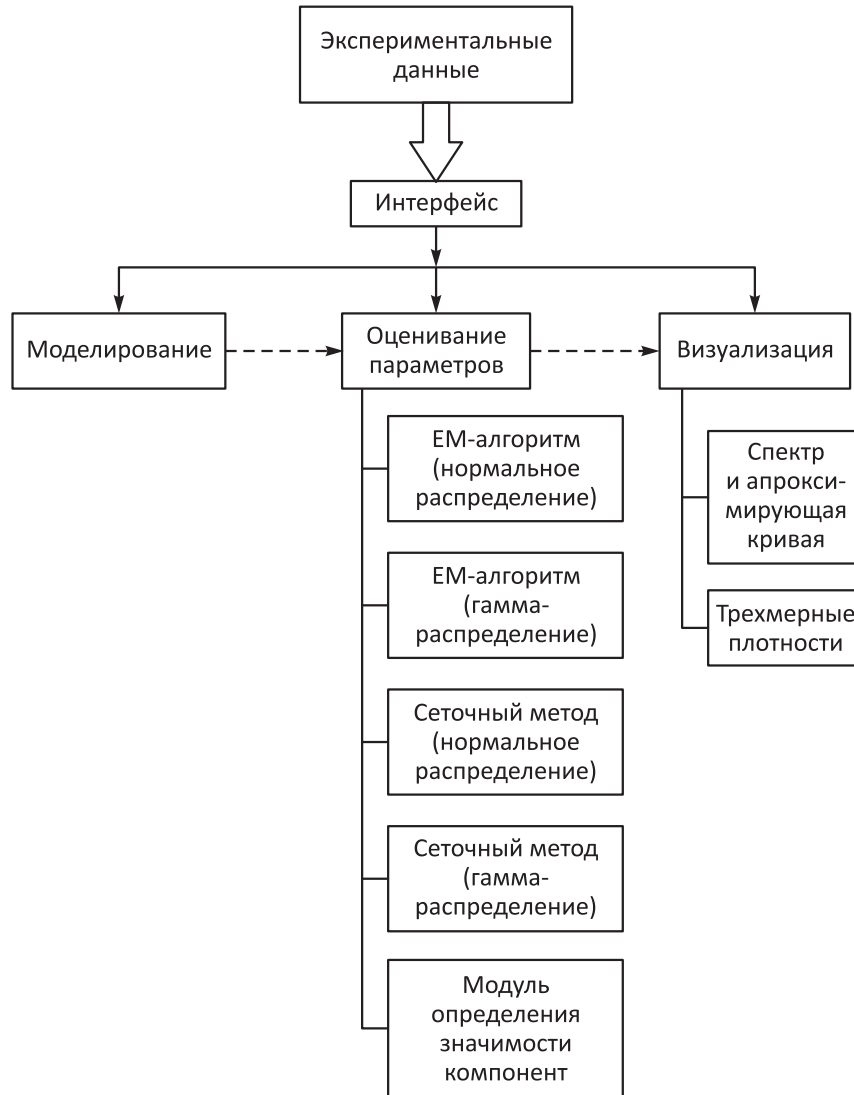


Рис. 1 Структура программного пакета

(3) блок визуализации, осуществляющий графическое отображение полученных результатов.

Очевидно, что каждый из этих этапов должен выполняться последовательно друг за другом (на рис. 1 это продемонстрировано с помощью пунктирных

стрелок), однако начать можно с любого пункта (на рис. 1 возможные пути выполнения изображены сплошными стрелками от блока «Интерфейс»). Дело в том, что в программе предусмотрена работа с данными, которые были ранее сохранены на диске, что позволяет избежать повторного моделирования или отыскания оценок.

Тестовая выборка строится с помощью описанной в предыдущем разделе процедуры. При оценивании параметров по выбору пользователя могут быть использованы следующие методы:

- EM-алгоритм для нормальных распределений;
- EM-алгоритм для гамма-распределений;
- сеточный метод для нормальных распределений;
- сеточный метод для гамма-распределений.

Здесь методы, основанные на нормальном распределении, можно использовать для одно- и двусторонних спектров, а основанные на гамма-распределении — для односторонних спектров. EM-алгоритм является достаточно универсальным методом отыскания оценок максимального правдоподобия. С его помощью можно добиться приемлемого баланса между скоростью и качеством вычислений. Однако в случае, когда есть некоторая дополнительная информация о параметрах, более эффективными с вычислительной точки зрения могут оказаться сеточные методы, поэтому они включены в итоговый пакет.

Известно, что при построении моделей типа конечных сдвиг-масштабных смесей вероятностных распределений возникает задача корректного определения числа компонент в приближающей смеси (см. [7]). Поэтому для анализа могут быть использованы асимптотически наиболее мощные критерии для моделей добавления и расщепления компоненты, а также расчета значений статистики для критерия Ло [8]. Программа проверяет по упомянутым выше трем критериям значимость компоненты с малым весом, выдавая в качестве ответа число компонент в смеси для каждого из них. Уровень значимости критерия служит параметром метода.

В блоке визуализации осуществляется графический вывод результатов вычислений. По полученным оценкам параметров строятся графики аппроксимирующей кривой и составляющих ее компонент для наглядной интерпретации получаемых результатов. Отображаемые графики позволяют наглядно определить значимое число компонент (т. е. процессов в турбулентной плазме) и установить величины ряда физических характеристик. Более того, анализируемые ряды обычно представляют собой серии спектров, полученных для различных этапов экспериментов. Достаточно наглядным и удобным представляется отображение эволюции спектров (и, соответственно, приближающих кривых) во времени. Для этого был создан специальный модуль отображения трехмерных поверхностей. В каждый момент времени сечение графика представляет собой спектр (аппроксимирующую кривую). Трехмерные графики можно вращать в пространстве с

помощью мыши и масштабировать. Полученные изображения могут быть сохранены в одном из распространенных графических форматов (например, JPEG или PNG).

Математические алгоритмы были протестированы на тестовых выборках с заранее известными характеристиками. В рамках разработки информационной технологии были созданы специальные процедуры для повышения эффективности и точности вычислений. Для пакета программ в целом, а также составляющих его модульных решений были получены свидетельства о государственной регистрации программ для ЭВМ. Использование различных алгоритмов и специальных техник в единой программе сделало возможным проведение детального анализа данных для устойчивых серий спектров, полученных для низкочастотной плазменной турбулентности на краю и в центре плазменного шнура в теллараторе L-2M.

Программный пакет представляет собой исполнимый файл, который может запускаться как на ЭВМ с установленной средой MATLAB, так и без нее (однако в этой ситуации потребуется установка свободно распространяемого компонента MATLAB Compiler Runtime).

4 Демонстрация полученных результатов

В данном разделе приведены наглядные образцы работы разработанного программного обеспечения, а также обсуждаются полученные физические результаты.

На рис. 2 изображены примеры графического вывода программы. На каждом из них представлены:

- исходный спектр;
- аппроксимирующая смесь;
- компоненты смеси.

По оси абсцисс отложены значения частот, по оси ординат — соответствующая величина спектра для конкретного временного интервала эксперимента. Стоит отметить, что несмотря на обычно весьма большое число процессов в турбулентной плазме, формирующими оказываются сравнительно немного процессов. На приведенных графиках представлена демонстрация этого факта — видно, что значимыми (видимыми при полученных значениях параметров, обладающими существенно нетривиальными весами) оказываются не более 4–5 компонент.

На рис. 2, а продемонстрировано приближение двустороннего спектра, а также образец англоязычного вывода программы. На рис. 2, б приведен пример для одностороннего спектра с русскоязычным выводом. Отметим, что на данных графиках изображено приближение кривой, представляющей собой конечную сдвиг-масштабную смесь нормальных законов.

Рассмотрим основные физические результаты, полученные в настоящий момент благодаря использованию разработанной информационной технологии:

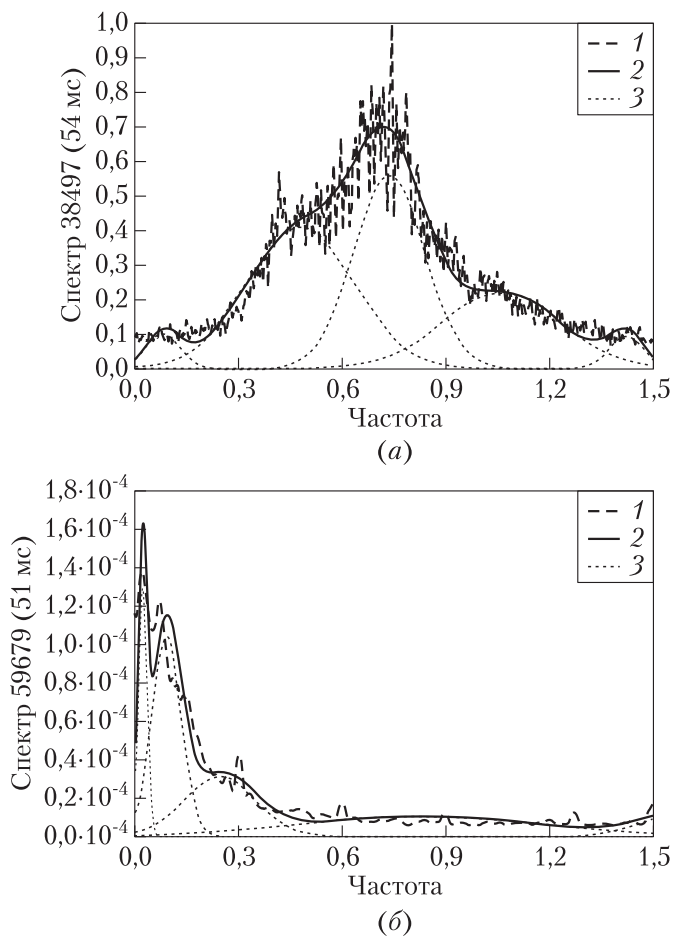


Рис. 2 Примеры приближения двустороннего (а) и одностороннего (б) спектров смесью нормальных распределений: 1 — сглаженный спектр; 2 — смесь; 3 — одиночные компоненты

- проведение идентификации спектра: удалось определить форму гармоник в амплитудном спектре;
- выделение корректного числа процессов в турбулентной плазме: установлено, что новые результаты подтверждают выводы, ранее сделанные на основании других методик исследования;
- выявление повторяемости стохастических процессов с характерными средними частотами полуширины спектра;

- определение таких физических показателей функционирования плазмы, как величина радиального электрического поля, фазовые скорости флуктуаций и др.

5 Заключение

Итак, в рамках создания информационной технологии был разработан эффективный программный инструмент исследования процессов в области физики плазмы (а значит, и в сфере альтернативных источников энергии), позволяющий получать содержательные результаты непосредственно в настоящий момент времени.

В силу информационных свойств вероятностных распределений, известных физических результатов из ряда смежных областей интересным представляется рассмотрение задачи моделирования односторонних спектров конечными смесями гамма-распределений. Для решения данной задачи был создан программный модуль анализа спектров с помощью смесей гамма-распределений, реализующий интерфейс и программные методы для разделения спектров произвольной природы на компоненты, представляющие собой гамма-распределения с различными параметрами и весами. Более того, в ряде работ (см., например, [9, 10]) показана эффективность использования значительно более сложных моделей типа дисперсионно-сдвиговых смесей для высокочастотных финансовых данных. Имеющаяся взаимосвязь между спектрами и частотами, а также схожий стохастический характер процессов в турбулентной плазме и на финансовых рынках [4] показывают необходимость проведения дополнительных исследований более сложных моделей и в физике плазмы. При этом создание эффективных вычислительных процедур для таких моделей должно составлять предмет отдельного исследования и включать в себя математические, программные и аппаратные решения.

Разработанные методы вносят вклад в область применения информационных технологий в физике турбулентной плазмы. Предложенная технология может использоваться для построения центров по обработке экспериментальных данных. Более того, возможно ускорение работы методов, значительное увеличение количества одновременно обрабатываемых данных за счет привлечения методик параллельного программирования и задействования новых вычислительных возможностей (проведение расчетов на графических процессорах, кластерах и т. п.). Данные улучшения в совокупности с продемонстрированным качеством результатов позволяют предположить эффективность потенциального применения предлагаемых процедур для обработки данных со стеллараторов принципиально более высокой сложности (вроде упоминавшегося выше термоядерного реактора ИТЕР).

Автор выражает глубокую признательность профессору *В. Ю. Королеву* (факультет вычислительной математики и кибернетики МГУ им. М. В. Ломоносова) за полезные обсуждения при работе над математическими моделями, *Н. Н. Скворцовой* и *Д. В. Малахову* (Институт общей физики РАН им.

А. М. Прохорова) за предоставленные экспериментальные данные и физическую интерпретацию полученных результатов.

Литература

1. Горшенин А. К., Королев В. Ю., Малахов Д. В., Скворцова Н. Н. Анализ тонкой стохастической структуры хаотических процессов с помощью ядерных оценок // Математическое моделирование, 2011. Т. 23. № 4. С. 83–89.
2. Горшенин А. К., Королев В. Ю., Малахов Д. В., Скворцова Н. Н. Об исследовании плазменной турбулентности на основе анализа спектров // Компьютерные исследования и моделирование, 2012. Т. 4. № 4. С. 793–802.
3. Pshenichnikov A. A., Kolik L. V., Malykh N. I., et al. The use of Doppler reflectometry in the L-2M stellarator // Plasma Phys. Rep., 2005. Vol. 31. No. 7. P. 554–561.
4. Королев В. Ю. Вероятностно-статистические методы декомпозиции волатильности хаотических процессов. — М.: Изд-во МГУ, 2011. 512 с.
5. Dempster A., Laird N., Rubin D. Maximum likelihood estimation from incompleting data // J. Roy. Stat. Soc. B, 1977. Vol. 39. No. 1. P. 1–38.
6. Gorshenin A., Korolev V., Kuzmin V., Zeifman A. Coordinate-wise versions of the grid method for the analysis of intensities of non-stationary information flows by moving separation of mixtures of gamma-distribution // 27th Conference (European) on Modelling and Simulation Proceedings. — Dudweiler, Germany: Digitaldruck Pirrot, 2013. P. 565–568.
7. Бенинг В. Е., Горшенин А. К., Королев В. Ю. Асимптотически оптимальный критерий проверки гипотез о числе компонент смеси вероятностных распределений // Информатика и её применения, 2011. Т. 5. Вып. 3. С. 4–16.
8. Lo Y., Mendell N. R., Rubin D. B. Testing the number of components in a normal mixture // Biometrika, 2001. Vol. 88. No. 3. P. 767–778.
9. Королев В. Ю., Черток А. В., Корчагин А. Ю., Горшенин А. К. Вероятностно-статистическое моделирование информационных потоков в сложных финансовых системах на основе высокочастотных данных // Информатика и её применения, 2013. Т. 7. Вып. 1. С. 12–21.
10. Gorshenin A. K., Korolev V. Yu., Zeifman A. I., Shorgin S. Ya., Chertok A. V., Evstafyev A. I., Korchagin A. Yu. Modelling stock order flows with non-homogeneous intensities from high-frequency data // AIP Conference Proceedings, 2013. Vol. 1558. P. 2394–2397.

Поступила в редакцию 27.12.13

**INFORMATION TECHNOLOGY TO RESEARCH THE FINE
STRUCTURE OF CHAOTIC PROCESSES IN PLASMA
BY THE ANALYSIS OF SPECTRA**

A. K. Gorshenin^{1,2}

¹Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

²Moscow Institute of Radio, Electronics, and Automation (MIREA), 78 Vernadskogo Ave., Moscow 119454, Russian Federation

Abstract: The paper deals with the information technology created for investigation of specific structures, forming turbulence in plasma, with the help of spectral analysis. The mathematical basis is the probabilistic approach using a novel bootstrap algorithm. The finite mixtures of various probability distributions are used as a basic model to describe the fine structure of chaotic processes. To find the statistical estimates of the unknown parameters of the model, specially created statistical procedures, improving the properties of the classical EM (expectation-maximization) algorithm, are included in the software. Considerable attention is devoted to the structure and general principles of software functioning. Practical results obtained in an important area of modern physics are discussed as well (e. g., identification of the number of forming processes, evaluation of several physical parameters, etc.). The vivid examples of performance of the created software are given.

Keywords: spectrum; information technology; turbulent plasma; bootstrap; probabilistic models

DOI: 10.14357/08696527140107

Acknowledgments

The research was supported by the President of Russian Federation Grant for Government Support of Young Russian Scientists (No. MK-4103.2014.9) and Russian Foundation for Basic Research (projects Nos. 12-07-00115a and 14-07-00041a). The author acknowledges Prof. V. Yu. Korolev (Faculty of Computational Mathematics and Cybernetics of M. V. Lomonosov Moscow State University) for fruitful discussions on mathematical models, N. N. Skvortsova and D. V. Malakhov (A. M. Prokhorov Institute of General Physics of the Russian Academy of Sciences) for experimental data and physical interpretation of the obtained results.

References

1. Gorshenin, A. K., V. Yu. Korolev, D. V. Malakhov, and N. N. Skvortsova. 2011. Analiz tonkoy stokhasticheskoy struktury khaoticheskikh protsessov s pomoshch'yu yadernykh otsenok [Analysis of fine stochastic structure of chaotic processes by kernel estimators]. *Matematicheskoe Modelirovanie* 23(4):83–89.
2. Gorshenin, A. K., V. Yu. Korolev, D. V. Malakhov, and N. N. Skvortsova. 2012. Ob issledovanii plazmennoy turbulentnosti na osnove analiza spektrov [On the investigation of plasma turbulence by the analysis of the spectra]. *Komp'yuternye issledovaniya i modelirovanie* [Computer Research and Modeling] 4(4):793–802.
3. Pshenichnikov, A. A., L. V. Kolik, N. I. Malykh, et al. 2005. The use of Doppler reflectometry in the L-2M stellarator. *Plasma Phys. Rep.* 31(7):554–561.
4. Korolev, V. Yu. 2011. *Veroyatnostno-statisticheskie metody dekompozitsii volatil'nosti khaoticheskikh protsessov* [Probabilistic and statistical methods of decomposition of volatility of chaotic processes]. Moscow: Moscow State University Publishing House. 512 p.
5. Dempster, A., N. Laird, and D. Rubin. 1977. Maximum likelihood estimation from incompleting data *J. Roy. Stat. Soc. B* 39(1):1–38.
6. Gorshenin, A., V. Korolev, V. Kuzmin, and A. Zeifman. 2013. Coordinate-wise versions of the grid method for the analysis of intensities of non-stationary information flows by moving separation of mixtures of gamma-distribution. *27th Conference (European) on Modelling and Simulation Proceedings*. Dudweiler, Germany. 565–568.
7. Bening, V. E., A. K. Gorshenin, and V. Yu. Korolev. 2011. Asimptoticheski optimal'nyy kriteriy proverki gipotez o chisle komponent smesi veroyatnostnykh raspredeleniy [An asymptotically optimal test for the number of components of a mixture of probability distributions]. *Informatika i ee Primeneniya — Inform. Appl.* 5(3):4–16.
8. Lo, Y., N. R. Mendell, and D. B. Rubin. 2001. Testing the number of components in a normal mixture. *Biometrika* 88(3):767–778.
9. Korolev, V. Yu., A. V. Chertok, A. Yu. Korchagin, and A. K. Gorshenin. 2013. Veroyatnostno-statisticheskoe modelirovanie informatsionnykh potokov v slozhnykh finansovykh sistemakh na osnove vysokochastotnykh dannykh [Probability and statistical modeling of information flows in complex financial systems based on high-frequency data]. *Informatika i ee Primeneniya — Inform. Appl.* 7(1):12–21.
10. Gorshenin, A. K., V. Yu. Korolev, A. I. Zeifman, S. Ya. Shorgin, A. V. Chertok, A. I. Evstafyev, and A. Yu. Korchagin. 2013. Modelling stock order flows with non-homogeneous intensities from high-frequency data. *AIP Conference Proceedings* 1558: 2394–2397.

Received December 27, 2013

Contributor

Gorshenin Andrey K. (b. 1986) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; associate professor, Moscow Institute of Radio, Electronics, and Automation (MIREA), Moscow 119454, Russian Federation; agorshenin@ipiran.ru

АЛГОРИТМ И ПРОГРАММНЫЙ КОМПЛЕКС ДЛЯ АНАЛИЗА ХАРАКТЕРИСТИК АКУСТИЧЕСКИХ ВОЛН ПЛАЗМЕННОГО АКТУАТОРА, ВСТРОЕННОГО В МОДЕЛЬНОЕ СОПЛО САМОЛЕТА*

А. Д. Черноусов¹, Д. В. Малахов², Н. Н. Скворцова³

Аннотация: В настоящее время остро стоит проблема по разработке новых технологий снижения шума, в том числе и по направленному воздействию на шум на основе взаимодействия плазменных возмущений и звукообразующих пульсаций. Одним из устройств, построенных на таком принципе, являются разрабатываемые в ИОФ РАН плазменные актуаторы (группа связанных друг с другом разрядников, встроенных по периметру сопла) различной конфигурации и формы. Цель данной работы — сравнительный анализ спектров акустических волн, получаемых в результате работы плазменных актуаторов, на воздушный поток, истекающий из сопел. Исследования проводились для актуаторов различной конструкции и различной формы сопел. Для этого были решены следующие задачи: сортировка и анализ экспериментальных данных; разработка алгоритма и создание программы для исследования спектральных и корреляционных характеристик акустических сигналов; оценка спектральных и корреляционных характеристик акустических сигналов в экспериментах с двумя типами актуаторов: с потоками газа и без таковых; сравнительный анализ спектральных и корреляционных характеристик акустических сигналов, получаемых на выходе из сопла при взаимодействии воздушного потока с плазмой актуаторов различной конфигурации. При помощи созданного в среде MatLab программного комплекса были исследованы спектральные и корреляционные характеристики акустических сигналов, что позволило построить физическую модель работы плазменного актуатора в сопле самолета. С помощью разработанной методики исследований показана перспективность использования влияния плазменных актуаторов на воздушный поток, проходящий через сопло.

Keywords: плазменный актуатор; акустические волны; программный комплекс; алгоритм; сопло; спектральные характеристики; робастные спектры

DOI: 10.14357/08696527140108

*Исследование выполнено при поддержке Гранта Президента Российской Федерации для государственной поддержки молодых российских ученых МК-5607.2013.2.

¹Институт общей физики им. А. М. Прохорова Российской академии наук; Акустический институт имени академика Н. Н. Андреева, chado2008@yandex.ru

²Институт общей физики им. А. М. Прохорова Российской академии наук, 89199945791@mail.ru

³Институт общей физики им. А. М. Прохорова Российской академии наук; Национальный исследовательский ядерный университет «МИФИ», mukudori@mail.ru

1 Введение

С начала 2000-х гг. активное управление воздушным потоком при помощи плазменных актуаторов стало набирать все большую популярность. Главное преимущество этого метода в том, что он направленно преобразует электрическую энергию в кинетическую без вовлечения движущихся механических частей. Таким образом, можно считать, что это очень простая микроэлектромеханическая система (МЭМС). Кроме того, время отклика такой системы очень мало, что позволяет осуществлять управление в реальном времени при высокой частоте. Недостатком метода является низкая эффективность в преобразовании энергии. Поверхностные разряды могут также изменять свойства газа на стенке, такие как плотность или вязкость [1].

Свои варианты плазменных актуаторов различной конфигурации и формы были разработаны в ИОФ РАН и названы RING-1 и RING-2 [2]. Эти устройства работают на принципе скользящего разряда и могут быть встроены по периметру сопла. Многоискровые актуаторы RING-1 и RING-2 работают от специального источника питания, генерирующего последовательность высоковольтных (ВВ) импульсов с частотой повторения до 1 кГц. Изучение характеристик акустических волн, образующихся в сопле в результате работы актуатора, является важным, потому что они могут понижать шумность истекающего из сопла потока. В данной работе представлены алгоритм и программный комплекс, позволяющие анализировать звуковые волны, получающиеся в результате работы актуатора, включенного на выходе макета сопла, как с воздушным потоком, так и без него. Разработанный алгоритм позволяет выделять из записи экспериментальных данных (различных форматов) составные части — от одного конкретного импульса из всей выборки записи до группы из произвольного числа импульсов. А программная оболочка (GUI — graphical user interface), созданная в среде MatLab, позволяет построить устойчивые робастные спектральные и автоковариационные функций акустических сигналов.

2 Информация об экспериментальных данных

Измерения акустических волн проводились с помощью профессиональной звуковой аппаратуры таким образом, чтобы ударная волна не проявлялась в сигнале с микрофона. На расстоянии $\sim 1,5$ м под углом 30° к оси сопла был установлен микрофон, регистрирующий акустические характеристики системы. С микрофона записывался электрический сигнал и выводились спектральные характеристики сигнала. Вне зависимости от типа эксперимента (с воздушным потоком через сопло или без него) структура выходных данных имела вид, представленный на рис. 1. Всплески амплитуды акустического сигнала (рис. 1, *a*–*1*, *в*) связаны с плазменными разрядами, генерируемыми актуатором. На рис. 1, *б* представлен выделенный фрагмент из полной записи (см. рис. 1, *a*), а на рис. 1, *в*, изображен одиночный импульс, вычлененный из фрагмента записи (см. рис. 1, *б*).

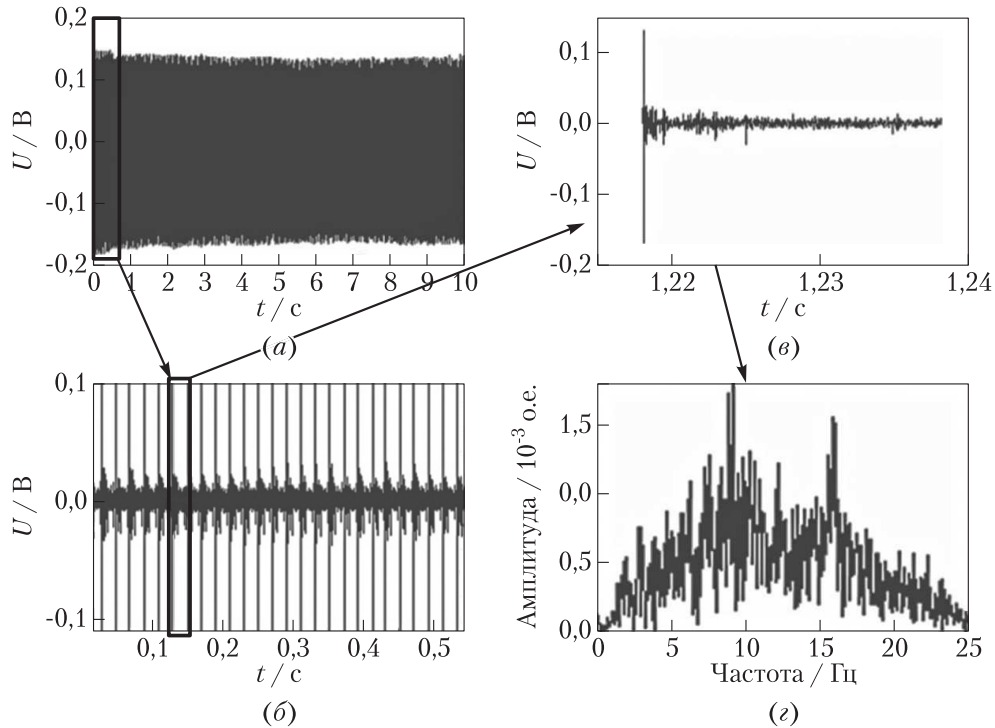


Рис. 1 Временная выборка амплитуды звукового сигнала при импульсной работе актуатора RING-2 с частотой 50 Гц: (а) полная выборка; (б) фрагмент выборки; (в) одиночный импульс; (г) спектр одиночного импульса

На рис. 1, г показан спектр единичного импульса, изображенного на рис. 1, в. Весь цикл последовательного разбиения полной выборки записи на фрагменты и их последующей обработки демонстрирует часть тех возможностей, которые предоставляет программный комплекс обработки экспериментальных данных.

3 Алгоритм обработки измерительной информации

Спектр сигнала, построенный по всей выборке (см. рис. 1, а), является суммой двух состояний и не дает информации о том, как ведут себя акустические волны, порождаемые актуатором, внутри сопла. Для получения представления о том, что происходит внутри сопла после одного или серии импульсов плазменного актуатора, было принято решение разбить всю временную область каждой выборки записанных данных на отдельные импульсы. Для этого был разработан специальный алгоритм и написана программа в среде программирования MatLab, позволяющая исследовать:

- зависимости сигнала или его фрагмента (одиночного импульса) от времени;
- спектры сигнала или его фрагмента, в том числе и робастных спектров для задаваемого количества фрагментов (одиночных импульсов) исходной выборки;
- автоковариационную функцию сигнала или его фрагмента (одиночного импульса);

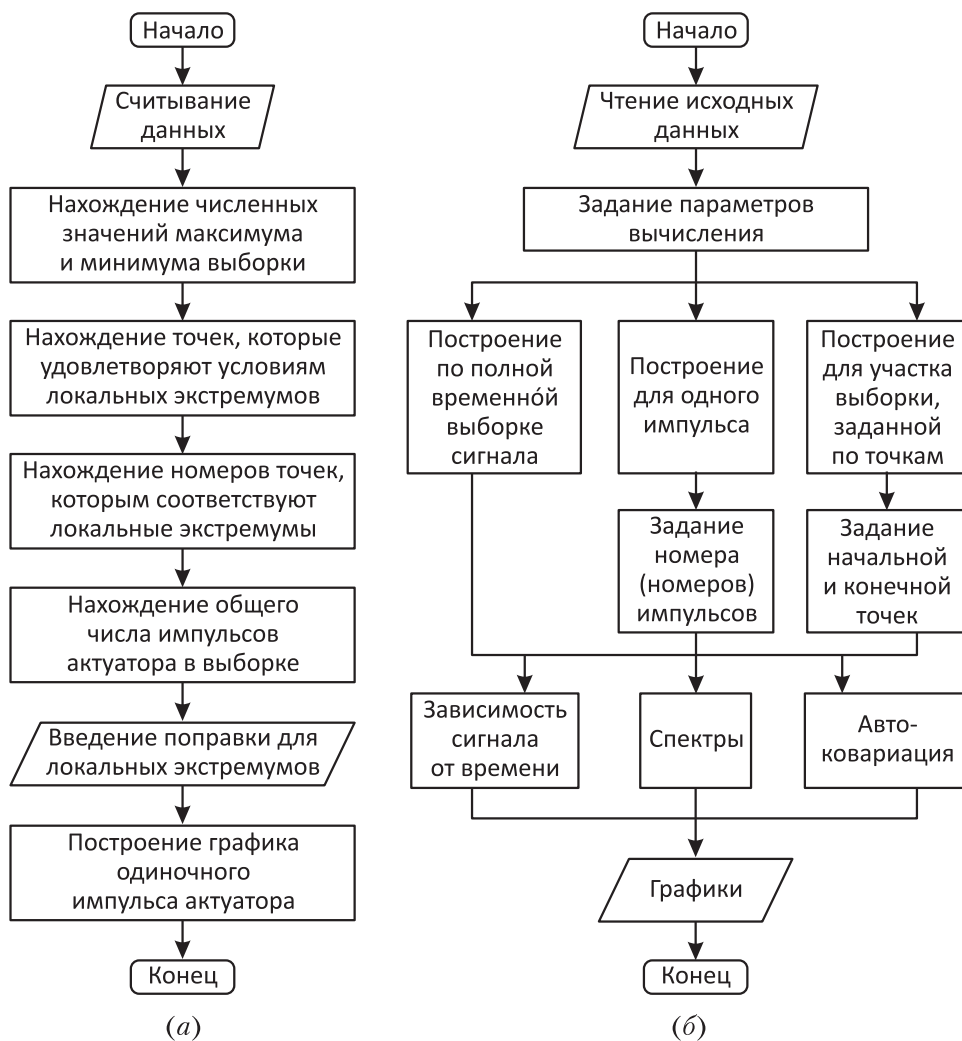


Рис. 2 Блок-схема алгоритма разбиения данных эксперимента на отдельные импульсы актуаторов (а) и порядка работы с комплексом обработки данных (б)

- зависимости от времени двух выбранных сигналов или их фрагментов (одиночных импульсов) на одних координатных осях;
- робастные спектры двух выбранных сигналов или их фрагментов (одиночных импульсов) на одних координатных осях;
- автоковариационные функции двух выбранных сигналов или их фрагментов (одиночных импульсов) на одних координатных осях.

Блок-схемы алгоритма и принципа работы разработанной программы изображены на рис. 2, а внешний вид программного комплекса на рис. 3.

Считывание данных и нахождение численных значений максимума и минимума выборки выполняются стандартными командами MatLab. После считывания данных необходимо вычислить максимум и минимум всей выборки, так как их значения будут служить ориентиром для нахождения локальных экстремумов одиночных импульсов. Наиболее сложной по реализации частью алгоритма является нахождение точек выборки, которым соответствуют локальные экстремумы, служащие признаками возникновения нового импульса. Принцип отбора таких точек проиллюстрирован на блок-схеме (рис. 4). Данная схема нахождения точек, удовлетворяющих условию локального экстремума, выбрана из тех

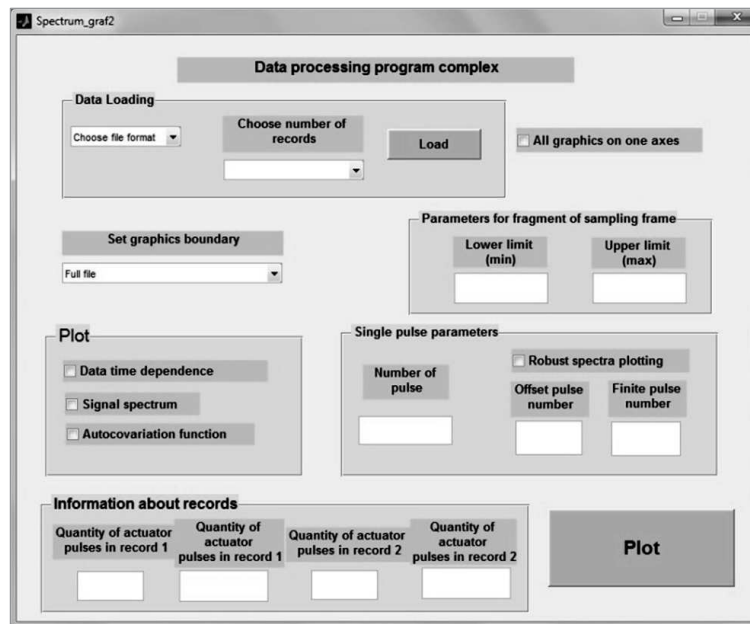


Рис. 3 Интерфейс пользователя программного комплекса обработки данных

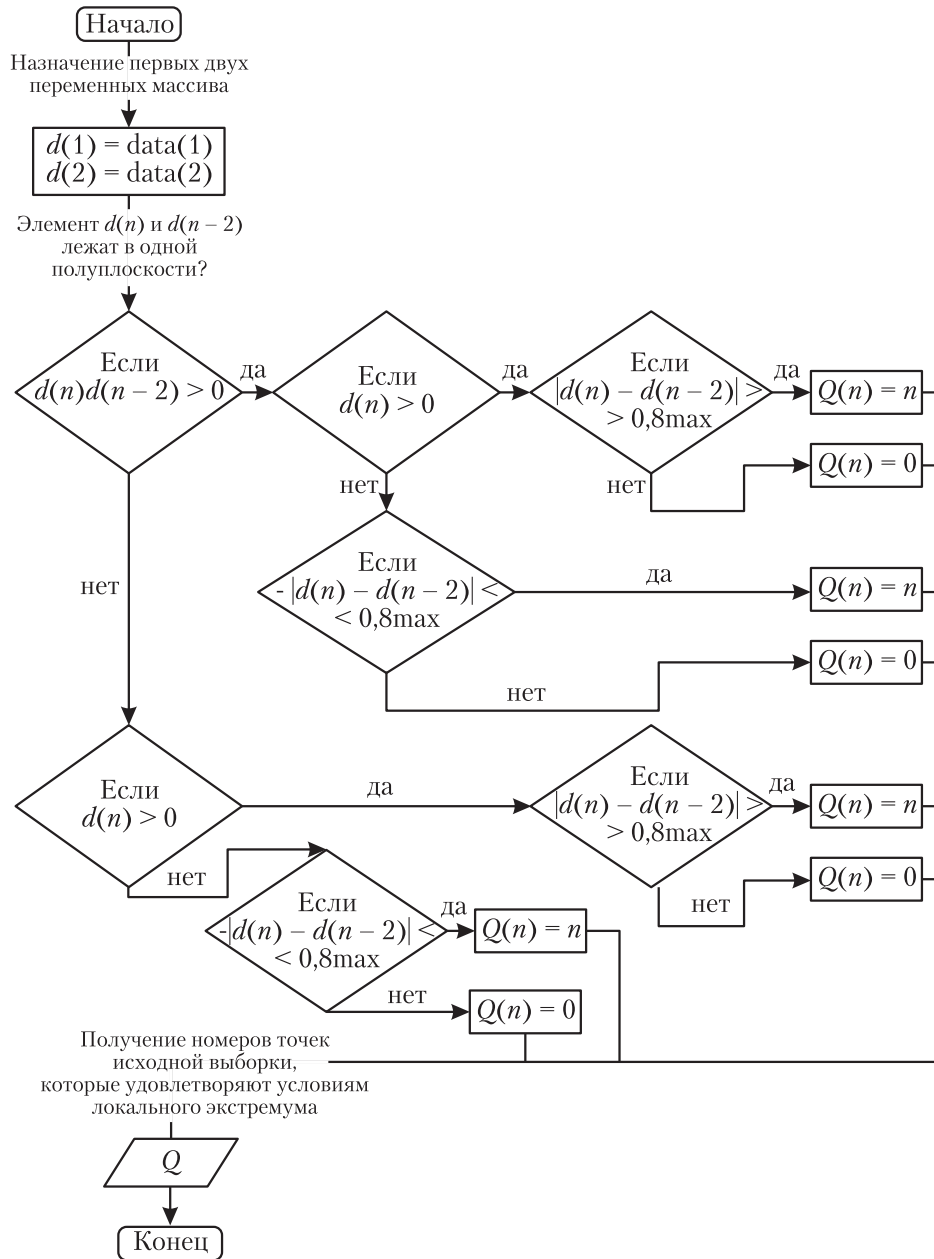


Рис. 4 Блок-схема нахождения номеров точек, которым соответствуют локальные максимумы

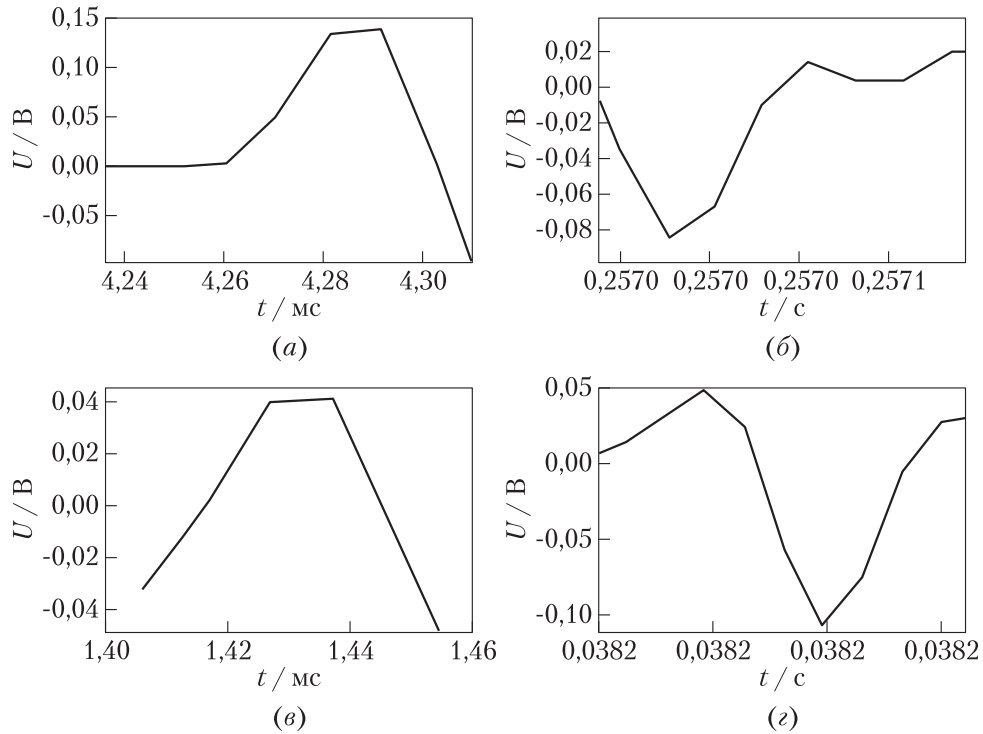


Рис. 5 Варианты возникновения локальных экстремумов во временной выборке данных

соображений, что возможно всего 4 варианта возникновения ярко выраженных экстремумов в записях сигнала:

- (1) точка экстремума и предшествующие ей несколько точек находятся в верхней полуплоскости (рис. 5, а);
- (2) точка экстремума и предшествующие ей несколько точек находятся в нижней полуплоскости (рис. 5, б);
- (3) точка экстремума находится в верхней полуплоскости, а предшествующие ей несколько точек находятся в нижней полуплоскости (рис. 5, в);
- (4) точка экстремума находится в нижней полуплоскости, а предшествующие ей несколько точек находятся в верхней полуплоскости (рис. 5, г).

Условия выявления локальных максимумов и минимумов таковы:

$$|d(n) - d(n - 2)| > 0,8 \max; \quad -|d(n) - d(n - 2)| > 0,8 \min, \quad (1)$$

где \max и \min — это максимальное и минимальное значения всей временной выборки сигнала соответственно, а $d(n)$ — текущее значение сигнала во временной

выборке. Эти условия были выбраны как оптимальные для всех проведенных измерений, в результате просмотра всех баз данных экспериментов и оптимизации решений нахождения локальных экстремумов одиночных импульсов.

Далее из массива данных Q , который состоит из нулей — номеров точек выборки, не удовлетворяющих рассмотренным выше условиям, — и значений номеров точек выборки, удовлетворяющих условиям, с помощью стандартных команд MatLab убираются нули. Из оставшихся значений Q отсеиваются все значения, отстоящие на 40 точек от локального экстремума. Это необходимо для того, чтобы убрать те значения локальных экстремумов, которые удовлетворяют условиям (1), но не являются ориентиром начала импульса, так как в зависимости от условий эксперимента импульсы актуатора могут иметь несколько достаточно больших экстремумов, а интерес представляет самый первый из них. Затем программа рассчитывает общее число импульсов, найденное в выборке. При построении графиков из значений экстремумов вычитается 5 точек, эта мера также выявлена в результате оптимизации баз данных экспериментов и необходима, так как зарождение импульса происходит до появления экстремума.

4 Выводы

Благодаря устойчивым характеристикам робастных спектров и автоковариационным функциям акустических сигналов, полученных с помощью разработанного программного комплекса, было установлено, что акустические волны, образующиеся в результате работы плазменного актуатора RING-2, встроенного в модельное сопло, проходят два этапа:

- (1) импульс актуатора с последующей реверберацией;
- (2) выход сигнала на квазистационарный уровень, сравнимый с уровнем шума.

Литература

1. *Moreau E.* Airflow control by non-thermal plasma actuators // *J. Phys. D: Appl. Phys.*, 2007. Vol. 40. No. 3. P. 605–608.
2. *Бархударов Э. М., Бережецкая Н. К., Косый И. А., Малахов Д. В., Тактакишвили М. И., Шакаров М. Р., Копьёв В. А.* Акустическая и ударно-волновые характеристики многоэлектродного кольцевого плазменного актуатора RING-1 // 2-я Всеросс. открытая конф. по авиационной акустике: Тез. докл. — М.: Центральный аэрогидродинамический институт им. проф. Н. Е. Жуковского (ЦАГИ), 2011. С. 147–148.

Поступила в редакцию 17.02.14

ALGORITHM AND PROGRAM COMPLEX FOR ANALYSIS OF THE CHARACTERISTICS OF ACOUSTIC WAVES OF PLASMA ACTUATOR FLUSH-MOUNTED IN THE MODEL PLANE NOZZLE

A. D. Chernousov^{1,2}, D. V. Malakhov¹, and N. N. Skvortsova^{1,3}

¹Prokhorov General Physics Institute, Russian Academy of Sciences, 38 Vavilov Str., Moscow 119991, Russian Federation

²Andreev Acoustic Institute, 4 Shvernik Str., Moscow 117036, Russian Federation

³National Research Nuclear University “MEPHI,” 31 Kashirskoye Shosse, Moscow 115409, Russian Federation

Abstract: Currently, the problem of developing new technologies of reducing the noise of aircraft engines, including directional impact on the noise on the basis of interaction between plasma disturbances and acoustic waves, stands acutely. One of the devices built on this principle is being developed in GPI RAS. These devices are plasma actuators (group of related to each other discharge gaps built on the perimeter of the nozzle) of various shapes and forms. The aim of this study was to perform a comparative analysis of spectra of acoustic waves produced as a result of work of a plasma actuator for air flowing out of the nozzle. The research was carried out for actuators of various designs and different shapes of nozzles. The following tasks have been solved: sorting and analysis of experimental data; development of the algorithm and creation of the program for studying spectral and correlation characteristics of acoustic signals; evaluation of spectral and correlation characteristics of acoustic signals in experiments with two types of actuators (with and without gas flow); comparative analysis of correlation and spectral characteristics of acoustic signals received at the outlet of the nozzle when an air flow interacts with plasma actuators of various configurations. With the help of the software program complex created in the MatLab environment, spectral and correlation characteristics of acoustic signals were investigated which allowed to construct the physical model of a plasma actuator in a plane nozzle. With the help of the developed research methodologies, the promising effect of using plasma actuators influence on air flow through the nozzle was proved.

Keywords: plasma actuator; acoustic waves; program complex; algorithm; nozzle; spectral characteristics; robust spectra

DOI: 10.14357/08696527140108

Acknowledgments

The research was supported by the President of Russian Federation Grant for Government Support of Young Russian Scientists No. MK-5607.2013.2.

References

1. *Moreau, E.* 2007. Airflow control by non-thermal plasma actuators. *J. Phys. D: Appl. Phys.* 40(3):605–608.
2. *Barhudarov, Je. M., N. K. Berezheckaja, I. A. Kossyj, D. V. Malahov, M. I. Tak-takishvili, M. R. Shakarov, and V. A. Kop'jov.* 2011. Akusticheskaya i udarno-volnovyye kharakteristiki mnogoelektrodnogo kol'tsevogo plazmennogo aktuatora RING-1 [Acoustic and shock-wave characteristics of multielectrode annular plasma actuator RING-1]. *2-ya Vserossiyskaya Otkrytaya Konferentsiya po Aviatsionnoy Akustike: Tezisy dokladov* [2nd All-Russian Public Conference on Aviation Acoustics: Abstracts]. Zvenigorod. 147–148.

Received February 17, 2014

Contributors

Chernousov Andrey D. (b. 1987) — PhD student, Prokhorov General Physics Institute, Russian Academy of Sciences, 38 Vavilov Str., Moscow 119991, Russian Federation; leading engineer, Andreev Acoustic Institute (JSC AAI), 4 Shvernika Str., Moscow 117036, Russian Federation; chado2008@yandex.ru

Malahov Dmitrii V. (b. 1985) — Candidate of Science (PhD) in physics and mathematics, scientist, Prokhorov General Physics Institute, Russian Academy of Sciences, 38 Vavilov Str., Moscow 119991, Russian Federation; 89199945791@mail.ru

Skvortsova Nina N. (b. 1956) — Doctor of Science in physics and mathematics, associate professor; leading scientist, Prokhorov General Physics Institute, Russian Academy of Sciences, 38 Vavilov Str., Moscow 119991, Russian Federation; professor, National Research Nuclear University “MEPHI,” 31 Kashirskoye Shosse, Moscow 115409, Russian Federation; mukudori@mail.ru

МЕТОД ФОРМИРОВАНИЯ СИСТЕМЫ ПОКАЗАТЕЛЕЙ ЖИВУЧЕСТИ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННЫХ СЕТЕЙ

А. А. Зацаринный¹, Н. Г. Буроменский², А. И. Гаранин³

Аннотация: Для объективного исследования свойств живучести информационно-телекоммуникационных сетей (ИТС) как сложных технических систем необходима количественная оценка этих свойств. При этом каждое свойство ИТС должно иметь количественную характеристику, которая определяется соответствующим показателем живучести. Для количественной характеристики полного перечня свойств ИТС необходимо сформировать систему показателей живучести. В статье предложены методические подходы к формированию системы показателей живучести ИТС, основанные на анализе условий их применения, оценке факторов, влияющих на живучесть системы, и определении свойств, которыми должна обладать система для выполнения требуемых функций. Для определения таких свойств рассматривается некоторая конфликтная ситуация, в которой две противоборствующие группировки решают задачи достижения предпочтительных для каждой из них состояний. Принято, что приоритетной задачей конфликтующей стороны является выведение из строя ИТС, функционирующей в составе критически важной информационной инфраструктуры в условиях преднамеренных поражающих действий. При этом никаких ограничений на применяемые средства поражения не накладывается. Выполнен анализ факторов, влияющих на живучесть системы, определены свойства живучести ИТС. Для определения количественных характеристик свойств живучести введено вероятностное пространство и на его основе предложены единичные и комплексные показатели живучести. Даны рекомендации по использованию единичных и комплексных показателей при проектировании, производстве и эксплуатации ИТС.

Ключевые слова: информационно-телекоммуникационные сети; условия эксплуатации; поражающие факторы; случайная величина; свойства; живучесть; показатели живучести

DOI: 10.14357/08696527140109

1 Введение

Задача разработки показателей живучести является составной частью методологии анализа и синтеза живучести сложных технических систем, в том числе и ИТС [1–15]. Это объясняется тем, что объективное исследование живучести

¹Институт проблем информатики Российской академии наук, azatsarinny@ipiran.ru

²Филиал ФГКУ «46 ЦНИИ» Минобороны России, г. Мытищи, buromienskii@mail.ru

³Институт проблем информатики Российской академии наук, algaranin@mail.ru

сложных технических систем невозможно без количественной оценки показателей живучести. Разнообразие ИТС, процессов их разрушения и восстановления делают проблему разработки методических основ формирования показателей живучести крайне востребованной.

В упомянутых выше публикациях предлагается большое число показателей живучести сложных систем. При этом в некоторых работах [6, 7, 16] дается достаточно содержательная классификация свойств и показателей живучести. Однако методики определения показателей в приведенных работах не в полной мере учитывают влияние на живучесть ИТС условий применения объекта исследования. Именно этим и объясняется большое число показателей живучести и отсутствие какой-либо взаимосвязи между ними. Рассмотрим кратко эволюцию понятия «живучесть» применительно к различным системам.

В большинстве публикаций [2, 3, 5] под живучестью сети понимаются различные свойства систем и сетей связи, характеризующие их устойчивость к отказам элементов.

В [7] под живучестью системы связи понимается «устойчивость системы связи к повреждению элементов стихийными факторами и преднамеренными воздействиями противника».

В [10, 11, 13–15] под живучестью (катастрофоустойчивостью) системы понимается «способность системы сохранять критически важные информационные и программные ресурсы и продолжать выполнение своих функций в условиях деградации архитектуры системы в результате стихийных бедствий, техногенных аварий и катастроф, целенаправленного воздействия людей».

В [1] «под живучестью системы понимается ее способность длительное время, с учетом вероятностей состояний системы, при которых она еще остается работоспособной, сохранять свои характеристики и обеспечивать выполнение своих функций при определенных методах и условиях ее эксплуатации».

В [6] под живучестью понимается «свойство системы связи сохранять и восстанавливать способность к выполнению основных функций в заданном объеме и в течение заданной наработки при изменении структуры системы и (или) алгоритмов и условий ее функционирования вследствие неблагоприятных воздействий».

И, наконец, в [12] дано определение живучести применительно к системе военной связи: «Способность системы военной связи обеспечивать управление войсками, силами и оружием в условиях воздействия обычного и ядерного оружия противника».

При анализе вышеприведенных определений обратим внимание на следующее. Во-первых, живучесть рассматривается применительно к сложным техническим системам [1, 6, 10, 13], системам и сетям связи [4, 5, 7, 8]. Во-вторых, живучесть рассматривается как внутреннее свойство системы, которым она обладает независимо от условий применения и проявляется только при воздействии неблагоприятных внешних факторов (НВ) и поражающих факторов (ПФ) оружия. И, наконец, в-третьих, живучесть определяется как свойство сложной системы сохранять (восстанавливать) способность к выполнению основных функций.

С учетом вышеизложенного живучесть ИТС определим как свойство, обеспечивающее ее устойчивость к неблагоприятным внешним воздействиям, факторам обычного и специального оружия, а также способность полного или минимально необходимого восстановления основных функций в заданных условиях применения.

В статье предлагаются методические подходы к определению показателей живучести ИТС.

2 Основные факторы, влияющие на устойчивость функционирования информационно-телекоммуникационных сетей

Повреждения и разрушения ИТС происходят в результате воздействия ПФ обычного, специального оружия и НВ. Под обычным оружием будем понимать средства поражения (СРП), применяющие различные виды боеприпасов в обычном снаряжении.

К специальному оружию относится ядерное оружие и так называемое оружие функционального воздействия на ИТС, осуществляемого с использованием мощных импульсов электромагнитных излучений (ИЭИ).

Ядерное оружие представляет собой наиболее разрушительный вид оружия с широким набором ПФ, основными из которых являются воздушная ударная волна, световое излучение, проникающая радиация и радиоактивное заражение местности [17–19].

Что касается оружия функционального воздействия, то ПФ являются интенсивность поля излучения на элементах ИТС и спектральный состав потока излучения. Импульсы электромагнитных излучений оказывают комплексное воздействие на ИТС за счет высокой проникающей способности через антенно-фидерные устройства. Это вызывает различные функциональные повреждения в радиоприемных, радиопередающих устройствах и в системе электроснабжения.

Формализуем условия применения (УП) ИТС. Для этого введем следующие обозначения:

$e_0 \in E_0$ — множество внешних воздействующих факторов (ВВФ);

$e_{\text{ПФ}} \in E_{\text{ПФ}}$ — множество ПФ.

Тогда $E = E_0 \cup E_{\text{ПФ}}$ — условия применения ИТС.

Далее в работе рассматриваются УП, связанные только с воздействием ПФ, т. е. $e_{\text{ПФ}}, e_{\text{ПФ}} \in E_{\text{ПФ}}$.

3 Основные свойства информационно-телекоммуникационных сетей

Наиболее общим свойством системы связи, отражающим влияние всех факторов, приводящих к отказам системы, является устойчивость.

Устойчивость — свойство системы связи, заключающееся в ее способности осуществлять своевременную передачу информации в необходимом объеме и с качеством, не хуже заданного при определенных условиях функционирования.

В зависимости от условий функционирования ИТС можно выделить следующие свойства:

- структура — свойство системы, обусловленное наличием совокупности элементов и устойчивой взаимосвязью между ними;
- стойкость элементов к ПФ — свойство элементов, обеспечивающее их стойкость к воздействию ПФ;
- восстанавливаемость — свойство системы, обусловленное возможностью устранения повреждений и восстановления ее работоспособности;
- управляемость — свойство системы, обеспечивающее сохранение требуемого уровня работоспособности за счет управления маршрутизацией обмена информацией между абонентами сети или ее реконфигурации;
- разведзащищенность — свойство системы, обеспечивающее невозможность идентификации ее элементов средствами разведки противника.

На рисунке приведена классификация (дерево) свойств, определяющих живучесть ИТС.

Живучесть ИТС	Структура
	Стойкость элементов к ПФ
	Восстанавливаемость
	Управляемость
	Разведзащищенность

Классификация (дерево) свойств, определяющих живучесть ИТС

Перечисленные выше свойства обуславливают соответствующие единичные показатели живучести ИТС.

Структурная живучесть — свойство системы сохранять выполнение своих функций хотя бы на минимально допустимом уровне в течение требуемого времени при пассивном противодействии повреждениям элементов системы.

Функциональная живучесть — свойство системы сохранять требуемый уровень работоспособности за счет принятого алгоритма управления маршрутизацией обмена информацией между абонентами системы или ее реконфигурации.

Техническая живучесть — свойство системы сохранять выполнение своих функций за счет восстановления работоспособности ее элементов.

Элементная живучесть — свойство элементов ИТС, характеризующее их стойкость к воздействию ПФ и способность восстанавливать работоспособность.

Разведзащищенность в данном случае также является одним из единичных свойств живучести ИТС.

Отметим, что набор свойств, представленных на рисунке, не является замкнутым, его можно расширить или сузить. Это определяется системой взглядов (принципов) на обеспечение живучести ИТС.

4 Пространство элементарных событий

Под пространством элементарных событий в данном случае понимается совокупность случайных событий, определяющих устойчивость функционирования ИТС в условиях применения.

Основой для построения пространства элементарных событий и системы показателей живучести ИТС является приведенное выше дерево свойств и их возможные варианты проявления в УП.

Предлагаемый подход базируется на стремлении совместить субъективные представления о живучести ИТС с достаточной степенью объективности ее оценки.

Для оценки живучести, а соответственно, и своевременности предоставления требуемых услуг пользователю предлагаются следующие стохастические параметры:

1. Случайная величина (СВ) $\xi_{\text{жив}}(e, t)$, характеризующая состояние работоспособности ИТС, которое отождествляется с живучестью, в качестве возможных элементарных событий принимает одно из трех значений:
 - (а) «ИТС находится в исправном состоянии», если в процессе информационного обмена, начавшегося в момент t и протекающего в условиях e , $e \in E$, ИТС обеспечивает пользователю возможность решения всех задач по управлению объектами группировки;
 - (б) «ИТС находится в неисправном, но работоспособном состоянии», если для информационного обмена, начавшегося в момент t в условиях e , $e \in E$, ИТС имеет повреждения и/или находится в состоянии восстановления при быстрой реакции системы технического обеспечения (СТО) ИТС и пользователю обеспечивается ограниченная (на уровне не ниже минимально допустимого) возможность решения задач по управлению объектами группировки;
 - (в) «ИТС находится в неработоспособном состоянии», если для информационного обмена, начавшегося в момент t в условиях e , $e \in E$, ИТС имеет повреждения, не позволяющие пользователю решать задачи по управлению объектами, даже на минимально допустимом уровне с учетом возможностей СТО ИТС по ее восстановлению.

Необходимо подчеркнуть, что если элементы ИТС имеют скрытые дефекты и в реальности система функционирует ненормально вследствие наличия скрытых дефектов (возможен, например, «сбой в работе оконечных устройств»), но пользователь не подозревает о сбоях в работе, все равно такое состояние ИТС будем относить к значению «ИТС находится в исправном состоянии».

2. Случайная величина $\xi_{\text{стр}}(e, t)$, характеризующая влияние структуры на живучесть (структурная живучесть) ИТС в УП e , $e \in E$, в течение времени t ,

$t \in T_{\Pi}$, при ограничениях на восстановление и живучесть элементов, принимает одно из трех значений:

- (а) «Структурная живучесть ИТС высокая», если в УП e , $e \in E$, в течение времени t , $t \in T_{\Pi}$, обеспечивается связность системы на уровне, достаточном для решения всех задач по управлению объектами группировки;
- (б) «Структурная живучесть ИТС допустимого уровня», если в УП e , $e \in E$, в течение времени t , $t \in T_{\Pi}$, обеспечивается связность на уровне, обеспечивающем решение минимально необходимого объема задач по управлению объектами группировки;
- (в) «Структурная живучесть ИТС низкая», если в УП e , $e \in E$, в течение времени t , $t \in T_{\Pi}$, связность системы не обеспечивает решение задач даже в минимально необходимом объеме.

3. Случайная величина $\xi_{гж}(e, t)$, характеризующая восстановление работоспособности ИТС (техническая живучесть), принимает одно из трех значений:

- (а) «Обеспечивается восстановление ИТС при быстрой реакции СТО ИТС», если время τ устранения повреждений, начавшегося в момент t , не превышает $\tau_{зад}$ и $\xi_{жив}(e, t) = \text{«ИТС находится в исправном состоянии»}$;
- (б) «Обеспечивается восстановление ИТС при медленной реакции СТО ИТС», если $t_{в} \leq t_{в доп}$ (где $t_{в доп}$ — допустимое время восстановления) и $\xi_{жив}(e, t) = \text{«ИТС находится в неисправном, но работоспособном состоянии»}$;
- (в) «Восстановление ИТС не обеспечивается», если $t_{в} > t_{в доп}$ при медленной реакции СТО ИТС и $\xi_{жив}(e, t) = \text{«ИТС находится в неработоспособном состоянии»}$.

При этом быстрая реакция СТО ИТС означает на практике устранение повреждений ИТС за установленное время. Восстановление при медленной реакции СТО ИТС означает устранение повреждений ИТС за минимально допустимое время, при превышении которого может возникнуть угроза невыполнения поставленной задачи. Методические вопросы построения СТО ИТС, обеспечивающие восстановление элементов системы за требуемое время, приведены в работе [20].

4. Для оценки функциональной живучести ИТС предлагается стохастический параметр $\xi_{фж}(e, t)$. Случайная величина $\xi_{фж}(e, t)$ характеризует степень управляемости ИТС и принимает одно из трех значений:

- (а) «Обеспечивается управление ИТС при быстрой реакции АСУ», если время t_y предоставления пользователю услуг в течение времени t , $t \in T_{\Pi}$, в УП e , $e \in E$, не превышает заданного, $t_y \leq t_{y зад}$, и $\xi_{жив}(e, t) = \text{«ИТС находится в исправном состоянии»}$;
- (б) «Обеспечивается управление ИТС при медленной реакции АСУ», если время t_y предоставления пользователю услуг не превышает заданного,

- $t_y \leq t_{y \text{ зад м}}$ (где $t_{y \text{ зад м}}$ и $t_{y \text{ зад б}}$ — задаваемые ограничения на время предоставления АСУ услуг — верхняя и нижняя границы), несмотря на то что $\xi_{\text{жив}}(e, t) = \langle \text{ИТС находится в неисправном состоянии} \rangle$;
- (в) «Управление ИТС не обеспечивается», $\xi_{\text{жив}} = \langle \text{ИТС находится в неисправном состоянии} \rangle$.
5. Случайная величина $\xi_{\text{эж}}(e, t)$ — характеризует живучесть элемента ИТС и принимает одно из следующих значений:
- (а) «Элементарная живучесть высокая», если за время $t, t \geq T_{\text{бд}}$, в условиях $e, e \in E$, обеспечивается стойкость к воздействию ПФ;
- (б) «Элементарная живучесть на допустимом уровне», если за время $t, t \geq T_{\text{п}}$, в условиях $e, e \in E$, стойкость элементов к ПФ обеспечивается на допустимом уровне;
- (в) «Элементарная живучесть не обеспечивается», если за время $t, t \geq T_{\text{п}}$, в условиях $e, e \in E$, стойкость элементов к ПФ не обеспечивается.
6. Случайная величина $\xi_{\text{рзв}}(e, t)$, характеризующая разведзащищенность ИТС, принимает одно из трех значений:
- (а) «Разведзащищенность элементов ИТС обеспечивается», если за время $t, t \geq T_{\text{бд}}$, в условиях $e, e \in E$, средствами разведки противника не могут быть идентифицированы элементы СВС;
- (б) «Разведзащищенность элементов ИТС на допустимом уровне», если за время $t, t \geq T_{\text{бд}}$, в УП $e, e \in E$, средствами разведки противника идентифицировано допустимое число элементов ИТС;
- (в) «Разведзащищенность элементов ИТС не обеспечивается», если за время $t, t \geq T_{\text{бд}}$, в УБП $e, e \in E$, средствами разведки противника идентифицировано число элементов ИТС больше допустимого.

При необходимости учета других факторов, влияющих на живучесть ИТС, например за счет безопасности связи, вероятностное пространство живучести ИТС может быть расширено и представлено в виде

$$\xi_{\text{бс}}(e, t) = \varphi(\xi_{\text{вир}}(e, t), \xi_{\text{умышл}}(e, t)),$$

где $\xi_{\text{бс}}(e, t)$, $\xi_{\text{вир}}(e, t)$ и $\xi_{\text{умышл}}(e, t)$ — случайные величины, характеризующие соответственно безопасность связи, наличие или отсутствие скрытых вирусных искажений, умышленные действия противника по искажению информации; φ — неизвестная функция.

Аналогичным образом пространство элементарных событий для введенных параметров может быть сужено. К примеру, для $\xi_{\text{тж}}(e, t)$ возможно оставить вместо трех событий два: «Обеспечивается восстановление ИТС» (объединяющее первые два) и «Восстановление ИТС не обеспечивается».

Все множество случайных величин, характеризующих живучесть ИТС, разделим на три подмножества, дающих возможность оценить в целом уровень ее живучести.

Живучесть ИТС предлагается характеризовать стохастическим параметром $\xi_{\text{жив}}(e, t)$:

- **«Живучесть ИТС высокая»**, если:
 - $\xi_{\text{стр}}(e, t) = \text{«Структурная живучесть ИТС высокая»}$;
 - $\xi_{\text{тж}}(e, t) = \text{«Обеспечивается восстановление ИТС при быстрой реакции СТО ИТС»}$;
 - $\xi_{\text{эж}}(e, t) = \text{«Живучесть элементов ИТС высокая»}$;
 - $\xi_{\text{фж}}(e, t) = \text{«Обеспечивается управление ИТС при быстрой реакции АСУ»}$;
 - $\xi_{\text{рзв}}(e, t) = \text{«Разведзащищенность высокая»}$;
- **«Живучесть ИТС допустимого уровня»**, если:
 - $\xi_{\text{стр}}(e, t) = \text{«Структурная живучесть ИТС допустимого уровня»}$;
 - $\xi_{\text{тж}}(e, t) = \text{«Обеспечивается восстановление работоспособности ИТС при медленной реакции СТО»}$;
 - $\xi_{\text{эж}}(e, t) = \text{«Живучесть элементов ИТС на допустимом уровне»}$;
 - $\xi_{\text{фж}}(e, t) = \text{«Обеспечивается управление ИТС при медленной реакции АСУ»}$;
 - $\xi_{\text{рзв}}(e, t) = \text{«Разведзащищенность на допустимом уровне»}$;
- **«Живучесть ИТС низкого уровня»**, если:
 - $\xi_{\text{стр}}(e, t) = \text{«Структурная живучесть ИТС низкая»}$;
 - $\xi_{\text{тж}}(e, t) = \text{«Восстановление ИТС не обеспечивается»}$;
 - $\xi_{\text{эж}}(e, t) = \text{«Живучесть элементов ИТС низкая»}$;
 - $\xi_{\text{фж}}(e, t) = \text{«Управление ИТС не обеспечивается»}$;
 - $\xi_{\text{рзв}}(e, t) = \text{«Разведзащищенность на низком уровне»}$;
- **«Живучесть ИТС сомнительного уровня»** при всех остальных сочетаниях параметров $\xi_{\text{стр}}(e, t)$, $\xi_{\text{тж}}(e, t)$, $\xi_{\text{эж}}(e, t)$, $\xi_{\text{фж}}(e, t)$ и $\xi_{\text{рзв}}(e, t)$.

5 Система показателей живучести информационно-телекоммуникационных систем

На основе разработанного выше вероятностного пространства предложим систему показателей живучести ИТС.

Пусть каждое из приведенных выше свойств в УП проявляется случайным образом. Такое предположение вполне естественно и не требует пояснений.

Тогда СВ $\xi_{\text{жив}}(e, t)$, характеризующую способность ИТС обеспечивать управление объектами группировки в УП, представим в виде множества, включающего следующие элементарные события:

$$\xi_{\text{жив}}(e, t) = \{\xi_{\text{стр}}(e, t), \xi_{\text{тж}}(e, t), \xi_{\text{эж}}(e, t), \xi_{\text{фж}}(e, t), \xi_{\text{рзв}}(e, t)\}, \quad (1)$$

где $\xi_{\text{стр}}(e, t)$ — СВ, характеризующая влияние структуры на живучесть ИТС в УП e , $e \in E$, в течение времени t , $t \in T_{\Pi}$; $\xi_{\text{тж}}(e, t)$ — СВ, характеризующая восстановление работоспособности ИТС; $\xi_{\text{эж}}(e, t)$ — СВ, характеризующая живучесть элемента ИТС; $\xi_{\text{фж}}(e, t)$ — СВ, характеризующая степень управляемости ИТС; $\xi_{\text{рзв}}(e, t)$ — СВ, характеризующая разведзащищенность ИТС.

На основе полученного вероятностного пространства предложим систему показателей живучести ИТС. Формирование системы показателей произведем с использованием следующих методических положений.

Введем вероятностное пространство (Q, S, P) , где Q — конечное пространство элементарных событий, введенное выше; S — класс всех подмножеств множества Q , удовлетворяющий свойствам σ -алгебры для сформированного пространства СВ; P — вероятностная мера на пространстве элементарных событий. При этом, поскольку $\Omega = (\omega_k)$ конечно, достаточно установить отображение $\omega_k \rightarrow p_k = P(\omega_k)$ такое, что $p_k \geq 0$ и $\sum_k p_k = 1$. Такое соответствие предлагается

на основе формализации процессов обеспечения живучести ИТС путем применения предельной теоремы для регенерирующих процессов и использования результатов анализа одно- и многолинейных систем массового обслуживания [4]. Выделим при этом систему единичных и комплексных показателей.

Под единичными показателями будем понимать вероятность свершения каждого элементарного события рассмотренного выше пространства.

$P_{\text{стр}}(e, t)$ — вероятность того, что в УП e , $e \in E$, в течение времени t , $t \in T_{\Pi}$, обеспечивается работоспособность ИТС на уровне, достаточном для решения требуемого объема задач по управлению объектами оборонного и народнохозяйственного назначения без учета восстановления (вероятность того, что $\xi_{\text{стр}}(e, t) = \text{«Структурная живучесть высокая»}$).

$\tilde{P}_{\text{стр.д}}(e, t)$ — вероятность того, что в УП e , $e \in E$, в течение времени t , $t \in T_{\Pi}$, обеспечивается работоспособность ИТС на уровне, позволяющем решать минимально необходимый (допустимый) объем задач по управлению объектами оборонного и народнохозяйственного назначения (вероятность того, что $\xi_{\text{стр}}(e, t) = \text{«Структурная живучесть ИТС на допустимом уровне»}$).

$\bar{P}_{\text{стр.н}}(e, t)$ — вероятность того, что работоспособность ИТС не обеспечивает решения задач по управлению объектами оборонного и народнохозяйственного назначения даже на минимально допустимом уровне в УП e , $e \in E$, в течение времени t , $t \in T_{\Pi}$ (вероятность того, что «Структурная живучесть ИТС низкая»).

$P_{\text{жф}}(e, t)$ — вероятность того, что время предоставления услуг должностному лицу для обеспечения управления объектами оборонного и народнохозяйственного назначения в УП e , $e \in E$, не превысит заданного t , $t \in T_{\Pi}$ (вероятность того, что $\xi_{\text{упр}}(e, t) = \text{«Обеспечивается управление ИТС при быстрой реакции АСУ»}$).

$\tilde{P}_{\text{уф}}(e, t)$ — вероятность того, что время предоставления должностному лицу услуг для решения задач управления объектами оборонного и народнохозяйствен-

ного назначения в УП $e, e \in E$, не превышает допустимого $t, t \in T_{\Pi}$ (вероятность того, что $\xi_{\text{фж}}^c(e, t) = \text{«Обеспечивается автоматизированное управление ИТС при медленной реакции АСУ»}$).

$\bar{P}_{\text{фж}}(e, t)$ — вероятность того, что время предоставления должностному лицу услуг не обеспечивает управление объектами оборонного и народнохозяйственного назначения в УП $e, e \in E$, в течение времени $t, t \in T_{\Pi}$ (вероятность того, что $\xi_{\text{упр}}(e, t) = \text{«Автоматизированное управление системой не обеспечивается»}$).

$P_{\text{тж}}(e, t)$ — вероятность того, что время восстановления заданного уровня работоспособности ИТС системой технического обеспечения ИТС за установленное время $t_{\text{в}}$ в УП $e, e \in E$, не превышает допустимого $t_{\text{доп}}$ (вероятность того, что $\xi(e, t) = \text{«Обеспечивается восстановление работоспособности ИТС при быстрой реакции СТО ИТС»}$).

$\tilde{P}_{\text{тж}}(e, t)$ — вероятность того, что время восстановления $t_{\text{в}}$ работоспособности ИТС системой технического обеспечения ИТС превышает допустимое, но меньше T_{Π} в УП $e, e \in E$ (вероятность того, что $\xi(e, t) = \text{«Обеспечивается восстановление работоспособности ИТС при медленной реакции СТО ИТС»}$).

$\bar{P}_{\text{тж}}(e, t)$ — вероятность того, что время восстановления работоспособности ИТС в УП $e, e \in E$, превышает время T_{Π} (вероятность того, что $\xi(e, t) = \text{«Восстановление ИТС не обеспечивается»}$).

$P_{\text{рзв}}(e, t)$ — вероятность того, что средствами обнаружения и разведки противника в УБП $e, e \in E$, в течение времени $t, t \in T_{\text{бд}}$, не будут идентифицированы элементы СВС (вероятность того, что $\xi(e, t) = \text{«Разведзащищенность СВС обеспечивается»}$).

$\tilde{P}_{\text{рзв}}(e, t)$ — вероятность того, что средствами обнаружения и разведки противника в УБП $e, e \in E$, в течение времени $t, t \in T_{\text{бд}}$, число вскрытых элементов СВС будет на допустимом уровне (вероятность того, что $\xi(e, t) = \text{«Разведзащищенность СВС на допустимом уровне»}$).

$\bar{P}_{\text{рзв}}(e, t)$ — вероятность того, что число вскрытых средствами обнаружения и разведки противника элементов СВС УБП $e, e \in E$, в течение времени $t, t \in T_{\text{бд}}$, превысит допустимое значение (вероятность того, что $\xi(e, t) = \text{«Разведзащищенность СВС не обеспечивается»}$).

Дифференциальное выражение показателей живучести ИТС представляет собой распределение вероятностей состояний ИТС и имеет вид:

$$\begin{aligned} & \{(P_{\text{стр}}(e, t), \tilde{P}_{\text{стр}}(e, t), \bar{P}_{\text{стр}}(e, t)); (P_{\text{фж}}(e, t), \tilde{P}_{\text{фж}}(e, t), \bar{P}_{\text{фж}}(e, t)); \\ & (P_{\text{тж}}(e, t), \tilde{P}_{\text{тж}}(e, t), \bar{P}_{\text{тж}}(e, t)); (P_{\text{эж}}(e, t), \tilde{P}_{\text{эж}}(e, t), \bar{P}_{\text{эж}}(e, t)); \\ & (P_{\text{рзв}}(e, t), \tilde{P}_{\text{рзв}}(e, t), \bar{P}_{\text{рзв}}(e, t))\}. \end{aligned}$$

Интегральная оценка показателей живучести представляет собой распределение мер живучести по обобщенным состояниям ИТС:

$$P_{\text{жив}}(e, t) = \{P_{\text{жив.в}}(e, t), P_{\text{жив.д}}(e, t), P_{\text{жив.н}}(e, t)\},$$

которая отражает распределение живучести ИТС при решении задач по управлению объектами группировки в УП e , $e \in E$, в течение времени t , $t \in T_{\Pi}$. Здесь:

$P_{\text{жив.в}}(e, t)$ — вероятность того, что ИТС способна обеспечить пользователю в любой момент t , $t \in T_{\Pi}$, все предусмотренные услуги по управлению объектами группировки в УП e , $e \in E$, в течение времени t , $t \in T_{\Pi}$ (вероятность того, что $\xi(e, t) = \text{«Живучесть ИТС высокая»}$);

$P_{\text{жив.д}}(e, t)$ — вероятность того, что система способна выполнять задачи по управлению объектами группировки в УП e , $e \in E$, в течение времени t , $t \in T_{\Pi}$, на минимально допустимом уровне (вероятность того, что $\xi(e, t) = \text{«Живучесть ИТС на допустимом уровне»}$);

$P_{\text{жив.н}}(e, t)$ — вероятность того, что система не обеспечивает выполнение задач по управлению объектами группировки в УП e , $e \in E$, в течение времени t , $t \in T_{\Pi}$ (вероятность того, что $\xi(e, t) = \text{«Живучесть ИТС низкая»}$).

Таким образом, построено пространство элементарных событий и предложена система показателей для оценки и анализа живучести ИТС.

Введение дифференциального и интегрального выражений для комплексной оценки живучести ИТС обусловлено наличием различных способов дальнейшего применения, разрабатываемых методов оценки живучести. Так, для более глубокого представления о различных свойствах ИТС, выбора менее дорогостоящих путей повышения живучести, а также выявления узких мест при применении ИТС целесообразно использовать дифференциальное выражение показателей живучести. В то же время, если предстоит сравнение нескольких ИТС или предлагаемых (реализованных) способов защиты, то зачастую только интегральные выражения показателей позволяют сделать окончательный выбор.

6 Заключение

В работе проведен анализ факторов, влияющих на живучесть системы. Проведена классификация основных свойств ИТС. Разработан метод формирования показателей живучести ИТС с учетом влияния условий применения ИТС.

Предложен методический подход к формированию системы показателей живучести ИТС, основанный на анализе условий их применения, оценке факторов, влияющих на живучесть системы, и определении свойств, которыми должна обладать система для выполнения требуемых функций.

Для определения количественных характеристик свойств живучести предложены единичные и комплексные показатели живучести. Даны рекомендации по использованию единичных и комплексных показателей при проектировании, производстве и эксплуатации ИТС.

Литература

1. *Кративин В. Ф.* О теории живучести сложных систем. — М.: Наука, 1978. 247 с.
2. *Коробов А. А., Шевчук Ю. Б.* Надежность сетей передачи данных // Вычислительные сети. — М.: Наука, 1981. С. 165–199.
3. *Нечепуренко М. И.* Модели структурного резервирования систем // Прикладные задачи на графах и сетях: Мат-лы Всесоюзн. совещания — Новосибирск, 1981. С. 57–86.
4. *Щербина Л. П.* Основы теории сетей военной связи. — Л.: ВАС, 1983. 65 с.
5. *Дудник Б. Я., Овчаренко В. Ф., Орлов В. К. и др.* Надежность и живучесть систем связи. — М.: Радио и связь, 1984. 215 с.
6. *Черкесов Г. Н.* Методы и модели оценки живучести сложных систем. — М.: Знание, 1987. 55 с.
7. *Попков В. К.* Математические модели живучести сетей связи. — Новосибирск: СО АН СССР, 1990. 235 с.
8. *Буроменский Н. Г.* О моделировании живучести систем связи военного назначения. — Мытищи: 16 ЦНИИИ МО РФ, 1996. Научно-технический сборник № 4. С. 8–13.
9. *Буроменский Н. Г.* Детерминированные характеристики живучести систем связи военного назначения. — Мытищи: 16 ЦНИИИ МО РФ, 1996. Научно-технический сборник № 4. С. 23–27.
10. *Будзко В. И., Беленков В. Г., Кейер П. А.* К выбору варианта построения катастрофоустойчивых информационно-телекоммуникационных систем // Системы и средства информатики, 2003. Вып. 13. С. 16–40.
11. *Будзко В. И., Соловьев А. Н., Киселев Э. В.* Вопросы защиты от угроз со стороны обслуживающего персонала в катастрофоустойчивых центрах обработки информации // Системы и средства информатики, 2003. Вып. 13. С. 41–63.
12. ГОСТ РВ 52216-2004 Связь военная. Термины и определения. 11 с.
13. *Кейер П. А., Будзко В. И., Беленков В. Г.* Катастрофоустойчивые решения в информационно-телекоммуникационных системах высокой доступности // Научно-емкие технологии, 2005. Т. 6. № 6. С. 38–47.
14. *Кейер П. А., Будзко В. И., Козлов А. Н.* Оптимизация построения и функционирования средств обеспечения катастрофоустойчивости // Системы высокой доступности, 2006. Т. 2. № 1. С. 30–45.
15. *Зацаринный А. А., Гаранин А. И., Козлов С. В.* Некоторые методические подходы к оценке надежности элементов информационно-телекоммуникационных сетей // Системы и средства информатики, 2011. Вып. 21. С. 21–33.
16. *Тараканов К. В.* Математика и вооруженная борьба. — М.: Воениздат, 1974. 240 с.
17. *Мырова Л. О., Попов В. Д., Верхотуров В. И. и др.* Анализ стойкости систем связи к воздействию излучений. — М.: Радио и связь, 1993. 186 с.
18. *Четиженко А. Э.* Радиоэлектронная аппаратура и ядерный взрыв. — М.: Воениздат, 1997. 210 с.
19. *Буроменский Н. Г.* Войска связи свою задачу выполнили // Москва — Чернобылю. — М.: Воениздат, 1998. Кн. 1. С. 290–304.

20. Зацаринный А. А., Буроменский Н. Г., Гаранин А. И. Методические вопросы формирования системы технического обеспечения информационно-телекоммуникационных сетей // Системы и средства информатики, 2013. Т. 23. № 2. С. 154–169.

Поступила в редакцию 14.02.14

METHOD OF FORMING THE SYSTEM OF INDICATORS OF INFORMATION AND TELECOMMUNICATION SYSTEMS VITALITY

A. A. Zatsarinnyy¹, N. G. Buromensky², and A. I. Garanin¹

¹Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

²Branch of the Federal State Government Institution “46 Central Research Institute of Defence Ministry of Russia,” Mytishchi 141006, Moscow Region, Russian Federation

Abstract: For objective investigation of the properties of survivability of information and telecommunication systems (ITS) as complex technical systems, a qualitative assessment of these properties is needed. Each property of ITS should have a quantitative characteristic, which is determined by the respective indicator of vitality. For quantitative characterization of the full list of properties of ITS, it is necessary to create a system of indicators of vitality. The article suggests methodical approaches used to form a system of indicators of survivability of ITS based on the analysis of the conditions of their application, evaluation of factors influencing survivability of the system, and definition of the properties that a system should have in order to perform the required functions. To determine these properties, the paper considers some conflict situation in which two opposing factions solve the task of achieving a preference for each of them. It was decided that the priority for each of the conflicting sides is the disabling of its functioning in the critical information infrastructure in conditions of intentional damaging effects. No limitation on the means used for defeat is imposed. The analysis of factors affecting vitality of the system defined the properties survivability ITS. To determine the quantitative characteristics of survivability properties, the paper introduces a probability area and on its basis single and comprehensive measures of survivability are offered. The paper provides recommendations for using single and complex indices to design, manufacture, and operate ITS.

Keywords: information and telecommunication systems; operating conditions; affecting factors; random variable; property; survivability; indicators of survivability

DOI: 10.14357/08696527140109

References

1. Krapivin, V. F. 1978. *O teorii zhivuchesti slozhnykh sistem* [The theory of complex systems survivability]. Moscow: Nauka [Science]. 247 p.
2. Korobov, A. A., and Yu. B. Shevchuk. 1981. Nadezhnost' setey peredachi dannykh [Reliability of data transmission networks]. *Vychislitel'nye Seti* [Computer Networks]. Moscow: Nauka [Science]. 199–165.
3. Nechepurenko, M. I. 1981. Modeli strukturnogo rezervirovaniya sistem [Model structural redundancy systems]. *Prikladnye zadachi na grafax i setyakh* [Applied problems on graphs and networks]. Novosibirsk. 57–86.
4. Shcherbina, L. P. 1983. *Osnovy teorii setey voennoy svyazi* [Fundamentals of the theory of military communication networks]. Leningrad: Military Academy of communications. 65 p.
5. Dudnik, B. Ya., V. F. Ovcharenko, V. K. Orlov, et al. 1984. *Nadezhnost' i zhivuchest' sistem svyazi* [Reliability and survivability systems]. Moscow: Radio i svyaz' [Radio and communications]. 215 p.
6. Cherkosov, G. N. 1987. *Metody' i modeli otsenki zhivuchesti slozhnykh sistem* [Methods and models for the assessment of complex systems survivability]. Moscow: Znanie [Knowledge]. 55 p.
7. Popkov, V. K. 1990. *Matematicheskie modeli zhivuchesti setey svyazi* [Mathematical models of communication systems survivability]. Novosibirsk: The Siberian Branch USSR Academy of Sciences. 235 p.
8. Buromenskii, N. G. 1996. O modelirovanii zhivuchesti sistem svyazi voennogo naznacheniya [About modeling of survivability of military-purpose communication systems]. Mytishchi: 16 CNIII MO RF [16 Central Research Institute of the Ministry of Defense of Russian Federation]. Scientific and technical collection. 4:8–13.
9. Buromenskii, N. G. 1996. Determinirovannye kharakteristiki zhivuchesti sistem svyazi voennogo naznacheniya [Deterministic characteristics survivability of military-purpose communication systems]. Mytishchi: 16 CNIII MO RF [16 Central Research Institute of the Ministry of Defense of Russian Federation]. Scientific and technical collection. 4:23–27.
10. Budzko, V. I., V. G. Belenkov, and P. A. Kajer. 2003. K vyboru varianta postroeniya katastrofoustoychivyykh informatsionno-telekommunikatsionnykh sistem [To the choice of variants of building the disaster-recovery information and telecommunication systems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 13:16–40.
11. Budzko, V. I., A. N. Solov'ev, and E. V. Kiselev. 2003. Voprosy zashchity ot ugroz so storony obsluzhivayushchego personala v katastrofoustoychivyykh tseentrakh obrabotki informatsii [Issues of protection against threats from the staff in resilient centers information]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 13:41–63.
12. GOST RV 52216-2004 Svyaz' voennaya. Terminy' i opredeleniya [Link military. Terms and definitions]. Moscow: Gosstandart of Russia. 11 p.
13. Kejer, P. A., V. I. Budzko, and V. G. Belenkov. 2005. Katastrofoustoychivye resheniya v informatsionno-telekommunikatsionnykh sistemakh vysokoy dostupnosti [Resilient solutions in information-telecommunication systems with high availability]. *Naukoemkie Texnologii* [Scientific Technologies] 6(6): 38–47.

14. Kejer, P. A., V. I. Budzko, and A. N. Kozlov. 2006. Optimizatsiya postroeniya i funktsionirovaniya sredstv obespecheniya katastrofoustoychivosti [Optimization of the construction and functioning of means for provision of disaster recovery]. *Sistemy Vysokoy Dostupnosti* [High Availability Systems] 2:30–45.
15. Zatsarinnyy, A. A., A. I. Garanin, and S. V. Kozlov. 2011. Nekotorye metodicheskie podkhody k otsenke nadezhnosti elementov informatsionno-telekommunikatsionnykh setey [Some methodical approaches to the estimation of the reliability of elements of information-telecommunication networks]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 21:21–33.
16. Tarakanov, K. V. 1974. *Matematika i vooruzhennaya bor'ba* [Mathematics and armed struggle]. Moscow: Voenizdat. 240 p.
17. Myrova, L. O., V. D. Popov, and V. I. Verkhoturov. 1993. *Analiz stoykosti sistem svyazi k vozdeystviyu izlucheniya* [The analysis of stability of systems due to the effects of radiation]. Moscow: Radio i svyaz' [Radio and Communication]. 186 p.
18. Chepizhenko, A. Z. 1997. *Radioelektronnaya apparatura y yadernyy vzryv* [Electronic equipment and nuclear explosion]. Moscow: Voenizdat. 210 p.
19. Buromenskii, N. G. 1998. Voyska svyazi svoyu zadachu vpolnili [Communication troops fulfilled their task]. *Moskva — Chernobylyu* [Moscow to Chernobyl]. I:290–304.
20. Zatsarinnyy, A. A., N. G. Buromenskii, and A. I. Garanin. 2013. Metodicheskie voprosy formirovaniya sistemy tekhnicheskogo obespecheniya informatsionno-telekommunikatsionnykh setey [The methodology of the development of the system of technical security of information and telecommunication networks]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics*. 23(2):154–169.

Received February 14, 2014

Contributors

Zatsarinnyy Alexander A. (b. 1951) — Doctor of Science in technology, professor, Deputy Director, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; azatsarinny@ipiran.ru

Buromenskii Nikolai G. (b. 1937) — Candidate of Science (PhD) in technology, leading scientist, Branch of the Federal State Government Institution “46 Central Research Institute of Defense Ministry of Russia,” Mytishchi 141006, Moscow Region, Russian Federation; buromienskii@mail.ru

Garanin Alexander I. (b. 1951) — Candidate of Science (PhD) in technology, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; algaranin@mail.ru

ФУНКЦИОНАЛЬНАЯ ГИБРИДНАЯ ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА ДЛЯ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПРИ ДИАГНОСТИКЕ АРТЕРИАЛЬНОЙ ГИПЕРТЕНЗИИ

И. А. Кириков¹, А. В. Колесников², С. Б. Румовская³

Аннотация: Рассматривается моделирование системы поддержки принятия решений сложной задачи диагностики артериальной гипертензии (СЗДАГ) в рамках проблемно-структурной (ПС) методологии функциональных гибридных интеллектуальных систем (ФГиИС). Представлены результаты разработки гетерогенного модельного поля СЗДАГ, спецификации областей релевантности автономных моделей подзадач и разработки алгоритмов синтеза интегрированного метода и модели решения сложной (гетерогенной) задачи. Интегрированный метод и модели проектируются в соответствии с рекомендациями по диагностике и лечению артериальной гипертензии (АГ) Всероссийского научного общества кардиологов. Результатом работы алгоритма синтеза метода решения гетерогенной задачи стала архитектура виртуального медицинского консилиума, диагностирующего АГ и позволяющего динамично синтезировать над неоднородным модельным полем новый метод выработки диагноза для каждого пациента индивидуально.

Ключевые слова: функциональная гибридная интеллектуальная система; гетерогенное модельное поле; интегрированный метод решения гетерогенной задачи; диагностика артериальной гипертензии; виртуальный медицинский консилиум

DOI: 10.14357/08696527140110

1 Введение

Артериальная гипертензия в Российской Федерации, как и во всех странах с развитой экономикой, — медико-социальная проблема с высоким риском осложнений, широкой распространенностью и недостаточным контролем артериального давления в масштабе человеческого общества. Анализ проблемной среды диагностики АГ на примере Калининградской клинической областной больницы (КОКБ) показал:

¹Калининградский филиал Института проблем информатики Российской академии наук, baltbipiran@mail.ru

²Балтийский федеральный университет им. И. Канта; Калининградский филиал Института проблем информатики Российской академии наук, avkolesnikov@yandex.ru

³Калининградский филиал Института проблем информатики Российской академии наук, sophiyabr@gmail.com

1. Диагностические ситуации и накопленные в теоретической и практической медицине знания об АГ гетерогенны, что проявляется в существенном разнообразии (кардиология, эндокринология, функциональная диагностика, офтальмология, терапия, неврология и др.), значительном объеме информации (около 800 параметров состояния больного), присутствии в рассуждениях НЕ-факторов А. С. Нариньяни (неполнота, неточность, недоопределенность анатомических, физиологических и патологических характеристик пациента) [1], количественных, качественных, графических данных (электрокардиограмм (ЭКГ), эхокардиограмм (ЭхоКГ), рентгеновских снимков и др.), которую врач с опытом работы до 5 лет и тем более начинающий врач не в состоянии изучить, оценить и использовать при постановке диагноза. Как следствие, сложная диагностическая задача релевантна модели «гетерогенная задача» [2] и требует коллективных усилий и междисциплинарных инструментов.
2. Отсутствует стандартизация в терминологии, формате и параметрических шкалах измерительного диагностического оборудования. Универсализм диагностического кодирования вступил в противоречие с актуальностью детализации номенклатуры признаков и симптомов, форматов для регистрации данных, а также организацией записей для каждого пациента индивидуально, заново, динамично. Поэтому не существует одного-единственного, применимого в любой ситуации метода диагностики. Он должен быть всякий раз предварительно синтезирован сообразно специфическим характеристикам пациента, а затем уже применен с учетом имеющейся аппаратуры, междисциплинарных знаний и опыта.
3. В широкой клинической практике не применяются мобильные комплексные диагностические системы, оснащенные современной измерительной аппаратурой и компьютерными системами поддержки принятия решений (КСППР), которые врач мог бы использовать индивидуально, сообразно своим теоретическим, профессиональным знаниям и опыту.
4. Несовершенны методы и средства медицинской информатики для консультаций врача, вырабатывающего диагноз АГ в условиях ограниченности ресурсов, недостатка времени, дефицита экспертов, неопределенности информации об окружающем мире и исследуемом пациенте.

Разработка КСППР позволит перейти к дифференциальной диагностике среди десятков, сотен и даже тысяч нозологических единиц, сократить число ошибок в ситуациях Р. Ригельмана [3]: (1) часто встречающихся заболеваний, но с атипичными симптомами; (2) наличием симптомов-миражей, увязываемых с патологией, но не имеющих к ней никакого отношения, и болезней-хамелеонов, маскирующихся под другую патологию; (3) при поиске зебр, комбинаций редких болезней.

Для создания КСППР предлагается использовать методы ФГиИС [2] — искусственных гетерогенных систем, моделирующих коллективные решения (кон-

силиумы), разнообразии хранимой и перерабатываемой информации в которых релевантно разнообразию информации в диагностических ситуациях. Такие системы, иными словами «виртуальные консилиумы», потенциально способны динамично синтезировать над неоднородным модельным полем новый метод выработки диагноза АГ для каждого пациента индивидуально и применять его. По мнению эксперта-кардиолога КОКБ, разработка подобных виртуальных медицинских консилиумов, в частности для диагностики АГ, и их эксплуатация актуальны в лечебно-профилактических учреждениях широкого профиля: фельдшерско-акушерских пунктах, амбулаториях, медико-санитарных частях, поликлиниках, — они повысят качество диагностических решений в подобных лечебных учреждениях, несмотря на возможное удорожание диагностики пациентов.

Настоящая работа посвящена результатам создания гетерогенного модельного поля СЗДАГ, спецификации областей релевантности автономных моделей подзадач СЗДАГ и разработки алгоритмов синтеза интегрированного метода и модели решения СЗДАГ, релевантной практике консилиумов в клинической медицине и в рамках ПС методологии ФГиИС.

2 Автономные модели подзадач и гетерогенное модельное поле

На этом этапе в соответствии с методологиями элементов ФГиИС [2] разрабатывается гетерогенное модельное поле MOD^a (возможно, и одноэлементное) для решения однородных задач из декомпозиции неоднородной СЗДАГ, с которой можно ознакомиться в [4]: задача «Распознавание и интерпретация электрокардиограммы» ($rgb_{РЭКГ}^h$, РЭКГ), задача «Анализ суточного мониторирования артериального давления» ($rgb_{СМАД}^h$, СМАД), группа из девяти технологических подзадач «Построение информативного набора признаков (симптомов) при диагностике заболеваний из областей 1–9» ($rgb_{ППС_{1-9}}^h$, ППС_{1–9}), группа из девяти функциональных подзадач «Диагностика критериев оценки сердечно-сосудистого риска и вторичной АГ у пациента» ($rgb_{ДАГ_{1-9}}^h$, ДАГ_{1–9}) и задача «Оценка степени и стадии артериальной гипертензии, степени риска сердечно-сосудистых заболеваний» ($rgb_{ССС}^h$, ССС). Области 1–9 — это области однородных параметров, полученные редукцией сложной задачи, отображенные в декомпозиции СЗДАГ и содержащие подзадачи диагностики: поражений органов мишеней (индекс 1), факторов риска (2), цереброваскулярных болезней (3), метаболического синдрома и сахарного диабета (4), заболеваний периферических артерий (5), ишемической болезни сердца (6), эндокринной АГ (7), паренхиматозной нефропатии (8) и реноваскулярной АГ (9) соответственно.

Кроме того, на этапе выбора методов решения однородных задач в [4] установлено наличие альтернативности (функциональной избыточности) инструментариев решения: подзадаче $rgb_{СМАД}^h$ было сопоставлено множе-

ство методов $\{met_{n1}^a, met_{e2}^a\}$, а подзадаче $prb_{ССС}^h$ — множество методов $\{met_{f1}^a, met_{e2}^a\}$ и для их решения необходимо разрабатывать по две модели — $\{mod_{nСМАД}^a, mod_{eСМАД}^a\}$ и $\{mod_{fССС}^a, mod_{eССС}^a\}$ соответственно. Остальным 19 подзадачам было сопоставлено одноэлементное множество методов решения: каждой подзадаче из группы $prb_{ППС1-9}^h$ — $\{met_g^a\}$, каждой подзадаче из группы $prb_{ДАГ1-9}^h$ — $\{met_f^a\}$ и подзадаче $prb_{РЭКГ}^h$ — $\{met_n^a\}$. Здесь n — искусственные нейронные сети (ИНС); e — экспертные системы; f — нечеткие системы; g — генетические алгоритмы (ГА).

Таким образом, СЗДАГ редуцирована на 21 подзадачу, для решения которых по результатам этапа выбора методов решения однородных задач необходимо разработать 23 модели, чтобы перейти к этапу синтеза метода решения СЗДАГ.

2.1 Автономные модели, решающие подзадачи РЭКГ и СМАД искусственными нейронными сетями

Модель ИНСЭКГ подзадачи РЭКГ обрабатывает ЭКГ-образы и сведена к задаче классификации, т. е. отнесения ЭКГ-образа к одному из семи классов: норма, гипертрофия левого желудочка (ГЛЖ) (патология 1), отклонение электрической оси сердца влево (патология 2), блокада левой ножки пучка Гиса (патология 3), рубцовое поражение миокарда — низкая сократимость, систолическая дисфункция (патология 4), признаки перегрузки левого предсердия — широкий зубец или двугорбый Р (патология 5), аритмия желудочковая и мерцательная (патология 6).

Электрокардиограмма (рис. 1) — это графическое отображение (регистрация) прохождения электрического импульса по проводящей системе сердца [5].

Прохождение импульса по проводящей системе сердца графически записывается по вертикали в виде пиков — подъемов и спадов кривой линии. Эти пики

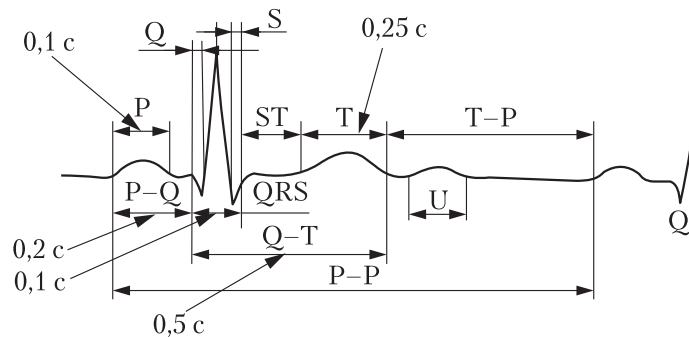


Рис. 1 PQRST-комплекс ЭКГ

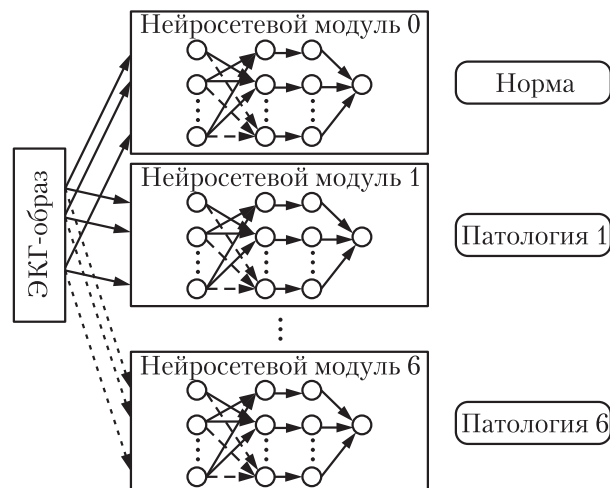


Рис. 2 Модульный вариант построения нейронной сети для распознавания патологий сердца по ЭКГ

принято называть зубцами ЭКГ и обозначать латинскими буквами Р, Q, R, S и Т. Помимо регистрации зубцов на ЭКГ по горизонтали записывается время, в течение которого импульс проходит по определенным отделам сердца. Отрезок на ЭКГ, измеренный по своей продолжительности во времени (в секундах), называют интервалом (на рис. 1 ширина зубца Р — 0,1 с, длина интервала Р–Q — 0,2 с, продолжительность QRS-комплекса — 0,1 с, ширина зубца Т — 0,25 с, длина интервала Q–Т — 0,5 с).

Результаты исследования мозга человека показывают, что он построен по функциональному принципу (зрение, моторные функции и др.) [3]. Аналогичный принцип положен и в основу построения ИНС для распознавания патологий сердца. На рис. 2 предлагается модульная структура нейросетевых блоков, отвечающих за распознавание только одной патологии.

Архитектуры модулей могут быть различны, а сами модули не связаны друг с другом. Сигнал образа ЭКГ поступает на все модули одновременно. Такая структура позволяет модернизировать сеть при необходимости повышения количества выявляемых патологий путем добавления новых модулей без переобучения всех модулей сети.

В качестве топологии модулей ИНСРЭКГ выбраны сети без обратных связей ввиду простоты реализации и гарантированной сходимости сети после обучения. Архитектура модулей — многослойные сети с прямым распространением сигнала, так как для них характерно последовательное выделение признаков из исходного образа, что способствует эффективному распознаванию. Алгоритм обучения нейросетей — «обратное распространение ошибки» — популярный

и реализованный в MATLABe алгоритм обучения сетей выбранной топологии. В качестве парадигмы обучения выбрано обучение с учителем [6], так как имеется достаточное количество входов и ассоциируемых с ними выходов.

В качестве функции активации выбрана логистическая функция $f^A(x)_P = 1/(1 + e^{-ax})$ [6], наиболее применяемая для многослойных сетей с прямым распространением сигнала в связи с ее гладкостью и непрерывностью, позволяющими обучать сеть методом обратного распространения ошибки. Кроме того, производная легко выражается через саму функцию, что ускоряет обучение. Из выражения для функции активации видно, что амплитуда выходных сигналов нейронов ИНС [6] — действительное число, лежащее в интервале $[0, 1]$.

Число нейронов входного слоя ИНСРЭКГ — $0,85 \cdot 150 = 128$ — обусловлено временным окном анализа электрокардиосигнала (ЭКС) (0,85 с) и частотой дискретизации ЭКГ, типичной для современных мониторов (150 Гц). Число нейронов в скрытом слое каждого из модулей изменяется от 1 до 50 с шагом 10 (при таком шаге существенно не снижается информативность исследования, а затраты вычислительных ресурсов остаются приемлемыми) и выбирается на основе значения ошибки обучения каждого модуля. На вход каждой сети передается один и тот же образ ЭКС, т. е. образ ЭКС, соответствующий той или иной патологии.

Для обучения с учителем и работы ИНСРЭКГ на вход подаются обучающая и тестирующая последовательности, состоящие из примеров $(X_1, \dots, X_O; Y_0, Y_1, \dots, Y_6)$, где X_1, \dots, X_O — это вектор образов ЭКС, состоящий из чисел интервала $[0, 1]$, а $O = 128$ входов ИНС; Y_0 — выход, соответствующий норме, Y_1, \dots, Y_6 — выходы, соответствующие шести диагностируемым на ЭКГ патологиям: Y_0, Y_1, \dots, Y_6 принимают значение 1, если норма или соответствующая патология имеют место, и 0 в противном случае.

От алгоритма подготовки таких образов зависит качество распознавания ЭКС. Технология предварительной обработки сигнала аналогична [7]. Подготовка образов ЭКГ-записей организована следующим образом: вначале осуществляется цикл перебора R-зубцов, как наиболее выраженных [5], доминирующих на большинстве записей; затем по зубцам выделяются кардиоциклы (см. рис. 2), затем осуществляется передискретизация ЭКГ до частоты 150 Гц, определение минимумов и максимумов кардиокомплекса и линейная нормализация до диапазона нейронов $[0; 1]$. В итоге сформирована база данных (БД), состоящая из 400 образов ЭКГ-записей.

Для подготовки БД тестовых и обучающих образов вначале из общего массива образов ЭКГ выбираются обучающие и тестовые образы для каждой из шести патологий по отдельности, а также для нормы ЭКГ пациента, затем обучающие образы объединяются в БД обучающих эталонов, а тестовые образы объединяются в БД эталонов. Для обучающих и тестовых БД подготовлено по 200 примеров. Для создания БД использован архив PhysioBank — структурированный массив цифровых записей ЭКГ Physikalish-Technische Bundesanstalt [8].

При реализации ИНС один из важных этапов процедур — нормализация, т. е. предварительная обработка данных обучающих и тестовых последовательностей, при которой значения признаков, образующих вектор, приводятся к некоторому заданному диапазону. Для предварительной обработки данных и приведения их к диапазону $[0, 1]$ выбрана линейная нормализация, не искажающая форму сигнала и не добавляющая в него новых признаков в отличие от экспоненциальной нормализации. Результат работы ИНС оценивается с мерами точности (ошибками обучения), вычисляемыми алгоритмом обратного распространения по формуле [6]:

$$\delta_i = y_i^3 - d_i, \quad (1)$$

где y_i^3 — полученное значение выхода i -го нейрона выходного третьего слоя; d_i — требуемое значение i -го выхода, $i = \overline{0, 6}$.

Модель ИНССМАД подзадачи СМАД решает задачу прогнозирования и контроля данных суточного мониторинга артериального давления пациента и состоит из двух нейросетей.

Обучающие и тестовые выборки сформированы на основе архива данных суточного мониторинга артериального давления КОКБ.

В качестве топологии модулей ИНССМАД выбраны сети без обратных связей. Архитектура нейросетей — это многослойные сети с прямым распространением сигнала с обратным распространением ошибки и логистической функцией активации. Число нейронов входного слоя обеих сетей ИНССМАД — два (пол и возраст пациента). В выходном слое — 24 нейрона, что обусловлено тем, что замеры систолического и диастолического давления у пациента производятся каждые полчаса, а затем данные за час усредняются и передаются на вход соответствующей нейросети. Исходное число нейронов в скрытом слое меняется в процессе экспериментов с моделью от 1 до 10 и подбирается по значению ошибки обучения (1). На основании данных о поле и возрасте больного в сети рассчитываются нормальные показатели суточного мониторинга артериального давления, и они сравниваются с показателями пациента.

Для обучения с учителем и работы ИНССМАД на вход подаются обучающая и тестирующая последовательности, состоящие из примеров $(X_1, X_2; Y_{11}, \dots, Y_{124}; Y_{21}, \dots, Y_{224})$, где X_1, X_2 — это вектор входов ИНС, содержащий данные пола (0 — мужской, 1 — женский) и возраста (нормализуются до значений в интервале $[0; 1]$); $Y_{11}, \dots, Y_{124}; Y_{21}, \dots, Y_{224}$ — выходы 1-й и 2-й нейросетей соответственно с данными по систолическому и диастолическому давлению соответственно (нормализуются до значений в интервале $[0; 1]$). Результат работы ИНС оценивается с помощью ошибки обучения ИНС по формуле (1) при $i = \overline{1, 24}$. Предварительная обработка данных обучающих и тестовых последовательностей ведется линейной нормализацией, при которой значения признаков, образующих вектор, приводятся к диапазону $[0, 1]$.

ИНССМАД и ИНСРЭКГ реализуются средствами Neural Network Toolbox пакета MATLAB поэтапно: (1) формируется последовательность входов и

выходов; (2) создается новая нейросеть (выбирается тип нейросети, задаются диапазоны входа, функция обучения, функция ошибки обучения, число слоев, число нейронов в слое и функция активации слоев); (3) инициализация сети (задание начальных весов и смещений); (4) моделирование; (5) обучение и адаптация. Выполнены четыре этапа и запланированы обучение и адаптация нейросетей, решающих подзадачи СМАД и РЭКГ.

Концептуальные модели i -го модуля ИНС подзадач РЭКГ и СМАД обобщены в схему, построенную на базе (3.24) [2] и расширенную свойствами (тип нормализации, архитектура ИНС и модель нейронов ИНС) и ресурсами (входы и выходы ИНС, обучающая и тестирующая последовательности):

$$\begin{aligned} \text{mod}_{nZ}^a = & R^{\text{met prb}} \left(\text{met}_{1Z}^h, \text{prb}_{nZ}^h \right) \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_{\text{kon}}^k \right) \circ \\ & \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_{\text{bbox}}^{\text{sh}} \right) \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_m^{\text{mlang}} \right) \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_{\text{eq}}^{\text{mlang}} \right) \circ \\ & \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_{\text{direct}}^{\text{solpr}} \right) \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_{\text{invdist}}^{\text{learnalg}} \right) \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_Z^{\text{arch}} \right) \circ \\ & \circ R^{\text{met res}} \left(\text{met}_{1Z}^h, \text{res}_Z^{\text{in}} \right) \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{EV}_{nZ} \right) \circ R^{\text{met res}} \left(\text{met}_{1Z}^h, \text{res}_Z^{\text{out}} \right) \circ \\ & \circ R^{\text{met res}} \left(\text{met}_{1Z}^h, \text{res}_Z^{\text{ed test}} \right) \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_{\text{lin}}^{\text{norm}} \right) \circ \\ & \circ R^{\text{met mod}} \left(\text{met}_{1Z}^h, N \right) | Z \in \{ \text{СМАД, РЭКГ} \} , \quad (2) \end{aligned}$$

где met_{1Z}^h — метод; pr_{kon}^k — классификатор метода: «коннекционистский»; $\text{pr}_{\text{bbox}}^{\text{sh}}$ — модель-схема метода: «черный ящик»; $\text{pr}_m^{\text{mlang}}$ — язык описания модели: «матрицы»; $\text{pr}_{\text{eq}}^{\text{mlang}}$ — язык описания модели: «уравнения»; $\text{pr}_{\text{direct}}^{\text{solpr}}$ — процедура получения решения: «прямое распространение в нейросетях»; $\text{pr}_{\text{invdist}}^{\text{learnalg}}$ — алгоритм обучения нейросетей: «обратное распространение»; prb_{nZ}^h — схемы ролевых концептуальных моделей подзадач СЗДАГ, решаемых нейросетями, которые определены схемой (2) в [4]; $\text{res}_Z^{\text{arch}}$ — архитектура ИНС; res_Z^{in} и $\text{res}_Z^{\text{out}}$ — входы и выходы ИНС (соответствуют $X_{\text{РЭКГ}}$ и $Y_{\text{РЭКГ}}$); $\text{res}_Z^{\text{ed test}}$ — обучающая и тестирующая последовательности; N — модель нейрона (см. (3) ниже); $\text{pr}_{\text{lin}}^{\text{norm}}$ — тип нормализации: «линейная»; EV_{nZ} — спецификатор, включающий знания о средней ошибке обучения (5.5.37)–(5.5.38) из [9] и знания о гибридных возможностях (табл. 7.2 и 8.6 из [9]).

Модель нейронов ИНСЭКГ и ИНССМАД может быть представлена как

$$N_P = \left\langle X_P^{\text{H}}, y_P, \text{Sw}_P, f^{\Sigma}(x)_P, f^A(x)_P \right\rangle | P \in \{ \text{РЭКГ, СМАД} \} , \quad (3)$$

где $X_P^{\text{H}} \subseteq X_P$, X_P^{H} — множество входов нейрона, $X_{\text{РЭКГ}}^n \subset X_{\text{РЭКГ}} = \{ x_{i\text{РЭКГ}}^{\text{in}} \} | i = \overline{1, 128}$ — множество нейронов входного слоя каждого мо-

дуля, $X_{\text{СМАД}}^{\text{in}} \subset X_{\text{СМАД}} = \{x_i^{\text{in}}\}_{i=2}$ — множество нейронов входного слоя каждой нейросети в ИНССМАД, $y_P \in Y_P$ — выходной сигнал, $Y_{\text{РЭКГ}}^{\text{out}} \subset Y_{\text{РЭКГ}} = \{y_j^{\text{out}}\}_{j=1}$ — множество нейронов выходного слоя каждого модуля, $Y_{\text{САД}}^{\text{out}} \subset Y_{\text{СМАД}} = \{y_j^{\text{out}}\}_{j=24}$ — множество нейронов выходного слоя каждого модуля, $f^{\Sigma}(x)_P = \sum_i w_{ij}x_{ij}$ — функция суммирования, $f^A(x)_P$ — функция активации, Sw_P — матрица-столбец синаптических весов нейрона (до обучения имеет произвольные ненулевые значения элементов).

2.2 Автономные модели, решающие девять подзадач из группы ППС_{1–9} посредством генетических алгоритмов

Модели ГАППС_{1–9} подзадач «Построение информативного набора признаков (симптомов) при диагностике заболеваний из областей 1–9» решают задачи безусловной оптимизации на множестве признаков, по которым диагностируются заболевания из областей 1–9 декомпозиции СЗДАГ (см. выше), и исключают из рассмотрения избыточные и взаимозависимые признаки заболеваний, что минимизирует общее число входных переменных (признаков, симптомов) на входе нечетких систем, решающих задачи классификации патологического состояния здоровья пациента НСДАГ₁, ..., НСДАГ₉. Это уменьшает объем информации о показателях состояния здоровья пациента, вводимых врачом в систему, что снижает время опроса, осмотра и обследования пациента врачом.

Каждый классический ГА [9] из ГАППС_{1–9} случайным образом генерирует популяцию в 20 особей. Размер индивидуума равен общему числу учитываемых признаков в соответствующей подзадаче из группы ДАГ_{1–9} для подзадач ППС_{1–9} и равна 50, 20, 80, 20, 70, 85, 200, 215, 30 генам соответственно, причем значение гена, равное 1, соответствует симптому (признаку), принадлежащему множеству «Учитываемые признаки», а 0 — признаку, не включенному в это множество.

Далее каждый из $N_{\text{КВ}}$ индивидуумов популяции накладывается как маска на соответствующую номеру ГА нечеткую систему, учитывающую только те симптомы, которые принимают значения 1 у индивидуума. Затем вычисляются $N_{\text{КВ}}$ среднеквадратичных ошибок постановки диагноза и функции приспособленности индивидуумов как минимум f_i среднеквадратичной ошибки классификации соответствующей нечеткой системы под номером Z на тестовой выборке признаков (симптомов) состояния здоровья пациента (составляет 15 наблюдений):

$$f_i = \frac{1}{N_k} \sqrt{\sum_{l=1}^{N_k} \left(\frac{1}{N} \sum_{j=1}^N \delta_{ij} \right)^2},$$

где $i = \overline{1, N_{\text{KB}}}$ — номер индивидуума в популяции, j — число наблюдений в тестовой выборке ($N = 15$); f_i — среднеквадратическая ошибка классификации k -й НС на тестовой выборке признаков (симптомов) состояния здоровья пациента (составляет 15 наблюдений); δ_{lj} — ошибка классификации по l -му выходу k -й НС (равна 1, если совпадает с эталоном и 0 — если нет), N_k — число выходов k -й НС.

Отбор родителей идет методом турнирной селекции (размер турнира — 2): выбираются лучшие особи, из которых порождаются новые особи-потомки с помощью операторов мутации (вероятность 0,01) и скрещивания (вероятность 0,5) и порождается новая популяция.

Имитация эволюции популяции идет в 100 итераций, затем выбирается наилучшая особь, оптимальный набор признаков (симптомов), — это и есть маскирующая информация, аргумент функции модификации соответствующей нейронной сети (по умолчанию содержит весь набор признаков, отобранный экспертами).

Концептуальные модели для ГАППС_{1–9} построены на базе (3.28) [2], обобщены в (4) и расширены свойствами (класс ГА, база знаний, число итераций алгоритма, функция оценки приспособленности индивидуумов, метод селекции):

$$\begin{aligned} \text{mod}_{gZ}^a = & R^{\text{met prb}} \left(\text{met}_{1Z}^h, \text{prb}_{gZ}^h \right) \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_{\text{ev}}^k \right) \circ \\ & \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_{\text{ev}}^{\text{sh}} \right) \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_{\text{kl}}^{\text{kl}} \right) \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_{\text{al}}^{\text{mlang}} \right) \circ \\ & \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_{\text{mexp}}^{\text{solpr}} \right) \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_Z^{\text{KB}} \right) \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, N \right) \circ \\ & \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, f_z \right) \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{EV}_{gZ} \right) \circ \\ & \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_{\text{rul}}^{\text{sel}} \right) \mid Z \in \{ \text{ППС}_i \mid i = \overline{1, 9} \}, \quad (4) \end{aligned}$$

где met_{1Z}^h — метод; pr_{ev}^k — классификатор метода: «эволюционный»; $\text{pr}_{\text{ev}}^{\text{sh}}$ — модель-схема метода: «эволюция»; $\text{pr}_{\text{kl}}^{\text{kl}}$ — классический генетический алгоритм; $\text{pr}_{\text{al}}^{\text{mlang}}$ — язык описания модели: «алгоритм»; $\text{pr}_{\text{mexp}}^{\text{solpr}}$ — процедура получения решения: «машинные эксперименты с моделирующим (генетическим) алгоритмом»; prb_{gZ}^h — схемы ролевых концептуальных моделей подзадач СЗДАГ, решаемых ГА, которые определены схемой (2) в [4]; pr_Z^{KB} — схема базы знаний $\text{KB} = \{ I_j \mid j = \overline{1, N_{\text{KB}}} \}$, $N_{\text{KB}} = 20$; $I_j = g_1 \circ \dots \circ g_q \circ \dots \circ g_{N_I}$ — микроуровневое представление индивидуума, а g_q — ген, принимающий значения 1 или 0; N_I — число генов в решении; N — число итераций; $\text{pr}_{\text{rul}}^{\text{sel}}$ — метод селекции (метод рулетки); f_z — функция приспособленности; EV_{gZ} — знания о гибридных возможностях (табл. 7.2 и 8.6 из [9]).

2.3 Автономные модели, решающие девять подзадач из группы ДАГ_{1–9} и подзадачу ССС посредством методологии нечетких систем

Модели нечетких систем НСДАГ₁, ..., НСДАГ₉ подзадач: диагностика поражений органов-мишеней (НСДАГ₁), диагностика факторов риска АГ (НСДАГ₂), диагностика цереброваскулярных болезней (НСДАГ₃), диагностика метаболического синдрома и сахарного диабета (НСДАГ₄), диагностика заболеваний периферических артерий (НСДАГ₅), диагностика ишемической болезни сердца (НСДАГ₆), диагностика эндокринной АГ (НСДАГ₇), диагностика паренхиматозной нефропатии (НСДАГ₈) и диагностика реноваскулярной АГ (НСДАГ₉), объединенных для описания в группу «Диагностика критериев оценки сердечно-сосудистого риска и вторичной АГ у пациента» (ДАГ_{1–9}) и решающих задачи классификации патологического состояния здоровья пациента.

Модель нечеткой системы НСССС подзадачи «Оценка степени и стадии артериальной гипертензии, степени риска сердечно-сосудистых заболеваний» классифицирует степень АГ по данным систолического и диастолического давления и/или данным от ИНССМАД, стадию эссенциальной АГ (или гипертонической болезни) в случае отсутствия вторичной АГ (диагностируется НСДАГ_{7–9}) по данным от НСДАГ_{1–3} и НСДАГ₅, НСДАГ₆ и классифицирует общий риск сердечно-сосудистых заболеваний у больного по данным от НСДАГ_{1–6} и формирует диагноз.

Для имитации работы эксперта над подзадачей для каждой из НС (НСДАГ₁, ..., НСДАГ₉, НСССС) выбран алгоритм нечетких рассуждений Сугено, так как посылки правил базы знаний заданы нечеткими множествами, а заключения — четкими числами: 1 (истина) или 0 (ложь). Определено пространство выходов $Y = Y_1 \times \dots \times Y_{N_Y}$ — четких констант, принимающих значение 0 или 1, размерность которого (число классов) $N_Y = 3, 7, 3, 2, 3, 3, 7, 3, 3, 15$ соответственно, и пространство входов $X = X_1 \times \dots \times X_{N_X}$ — нечетких лингвистических переменных (например, «загрудинные боли», «боли в животе», «вариабельные боли» с термами «интенсивные», «умеренные», «слабые» или «отсутствуют», функции принадлежности которых отражены на рис. 3, а, или «частота дыхательных движений» с термами «редкая», «нормальная», «повышенная», «удушье», функции принадлежности которых показаны на рис. 3, б). В нечетких системах используется два типа функций (см. рис. 3) — сигмоидная $\mu(u) = 1/(1 + \exp(-a(u - c)))$, где a — коэффициент крутизны, c — координата перехода через 0,5, и треугольная

$$\mu(u) = \begin{cases} 0, & \text{если } u \leq a \text{ или } u \geq c; \\ \frac{u - a}{b - a}, & \text{если } a < u \leq b; \\ \frac{c - u}{b - c}, & \text{если } b < u < c, \end{cases}$$

где (a, c) — носитель нечеткого множества; b — координата максимума.

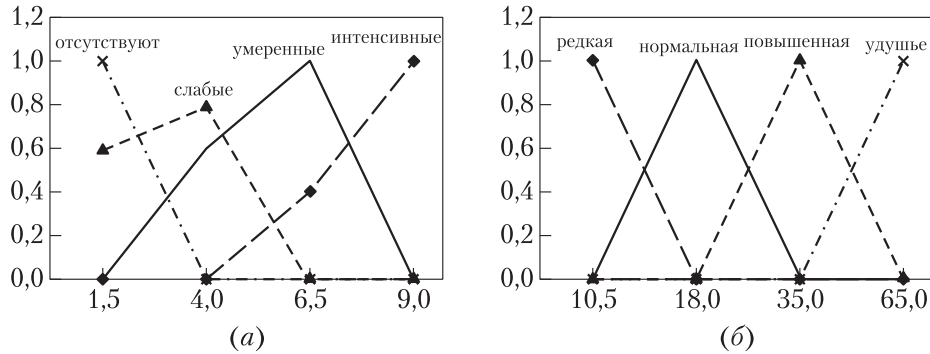


Рис. 3 Функции принадлежности термов нечетких переменных «загрудинные боли», «боли в животе», «вариабельные боли», «частота дыхательных движений»

Параметры функций принадлежности, представленных на рис. 3: терм «интенсивные» — треугольная с параметрами $a = 4$, $c = 9$, $b = 9$; терм «умеренные» — треугольная с параметрами $a = 1,5$, $c = 4$, $b = 9$; терм «слабые» — функция принадлежности в виде разности двух сигмоидных функций с параметрами $a_1 = 1,5$, $c_1 = 1,124$, $a_2 = 2,75$, $c_2 = 5,25$; терм «отсутствуют» — треугольная с параметрами $a = 1,5$, $c = 1,5$, $b = 4$; терм «редкая» — треугольная с параметрами $a = 10,5$, $c = 10,5$, $b = 18$; терм «нормальная» — треугольная с параметрами $a = 10,5$, $c = 18$, $b = 35$; терм «повышенная» — треугольная с параметрами $a = 18$, $c = 35$, $b = 65$; терм «удушье» — треугольная с параметрами $a = 35$, $c = 65$, $b = 65$. Функции принадлежности всех входных лингвистических переменных построены методом экспертного опроса [10].

Пространства входов и выходов X, Y и база знаний (БЗ) КВ для каждой нечеткой системы строятся по рекомендациям комитета экспертов Всероссийского научного общества кардиологов (ВНОК) по диагностике и лечению АГ [11], исследований Staessen *et al.* [12] и результатам анализа бесед с экспертом ККОБ. Причем размеры баз знаний для НСДАГ₁, ..., НСДАГ₉, НСССС равны 10, 14, 15, 10, 30, 20, 30, 20, 24, 70 соответственно и содержат нечеткие правила вида $(x_{1j} = a_{1j} \Theta_j \dots \Theta_j x_{nj} = a_{nj}) \Rightarrow y = d_j$, где j — номер правила в базе знаний, Θ_j — логическая операция, а d_j — четкое число (1 — истина, 0 — ложь), например: «Если боли за грудиной = интенсивные И на ЭКГ рубцовые изменения = есть И на ЭхоКГ зоны акинеза миокарда = есть, то инфаркт миокарда = истина». Для дефаззификации используется модифицированный вариант метода центра тяжести для одноточечных множеств:

$$y = \frac{\sum_{i=1}^n x_i \mu(x_i)}{\left(\sum_{i=1}^n \mu(x_i) \right)^{-1}},$$

применяемый при нечетком логическом выводе Сугено (7.15) в [13].

Концептуальные модели для НСДАГ_{1–9} и НСССС построены на базе (3.22) [2], обобщены ниже и расширены ресурсами (пространства входных и выходных переменных, база знаний системы, множество функций принадлежности, метод дефаззификации):

$$\begin{aligned}
 \text{mod}_f^a Z = & R^{\text{met prb}} \left(\text{met}_{1Z}^h, \text{prb}_f^h Z \right) \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_{1\text{gl}}^k \right) \circ \\
 \circ & R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_{\text{prod}}^{\text{mlang}} \right) \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_{\text{fuzz}}^{\text{solpr}} \right) \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{EV}_f Z \right) \circ \\
 & \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_Z^{\text{KB}} \right) \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, X \right) \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, Y \right) \circ \\
 \circ & R^{\text{met pr}} \left(\text{met}_{1Z}^h, F^\mu \right) \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{MET}^{\text{Def}} \right) \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_{\text{gbox}}^{\text{sh}} \right) \circ \\
 & \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_{\text{cyg}}^{\text{alg fuzz}} \right) \Big| Z \in \{ \text{ДАГ}_i, \text{ССС} | i = \overline{1, 9} \}, \quad (5)
 \end{aligned}$$

где $\text{met}_{1,Z}^h$ — метод; $\text{pr}_{1\text{gl}}^k$ — классификатор метода: «логико-лингвистический (символьный)»; $\text{pr}_{\text{gbox}}^{\text{sh}}$ — модель-схема метода: «серый ящик»; $\text{pr}_{\text{prod}}^{\text{mlang}}$ — язык описания модели: «продукции»; $\text{pr}_{\text{fuzz}}^{\text{solpr}}$ — процедура получения решения: «нечеткий вывод»; $\text{prb}_f^h Z$ — схема ролевых концептуальных моделей подзадачи СЗДАГ, решаемых нейронной сетью, которая определена схемой (2) в [4]; $\text{EV}_f Z$ — знания о гибридных возможностях (табл. 7.2 и 8.6 из [9]); pr_Z^{KB} — схема БЗ KB как множество из $k = 1, \dots, N_{\text{KB}}$ правил; F^μ — множество функций принадлежности пространства входов $X = X_1 \times \dots \times X_{N_X}$ и пространства выходов $Y = Y_1 \times \dots \times Y_{N_Y}$; $\text{pr}_{\text{cyg}}^{\text{alg fuzz}}$ — алгоритм нечеткого вывода (выбран алгоритм Сугено); MET^{Def} — метод дефаззификации.

2.4 Автономная модель, решающая подзадачи ССС и СМАД посредством методологии экспертных систем

Модель продукционной ЭСССС подзадачи «Оценка степени и стадии артериальной гипертензии, степени риска сердечно-сосудистых заболеваний» классифицирует степень АГ (7 классов), стадию АГ (эссенциальной) при отсутствии вторичной АГ у больного (3 класса) и общий риск сердечно-сосудистых заболеваний (5 классов).

Модель ЭССМАД подзадачи «Анализ суточного мониторинга артериального давления» прогнозирует показатели суточного мониторинга АД и контролирует их отклонение от нормы.

Для обеих экспертных систем по результатам анализа рекомендаций комитета экспертов ВНОК [11], исследований *Staessen et al.* [12] и результатам анализа бесед с экспертом ККОБ разрабатывается база фактов (БФ), содержащая символьное описание ресурсов предметной области (размерность БФ

ЭСССС — 30 символьных описаний ресурсов подзадачи ССС, размерность БФ ЭССМАД — 60 символьных описаний ресурсов подзадачи СМАД), и БЗ как совокупность символьных правил-продукций, имитирующих минимально возможную конструкцию точных рассуждений эксперта-врача (размерность БЗ ЭСССС — 45 правил, БЗ ЭССМАД — 50 правил).

Концептуальные модели для ЭСССС и ЭССАДДАД объединены в (6), построены на базе (3.14) [2] и расширены ресурсами (БФ и БЗ системы):

$$\begin{aligned} \text{mod}_{eZ}^a = & R^{\text{met prb}} \left(\text{met}_{1Z}^h, \text{pr}_{eZ}^h \right) \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_{\text{igl}}^k \right) \circ \\ \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_{\text{prod}}^{\text{mlang}} \right) \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_{\text{forw}}^{\text{solpr}} \right) \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{EV}_{eZ} \right) \circ \\ \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_Z^{\text{KB}} \right) \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_{\text{ifthen}}^{\text{sh}} \right) \circ \\ \circ R^{\text{met pr}} \left(\text{met}_{1Z}^h, \text{pr}_Z^{\text{FB}} \right) \Big| Z \in \{ \text{СМАД, ССС} \}, \quad (6) \end{aligned}$$

где met_{1Z}^h — метод; pr_{igl}^k — классификатор метода — «логико-лингвистический (символьный)»; $\text{pr}_{\text{ifthen}}^{\text{sh}}$ — модель-схема метода: «условие–действие»; $\text{pr}_{\text{prod}}^{\text{mlang}}$ — язык описания модели: «продукции»; $\text{pr}_{\text{forw}}^{\text{solpr}}$ — процедура получения решения: «прямой вывод»; prb_{eZ}^h — схема концептуальной модели подзадачи СЗДАГ, решаемой экспертной системой, которая определена схемой (2) в [1]; EV_{eZ} — знания о гибридных возможностях (табл. 7.2 и 8.6 из [9]); pr_Z^{KB} — схема БЗ KB; pr_Z^{FB} — БФ экспертной системы.

Таким образом, по спроектированным однородным моделям (2)–(6) можно сделать заключение, что задаче $\text{prb}_{\text{СЗДАГ}}^u$ — соответствует гетерогенное модельное поле M^a размерности $N_m = 23$ (рис. 4).

На рис. 4 для удобства восприятия технологические подзадачи из группы «Построение информативного набора признаков (симптомов) при диагностике заболеваний из областей 1–9» ($\text{prb}_{\text{ППС}_{1-9}}^h$) разбиты на 2 подгруппы $\text{prb}_{\text{ППС}_{1-3}}^h$ и $\text{prb}_{\text{ППС}_{4-9}}^h$, а подзадачи диагностики из группы «Диагностика критериев оценки сердечно-сосудистого риска и вторичной АГ у пациента» ($\text{prb}_{\text{ДАГ}_{1-9}}^h$) — на $\text{prb}_{\text{ДАГ}_{1-3}}^h$ и $\text{prb}_{\text{ДАГ}_{4-9}}^h$. Аналогично на рис. 4 были объединены и соответствующие модели из гетерогенного модельного поля.

Из рис. 4 видно, что для решения подзадачи $\text{prb}_{\text{СМАД}}^h$ разработаны две модели: нейровычислений и рассуждений на БЗ. Для решения подзадачи $\text{prb}_{\text{ССС}}^h$ также разработаны две модели: нечетких рассуждений и рассуждений на БЗ. Для остальных подзадач, в том числе и для решения каждой из технологических задач, которые обозначены на рис. 4 пунктирными овалами, разработано по одной модели.

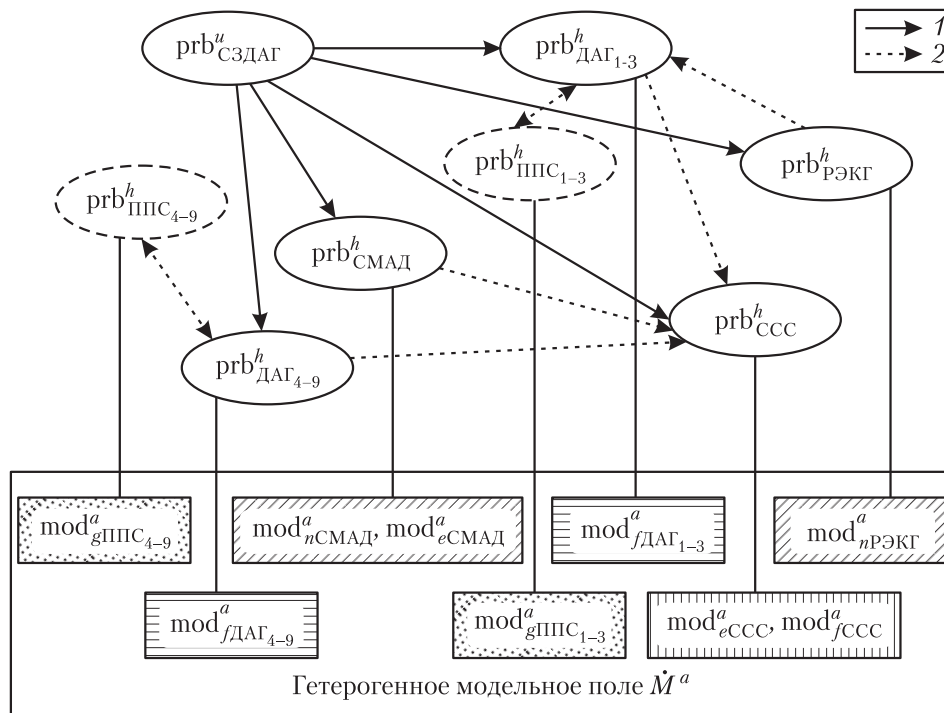


Рис. 4 Декомпозиция СЗДАГ и гетерогенное модельное поле: 1 — отношения «целое–часть» $R_2^{\text{prb}^{\text{prb}}}$; 2 — отношения информационно-управляющие $R_3^{\text{prb}^{\text{prb}}}$

Альтернативность инструментариев применительно к решению одной и той же подзадачи (функциональная избыточность), имеющая место для подзадач $\text{prb}_{\text{СМАД}}^h$ и $\text{prb}_{\text{ССС}}^h$, указывает на необходимость спецификации областей релевантности — множеств значений параметров внешней среды и характеристик ресурсов, учитываемых в описании состояния СППР, которые впоследствии при синтезе метода моделирования для неоднородной задачи влияют на выбор одного метода для каждой однородной подзадачи из альтернативного множества методов.

Таким образом, для задачи $\text{prb}_{\text{СМАД}}^h$ сформировано множество из двух пар «модель–интервалы»: $\{\text{mod}_n^a \{\text{int}_{\text{sex}}^n, \text{int}_{\text{age}}^n\}, \text{mod}_e^a \{\text{int}_{\text{sex}}^e, \text{int}_{\text{age}}^e, \text{int}_{\text{SAD}}^e, \text{int}_{\text{DAD}}^e, \text{int}_{\text{time}}^e\}\}$, где $\text{int}_{\text{sex}}^n = \{1, 0\}$ — пол мужской (1) или женский (0); $\text{int}_{\text{age}}^n = [18, 100]$ — возраст; $\text{int}_{\text{sex}}^e = \{\text{«мужской»}, \text{«женский»}\}$, $\text{int}_{\text{age}}^e = \text{int}_{\text{age}}^n$, $\text{int}_{\text{SAD}}^e = [50, 250]$ — систолическое артериальное давление; $\text{int}_{\text{DAD}}^e = [0, 200]$ — диастолическое артериальное давление; $\text{int}_{\text{time}}^e = \{\text{«утро»}, \text{«день»}, \text{«ночь»}, \text{«сутки»}\}$.

Для задачи $\text{prb}_{\text{ССС}}^h$ множество из двух пар «метод–интервалы»: $\{\text{mod}_f^a \{\text{int}_{\text{ДАГ}_{1-9}}^f, \text{int}_{\text{СМАД}}^f, \text{int}_{\text{САД}}^f, \text{int}_{\text{ДАД}}^f\}, \text{mod}_e^a \{\text{int}_{\text{ДАГ}_{1-9}}^e, \text{int}_{\text{СМАД}}^e, \text{int}_{\text{САД}}^e, \text{int}_{\text{ДАД}}^e\}\}$, где $\text{int}_{\text{ДАГ}_{1-9}}^f = \text{int}_{\text{ДАГ}_{1-9}}^e = \{[0, 1], [0, 10]\}$ — информация о факторах риска АГ, поражениях органов-мишеней, ассоциативных клинических состояниях, вторичной эндокринной или почечной АГ; $\text{int}_{\text{СМАД}}^f = \text{int}_{\text{СМАД}}^e = [0, 100]$ — степень отклонения данных суточного мониторинга АД от нормы; $\text{int}_{\text{САД}}^e = [50, 250]$; $\text{int}_{\text{ДАД}}^e = [0, 200]$.

После построения гетерогенного модельного поля и спецификации областей релевантности можно перейти к формированию таблицы гибридных стратегий и конструированию метода решения неоднородной задачи СЗДАГ.

3 Алгоритм синтеза метода решения неоднородной задачи

Для разработки алгоритма введем отношения соответствия между исходными данными и целями неоднородной задачи $\text{prb}_{\text{СЗДАГ}}^u$, а также целями и исходными данными однородных подзадач соответственно $R_4^{\text{prb prb}}$ [2].

Отображения $(f_1^{\text{GL}}, f_2^{\text{GL}}, f_3^{\text{GL}}, f_4^{\text{GL}}, f_5^{\text{GL}}$ соответственно) $\text{GL}_{\text{СЗДАГ}}^u$ (диагноз — заключение о сущности болезни и состоянии пациента) на $\text{GL}_{\text{РЭКГ}}^h$ (диагностирование патологических состояний на ЭКГ, установление соответствия состояния пациента норме), $\text{GL}_{\text{СМАД}}^h$ (диагностирование отклонения показаний СМАД пациента от нормы для последующей оценки степени АГ), $\text{GL}_{\text{ППС}_{1-9}}^h$ (оптимизация набора симптомов для диагностики заболеваний и рисков АГ), $\text{GL}_{\text{ДАГ}_{1-9}}^h$ (выявление факторов риска АГ, поражений органов-мишеней, ассоциативных клинических состояний, вторичной эндокринной или почечной АГ либо оценка состояния пациента как «норма») и $\text{GL}_{\text{ССС}}^h$ (оценка степени и стадии АГ, риска сердечно-сосудистых заболеваний) являются сюръекциями, так как область значений $\text{GL}_{\text{СЗДАГ}}^u$ совпадает со всем множеством $\text{GL}_{\text{РЭКГ}}^h$, $\text{GL}_{\text{СМАД}}^h$, $\text{GL}_{\text{ППС}_{1-9}}^h$, $\text{GL}_{\text{ДАГ}_{1-9}}^h$ и $\text{GL}_{\text{ССС}}^h$ соответственно.

Введем обозначения: $\text{DAT}_{\text{СЗДАГ}}^u$ — это показатели состояния здоровья пациента и их значения (x_1), а также множества субъектов (x_2), объектов (x_3) и задач диагностики (x_4); $\text{DAT}_{\text{РЭКГ}}^h$ — образ электрокардиосигнала пациента (x_5); $\text{DAT}_{\text{СМАД}}^h$ — пол (x_6), возраст (x_7) и данные СМАД пациента (x_8); $\text{DAT}_{\text{ППС}_{1-9}}^h$ — множество $S_{\text{simp}} = \{S_{\text{simp}} | i = \overline{1, 9}\}$ подмножеств симптомов заболеваний ($\{x_{9i}\}$); $\text{DAT}_{\text{ДАГ}_{1-9}}^h$ — множество симптомов из областей 1–9 ($\{x_{10i}\}$); $\text{DAT}_{\text{ССС}}^h$ — отклонение от нормы $\Delta_{\text{СМАД}} \in \{0; 100\}$ (x_{11}); результирующее множество диагноза Rez (x_{12}) и множество показателей здоровья $S_{\text{simp}10}$ для подзадачи ССС (x_{13}).

Тогда имеют место отображения

$$\begin{aligned} f_1^{\text{DAT}} &: \text{DAT}_{\text{СЗДАГ}}^u \rightarrow \text{DAT}_{\text{РЭКГ}}^h; \\ f_2^{\text{DAT}} &: \text{DAT}_{\text{СЗДАГ}}^u \rightarrow \text{DAT}_{\text{СМАД}}^h; \\ f_3^{\text{DAT}} &: \text{DAT}_{\text{СЗДАГ}}^u \rightarrow \text{DAT}_{\text{ППС}_{1-9}}^h; \\ f_4^{\text{DAT}} &: \text{DAT}_{\text{СЗДАГ}}^u \rightarrow \text{DAT}_{\text{ДАГ}_{1-9}}^h; \\ f_5^{\text{DAT}} &: \text{DAT}_{\text{СЗДАГ}}^u \rightarrow \text{DAT}_{\text{ССС}}^h, \end{aligned}$$

которые инъективны, так как образованы следующими парами: $\{(x_1, x_5)\}$, $\{(x_1, x_6), (x_1, x_7), (x_1, x_8)\}$, $\{(x_1, x_{91}), \dots, (x_1, x_{99})\}$, $\{(x_1, x_{101}), \dots, (x_1, x_{109})\}$ и $\{(x_1, x_{11}), (x_1, x_{13}), (x_4, x_{12})\}$ соответственно, т. е. каждый элемент из области их значений имеет единственный прообраз.

После установления отношений соответствия заменим отношения декомпозиции $R_3^{\text{prb}^h \text{prb}^b}$, которые установлены между однородными подзадачами из декомпозиции задачи $\text{reb}_{\text{СЗДАГ}}^u$, отношениями интеграции ${}^{\psi\phi}R_{\text{int}}^t$ [2], где ψ и ϕ — это подзадачи, знания которых участвуют в интеграции, а t — тип отношения интеграции, и введем интерфейсы автономных моделей [2]:

- (1) отношение декомпозиции $R_3^{\text{prb}^h \text{РЭКГ} \text{prb}^h \text{ДАГ}_1}$ заменяется отношением интеграции ${}^{\text{prb}^h \text{РЭКГ} \text{prb}^h \text{ДАГ}_1} R_{\text{int}}^{\text{dop}}$ дополнения одних знаний другими — логические знания готовят информацию о патологиях сердца человека в подзадаче распознавания и интерпретации ЭКГ и передают посредством интерфейса ${}^{\text{prb}^h \text{РЭКГ} \text{prb}^h \text{ДАГ}_1} \zeta_{\text{СЗДАГ}}^u$ для работы лингвистических знаний в задаче диагностики поражений органов-мишеней;
- (2) отношение декомпозиции $R_3^{\text{prb}^h \text{ППС}_{1-9} \text{prb}^h \text{ДАГ}_{1-9}}$ заменяется отношением интеграции «препроцессия» ${}^{\text{prb}^h \text{ППС}_{1-9} \text{prb}^h \text{ДАГ}_{1-9}} R_{\text{int}}^{\text{upr}}$ — «препроцессия» генетическими знаниями группы подзадач «Построение информативного набора признаков (симптомов) при диагностике заболеваний из областей 1–9» лингвистических знаний соответствующих подзадач из группы «Диагностика критериев оценки сердечно-сосудистого риска и вторичной АГ у пациента» (посредством соответствующих девяти однотипных интерфейсов ${}^{\text{prb}^h \text{ППС}_{1-9} \text{prb}^h \text{ДАГ}_{1-9}} \zeta_{\text{СЗДАГ}}^u$);
- (3) отношение декомпозиции $R_3^{\text{prb}^h \text{ДАГ}_{1-9} \text{prb}^h \text{ССС}}$ заменяется отношением интеграции ${}^{\text{prb}^h \text{ДАГ}_{1-9} \text{prb}^h \text{ССС}} R_{\text{int}}^{\text{dop}}$ дополнения одних знаний другими — лингвистические знания готовят информацию о патологиях организма человека и данные о признаках состояния здоровья в подзадачах из группы «Диагностика критериев оценки сердечно-сосудистого риска и вторичной АГ у

пациента» для работы логических знаний в подзадаче «Оценка степени и стадии артериальной гипертензии, степени риска сердечно-сосудистых заболеваний» (посредством соответствующих девяти однотиповых интерфейсов $\text{prb}_{\text{ДАГ1-9}}^h \text{prb}_{\text{ССС}}^h \xi_{\text{СЗДАГ}}^u$);

- (4) отношение декомпозиции $R_3^{\text{prb}_{\text{СМАД}}^h \text{prb}_{\text{ССС}}^h}$ заменяется отношением интеграции $\text{prb}_{\text{СМАД}}^h \text{prb}_{\text{ССС}}^h R_{\text{int}}^{\text{dop}}$ дополнения одних знаний другими — лингвистические знания готовят информацию о данных по СМАД пациента в подзадаче «Анализ суточного мониторингирования артериального давления» для работы логических знаний в подзадаче «Оценка степени и стадии артериальной гипертензии, степени риска сердечно-сосудистых заболеваний» (посредством интерфейса $\text{prb}_{\text{СМАД}}^h \text{prb}_{\text{ССС1}}^h \xi_{\text{СЗДАГ}}^h$).

Далее для того, чтобы отобразить вышеопределенные связи между однородными подзадачами с учетом априорных знаний о наилучших гибридных стратегиях [9, табл. 8.6], чтобы показать изменчивый состав и структуру модели неоднородной задачи СЗДАГ, а также варианты стратегий гибридизации для последующего синтеза ФГиИС, формируется таблица гибридных стратегий (ТГС) (см. таблицу).

Таблица гибридных стратегий задается матрицей «однородная задача/модели однородных задач — отношения интеграции — однородная задача/модели однородных задач» и впоследствии вместе с декомпозицией сложной задачи, гетерогенным модельным полем и отношениями интеграции и соответствия используется в качестве входных данных для алгоритма синтеза решения неоднородной задачи (алгоритм см. ниже).

В ячейках ТГС указаны отношения интеграции подзадач СЗДАГ на моделях, посредством которых они решаются, а в скобках — экспертные оценки качества соответствующих гибридных систем [9, табл. 8.6].

Функциональная гибридная интеллектуальная система автоматизированной диагностики АГ (ФГиИСАДАГ) — это система, использующая для решения сложной задачи диагностики АГ несколько методов имитации интеллектуальной деятельности человека (методы нейросетей, нечетких систем, ГА), объединенных в гетерогенное модельное поле, и моделирующая неоднородность процесса диагностики АГ.

Зададим алгоритм синтеза стратегии решения неоднородной задачи, разработанный на основе алгоритма 6.4 из [2], инициализирующего архитектуру ФГиИСАДАГ, интерпретация которой рассматривается как метод решения неоднородной задачи. Данный алгоритм моделирует принятие решений медицинским консилиумом (*лат. consilium* — совещание, обсуждение) — совещанием врачей разных специальностей с целью выработки заключения о состоянии здоровья обследуемого, установления диагноза болезни, определения ее прогноза, тактики обследования и лечения больного [14], т. е. моделирует деятельность ЛПР (лица,

Гибридные стратегии задачи СЗДАГ, T^u

Подзадачи СЗДАГ		$prb_{ДАГ_{1-9}}^h$	$prb_{ССС}^h$	
Подзадачи СЗДАГ	Модели подзадач	$mod_f^a ДАГ_{1-9}$	$mod_e^a ССС$	$mod_f^a ССС$
$prb_{ППС_{1-9}}^h$	$mod_g^a ППС_{1-9}$	$prb_{ППС_{1-9}}^h prb_{ДАГ_{1-9}}^h R_{int}^{upr} (7)$	—	—
$prb_{СМАД}^h$	$mod_n^a СМАД$	—	$prb_{СМАД}^h prb_{ССС}^h R_{int}^{dop} (3)$	$prb_{СМАД}^h prb_{ССС}^h R_{int}^{dop} (4)$
	$mod_e^a СМАД$	—	—	$prb_{СМАД}^h prb_{ССС}^h R_{int}^{dop} (6)$
$prb_{ДАГ_{1-9}}^h$	$mod_f^a ДАГ_{1-9}$	—	$prb_{ДАГ_{1-9}}^h prb_{ССС}^h R_{int}^{dop} (6)$	—
$prb_{РЭКГ}^h$	$mod_n^a РЭКГ$	$prb_{РЭКГ}^h prb_{ДАГ_1}^h R_{int}^{dop} (4)$	—	—

принимающего решение) (кардиолога) по организации консилиума, взаимодействию с экспертами (врачами узких специализаций) и принятию окончательного решения. При этом на вход алгоритма поступает гетерогенное модельное поле (см. разд. 2) — модель разнородных знаний экспертов, врачей смежных специальностей, профессионально лучше, чем ЛПР, знающих отдельные аспекты задачи, обладающих специальными знаниями и опытом для спецификации областей релевантности существенных параметров внешней среды, влияющей на СППР, и характеристик ресурсов, учитываемых при описании СППР, а также данными для выработки и принятия решения, которые высказывают свое личное, субъективное мнение, что помогает ЛПР в окончательном решении.

На вход алгоритма подаются:

- (1) декомпозиция $\overline{\text{prb}}_{\text{СЗДАГ}}^u = \left(\text{PRB}^h, r_3^{\text{prb prb}} \right)$ в виде сети (ориентированного графа без петель), гетерогенное модельное поле M^a и таблица гибридных стратегий, которые задают взаимно однозначные соответствия:

$$\begin{aligned} R_3^{\text{prb prb}} \left(\text{prb}_{\text{ППС}_{1-9}}^h, \text{prb}_{\text{ДАГ}_{1-9}}^h \right) &\leftrightarrow R_{\text{int}}^{\text{upr}} ; \\ R_3^{\text{prb prb}} \left(\text{prb}_{\text{СМАД}}^h, \text{prb}_{\text{ССС}}^h \right) &\leftrightarrow R_{\text{int}}^{\text{dop}} ; \\ R_3^{\text{prb prb}} \left(\text{prb}_{\text{РЭКГ}}^h, \text{prb}_{\text{ДАГ}_1}^h \right) &\leftrightarrow R_{\text{int}}^{\text{dop}} ; \\ R_3^{\text{prb prb}} \left(\text{prb}_{\text{ДАГ}_{1-9}}^h, \text{prb}_{\text{ССС}}^h \right) &\leftrightarrow R_{\text{int}}^{\text{dop}} ; \end{aligned}$$

- (2) множество отношений соответствия целей и исходных данных однородных задач из декомпозиции PRB^h , а также целей и исходных данных $\text{prb}_{\text{СЗДАГ}}^u$ — $f_1^{\text{GL}}, f_2^{\text{GL}}, f_3^{\text{GL}}, f_4^{\text{GL}}, f_5^{\text{GL}}, f_1^{\text{DAT}}, f_2^{\text{DAT}}, f_3^{\text{DAT}}, f_4^{\text{DAT}}, f_5^{\text{DAT}}$;
- (3) множество интерпретаторов автономных моделей $I^{M^a} = \{I^e, I^g, I^f, I^n\}$; множество межмодельных интерфейсов $\left\{ \text{prb}_{\text{СМАД}}^h \text{prb}_{\text{ССС}_1}^h \zeta_{\text{СЗДАГ}}^u, \text{prb}_{\text{ДАГ}_{1-9}}^h \text{prb}_{\text{ССС}}^h \zeta_{\text{СЗДАГ}}^u, \text{prb}_{\text{ППС}_{1-9}}^h \text{prb}_{\text{ДАГ}_{1-9}}^h \xi_{\text{СЗДАГ}}^u, \text{prb}_{\text{РЭКГ}}^h \text{prb}_{\text{ДАГ}_1}^h R_{\text{int}}^{\text{dop}} \right\}$ и множество моделей, решающих задачи из декомпозиции $\text{prb}_{\text{СЗДАГ}}^u$, $\{ \text{mod}_g^a \text{ППС}_{1-9}, \text{mod}_n^a \text{СМАД}, \text{mod}_e^a \text{СМАД}, \text{mod}_f^a \text{ДАГ}_{1-9}, \text{mod}_n^a \text{РЭКГ}, \text{mod}_e^a \text{ССС}, \text{mod}_f^a \text{ССС} \}$.
- (4) порядок решения подзадач СЗДАГ при помощи индекса $\text{In}(\text{In prb}^h)$, определяющего последовательность решения подзадач из декомпозиции PRB^h : ${}^1\text{prb}_{\text{ППС}_{1-9}}^h, {}^2\text{prb}_{\text{СМАД}}^h, {}^2\text{prb}_{\text{РЭКГ}}^h, {}^2\text{prb}_{\text{ДАГ}_{2-9}}^h, {}^3\text{prb}_{\text{ДАГ}_1}^h, {}^4\text{prb}_{\text{ССС}}^h$. Таким образом, вначале решаются задачи из группы ППС_{1–9}, затем задачи СМАД, РЭКГ и из группы ДАГ_{2–9}. Подзадача ДАГ₁ решается только третьей по

очередности, так как для ее решения требуется информация о решении подзадачи РЭКГ. Последней решается подзадача ССС, обрабатывающая информацию, переданную решением подзадач СМАД и ДАГ₁₋₉. Подзадачи с $\text{In} = 1$ запишем в массив M_1 , с $\text{In} = 2$ — в массив M_2 и т. д.

На выходе алгоритма — список $\text{Ls}^m = \left\{ r_{\text{int}}(\text{mod}_x^a, \text{mod}_y^a) \Big|_{\gamma}^{\xi} \right\}$, в котором знаки однородных задач заменены на знаки выбранных в соответствии с областями релевантности моделей, каждой модели поставлен в соответствие интерпретатор для имитации рассуждений и определены интерфейсы для обмена информацией; x, y — задачи из декомпозиции $\text{prb}_{\text{СЗДАГ}}^u$, причем каждой модели сопоставлен интерпретатор I^v , а каждому отношению интеграции сопоставлен интерфейс ${}^x\zeta_{\text{СЗДАГ}}^u$ или ${}^y\zeta_{\text{СЗДАГ}}^u$. Список Ls^m — исходная информация для интерпретации ФГИИС, которая рассматривается как метод решения сложной задачи, это множество, описывающее архитектуру ФГИИСАДАГ.

Введем обозначения:

- $E^l = \left\{ E_j^l \right\}$, $j = \overline{1, N_{\text{ТР}}}$ — множество матриц «модель–модель», строки которых соответствуют моделям задачи prb_{ξ}^h , а столбцы — моделям задачи prb_{γ}^h для множества троек $\left\{ r_{\text{int}}(\text{prb}_{\xi}^h, \text{prb}_{\gamma}^h) \right\}$, обрабатываемое на l -м шаге;
- N_M — число l -х шагов, соответствующее числу массивов, хранящих порядок решения подзадач из декомпозиции СЗДАГ (в данном случае $N_M = 4$, $l = \overline{1, 4}$ и на первом шаге алгоритма будут обработаны 9 троек, на втором — 10 троек, на третьем — 1, а $l = 4$ является признаком завершения работы алгоритма).

Причем введем ограничение: из подзадач prb_{γ}^h берутся те, In которых больше In подзадачи prb_{ξ}^h .

Алгоритм:

1. Начало, $l = 1$, $\text{In} = l$.
2. Выбрать из декомпозиции $\overline{\text{prb}_{\text{СЗДАГ}}^u} = \left(\text{PRB}^h, r_3^{\text{prb prb}} \right)$ подмножество троек $\text{PRB}_{\beta}^h = \left\{ r_3^{\text{prb prb}}(\text{In prb}_{\xi}^h, \text{In} > l \text{ prb}_{\gamma}^h) \right\} \Big|_{\beta = 1, N_{\text{ПН}}}$.
3. $\beta = 1$, $j = \beta$, $N_{\text{ТР}} = 0$.
4. Для β -й пары из PRB_{β}^h инициализировать матрицу E_j^l , помечая ее строки q и столбцы ψ моделями mod^a подзадач prb_{ξ}^h и prb_{γ}^h . Добавить матрицу E_j^l в E^l . $N_{\text{ТР}} = N_{\text{ТР}} + 1$.
5. $N_{\text{ТР}} = N_{\text{ПН}}$? Если «нет», то положить $\beta = \beta + 1$ и вернуться к п. 3. Если «да», то перейти к п. 6.
6. Начало цикла обработки множества матриц E^l , $j = 1$.

7. Берется E_j^l и в соответствии с T^u дается оценка только тем моделям $\{\text{mod}_q^a\}$, у которых есть отношения интеграции с $\{\text{mod}_\psi^a\}$. Если такая оценка уже есть, то занести ее в знаменатель и перейти к п. 8, иначе, используя нечеткую или экспертную систему (алгоритм 1 и табл. 4.1–4.4 приложения 4 в [9]), получить оценки моделей $\{\text{mod}_q^a\}$ и занести их в знаменатели дробей-элементов $\varepsilon_{q\psi}^j$ матрицы E_j^l .
8. Моделировать решения prb_ξ^h на каждой из $\{\text{mod}_q^a\}$ и полученные значения выходов подать на входы моделей $\{\text{mod}_\psi^a\}$. Используя нечеткую или экспертную системы (алгоритм 1 и табл. 4.1–4.4 приложения 4 в [9]), получить оценки моделей $\{\text{mod}_\psi^a\}$ задачи prb_γ^h , заносая оценки в числители дробей-элементов $\varepsilon_{q\psi}^j$ матрицы E_j^l . Выбрать $\max_{q\psi}\{\varepsilon_{q\psi}^j\}$ пару моделей и занести ее в список Ls^m .
9. $j = N_{\text{ПН}}$? Если «нет», то $j = j + 1$ и перейти к п. 7, иначе — к п. 10.
10. $l = N_M$? Если «нет», то $l = l + 1$ и перейти к п. 2, иначе — к п. 11.
11. Дополнить список Ls^m интерпретаторами моделей и межмодельными интерфейсами. Инициализировать ФГиИС в соответствии с Ls^m .
12. Конец.

Алгоритм легко программируем, так как явно показана связь очередности решения задачи с формированием (внешний цикл, переменная l и введено ограничение на формирование троек) и обработкой троек (внутренний цикл, переменная β) $r_3^{\text{prb prb}}(\text{In prb}_\xi^h, \text{In} > l \text{ prb}_\gamma^h)$.

Таким образом, определен алгоритм выбора стратегии решения сложной диагностической задачи, определены его входы и выход и можно перейти к синтезу архитектуры ФГиИС с помощью алгоритма, описанного выше, и имитировать работу ФГиИС на ЭВМ, т. е. в определенном декомпозицией порядке смоделировать решение однородных задач и, в соответствии с отношениями между целями и исходными данными однородных задач из декомпозиции prb^h , а также целями и исходными данными СЗДАГ, $\text{prb}_{\text{СЗДАГ}}^u$, получить результирующую информацию (диагноз).

4 Архитектура функциональной гибридной интеллектуальной системы

Результат работы алгоритма синтеза метода решения задачи СЗДАГ — описание архитектуры ФГиИСАДАГ (виртуального медицинского консилиума по диагностике АГ). Система реализуется в пакете MATLAB 7 и Simulink.

Дополним описание архитектуры ФГиИСАДАГ интерфейсами пользователя (рис. 5). Архитектура ФГиИСАДАГ — модульная и гибкая для «сборки» гибридной интеллектуальной системы с учетом запроса пользователя. В Simulink создается библиотека ФГиИСАДАГ с блоками всех ИНС, НС и ГА, что дает

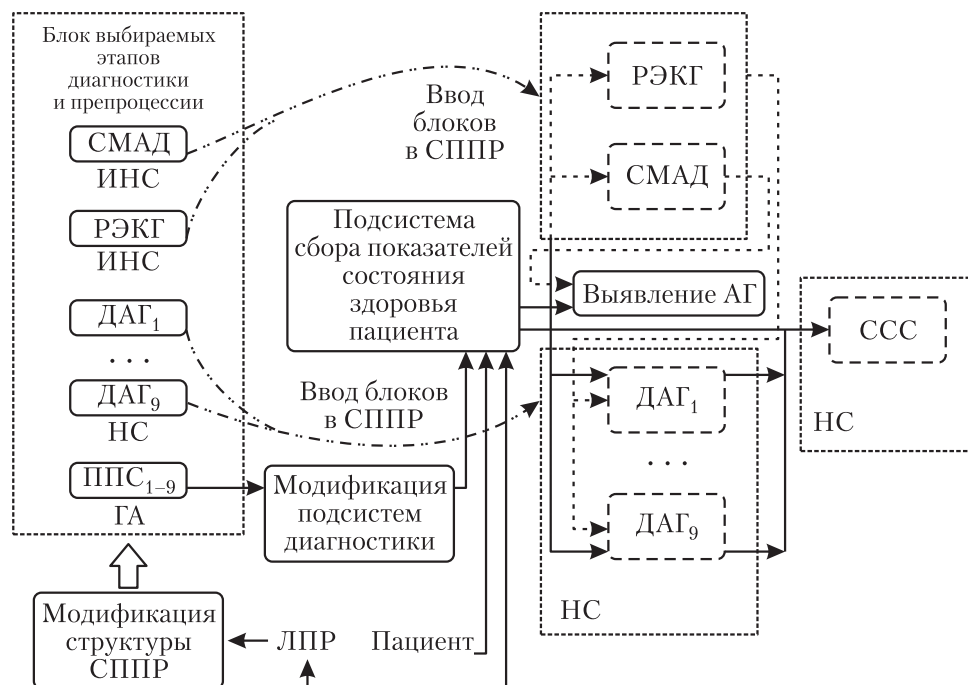


Рис. 5 Архитектура ФГиИСАДАГ

возможность методом Drag&Drop создавать и алгоритмически изменять состав гибрида, например, исключив из него ИНССМАД, если нет необходимости в анализе СМАД; или не включать ИНСРЭКГ, если данные по ЭКГ уже есть и нет необходимости в автоматизированном анализе образа ЭКГ; или не включать ГА (все или некоторые из них), если, по мнению разработчика (пользователя) гибрида, нет необходимости в модификации множеств входных признаков (симптомов), по которым диагностируется состояние пациента.

На рис. 5 модуль «Модификация структуры СППР» открывает интерфейс пользователя (ЛПР, врача-кардиолога), с помощью которого можно определить, будет ли ФГиСАДАГ решать подзадачи СМАД (решается посредством ИНС), РЭКГ (решается посредством ИНС), ППС₁₋₉ (решается посредством ГА) и все ли ДАГ₁₋₉ (решается посредством НС).

Если ППС₁₋₉ выбраны для решения, то активируется блок «ППС₁₋₉», с которого передается информация на блок «Модификация подсистем диагностики» в соответствии с результатами решения технологических подзадач «Построение информативного набора признаков (симптомов) при диагностике заболеваний», который корректирует работу блоков диагностики поражений органов-мишеней,

факторов риска, цереброваскулярных болезней, метаболического синдрома и сахарного диабета, заболеваний периферических артерий, ишемической болезни сердца, эндокринной АГ, паренхиматозной нефропатии и реноваскулярной АГ (блоки ДАГ₁, ..., ДАГ₉ соответственно) и открывает интерфейс пользователя, с помощью которого вводятся показатели состояния здоровья пациента (симптомы). Блоки ДАГ₁, ..., ДАГ₉ передают блоку ССС (решающему подзадачу «Оценка степени и стадии артериальной гипертензии, степени риска сердечно-сосудистых заболеваний») информацию о выявленных у пациента патологиях, являющуюся результатом моделирования принятия решения экспертами (врачами смежных специальностей). Подзадачи ДАГ₁, ..., ДАГ₉, ССС решаются с помощью нечетких систем.

5 Заключение

В работе даны результаты, полученные в КФ ИПИ РАН применительно к задаче диагностики АГ в рамках одного из классов искусственных гетерогенных систем — функциональных гибридных интеллектуальных систем — и связанные с синтезом метода решения СЗДАГ в рамках ПС-методологии. Для решения 21 подзадачи из декомпозиции сложной задачи $rgb_{СЗДАГ}^u$ спроектировано гетерогенное модельное поле из 23 однородных моделей, число которых превышает число подзадач в декомпозиции, что свидетельствует о функциональной избыточности инструментариев решения одной и той же подзадачи (подзадачи $rgb_{СМАД}^h$ и подзадачи $rgb_{ССС}^h$). Вследствие функциональной избыточности были специфицированы области релевантности для множеств моделей подзадач $rgb_{СМАД}^h$ и $rgb_{ССС}^h$ и построена ТГС — результат ввода отношений соответствия между исходными данными и целями неоднородной задачи $rgb_{СЗДАГ}^u$, а также исходными данными и целями однородных подзадач соответственно, ввода интерфейсов автономных моделей и замены отношений между подзадачами на отношения интеграции.

Для инициализации ФГиИСАДАГ, моделирующей гетерогенность процесса диагностики АГ, разработан алгоритм синтеза стратегии решения неоднородной задачи, моделирующий виртуальное принятие решений медицинским консилиумом при диагностике АГ. Результатом работы алгоритма синтеза метода решения гетерогенной задачи СЗДАГ является описание архитектуры ФГиИСАДАГ.

Продолжаются работы по имитации ФГиИС на ЭВМ средствами пакета MATLAB 7 – Simulink, модельные эксперименты и интерпретация их результатов.

Литература

1. Нариньяни А. С. Неопределенные множества — новый тип данных для представления Знаний. Препринт ВЦ СОАН СССР, 1980. Вып. 232. 28 с.

2. Колесников А. В., Кириков И. А. Методология и технология решения сложных задач методами функциональных гибридных интеллектуальных систем. — М.: ИПИ РАН, 2007. 387 с.
3. Окороков А. Н. Диагностика болезней внутренних органов. Диагностика болезней сердца и сосудов. Т. 6. — М.: Мед. литература, 2002. 464 с.
4. Кириков И. А., Колесников А. В., Румовская С. Б. Исследование сложной задачи диагностики артериальной гипертензии в методологии искусственных гетерогенных систем // Системы и средства информатики, 2013. Т. 23. № 2. С. 81–99.
5. Зудбинов Ю. И. Азбука ЭКГ. — 3-е изд. — Ростов-на-Дону: Феникс, 2003. 169 с.
6. Хайкин С. Нейронные сети: Полный курс / Пер. с англ. — Изд. 2-е, испр. — М.: Вильямс, 2006. 1110 с. (*Haykin S. Neural networks: Comprehensive foundation. — 2nd ed. — Singapore: Pearson Prentice Hall, Inc., 1999. 823 p.*)
7. Аль Мабрук М. Аппаратно-программные средства и алгоритмы распознавания патологий сердца на основе перцептронных сетей: Дисс. . . . канд. техн. наук. — Владимир, 2011. 203 с.
8. The research source for complex physiologic signals // PhysioNet. <http://physionet.org>.
9. Колесников А. В. Гибридные интеллектуальные системы. Теория и технология разработки / Под ред. А. М. Яшина. — СПб.: СПбГТУ, 2001. 711 с.
10. Штовба С. Д. Проектирование нечетких систем средствами Matlab. — М.: Горячая линия — Телеком, 2007. 288 с.
11. Комитет экспертов ВНОК. Диагностика и лечение артериальной гипертензии. Российские рекомендации // Системные гипертензии, 2010. Вып. 3. С. 5–26.
12. Staessen J., O'Brien E., Atkins N., Bulpitt C., Cox J., Fagard R., O'Malley K., Thijs L., Amery A. The increase in blood pressure with age and body mass index is overestimated by conventional sphygmomanometry // Am. J. Epidemiol., 1992. Vol. 136. P. 450–459.
13. Леоненков А. В. Нечеткое моделирование в среде MATLAB и fuzzyTECH. — СПб.: БХВ-Петербург, 2005. 736 с.
14. Малая медицинская энциклопедия. Т. 2. — М.: Сов. энциклопедия, 1991. С. 488.

Поступила в редакцию 11.09.13

FUNCTIONAL HYBRID INTELLIGENT DECISION SUPPORT SYSTEM FOR DIAGNOSTICS OF ARTERIAL HYPERTENSION

I. A. Kirikov¹, A. V. Kolesnikov^{1,2}, and S. B. Rumovskaya¹

¹Kaliningrad Branch of the Institute of Informatics Problems, Russian Academy of Sciences, 5 Gostinaya Str., Kaliningrad 236000, Russian Federation

²Immanuel Kant Baltic Federal University, 14 Nevskogo Str., Kaliningrad 236041, Russian Federation

Abstract: The paper introduces an approach to modeling decision support system for diagnostics of the arterial hypertension within the problem-structured methodology of functional artificial heterogeneous systems. The results of integrated method design are presented, including the following tasks: development of the heterogeneous model field for the complex problem of arterial hypertension diagnostics, region specification of subtasks autonomous model pertinence, and development of the algorithm for designing the integrated method and the model for solving the complex (heterogeneous) task. The models are designed in accordance with the recommendations of the Society of Cardiology of the Russian Federation for diagnostics and treatment of arterial hypertension. The architecture of the virtual physician's consultation diagnostics of arterial hypertension is the result of the algorithm synthesizing the method for solving the heterogeneous task. Such systems allow to synthesize dynamically a new method for elaborating diagnosis for each patient individually.

Keywords: functional hybrid intelligent system; heterogeneous model field; integrated method for solving the heterogeneous task; diagnostics of arterial hypertension

DOI: 10.14357/08696527140110

References

1. Narin'jani, A. S. 1980. *Nedopredelennye mnozhestva — novyy tip dannykh dlya predstavleniya Znaniy* [Underdetermined sets — a new type of data for knowledge representing]. Novosibirsk: CC SBAS USSR. Preprint No. 232. 28 p.
2. Kolesnikov, A. V., and I. A. Kirikov. 2007. *Metodologiya i tekhnologiya resheniya slozhnykh zadach metodami funktsional'nykh gibridnykh intellektual'nykh sistem* [Methodology and technology for solving of complex problems using the methodology of functional hybrid artificial systems]. Moscow: IPI RAN. 387 p.
3. Okorokov, A. N. 2002. *Diagnostika bolezney vnutrennikh organov. Diagnostika bolezney serdtsa i sosudov* [Diagnostics of visceral organs diseases. Diagnostics of heart and vessels diseases]. Moscow: Medical Literature. Vol. 6. 464 p.
4. Kirikov, I. A., A. V. Kolesnikov, and S. B. Rumovskaya. 2013. Issledovanie slozhnoy zadachi diagnostiki arterial'noy gipertenzii v metodologii iskusstvennykh geterogennykh sistem [Research of the complex problem at diagnosing the arterial hypertension within the methodology of artificial heterogeneous systems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 23(2):96–114.
5. Zudbinov, Ju. I. 2003. *Azbuka EKG* [Alphabet of ECG]. 3rd ed. Rostov-na-Donu: Phoenix Publ. 169 p.
6. Haykin, S. 1999. *Neural networks: Comprehensive foundation*. 2nd ed. Singapore: Pearson Prentice Hall, Inc. 823 p.
7. Al' Mabruk M. 2011. *Apparatno-programmnye sredstva i algoritmy raspoznavaniya patologiy serdtsa na osnove perseptronnykh setey* [Hardware, software and algorithms for recognition of heart pathology based on perceptron-type networks]. D.Sc. Diss. Vladimir. 203 p.
8. The research source for complex physiologic signals. Available at: <http://physionet.org> (accessed July 7, 2013).

9. Kolesnikov, A. V. 2001. *Gibridnye intellektual'nye sistemy. Teoriya i tekhnologiya razrabotki* [Hybrid artificial systems. Theory and development technology]. St. Petersburg: SPbGTU. 710 p.
10. Shtovba, S. D. 2007. *Proektirovanie nechetkikh sistem sredstvami Matlab* [Designing of fuzzy systems by Matlab features]. Moscow: Goryachaya liniya — Telekom. 288 p.
11. Komitet ekspertov VNOK. 2010. Diagnostika i lechenie arterial'noy gipertenzii. Rossiyskie rekomendatsii [Diagnosing and treatment of arterial hypertension. Russian recommendation]. *Sistemnye gipertenzii* [Systematic hypertension] 3:5–26.
12. Staessen, J., E. O'Brien, N. Atkins, C. Bulpitt, J. Cox, R. Fagard, K. O'Malley, L. Thijs, and A. Amery. 1992. The increase in blood pressure with age and body mass index is overestimated by conventional sphygmomanometry. *Am. J. Epidemiol.* 136(4):450–459.
13. Leonenkov, A. V. 2005. *Nechetkoe modelirovanie v srede MATLAB i fuzzyTECH* [Fuzzy modeling in MATLAB and fuzzyTECH]. St. Petersburg: BHV-Peterburg. 736 p.
14. *Malaya meditsinskaya entsiklopediya* [Small medical encyclopedia]. 1991. Moscow: Sov. Enciclopedia. 2:488.

Received September 11, 2013

Contributors

Kirikov Igor A. (b. 1955) — Candidate of Science (PhD) in technology, Director, Kaliningrad Branch of the Institute of Informatics Problems, Russian Academy of Sciences, 5 Gostinaya Str., Kaliningrad 236000, Russian Federation; baltbipiran@mail.ru

Kolesnikov Alexander V. (b. 1948) — Doctor of Science in technology, Director, professor, Immanuel Kant Baltic Federal University, 14 Nevskogo Str., Kaliningrad 236041, Russian Federation; senior scientist, Kaliningrad Branch of the Institute of Informatics Problems, Russian Academy of Sciences, 5 Gostinaya Str., Kaliningrad 236000, Russian Federation; avkolesnikov@yandex.ru

Rumovskaya Sophiya B. (b. 1985) — programmer I, Kaliningrad Branch of the Institute of Informatics Problems, Russian Academy of Sciences, 5 Gostinaya Str., Kaliningrad 236000, Russian Federation; sophiyabr@gmail.com

ВОПРОСЫ ВЫБОРА МЕДИЦИНСКОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ И ОБЛАЧНЫЕ ТЕХНОЛОГИИ

Г. Я. Илюшин¹, В. И. Лиманский²

Аннотация: Рассмотрены вопросы выбора медицинских информационных систем (МИС) для автоматизации деятельности лечебно-профилактических учреждений (ЛПУ) на основании их характеристик и соответствия МИС требованиям закона РФ «О персональных данных» и ведомственным приказам по взаимодействию с единой государственной информационной системой в сфере здравоохранения (ЕГИСЗ). Наряду с традиционными критериями выбора МИС по их функциональным характеристикам, по критериям клинической, организационной и экономической эффективности рассмотрены вопросы выбора МИС по критериям эффективности применения облачных реализаций МИС для конкретных видов ЛПУ. Приведен анализ проблем внедрения в ЛПУ МИС облачных реализаций, показаны их недостатки и преимущества, даны рекомендации о целесообразности применения облачных реализаций в зависимости от назначения и структуры лечебного учреждения. Рекомендовано применение облачных МИС для автоматизации лечебных учреждений поликлинического типа, где их применение может дать заметный экономический эффект и повысить качество обслуживания населения.

Ключевые слова: медицинские информационные системы; критерии выбора МИС; ЛПУ; облачные технологии; единая государственная информационная система в сфере здравоохранения; ЕГИСЗ

DOI: 10.14357/08696527140111

1 Рынок медицинских информационных систем для автоматизации деятельности медицинских организаций

Динамика рынка МИС показывает постоянно возрастающий интерес медицинских организаций (МО) лечебно-профилактического типа к автоматизации процессов их деятельности [1], что обусловлено желанием обеспечить максимальную эффективность лечебного процесса с одновременным сокращением затрат. По данным Министерства здравоохранения РФ [2] число только государственных ЛПУ составляло на конец 2010 г. 8960 единиц, из них больниц и диспансеров, имеющих стационары, — 5705. По информации, приведенной в [3], на начало 2012 г. в России выполнены поставки МИС в 4930 ЛПУ всех форм собственности. Динамика поставок МИС за 2005–2011 гг., по данным того же источника,

¹Институт проблем информатики Российской академии наук, ilushin@ipiran.ru

²Институт проблем информатики Российской академии наук, vlimansky@ipiran.ru

Динамика поставок МИС за 2005–2011 г. [3]

Год наблюдения	Количество ЛПУ, в которые были поставлены МИС	Число автоматизированных рабочих мест
2005	206	6 тыс.
2007	1000	34,8 тыс.
2009	2800	53,9 тыс.
2010	2440	52,3 тыс.
2011	4930	104,2 тыс.

отражена в таблице. Приведенные цифры свидетельствуют о нарастании темпа автоматизации ЛПУ и большом потенциале рынка МИС.

Число систем автоматизации деятельности МО, предлагаемых на рынке, достаточно велико. В каталоге ассоциации развития медицинских информационных технологий (АРМИТ) [4] в разделе «Медицинские информационные системы (МИС)» в рубриках «МИС многопрофильных медучреждений (поликлиники, диагностические центры, стационары, клиники и др.)», «МИС для стоматологических медучреждений» и «МИС для специализированных медучреждений и подразделений» на данный момент зарегистрировано 90, 14 и 16 медучреждений соответственно.

В публикации [5] приведены результаты классификации 190 МИС всех типов и назначений по видам решаемых задач (рис. 1). Сведения о МИС,

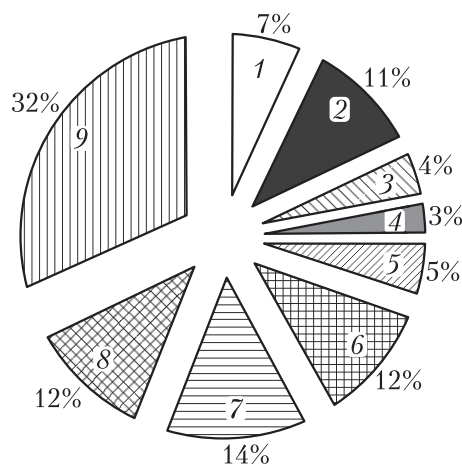


Рис. 1 Классификация МИС по видам решаемых задач [5]: 1 — поликлинические; 2 — лабораторные; 3 — стоматологические; 4 — обработка изображений; 5 — интернет-порталы и разработка МИС; 6 — управление здравоохранением; 7 — организационно-экономические; 8 — комплексные; 9 — лечебно-диагностические

изученных и проанализированных в ходе указанного исследования, размещены в виде таблицы [6] на сайте телемедицинского центра Тамбовского государственного технического университета, где указаны названия МИС, их разработчики, адреса в Интернете и краткие описания каждой МИС.

Приведенные сведения показывают разнообразие и большое число МИС, предлагаемых на рынке для автоматизации МО. Помимо перечисленных традиционных реализаций МИС на рынке автоматизации ЛПУ появился новый игрок в виде облачных МИС, реализуемых в рамках создания ЕГИСЗ и предлагаемых Министерством здравоохранения для внедрения в ЛПУ. Облачные МИС могут размещаться как на площадке федерального центра обработки данных (ЦОД), так и на площадках региональных ЦОД ЕГИСЗ. Выбор конкретной МИС для размещения на региональном уровне ЕГИСЗ осуществляется органами управления здравоохранением соответствующих регионов.

2 Критерии оценки традиционных реализаций медицинских информационных систем

Для выбора МИС можно, на первый взгляд, воспользоваться рекламной информацией, размещаемой разработчиками на своих сайтах и в рекламных публикациях. Однако представленные в таких материалах данные носят поверхностный характер, в мажорных тонах описывают достоинства систем и не раскрывают их реальные характеристики.

Другим подходом к выбору МИС является сбор и анализ информации о результатах внедрения МИС в других ЛПУ. Для оценки эффективности применения МИС предлагается использовать интегральные критерии — клиническую, организационную и экономическую эффективность внедрения [7]. В качестве критериев **клинической эффективности** эксперты называют такие показатели, как уменьшение количества врачебных ошибок при назначении препаратов и направлении на обследование, повышение точности, оперативности и информативности диагностических исследований, уменьшение числа обострений хронических заболеваний за определенный период времени, общее снижение заболеваемости, повышение степени соответствия лечения установленным стандартам. Наиболее адекватным показателем **организационной эффективности** системы, по данным того же источника, большинство экспертов считают уменьшение затрат рабочего времени медицинского персонала при подготовке отчетной документации. **Экономическая эффективность** выражается в росте доходов ЛПУ, в экономии фонда заработной платы, оптимизации расходов на медикаменты, в возможности руководства ЛПУ контролировать работу учреждения в реальном времени. Описанный подход интуитивно понятен, ведь в конечном счете внедрение МИС в ЛПУ преследует цели, отраженные в интегральных критериях эффективности. Однако цифровое выражение перечисленных критериев получить довольно трудно.

Более простым и доступным способом выбора предпочтительной МИС является оценка МИС по определенному набору критериев. Такой подход к выбору МИС позволяет при сравнительно небольших затратах времени и средств выбрать из множества информационных систем несколько наиболее подходящих. Затем следует провести более детальный анализ систем-претендентов, применяя критерии интегрального анализа. При анализе систем следует обращать внимание не на перечень функций и возможностей, а на качество, глубину и гибкость реализованных решений.

Набор критериев для оценки МИС достаточно широко освещается в различных публикациях [7–9]. Среди множества предлагаемых критериев можно выделить следующие группы:

1. Функциональные критерии — наличие и качество реализации бизнес-процессов в ЛПУ, в том числе амбулаторное и стационарное лечение, медикаментозное обеспечение, учет предоставленных по различным видам оплаты медицинских услуг, отчетность и т. д.
2. Стоимость МИС — затраты на приобретение, в том числе общесистемного программного обеспечения, на внедрение, эксплуатацию и техническую поддержку системы.
3. Взаимодействие с другими системами — системами автоматизации других ЛПУ, органов здравоохранения, поставщиками лекарственных средств и медицинских препаратов.
4. Администрирование МИС — управление системой, в том числе правами доступа пользователей к данным и функциям системы, аудит действий пользователей, резервное копирование и восстановление, обновление версий серверного и клиентского программного обеспечения.

Помимо оценки МИС по критериям ожидаемой эффективности внедрения и функциональным характеристикам следует обращать внимание на соответствие МИС требованиям ряда нормативных документов, в частности Федеральному закону № 152 «О персональных данных», и интеграции МИС с ЕГИСЗ.

Федеральный закон «О персональных данных» регулирует отношения, связанные с обработкой персональных данных, осуществляемой, в частности, юридическими лицами с использованием средств автоматизации. Все информационные системы, разработанные до вступления данного закона в силу, должны были быть приведены в соответствие с законом до 1 июля 2011 г. Необходимым условием соответствия системы требованиям защиты персональных данных является наличие модели угроз безопасности персональных данных [10]. Модель угроз формируется и утверждается оператором персональных данных в соответствии с методическими документами, разработанными в соответствии с п. 2 Постановления Правительства Российской Федерации от 17 ноября 2007 г. № 781 «Об утверждении Положения об обеспечении безопасности персональных данных при их обработке в информационных системах персональных данных».

В процессе внедрения системы в ЛПУ должна быть сформирована и утверждена Частная модель угроз [11, 12] и совместно с поставщиком МИС разработана документация, необходимая для эксплуатации систем обработки персональных данных.

Концепция создания ЕГИСЗ [13] предусматривает создание в ее составе ряда федеральных и региональных сервисов, предназначенных для поддержки деятельности ЛПУ. Медицинская информационная система любого ЛПУ должна в обязательном порядке взаимодействовать с рядом централизованных систем и сервисов ЕГИСЗ, в числе которых:

- централизованная система электронной записи на прием к врачу (электронная регистратура — ЭР);
- централизованная система ведения интегрированной электронной медицинской карты (ИЭМК);
- централизованная система выдачи и учета выписанных льготных рецептов, обеспечивающая персонифицированный учет оказанной медицинской помощи и лекарственного обеспечения;
- подсистема административно-хозяйственной деятельности (АХД);
- подсистема доступа к каталогу пользователей системы;
- подсистема доступа к нормативно-справочной информации и словарям медицинских терминологий;
- инфраструктура открытых ключей и электронной цифровой подписи.

По мере разработки и внедрения перечисленных, а также других систем и сервисов ЕГИСЗ все МИС должны быть модернизированы в соответствии с требованиями ЕГИСЗ. Поэтому при выборе МИС следует обратить внимание на текущее состояние МИС по обеспечению взаимодействия с ЕГИСЗ и на возможности поставщика МИС реализовать эти требования ЕГИСЗ в дальнейшем.

3 Выбор медицинской информационной системы облачной реализации

Оценка функциональных характеристик МИС облачной реализации и степени ее соответствия нуждам ЛПУ может осуществляться по критериям оценки традиционных МИС, размещаемых на площадках ЛПУ. Особое внимание при выборе для внедрения в ЛПУ МИС облачной реализации следует уделять тем преимуществам, а также тем проблемам и рискам, которые присущи облачным решениям. В широком смысле, переход к облачным технологиям означает аутсорсинг традиционных процессов управления информационно-технологическими (ИТ) инфраструктурами профессиональными внешними поставщиками. В качестве основных преимуществ использования облачных решений обычно называют:

- сокращение сроков внедрения решений на предприятии;
- снижение расходов на инфраструктуру;
- снижение затрат на программное обеспечение;
- сокращение эксплуатационных расходов;
- масштабируемость и динамическое выделение необходимых для функционирования решения ресурсов.

В случае с облачными реализациями МИС все перечисленные преимущества также имеют место, что делает, по крайней мере на первый взгляд, их внедрение в ЛПУ довольно привлекательным. Однако при этом внедрение облачных решений ставит перед ЛПУ ряд новых проблем и дополнительных рисков. Как было сказано выше, облачные технологии означают аутсорсинг процессов управления ИТ-инфраструктурой профессиональными внешними поставщиками, которые являются специалистами в своей области, но не в области предлагаемых ими прикладных решений.

Процесс внедрения МИС в ЛПУ предполагает, в частности, выполнение работ по настройке системы для данного конкретного ЛПУ. Эти работы включают: описание инфраструктуры ЛПУ и профилей методов диагностики и лечения, ввод списочного состава ЛПУ, создание и настройку рабочих мест персонала, описание прав их доступа и множество других подготовительных действий и манипуляций, необходимых для запуска МИС в эксплуатацию. Запуск МИС в эксплуатацию в любом ЛПУ предполагает также обучение персонала ЛПУ навыкам работы с данной системой. В процессе эксплуатации часто возникает необходимость в перенастройке или восстановлении работоспособности системы, нарушенной в результате эксплуатационных ошибок, изменений внешних требований или конфигурации ЛПУ. Все эти работы обычно выполнялись разработчиками системы или специально подготовленными ими группами по внедрению и сопровождению данной МИС в процессе эксплуатации, поскольку выполнение указанных работ требует знания особенностей реализации конкретной МИС и прикладной области, а не ИТ-технологий. При внедрении в ЛПУ МИС облачной реализации для выполнения перечисленных работ ЛПУ должно будет заключать дополнительные договоры на внедрение и сопровождение облачной МИС с ее разработчиками или со специальными структурами, которые должны создаваться при соответствующих ЦОД, предлагающих облачные реализации МИС для внедрения. Это приведет к удорожанию внедрения и эксплуатации МИС и в значительной степени нивелирует экономический эффект от применения облачных технологий.

Однако все эти организационные проблемы все же могут быть решены при должном внимании со стороны Министерства здравоохранения и его региональных структур. Большие опасения вызывает тот факт, что на пути внедрения облачных МИС в ЛПУ наряду с организационными проблемами существуют и технические трудности — главным образом это ограниченная пропускная способ-

ность и потенциальная ненадежность линий связи глобальной сети, связывающей ЛПУ с соответствующим облаком.

Отсутствие технических проблем при внедрении облачных реализаций МИС в ЛПУ в значительной мере зависит от предполагаемой степени автоматизации учреждения, а также от его назначения и структуры. При осуществлении начальной фазы автоматизации ЛПУ поликлинического типа, предусматривающей ведение расписания приема пациентов и автоматизацию работы регистратуры, никаких технических проблем возникать не должно, поскольку в данном случае предполагается интерактивный ввод и получение сравнительно небольших объемов данных и не требуется особо высокой надежности каналов связи. По тем же самым соображениям не вызовет осложнений и автоматизация рабочих мест врачей, если их функции ограничены ведением электронной истории болезней пациентов и/или выполнением тривиальных диагностических манипуляций (ультразвуковое исследование, цифровой рентген и т. п.). Однако даже в этих случаях на площадке ЛПУ помимо компьютеров рабочих станций все же должны быть установлены некоторые технические и программные средства, предназначенные для создания локальной сети ЛПУ и организации подключения сотрудников ЛПУ к облаку со своих рабочих мест, т. е. совсем без ИТ-инфраструктуры — сетевого и серверного оборудования — в ЛПУ обойтись не удастся, а значит, какие-то силы и средства на поддержку инфраструктуры в самом ЛПУ потратить все же придется.

Рассмотрим теперь на примере рабочих мест врачей-диагностов, чья деятельность связана с применением высокотехнологичного цифрового оборудования, возможные проблемы автоматизации их работы в рамках облачных МИС. На рис. 2 представлена структурная схема комплекса аппаратно-программных средств (АПК) радиологической информационной системы (РИС) «ГАММА МУЛЬТИВОКС» [14]. Этот или другие подобные ему комплексы применяются в крупных центрах и клиниках не только и не столько для получения и хранения отдельных изображений. Главной целью применения таких комплексов является использование компьютерных технологий для обработки, визуализации и системного анализа медицинских диагностических изображений, результатов функциональной диагностики, биофизических и биохимических методов исследования для построения виртуальной трехмерной динамической математической модели исследуемых структур и органов пациента для диагностики, планирования и оценки результатов консервативного или хирургического лечения [15].

Например, в процессе выполнения задач по диагностике и подготовке к операциям врачом-диагностом обрабатываются и сохраняются на сервере огромные объемы данных, предназначенных сугубо для внутрибольничного использования в процессе лечения больного, и лишь незначительная их часть должна впоследствии аккумулироваться в истории болезни и передаваться в сервисы ЕГИСЗ. Если серверы и программное обеспечение подобного комплекса будут вынесены в составе МИС в облако, тогда как рабочие станции и диагностическое оборудование останутся на «земле», в ЛПУ, то возникнет огромный трафик, с

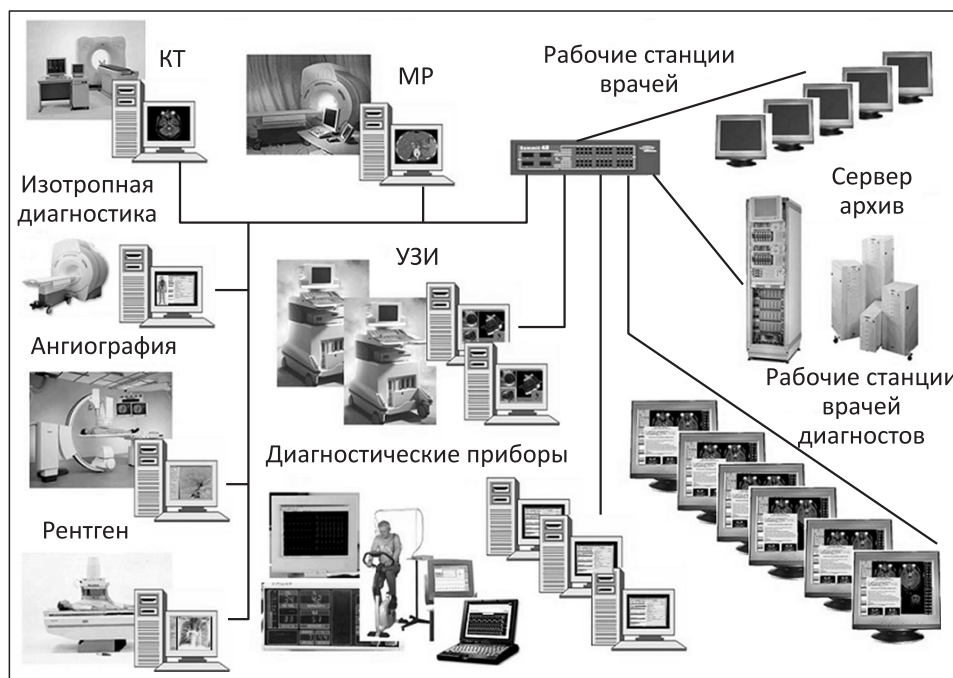


Рис. 2 Структурная схема РИС «ГАММА МУЛЬТИВОКС»

которым едва ли справятся линии связи глобальной сети, а обработка изображений резко замедлится, что абсолютно неприемлемо и неоправданно. Если же такой комплекс вместе с программным обеспечением и серверами вывести из облака и полностью разместить на площадке ЛПУ, то встанет вопрос о его сопряжении с базовой МИС в облаке. Решение такой задачи вполне реально, если надо выполнить сопряжение одной МИС с одним АПК и одной РИС. Но проблема заключается в том, что типовая облачная МИС одна, а таких АПК и соответствующих РИС может быть несколько, и в каждом конкретном случае закупку диагностического оборудования и РИС осуществляет медицинский центр, исходя из своего понимания задач, потребностей и с учетом выделенных финансовых средств. А разработать типовую МИС, которая сможет взаимодействовать с любой РИС без адаптации к конкретному оборудованию и специфике медицинской деятельности стационара, на данном этапе нереально. Кроме того, такой комплекс, размещенный в ЛПУ, надо обслуживать и им управлять, что опять-таки возвращает к необходимости аутсорсинга силами разработчиков, причем как МИС, так и РИС.

Другими примерами высокотехнологичных комплексов, в состав которых входит различное оборудование, программное и аппаратное обеспечение, явля-

ются «Цифровая операционная» — оснащение операционных блоков — и лабораторные комплексы. Программное обеспечение таких компонентов по соображениям надежности и высокого внутрикомпонентного трафика ни при каких условиях не может быть поднято в облако и размещено на серверах ЦОД в составе базовой МИС. Эти высокотехнологичные компоненты наряду с другими, в частности «хозяйственными», такими как «Бухгалтерия», «Аптека», «Питание», предлагаются различными поставщиками и также встраиваются в базовую МИС в процессе ее внедрения, т. е. для одной и той же базовой МИС набор таких компонентов в различных ЛПУ может отличаться или они вообще могут отсутствовать. Указанные компоненты как высокотехнологичные, так и хозяйственные, должны достаточно тесно взаимодействовать с находящейся в облаке базовой медицинской системой и с другими компонентами как по входу, так и по выходу, т. е. осуществлять с ними обмен данными.

Итак, при внедрении МИС в крупных медицинских центрах и клиниках в составе МИС будут присутствовать высокотехнологичные компоненты, которые по соображениям надежности и/или риска создания неприемлемого трафика ни при каких условиях не могут быть подняты в облако и должны размещаться на площадке ЛПУ в его локальной сети. При этом в локальной сети на территории ЛПУ будет установлено большое количество рабочих станций, серверов с их программным обеспечением, для сопровождения и поддержания работоспособности которого потребуются и специалисты в области ИТ-технологий. В таких условиях экономический эффект от размещения базовой части МИС в облаке на оборудовании соответствующего ЦОД практически исчезает. При этом возникают, возможно, неразрешимые проблемы по интеграции отдельно стоящих компонентов с облачной реализацией МИС, общей, по крайней мере в данном регионе, для всех ЛПУ.

Учитывая сказанное, можно утверждать, что применение облачных реализаций МИС для автоматизации деятельности крупных клиник и медицинских центров, использующих в своей работе высокотехнологичные комплексы, нецелесообразно как по экономическим, так и по техническим соображениям. Вместе с тем применение облачных технологий при автоматизации большинства лечебных учреждений поликлинического типа и больниц районного уровня, не оснащенных радиологическими комплексами и цифровыми операционными, может дать заметный экономический эффект и существенно сократить сроки их автоматизации.

4 Заключение

В статье изложены подходы к выбору МИС для внедрения в ЛПУ на основе оценки их функциональных характеристик и эффективности внедрения МИС в других ЛПУ по интегральным критериям с учетом реализации требований Федерального закона № 152 «О персональных данных» и интеграции МИС с ЕГИСЗ.

Рассмотрены преимущества, а также проблемы и риски, возникающие при выборе для внедрения в ЛПУ МИС облачной реализации. На основе анализа свойств МИС, имеющих в своем составе компоненты, управляющие функционированием высокотехнологичного медицинского оборудования, сделан вывод о нецелесообразности использования для таких МИС облачных технологий. Рекомендовано применение облачных МИС для автоматизации лечебных учреждений поликлинического типа и больниц районного уровня, где их применение может дать заметный экономический эффект, сократить сроки автоматизации МО и повысить качество обслуживания населения.

Литература

1. *Гусев А. В.* Обзор рынка комплексных медицинских информационных систем // Врач и информационные технологии, 2009. № 6. С. 4–17.
2. Ресурсы и деятельность учреждений здравоохранения. 2011 г. <http://www.rosminzdrav.ru/documents/6996-statisticheskaya-informatsiya/>.
3. *Гусев А. В.* Рынок медицинских информационных систем: обзор, изменения, тренды // Врач и информационные технологии, 2012. № 3. С. 6–15.
4. *Эльянов М. М.* Медицинские информационные технологии: Каталог. — М.: Capital Press, 2011. Вып. 11. 320 с.
5. *Фролов С. В., Маковеев С. Н., Семенова С. В., Фареа С. Г.* Современные тенденции развития рынка медицинских информационных систем // Вестник ТГТУ, 2010. Т. 16. № 2. С. 267–272.
6. Перечень основных МИС. http://www.telemed.tstu.ru/images/my/Tabl_MIS.doc.
7. *Лапрун И.* Эффективность внедрения медицинских информационных систем // PC Week Doctor, 2008. № 1(1). <http://www.pcweek.ru/idea/article/detail.php?ID=108436>.
8. *Гусев А. В.* Обзор функциональных возможностей российских медицинских информационных систем // Менеджер здравоохранения, 2006. № 12. С. 22–30.
9. *Харитонов А. А., Фатин П. А.* Критерии оценки медицинских информационных систем // Врач и информационные технологии, 2008. № 5. С. 15–19.
10. Модель угроз типовой медицинской информационной системы (МИС) типового лечебно-профилактического учреждения (ЛПУ) // Информационное письмо Минздрава № 328-29 от 5 марта 2010 г.
11. Методические рекомендации по составлению Частной модели угроз безопасности персональных данных при их обработке в информационных системах персональных данных учреждений здравоохранения, социальной сферы, труда и занятости // Рекомендации Минздрава от 24 декабря 2009 г. 11 с.
12. Методические рекомендации для организации защиты информации при обработке персональных данных в учреждениях здравоохранения, социальной сферы, труда и занятости // Рекомендации Минздрава от 24 декабря 2009 г.
13. Концепция создания Единой государственной информационной системы в сфере здравоохранения // Приложение к приказу Министерства здравоохранения и социального развития Российской Федерации от 28 апреля 2011 г. № 364.

14. Автоматизированная радиологическая информационная система MULTIVOX (АРИС MULTIVOX) // Медицинская техника для лучевой диагностики: Справочник. — М.: Интелфорум, 2004. С. 124–126.
15. Гаврилов А. В. 3D цифровой пациент. <http://i-russia.ru/media/files/41d3304b0f52b2eabbdf.pdf>.

Поступила в редакцию 30.09.13

QUESTIONS OF CHOOSING A MEDICAL INFORMATION SYSTEM AND CLOUD TECHNOLOGIES

G. I. Ilyushin and V. I. Limansky

Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str.,
Moscow 119333, Russian Federation

Abstract: The article covers the questions of choosing medical information systems (MIS) for automation of medical-prophylactic institutions work on the basis of their characteristics and their compliance with the MIS statutory requirements of the Russian Federation “About personal data” and their compliance with departmental orders regarding interaction with the unified state information system in the field of healthcare. Along with the traditional criteria of MIS chosen based on their functional characteristics and clinical, organizational, and economic efficacy criteria, the problems of choosing MIS in accordance with criteria of efficient use of “cloud” MIS realization for the specific kinds of medical-prophylactic institutions are considered. The article gives the analysis of application of cloud realizations in MIS of medical-prophylactic institutions, shows their merits and demerits, and gives recommendations on expediency of cloud realization depending on medical institution structure and specialization. This article recommends to use cloud MIS for automation of medical institutions like polyclinics where their use can give visible economic effect and raise the quality of medical service.

Keywords: medical information systems; criteria of MIS choice; medical-prophylactic institutions; cloud technologies; unified state information system in the field of healthcare

DOI: 10.14357/08696527140111

References

1. Gusev, A. V. 2009. Obzor rynka kompleksnykh meditsinskikh informatsionnykh sistem [Review of the market of complex medical information systems]. *Vrach i Informatsionnye Tekhnologii* [Doctor and Information Technologies] 6:4–17.
2. Resursy i deyatel'nost' uchrezhdeniy zdravookhraneniya [Resources and activity of healthcare institutions]. Available at: <http://www.rosminzdrav.ru/documents/6996-statisticheskaya-informatsiya/> (accessed February 12, 2014).

3. Gusev, A. V. 2012. Rynok meditsinskikh informatsionnykh sistem: Obzor, izmeneniya, trendy [Market of medical information systems: Review, changes, trends]. *Vrach i Informatsionnye Tekhnologii* [Doctor and Information Technologies] 3:6–15.
4. El'yanov, M. M. 2011. *Meditsinskie informatsionnye tekhnologii: Katalog* [Medical information technologies: Catalog]. Moscow: CapitalPress. Release 11. 320 p.
5. Frolov, S. V. S. N. Makoveev, S. V. Semenova, and S. G. Farea. 2010. Sovremennye tendentsii razvitiya rynka meditsinskikh informatsionnykh sistem [Current trends of development of the market of medical information systems]. *Vestnik TGTU* [TGTU Bulletin] 16:267–272.
6. Perechen' osnovnykh MIS [List of the main MIS]. Available at: http://www.telemed.tstu.ru/images/my/Tabl_MIS.doc (accessed February 12, 2014).
7. Laprun, I. 2008. Effektivnost' vnedreniya meditsinskikh informatsionnykh sistem [Efficiency of introduction of medical information systems]. *PC Week Doctor* 1(1). Available at: <http://www.pcweek.ru/idea/article/detail.php?ID=108436> (accessed February 12, 2014).
8. Gusev, A. V. 2006. Obzor funktsional'nykh vozmozhnostey rossiyskikh meditsinskikh informatsionnykh sistem [Review of functionality of the Russian medical information systems]. *Menedzher Zdravookhraneniya* [Manager of Health Care] 12:22–30.
9. Kharitonov, A. A., and P. A. Fatin. 2008. Kriterii otsenki meditsinskikh informatsionnykh sistem [Criteria of an assessment of medical information systems]. *Vrach i Informatsionnye Tekhnologii* [Doctor and Information Technologies] 5:15–19.
10. Model' ugroz tipovoy meditsinskoy informatsionnoy sistemy (MIS) tipovogo lechenno-profilakticheskogo uchrezhdeniya (LPU) [Model of threats of the standard medical information system (MIS) of the standard treatment-and-prophylactic establishment (TPE)]. Informatsionnoe pis'mo Minzdrava No.328-29 ot 5 marta 2010 g. [The information letter of Ministry of Health 328-29 of March 5, 2010].
11. Metodicheskie rekomendatsii po sostavleniyu Chastnoy modeli ugroz bezopasnosti personal'nykh dannykh pri ikh obrabotke v informatsionnykh sistemakh personal'nykh dannykh uchrezhdeniy zdravookhraneniya, sotsial'noy sfery, truda i zanyatosti [Methodical recommendations about drawing up Private model of threats of safety of personal information at their processing in information systems of personal information of healthcare institutions, social sphere, work and employment]. Rekomendatsii Minzdrava ot 24 dekabrya 2009 g [Recommendations of Ministry of Health of December 24, 2009].
12. Metodicheskie rekomendatsii dlya organizatsii zashchity informatsii pri obrabotke personal'nykh dannykh v uchrezhdeniyakh zdravookhraneniya, sotsial'noy sfery, truda i zanyatosti [Methodical recommendations for the information security organization when processing personal information in healthcare institutions, social sphere, work and employment]. Rekomendatsii Minzdrava ot 24 dekabrya 2009 g. [Recommendations of Ministry of Health of December 24, 2009].
13. Kontseptsiya sozdaniya edinoy gosudarstvennoy informatsionnoy sistemy v sfere zdravookhraneniya [The concept of national information system creation in the health care sphere]. Prilozhenie k prikazu Ministerstva zdravookhraneniya i sotsial'nogo razvitiya Rossiyskoy Federatsii ot 28 aprelya 2011 g. No. 364 [The annex to the order of the Ministry of Health and Social Development of the Russian Federation of April 28, 2011. No. 364].

14. Avtomatizirovannaya radiologicheskaya informatsionnaya sistema MULTIVOX (ARIS MULTIVOX) [The automated radiological information MULTIVOX system (ARIS MULTIVOX)]. 2004. *Meditsinskaya Tekhnika dlya Luchevoy Diagnostiki* [Medical Equipment for Radiodiagnostics]. Moscow: Intelforum. 124–126.
15. Gavrilov, A. V. 3D tsifrovoy patsient [3D digital patient]. Available at: <http://i-russia.ru/media/files/41d3304b0f52b2eabbdf.pdf> (accessed February 12, 2014).

Received September 30, 2013

Contributors

Ilyushin Gennadiy Y. (b. 1947)— Candidate of Science (PhD) in technology, Head of Laboratory, Institute of Information Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; ilushin@ipiran.ru

Limansky Valery I. (b. 1952)— Candidate of Science (PhD) in technology, senior scientist, Institute of Information Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Moscow, Russian Federation; vlimansky@ipiran.ru

ВНЕДРЕНИЕ И ИСПОЛЬЗОВАНИЕ В УНИВЕРСИТЕТЕ ТЕХНОЛОГИИ ВИРТУАЛИЗАЦИИ РАБОЧИХ МЕСТ

*М. А. Вышиванов¹, Д. В. Гмарь², В. В. Крюков³, К. С. Нагорняк⁴,
К. И. Шахгельдян⁵*

Аннотация: Рассмотрено решение задачи виртуализации компьютерных рабочих мест университета. Сделан анализ и обоснован выбор технологии нулевых клиентов с микросхемой Teradici и программного обеспечения (ПО) VMware View. Описана облачная инфраструктура университета, представлены результаты реализации проекта по внедрению технологии виртуализации рабочих мест на основе VMware View и нулевых клиентов (330 рабочих мест), показаны преимущества применения этих решений, сделана оценка эффективности использования нулевых клиентов в университете. При использовании технологии нулевых клиентов стоимость владения рабочим местом за 5 лет снижается в 1,5 раза, при этом значительно уменьшаются энергопотребление и уровень шума, повышаются сохранность данных, скорость развертывания новых рабочих мест и обновления ПО, увеличивается эффективность загрузки аудиторного фонда университета.

Ключевые слова: виртуализация рабочих мест; нулевые клиенты; облачные технологии

DOI: 10.14357/08696527140112

1 Введение

Современный университет, так же как и любая крупная организация, рано или поздно сталкивается с ситуацией, когда информационная инфраструктура (ИТ-инфраструктура) достигает таких размеров и такой сложности, что на ее поддержку уходит большой объем временных, материальных и трудовых ресурсов. Проблемы, с которыми сталкивается крупная организация, имеющая развитую ИТ-инфраструктуру, при поддержании жизненного цикла оборудования и ПО можно сформулировать следующим образом:

¹Владивостокский государственный университет экономики и сервиса, maksim.vyshivanov@vvsu.ru

²Владивостокский государственный университет экономики и сервиса, dimer@vvsu.ru

³Владивостокский государственный университет экономики и сервиса, kryukov@vvsu.ru

⁴Владивостокский государственный университет экономики и сервиса, konstantin.nagornyak@vvsu.ru

⁵Владивостокский государственный университет экономики и сервиса, carinash@vvsu.ru

- значительные расходы на обновление и сопровождение применения ПК (ремонт, модернизацию, электроэнергию, поддержку климата и т. п.);
- рост количества ПК (что особенно свойственно университетам) приводит к увеличению штата ИТ-специалистов;
- ИТ-специалисты вынуждены тратить большую часть своего рабочего времени на поддержку существующей инфраструктуры, а не на ее развитие;
- невысокая эффективность использования вычислительных ресурсов на рабочих местах (вычислительные ресурсы ПК используются не более чем на 10%);
- необходимость оперативного оснащения компьютерами и приложениями большого количества новых рабочих мест;
- необходимость копирования данных из разных источников на переносной носитель для работы вне основного рабочего места;
- проблема обеспечения корпоративной мобильности лицензионного ПО — оперативного переноса ПО на определенные компьютеры (для университета, например, — в определенные учебные классы, которые указаны в расписании занятий);
- несанкционированное разукomплектование (или подмена комплектующих) компьютеров на рабочих местах;
- периодические потери файлов на ПК из-за выхода из строя жестких дисков.

Высокие темпы роста системных требований к персональному рабочему месту, увеличение числа рабочих мест, тарифов на электроэнергию, стоимости труда ИТ-специалистов ставят вопрос о таких изменениях в ИТ-инфраструктуре, которые обеспечивали бы не только сокращение издержек, но и повышение эффективности применения информационных технологий на корпоративном уровне в длительной перспективе.

Решение вышеописанных проблем возможно на основе применения технологий виртуализации применения персональных компьютеров на рабочих местах сотрудников и учебных местах в компьютерных классах (Virtual Desktop Infrastructure, VDI). Внедрение и использование технологии VDI, с одной стороны, ставит перед университетами сложные организационные и технические задачи, с другой — обеспечивает эффективность развития и эксплуатации ИТ-инфраструктуры. Авторами в 2012–2013 гг. осуществлено внедрение VDI во Владивостокском государственном университете экономики и сервиса (ВГУЭС). Проблемы и результаты проекта внедрения и использования VDI обсуждаются в этой работе.

2 Внедрение виртуальных рабочих мест в университетах и используемые технологии

Университет Беркли (США) одним из первых в 2008–2009 гг. реализовал пилотный проект по тестированию технологии виртуализации VDI [1]. Технология VDI является серверо-центричной вычислительной моделью, естественным продолжением модели тонкого клиента. Основные преимущества получены там, где рабочие места могут быть стандартизованы при соблюдении баланса между ценой и производительностью. В проекте университета Беркли выбор осуществлялся между технологиями VMware, Citrix, MokaFive. Проект реализовывался на основе VMware View. В качестве клиента использовались различные тонкие клиенты от Wyse с операционными системами Wyse ThinOS или Windows XP Embedded. Основные проблемы, выявленные при реализации проекта, состояли в следующем:

- сложности при создании новых виртуальных машин в пуле;
- потери профайлов после рекомпозиции;
- медленный вход в систему, что вело к разрыву сессий;
- удаление виртуальных машин иногда требовало слишком много времени;
- необходимость устанавливать VMware Infrastructure Client (VIC) на каждое рабочее место для административного доступа к виртуальной машине, что требует дополнительного обучения, а также более аккуратного применения учетных записей Active Directory (AD), чтобы избежать нерегламентированного доступа;
- проблемы с производительностью в части развертывания пулов и их изменений;
- проблемы связи с диском виртуальных машин;
- проблемы с флеш-драйверами, не все из которых работали с VIC;
- проблемы входа в систему пользователей, когда около 10% пользователей не могли войти в систему по неопределенным причинам;
- проблемы выбора пула, если у пользователя были права на несколько пулов;
- нехватка памяти и проблемы подключения оптического диска у тонких клиентов Wyse.

В заключении работы [1] сделан вывод о недостаточной зрелости технологии и предложено отложить внедрение.

В Стэнфордском университете рассматривают технологию VDI как механизм сокращения стоимости владения оборудованием и ПО, как средство обеспечения доступности рабочего места из любой точки [2]. В то же время отмечается высокая стоимость серверного оборудования, возможность использования одного сервера лишь для 30–80 виртуальных рабочих мест.

В Южно-Уральском государственном университете завершен проект по созданию инфраструктуры виртуальных рабочих мест на базе Microsoft Hyper-V Server [3]. Инфраструктура виртуальных рабочих столов развернута на базе Citrix XenDesktop VDI Edition и продуктов Microsoft System Center, что используется в нескольких учебных компьютерных классах, планируется осуществить переход на эту платформу за 2–3 года.

В Российском государственном торгово-экономическом университете реализован проект виртуализации серверных мощностей и рабочих мест пользователей также на базе Microsoft Hyper-V Server [4]. Внедрение выполнено в главном корпусе на половине рабочих мест и в компьютерных классах. Для организации автоматизированных рабочих мест использовалась терминальная технология на основе тонких клиентов Sun Ray 2. В центральной серверной построена ферма из восьми серверов SunFire X2200M2. Система хранения данных собрана с использованием как классических дисковых массивов Sun StorageTek 2500, так и новых гибридных хранилищ SunStorage 7000. Опыт внедрения показал значительную экономию электроэнергии [5].

На текущий момент для доступа к виртуальным рабочим столам VDI существует два типа клиентов: нулевые и тонкие. Все вышеописанные примеры касались технологии тонких клиентов. В тонких клиентах прошита облегченная операционная система [6]. Для нулевых клиентов не требуется операционной системы, так как используется протокол передачи данных PCoIP (PC over IP — ПК по протоколу IP) между сервером виртуализации и терминальным устройством. Требования к пропускной способности вычислительной сети для протокола PCoIP варьируются от 64 Кбит/с до 300 Мбит/с в зависимости от решаемой на терминалах с нулевыми клиентами задачи. Это позволяет ограничиться на терминальном устройстве программируемой логической интегральной схемой типа FPGA (field-programmable gate array) или микропроцессором типа Teradici [7].

Нулевые клиенты на основе Teradici поддерживают аппаратный обработчик протокола PCoIP, поэтому они обладают высокой производительностью. Основным недостатком таких устройств является жесткая ориентированность на инфраструктуру VMware View. Устройства этого типа выпускают Wyse, LG, Samsung, Cisco. Нулевые клиенты на основе FPGA содержат только обработчик собственного протокола передачи видеоизображения. Основным недостатком является невозможность перепрограммирования FPGA и, соответственно, отсутствие обновления при выходе новых версий протоколов.

Тонкие клиенты представляют собой миникомпьютеры со встроенной операционной системой и поддерживают несколько протоколов (PCoIP, RDP (Remote Desktop Protocol) и др.). В качестве недостатка тонких клиентов нужно отметить низкую по сравнению с нулевыми клиентами производительность (особенно в части графики и воспроизведения видеопотока) из-за того, что тонкие клиенты выполняют программное декодирование, а нулевые — аппаратное [8]. Производителями тонких клиентов являются такие компании, как Wyse, HP и др.

Помимо VMware на рынке виртуализации рабочих мест есть решения Citrix, Microsoft, NComputing. Для управления виртуальными рабочими местами можно использовать, например, Citrix Xenon Desktop [9]. К преимуществам Citrix следует отнести возможность работы с разными гипервизорами (VMware Virtual Infrastructure и Microsoft Hyper-V). К недостаткам — сложность установки и настройки, отсутствие встроенной поддержки пулов, в которых виртуальная машина автоматически закрепляется за пользователем, высокую стоимость ПО. Недостатком решений Microsoft является весьма ограниченный список поддерживаемых операционных систем, отсутствие полноценной «живой» миграции виртуальных машин и в целом менее готовый по сравнению с VMware продукт [9].

К указанным выше преимуществам VMware можно добавить значительно меньший размер гипервизора по сравнению с Microsoft, большое число поддерживаемых гостевых операционных систем, полноценную «живую» миграцию, простое администрирование и то, что VMware — надежно функционирующий продукт. К недостаткам VMware нужно отнести достаточно высокую стоимость ПО.

3 Виртуализация рабочих мест в университете

3.1 Информационная инфраструктура университета

В 2012 г. во ВГУЭС была разработана концепция Электронного кампуса, представленная в работе [10]. На основе разработанной концепции осуществлен переход к частной облачной инфраструктуре, что вызвано необходимостью модернизации вычислительной и сетевой инфраструктуры в соответствии с новыми требованиями: повышением качества сервиса при эксплуатации информационных систем и сервисов Электронного кампуса (надежности, производительности выполнения запросов), снижением эксплуатационных затрат на использование серверного оборудования и клиентских рабочих мест, уменьшением временных затрат на развертывание новых рабочих станций и прикладного ПО, упрощением управления лицензиями на ПО, увеличением пропускной способности межсерверных коммуникаций в центре обработки данных (ЦОД), а также между ЦОД и коммутаторами уровня распределения, улучшением управляемости и безопасности вычислительной сети.

Технологической основой Электронного кампуса является ЦОД, обеспечивающий надежность функционирования и требуемую производительность систем и сервисов корпоративной информационной среды (КИС) вуза.

К концу 2012 г. во ВГУЭС выполнена модернизация транспортной сети передачи данных, что потребовало внедрения нового коммутатора ядра сети (Cisco 6509) и частичного обновления оптических линий связи. После модернизации ЦОД все физические серверы включены непосредственно в ядро сети, что обеспечивает неблокируемую коммутацию на скорости 1 Гбит/с и позво-

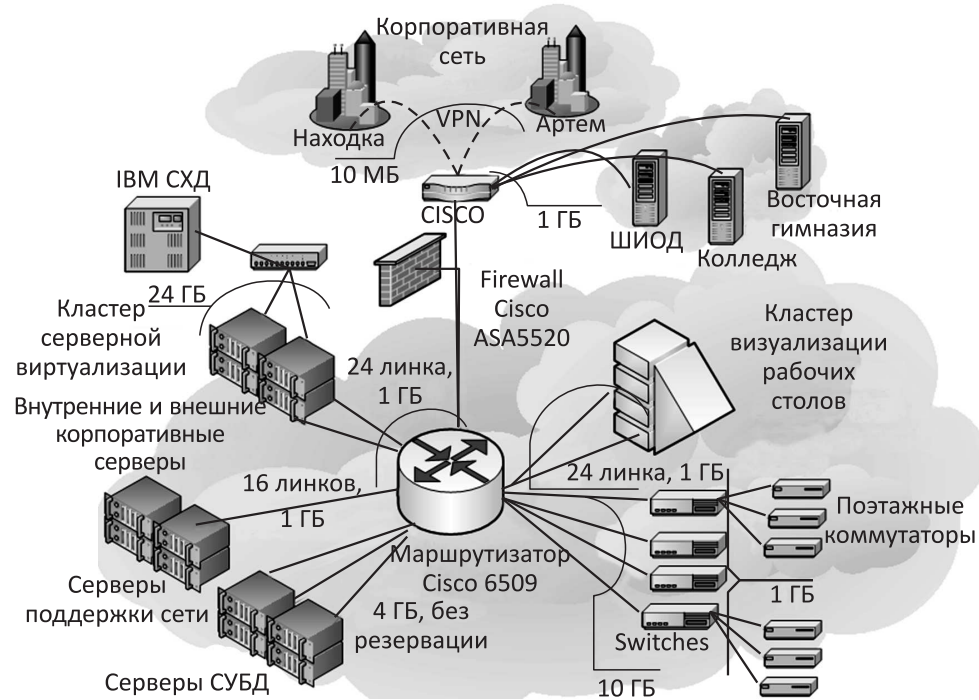


Рис. 1 Архитектура ЦОД ВГУЭС

ляет расширяться до 10 Гбит/с. Построены два кластера: кластер серверной виртуализации и кластер виртуализации рабочих мест (рис. 1).

В вычислительной инфраструктуре осуществлен переход от виртуализации серверов к созданию частного облака на основе технологии VMware. В результате обеспечено повышение надежности функционирования и производительности всех сервисов и систем КИС, повышение эффективности потребления ресурсов и экономия за счет рационального распределения вычислительных мощностей по задачам в соответствии с их реальной потребностью, уменьшение времени простоя оборудования, минимизация времени развертывания систем и затрат на их сопровождение.

3.2 Выбор технологии виртуализации рабочих мест

Для развертывания виртуализации рабочих мест необходимо прежде всего выбрать технологию тонкого или нулевого клиента. Серьезным недостатком тонкого клиента является его более низкая производительность по сравнению с нулевым клиентом. Наличие встроенной операционной системы также уменьшает привлекательность технологии при использовании в университете, где предпо-

Таблица 1 Сравнительные характеристики для выбора между тонким и нулевым клиентом

Характеристика	Тонкий клиент	Нулевой клиент
1. Производительность (количество воспроизводимых пикселей на экране в секунду), МРх	5–8 (программное декодирование)	50 (аппаратное декодирование)
2. Встроенная операционная система	Обязательна	Отсутствует
3. Приемлемый срок эксплуатации	5 лет	10 лет
4. Потребление электроэнергии	65 Вт/ч + монитор	6 Вт/ч (29 Вт/ч с монитором)
5. Наличие заменяемых комплектующих	Да	Нет
6. Риск выхода из строя	Средний	Низкий
7. Скорость развертывание новых рабочих мест	Средняя	Высокая
8. Поддержка протоколов	Любые	RSoIP
9. Возможность получения питания через коммутатор (технология POE)	Нет	Есть
10. Стоимость, руб.	25 000 (с монитором)	19 500 (с монитором)

чительнее полное централизованное управление. Сопоставление характеристик для выбора тонкого и нулевого клиента приведено в табл. 1.

Несмотря на то что тонкие клиенты имеют большую гибкость по сравнению с нулевыми клиентами в части доступных протоколов, выбор в пользу нулевого клиента очевиден.

Выбор между нулевыми клиентами Teradici и Pano Logic (FPGA) осуществлялся с учетом производительности и возможности обновления. Эксперименты проводились с устройствами обоих типов и позволили сделать однозначный выбор в пользу технологии на микропроцессорах типа Teradici, несмотря на более низкое энергопотребление устройств, работающих на базе FPGA (табл. 2).

Таблица 2 Сравнительные характеристики для выбора между Teradici и Pano Logic

Характеристики	Teradici	Pano Logic
1. Производительность	Высокая	Средняя
2. На какой базе построены	Процессор Tera 1	FPGA
3. Обновление прошивки	Возможно	Невозможно
4. Привязка к определенной технологии виртуализации	Есть (VMware)	Нет
5. Энергопотребление, Вт/ч	6,5	5
6. Плавность в отображении	Высокая	Низкая

В целом по результатам анализа и экспериментов выбрана технология VDI с заменой ПК нулевыми клиентами на микропроцессах типа Teradici.

Выбор технологии гипервизора и управления виртуальными рабочими местами основывался, во-первых, на том, что виртуализация в ЦОД ВГУЭС построена на основе VMware, во-вторых, на сравнении гипервизоров (табл. 3).

Выбор технологии VMware View определялся, во-первых, поддержкой технологии нулевых клиентов (на момент внедрения — только у VMware View). Во-вторых, использование VMware протоколов RCoIP и HTML Access позволяет существенно снизить требования к пропускной способности каналов связи и обеспечить работу в корпоративной сети из удаленных точек. В-третьих, инфраструктура Электронного кампуса вуза объединяет большое число серверов с различными операционными системами, для чего лучше приспособлена технология VMware, которая поддерживает большее число гостевых операционных систем. В-четвертых, VMware View удобнее в администрировании. Кроме того, поддержка передачи и обработки видеоинформации, автоматизация развертывания рабочих станций у VMware реализована лучше.

Таким образом, для внедрения во ВГУЭС облачной инфраструктуры выбрана технология VMware View с нулевыми клиентами на основе микропроцессора Teradici.

3.3 Внедрение технологии виртуализации рабочих мест

Для внедрения в университете виртуальных рабочих мест на основе нулевых клиентов необходимо определиться с пилотными площадками. В университете можно выделить следующие подразделения для виртуализации рабочих мест:

- компьютерный центр с несколькими аудиториями, в которых студенты различных специальностей выполняют лабораторные и практические задания на компьютерах;
- компьютерные классы кафедр и других учебных подразделений, в которых студенты близких направлений и специальностей выполняют лабораторные и практические задания на компьютерах;
- библиотека, в которой рабочие места нужны для доступа к информационным ресурсам;
- рабочие места преподавателей и сотрудников кафедр, где рабочее место может иметь многопользовательский доступ;
- рабочие места сотрудников университета, которые в однопользовательском режиме реализуют процессы с помощью систем и сервисов Электронного кампуса.

Наибольшую выгоду нулевые клиенты дают там, где используются однотипные рабочие места. Это, прежде всего, компьютерные классы в компьютерном центре, библиотека и компьютерные классы на кафедрах. Компьютерный

Таблица 3 Сравнительные характеристики для выбора между VMware, Microsoft и Citrix

Характеристика	VMware	Microsoft	Citrix
1. Используемый гипервизор в ЦОД	VMware ESX VMware ESXi	Microsoft Hyper-V	Гипервизор VMware или Microsoft
2. Объем памяти, выделяемый гипервизору, ГБ	2	10 — для Windows Server 2008 2,6 — для Windows Server Core	Зависит от выбранного гипервизора
3. Операционная система гипервизора	Встроенная	Windows 2008 x64 Standard/Enterprise/Datacenter Editions	Зависит от выбранного гипервизора
4. Поддержка нулевых клиентов	Да	Нет	Да
5. Поддерживаемые операционные системы	Microsoft Windows, Red Hat, SUSE Linux, FreeBSD, NetWare, Solaris 10	Microsoft Windows, SUSE Enterprise Linux	Microsoft Windows
6. Удобство администрирования	Высокое	Низкое	Зависит от выбранного гипервизора
7. Живая миграция*	Есть	Частично (во время миграции нельзя пользоваться виртуальной машиной, но данные не теряются)	Есть
8. Протоколы соединения клиента с сервером	RDP, PCoIP, HTML Access	RDP	HDX
9. Требования к пропускной способности сети	Зависит от выбранного протокола (для PCoIP — невысокие, для HTML Access — низкие)	Высокие	Невысокие

*Живая миграция — перенос виртуальной машины с одного физического сервера на другой без прекращения работы виртуальной машины и остановки сервисов.

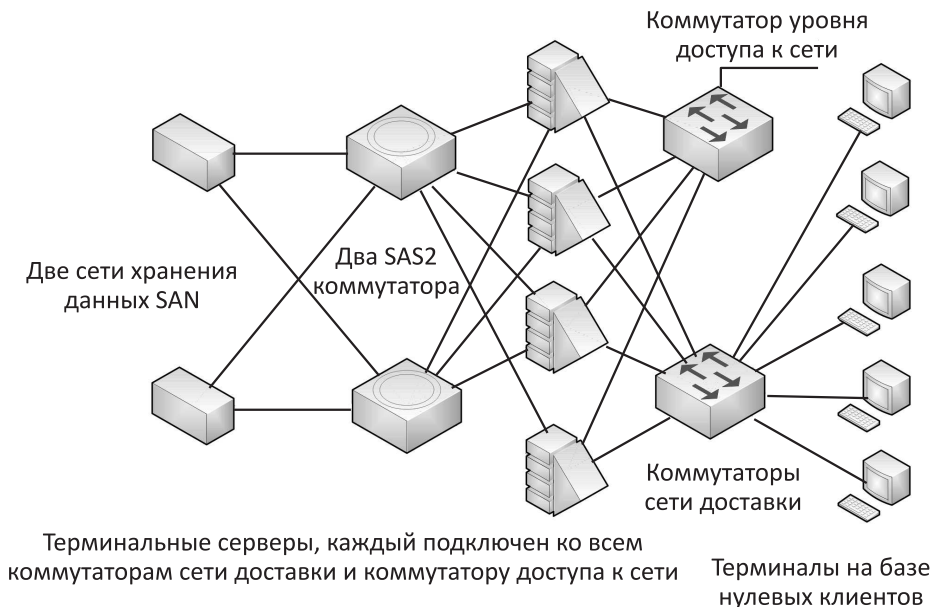


Рис. 2 Схема подключения терминалов для виртуализации рабочих мест

центр — это место, где требуется большое количество однотипных рабочих мест с часто меняющимся прикладным ПО и повышенным риском потери работоспособности приложений из-за неумелых действий пользователей. Поэтому на первом этапе реализации проекта в 2012 г. во ВГУЭС переоборудованы 4 компьютерных класса, где установлены нулевые клиенты с процессором Tera 1, встроенным в терминалы Samsung NC240 с диагональю 23,6" (90 учебных мест). Руководствуясь этим успешным опытом виртуализации учебных мест, в 2013 г. внедрено еще 240 нулевых клиентов (терминалы LG на базе процессора Teradici 2: 23CAV42K LG CAV42K V-Series Zero Client Cloud Monitor). Терминалы LG отличаются от Samsung наличием большего числа USB-портов, IPS-матрицей и новым, в несколько раз более производительным, чем Tera 1, процессором Tera 2.

Терминалы, поддерживающие технологию нулевых клиентов, установлены не только в компьютерных классах компьютерного центра и кафедр, но и в библиотеке и на рабочих местах персонала университета. Терминалы подключены к кластеру виртуализации рабочих мест. Кластер виртуализации рабочих мест подключен к двум хранилищам данных выделенной сети хранения в соответствии с архитектурой SAN через два коммутатора SAS2 (рис. 2).

Надежность функционирования нулевых клиентов обеспечивается дублированием связей и узлов (кластер подключен к двум коммутаторам, к которым с другой стороны подключены два хранилища), производительность обеспечивается дублированием каналов связи. Выход из строя любого узла в используемой

Таблица 4 Развертывание нулевых клиентов в университете

Место развертывания	Количество
1. Компьютерный центр	220
2. Классы кафедр и подразделений	50
3. Библиотека	30
4. Рабочие места сотрудников	30
Итого	330

схеме подключения не приведет к перебоям в функционировании инфраструктуры Электронного кампуса университета.

На серверной стороне виртуализацию рабочих мест обеспечивают четыре гипервизора, ОЗУ 2 ТБ, 256 ядер (AMD), 4 карты разгрузки процесса APEx2800, две системы хранения данных IBM DS3524 (48 × 900 ГБ), LSI SAS6160, сервер хранения профилей пользователей с дисковым массивом 72 ТБ.

На текущий момент во ВГУЭС функционирует 330 терминалов с поддержкой технологии нулевых клиентов под управлением технологии WMWare (табл. 4).

3.4 Опыт развертывания и использования технологии виртуализации рабочих мест

В компьютерном центре нулевые клиенты используются большим числом студентов при высокой нагрузке. Нулевые клиенты, развернутые на первом этапе, прошли апробацию в течение года, нулевые клиенты, развернутые на втором этапе, — 1 семестра. Пользователи нулевых клиентов работают в операционной системе Windows 7 x86, используют пакет Microsoft Office 2013 и различное прикладное ПО.

Первый вопрос, который необходимо решить при развертывании нулевых клиентов, — создание базового (эталонного) образа для большого числа рабочих мест. Для учебных классов компьютерного центра и кафедр могут быть использованы следующие стратегии:

1. Один образ — одна аудитория, это позволяет обеспечить отдельные настройки и ПО для каждого учебного класса.
2. Один образ на все аудитории, это позволяет значительно упростить управление ПО в компьютерном центре и учебных классах кафедр.
3. Промежуточный вариант: два и более образов со своим набором ПО, этим обеспечивается как высокая эффективность управления ПО, так и защита от несанкционированного запуска.

Первая стратегия не удовлетворяет требованию по эффективному использованию аудиторий при составлении расписания, так как ограничивает использование специализированного ПО в определенной аудитории компьютерного центра или

кафедры. Если же не делать жесткую привязку пула к аудитории, то большой выбор пулов при загрузке значительно усложняет корректную загрузку большого числа пользователей (студенты часто ошибаются при выборе пула). Кроме того, обслуживание большого числа пулов снижает эффективность ИТ-поддержки, так как требуется много времени на обновление ПО в эталонных образах.

Второй подход, будучи свободным от этих недостатков, может создавать проблемы при использовании лицензионного ПО с конкретными лицензиями, когда студенты несанкционированно запускают такое лицензионное ПО, что приводит к проблемам его использования другими студентами. Решение этой проблемы состоит в создании второго эталонного образа (или еще нескольких эталонных образов) со специализированным ПО и удалении такого специализированного ПО с конкурентными лицензиями из базового эталонного образа. Пул виртуальных машин, созданных на основе эталонного образа со специализированным ПО, подается во время перемены перед занятием в компьютерные классы, где по расписанию проводятся занятия со специализированным ПО. Использование единственного пула, бесспорно, значительно упрощает ИТ-поддержку, но усложняет работу отдельных пользователей и не может обеспечить необходимой индивидуализации обслуживания.

Наиболее удобным является третий вариант, когда создается более одного образа, но намного меньшим числом, чем количество компьютерных классов. Например, можно обойтись 2–4 образами. Для первого базового образа предлагается использовать только базовое ПО, включая в него помимо операционной системы такие пакеты, как MS Office, и некоторые другие. Для второго и третьего образов можно выделить помимо базового ПО специализированное ПО (ПО для дизайнеров, ПО для математиков и т. п.) или использовать иной принцип формирования базового образа.

Пулы виртуальных машин создаются по технологии Linked clone (связанных клонов). Эта технология позволяет разворачивать виртуальные машины из базового образа, на который изначально ставится все необходимое ПО, и при этом экономится дисковое пространство. Без использования этой технологии созданные клоны (виртуальные машины) используют столько же дискового пространства, сколько использует эталонная машина. В рассматриваемом случае используется дополнительный диск, на котором хранится только разница между диском эталонного образа и клона. Важно также использовать ресурсные пулы, т. е. пулы, которые позволяют ограничивать потребности виртуальных машин в ресурсах в рамках одного пула. Это значительно уменьшает вероятность того, что отдельные задачи мешают работе множества виртуальных машин кластера.

Для нулевых клиентов обеспечена возможность подключения ко всей необходимой периферии: принтерам, сканерам, многофункциональным устройствам, проекторам, смарт-доскам и экранам, веб-камерам, колонкам, наушникам, съемным жестким дискам.

До последнего времени технология нулевых клиентов не позволяла в полной мере применять ее для сложных графических программ, особенно для 3D-графи-

ки. Поэтому один из классов компьютерного центра, где используется программа AutoCAD, не переведен на технологию нулевых клиентов. Такая возможность появится в ближайшее время за счет использования в кластере виртуализации рабочих мест сервера с акселераторами NVIDIA GRID K2 для пула виртуальных рабочих мест с графическими пакетами и видеообработкой. Кроме того, в учебном процессе используется несколько устаревших пакетов прикладных программ (например, Statistica), работа которых на нулевых клиентах невозможна, в то же время стоимость версии, способной работать на новой технологии, слишком велика. В данном случае рассматривается вопрос о замене на пакет с лучшим соотношением цена/качество (где в характеристики качества включено требование работы на нулевых клиентах).

Для работы в библиотеке оптимально подходит первый базовый образ, так как все необходимые специализированные библиотечные программы имеют веб-интерфейс.

Установка нулевых клиентов у сотрудников университета потребовала дополнительных настроек. Во-первых, несмотря на то что сотрудники используют различное ПО, необходимо объединить их в максимально большие группы и для этих групп развернуть пулы виртуальных машин. Основным критерием оптимизации групп является их число, а это может привести к тому, что некоторые сотрудники получают доступ к ПО, которое им не нужно. Здесь может помочь то, что персонифицированным доступом к ПО можно управлять на уровне прав учетных записей службы каталогов сервиса типа Identity Management или иными средствами управления доступом, имеющимися в наличии у ПО. Если этого недостаточно, то можно воспользоваться тем же подходом, что и для компьютерного центра: созданием нескольких эталонных образов. Во-вторых, при развертывании нулевых клиентов у сотрудников требуется дополнительное время для переноса данных пользователя и подключения периферийных устройств. Например, для установки принтера необходимо подключиться к нему с эталонного образа, установить драйверы и после этого снова развернуть виртуальные машины, причем, даже если во всей организации установлены одинаковые принтеры, подключаться придется к каждому устройству. Для повышения уровня безопасности в ближайшее время планируется сделать различные пулы для сотрудников не только по набору ПО, но и по сети, т. е. у сотрудника будет возможность получать доступ только к тем ресурсам, для использования которых у него достаточно прав.

Установка терминалов позволила сотрудникам иметь доступ к своему рабочему месту с любого терминала или ПК (через программу View Client или любой браузер вне университета). Появилась возможность доступа к своему рабочему столу с мобильных устройств на платформах iOS и Android.

Для проведения работ на сервере гипервизоры переводятся в режим поддержки (maintenance), виртуальные машины мигрируют с такого гипервизора незаметно для пользователя (возможна заморозка экрана не более чем на 5 с) и пользователь продолжает работать на другом сервере. Ресурсы серверной

группы планируются по схеме $N + 1$, и выход одного сервера не влияет на работу пользователей.

3.5 Анализ результатов внедрения и использования технологии виртуализации рабочих мест

Внедрение нулевых клиентов в университете дает следующие преимущества:

- снижается уровень шума в аудиториях и как следствие создаются гораздо более комфортные условия для обучения студентов;
- экономия электроэнергии и высвобождение источников бесперебойного питания. Например, на класс с 30 ПК требовалось 3 источника бесперебойного питания мощностью 3000 Вт, для класса же, в котором установлено 30 нулевых клиентов, достаточно одного такого источника бесперебойного питания, причем его загрузка будет около 50%–70%;
- надежность рабочих мест: терминалы практически не выходят из строя, все учебные места в компьютерном классе используются;
- надежность работы ПО, так как выполняется полное обновление виртуальных машин до исходного состояния после завершения сеанса пользователями, отсутствуют сбои в работе ПО, возникающие по вине пользователя;
- высокая скорость установки и обновления ПО. Программа устанавливается один раз на эталонный образ, а потом образ разворачивается на все необходимые виртуальные машины;
- возврат рабочей станции до эталонного образа после каждого выхода пользователя из системы (при этом документы и настройки сохраняются);
- централизованное управление инфраструктурой (терминалами и виртуальными машинами), включая настройку, обновление ПО и контроль работоспособности;
- быстрое развертывание нового рабочего места (2 человеко-часа на класс из 30 новых рабочих мест);
- независимость составления расписания от типа необходимого ПО (нужное ПО будет подано в ту аудиторию, которая указана в расписании);
- единый профиль пользователя во всех аудиториях, т. е. с какого бы терминала и к какой бы виртуальной машине в пределах инфраструктуры ни подключился пользователь, все его документы и настройки сохраняются;
- централизация хранения данных — сохранность всей информации обеспечивается серверным оборудованием, а не отдельным компьютером; выход из строя нулевого клиента не повлечет за собой потерю данных пользователя. Применяются технологии избыточного хранения: средства централизованного резервного и архивного копирования.

Установка нулевых клиентов позволила оптимизировать кадровый состав ИТ-подразделений университета. Например, число ставок низкоквалифицированного персонала компьютерного центра сокращено на 30%, при этом повышены требования к уровню профессионализма инженеров компьютерного центра, расширена их зона ответственности, так как они занимаются поддержкой работы всех нулевых клиентов университета. Количество обслуживаемых рабочих мест на одного сотрудника увеличилось в 2 раза, но на текущий момент не покрывает 100% рабочего времени ИТ-специалистов, что позволяет увеличивать производительность труда при расширении внедрения нулевых клиентов и дальше. В соответствии с планами в 2014 г. в университете будет установлено 500 новых терминалов.

Нужно отметить, что внедрение нулевых клиентов в первую очередь влечет за собой сокращение низкоквалифицированного персонала — операторов, техников. Это объясняется тем, что не требуется непосредственно управлять каждым устройством и уменьшаются перебои в работе оборудования на рабочих местах (терминал более надежен, чем ПК).

Отдельно хотелось бы выделить финансовую составляющую процесса замены ПК на нулевые клиенты. Конечно, это влечет за собой значительные затраты на закупку серверного оборудования, терминалов и ПО, но эти затраты достаточно быстро окупаются за счет экономии. Авторам удалось добиться значительной экономии за счет внедрения нулевых клиентов по следующим причинам:

- зарплата персонала — сокращение ставок дало экономию около 1,2 млн руб. в год;
- экономия на стоимости владения оборудованием, включая расходы на электроэнергию.

В табл. 5 представлены приблизительные расчеты на 5 лет.

Экономия с учетом уже установленных 330 нулевых клиентов составит около 1,3 млн руб. в год (без учета экономии на зарплате). При этом нужно учитывать, что срок службы нулевых клиентов, скорее всего, составит больше 5 лет, так

Таблица 5 Стоимость владения ПК и нулевым клиентом за 5 лет эксплуатации

Статьи расходов	Расходы за 5 лет	
	ПК	Нулевой клиент
Клиентское устройство	19 000	19 500
Windows + SA/VDA	3500 + 3000	4500
VMware View 5.2	0	5600
Ресурсы ЦОД	0	4000
Зарплата с налогами	25 000 (125 ч)	5000 (25 ч)
Электроэнергия	8400	1200
Источник бесперебойного питания	1200	200
Итого	6 100	40 500

Таблица 6 Сравнение эксплуатационных характеристик ПК и нулевых клиентов

Характеристика	1 ПК	330 ПК	1 терминал	330 терминалов
Энергопотребление, Вт/ч	320–420 (в среднем 370)	122 100	29	9570
Расход электроэнергии за 6 мес., кВт	260	85 800	21	6930
Первичная установка ПО, человеко-часов	6	54	7	16
Обслуживание ПО, человеко-часов за 6 мес.	24	7920	24	152
Временные затраты на устранение неполадок за 6 мес., ч	—	9900	—	632
Количество поломок устройств, устраненных при помощи службы ремонта за 6 мес., шт.	—	64	—	4 (брак при поставке)

как они устаревают значительно медленнее ПК (это предположение основано на опыте применения в университете терминалов Sun Ray в течение 10 лет).

Преимущества применения нулевых клиентов сведены в табл. 6.

Использование технологии виртуализации рабочих мест на основе нулевых клиентов, по сравнению с использованием ПК, невозможно при отсутствии сетевых соединений. Но так как в большинстве случаев линии связи до опорных технических центров (сети распределения) резервируются, то проблема может возникнуть только на участке сети доступа. Такая авария может повлиять на работу не более 24–48 нулевых клиентов.

Опыт внедрения нулевых клиентов показал, что основные проблемы возникают при подключении принтеров к виртуальным рабочим местам. В случае, когда принтер сетевой, для подключения необходимо установить его драйверы на эталонный образ. Пул виртуальных машин может быть использован большим числом пользователей с разными устройствами печати, что требует значительных временных затрат на установку разных драйверов, соответствующих разным устройствам печати, и пересоздания всех виртуальных машин пула.

Еще одна проблема связана с тем, что при запуске нулевого клиента может происходить выбор пула виртуальных машин и пользователь может выбрать пул, в котором не установлены драйверы для его принтера.

Для решения этих проблем с установкой драйверов может быть использован программный комплекс ThinPrint, который представляет собой решение для оптимизации печати в серверных окружениях. Эта программа дает следующие преимущества:

- исключается необходимость установки родных драйверов для каждого принтера на эталонный образ;

- обеспечивается возможность автоматического подключения ближайшего к пользователю принтера.

В ситуации, когда принтер подключен непосредственно к рабочему месту, использование нулевого клиента требует дополнительных действий по подключению. К вышеописанным процедурам добавляется необходимость проброса подключения принтера к эталонному образу, но даже после этого не все принтеры будут корректно работать. Эта проблема обусловлена ограничением максимальной скорости передачи данных через USB-порт нулевого клиента (до 15 Мбит/с), что приводит к сбоям печати графической информации на некоторых моделях принтеров. Выход из описанной ситуации — использование сетевых принтеров или принт-серверов (стоимость одного устройства принт-сервера 1000–1500 руб.).

4 Выводы

В течение года в университете авторами статьи реализован проект по внедрению виртуальных рабочих мест на основе технологии нулевых клиентов. За это время нулевыми клиентами воспользовались 4042 пользователя, одновременно работало максимально 300 параллельных сессий. Опыт внедрения и использования показал следующие преимущества при переходе к технологии нулевых клиентов на базе VMware: значительно повысилась эффективность ИТ-поддержки учебных и рабочих мест пользователей, сократились издержки на техническое сопровождение, значительно уменьшилось энергопотребление, улучшилась эргономика, повысилось качество предоставляемых ИТ-услуг.

Реализация серверной части виртуализации рабочих мест на основе отказоустойчивого кластера (см. рис. 1) и резервирование брокера соединений (см. рис. 2) обеспечивают высокую надежность работы клиентских мест.

Успех проекта позволяет продолжить внедрение этого решения, и в 2014 г. планируется внедрить еще 500 терминалов. Кроме этого, в 2014 г. предполагается начать использовать нулевые клиенты для работы с 3D-пакетами.

Литература

1. *Holub T., Means R.* Desktop virtualization pilot project. — Berkeley: Technology Program Office, 2009. http://technology.berkeley.edu/cio/presentations/Desktop_Virtualization_Pilot.pdf.
2. Desktop Virtualization. Stanford University: IT Services. https://itservices.stanford.edu/strategy/platforms/desktop_virtualization.
3. Проект создания инфраструктуры виртуальных рабочих мест в Южно-Уральском государственном университете // Образование России. РФ, 2011. <http://образованиероссии.рф/blog/проект-создания-инфраструктуры-вирт>.
4. РГТЭУ меняет общую философию управления ИТ // ЮБиТек («Универсальные бизнес-технологии»), 2010. http://www.ubtec.ru/index.php?ELEMENT_ID=405.
5. *Корепанов Э. Р.* Некоторые подходы к разработке технологий тонкого клиента для защищенных информационных систем // Системы и средства информатики, 2013. Т. 23. № 21. С. 105–112.

6. Типы «тонких» клиентов для VMware Horizon View // V-GRADE, 2013. <http://www.vsphere5.ru/doku.php?id=thin-client-types>.
7. Выбор нулевого клиента для VMware Horizon View // V-GRADE, 2013. <http://www.vsphere5.ru/doku.php?id=choose-zero-client-for-vmware-view>.
8. Самойленко А. Сравнение технологий для виртуализации настольных ПК: VMware VDI и Citrix XenDesktop // VM Guru, 28/10/2008. <http://www.vmgu.ru/articles/vmware-vdi-comparison>.
9. Торицын Н. Техническое сравнение платформ виртуализации VMware ESX Server и Microsoft Hyper-V // VM Guru, 23/07/2008. <http://www.vmgu.ru/articles/vmware-esx-vs-microsoft-hyper-v>.
10. Крюков В. В., Шахгельдян К. И. Информационные технологии в университете: стратегия, тенденции, опыт // Университетское управление: практика и анализ, 2012. № 4. С. 101–112.

Поступила в редакцию 28.09.13

DEPLOYMENT AND USAGE OF DESKTOP VIRTUALIZATION AT UNIVERSITY

M. A. Vyshivanov, D. Gmar, V. V. Kryukov, K. S. Nagorniak, and C. I. Shakhgeldyan

Vladivostok State University of Economics and Service, 41 Gogol Str., Vladivostok 690014, Russian Federation

Abstract: Deployment of desktop virtualization at university is the main subject of the paper. The authors have analyzed and justified the choice of the zero client technology with the Teradici chip and the VMWare View software. They describe the cloud infrastructure of the university, present the results of the project to deploy the virtualization technology based on VMWare View and zero clients (330 zero clients), the advantages of using these solutions, and assessment of effectiveness of using zero clients at the university. Using the zero client technology reduces the cost of workplace ownership for 5 years by 1.5 times, minimizing noise and power consumption, increasing data integrity, speed of deployment of new workplaces and software update, efficiency of loading the university's classrooms.

Keywords: desktop virtualization; zero client; cloud computing

DOI: 10.14357/08696527140112

References

1. Holub, T., and R. Means. 2009. Desktop virtualization pilot project. Technology Program Office, Berkeley. Available at: http://technology.berkeley.edu/cio/presentations/Desktop_Virtualization_Pilot.pdf (accessed 2009).
2. Desktop Virtualization. Available at: https://itservices.stanford.edu/strategy/platforms/desktop_virtualization.

3. Proekt sozdaniya infrastruktury virtual'nykh rabochikh mest v Yuzhno-Ural'skom gosudarstvennom universitete [The project to create a Virtual Desktop Infrastructure in South Ural State University]. Available at: <http://xn--80aabfjba0cfdftira.xn--p1ai/blog/%D0%BF%D1%80%D0%BE%D0%B5%D0%BA%D1%82-%D1%81%D0%BE%D0%B7%D0%B4%D0%B0%D0%BD%D0%B8%D1%8F-%D0%B8%D0%BD%D1%84%D1%80%D0%B0%D1%81%D1%82%D1%80%D1%83%D0%BA%D1%82%D1%83%D1%80%D1%8B-%D0%B2%D0%B8%D1%80%D1%82/> (accessed August 17, 2011).
4. RGTJeU menyaet obshchuyu filosofiyu upravleniya IT [RGTEU change the overall IT management philosophy]. Available at: http://www.ubtec.ru/index.php?ELEMENT_ID=405 (accessed April 29, 2010).
5. Korepanov, E. R. 2013. Nekotorye podkhody k razrabotke tekhnologiy tonkogo klienta dlya zashchishchennykh informatsionnykh sistem [Some approaches to developing thin client technologies for secure information systems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 21:105-112.
6. Tipy “tonkikh” klientov dlya VMware Horizon View [Types of “thin” clients for VMware Horizon View]. Available at: <http://www.vsphere5.ru/doku.php?id=thin-client-types> (accessed March 21, 2014).
7. Vybory nulevogo klienta dlya VMware Horizon View [Selecting zero client for VMware Horizon View]. Available at: <http://www.vsphere5.ru/doku.php?id=choose-zero-client-for-vmware-view> (accessed March 21, 2014).
8. Sravnenie tekhnologiy dlya virtualizatsii nastol'nykh PK: VMware VDI i Citrix XenDesktop [Comparison of technologies for desktop virtualization: VMware VDI with Citrix XenDesktop]. Available at: <http://www.vmg.ru/articles/vmware-vdi-comparison> (accessed October 28, 2008).
9. Tekhnicheskoe sravnenie platform virtualizatsii VMware ESX Server i Microsoft Hyper-V [Technical comparison of virtualization platforms VMware ESX Server with Microsoft Hyper-V]. Available at: <http://www.vmg.ru/articles/vmware-esx-vs-microsoft-hyper-v> (accessed July 23, 2008).
10. Kryukov, V. V., and K. J. Shakhgelyan. 2012. Informatsionnye tekhnologii v universitete: Strategiya, tendentsii, opyt [Information technology at university: Strategy, trends, and experience]. *Universitetskoe Upravlenie: Praktika i Analiz* [University Management: Practice and Analysis] 4: 101-112.

Received September 28, 2013

Contributors

Vyshivanov Maxim A. (b. 1981) — leading software engineer, Vladivostok State University of Economics and Service, 41 Gogol Str., Vladivostok 690014, Russian Federation; maksim.vyshivanov@vvsu.ru

Gmar Dmitriy V. (b. 1980) — Head of Department, Vladivostok State University of Economics and Service, 41 Gogol Str., Vladivostok 690014, Russian Federation; dimer@vvsu.ru

Kryukov Vladimir V. (b. 1956) — Doctor of Science in economics, professor, Vice-President, Vladivostok State University of Economics and Service, 41 Gogol Str., Vladivostok 690014, Russian Federation; kryukov@vvsu.ru

Nagorniak Konstantin S. (b. 1985) — PhD student, Head of the Computer Center, Vladivostok State University of Economics and Service, 41 Gogol Str., Vladivostok 690014, Russian Federation; konstantin.nagornyak@vvsu.ru

Shakhgeldyan Carina I. (b. 1967) — Doctor of Science in technology, associate professor, Head of Information Technology Department, Vladivostok State University of Economics and Service, 41 Gogol Str., Vladivostok 690014, Russian Federation; carinash@vvsu.ru

МЕТОДИКА ПОСТРОЕНИЯ ВЕБ-СЕРВИСОВ РАСПРЕДЕЛЕННЫХ ПЛАТФОРМЕННО-НЕЗАВИСИМЫХ СИСТЕМ АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ

В. Н. Гридин¹, Г. Д. Дмитриевич¹, Д. А. Анисимов²

Аннотация: Рассматриваются вопросы внедрения интернет-технологий в системы автоматизированного проектирования (САПР) путем перехода к распределенным системам, построенным на основе реализации задач коммуникации и информационного обмена между приложениями. Такие независимо управляемые приложения являются автономными и могут взаимодействовать друг с другом в процессе выполнения общей задачи. Протоколы интернет-технологий представляют надежную базу для связывания подсистем и не требуют согласованного использования ресурсов, находящихся в разных узлах сети, что существенно упрощает процесс построения и эксплуатации распределенной САПР. Основным требованием для возможности реализации такой распределенной системы является согласованность интерфейсов, посредством которых связаны отдельные подсистемы. Приводится сравнительная оценка возможных способов построения программного обеспечения распределенных платформенно-независимых САПР на основе технологии веб-сервисов. Дается описание взаимодействия основных протоколов функционирования веб-сервисов в распределенных системах автоматизированного проектирования.

Ключевые слова: системы автоматизированного проектирования; веб-технологии; веб-сервисы; сервис-ориентированная архитектура; распределенные системы

DOI: 10.14357/08696527140113

1 Введение

Одной из актуальных задач в области информационных технологий является внедрение в системы автоматизированного проектирования веб-технологий, что позволяет реализовать построение систем с сервисно-ориентированной архитектурой, при которой информационные ресурсы доставляются потребителям посредством веб-сервисов [1, 2]. Технология веб-сервисов позволяет получить

¹Центр информационных технологий в проектировании Российской академии наук, info@ditc.ras.ru

²Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В. И. Ульянова (Ленина) (СПбГЭТУ), anisimovdenis2009@yandex.ru

достаточную абстракцию программного обеспечения САПР, при которой приложения могут быть всегда использованы пользователями, работающими только на уровне подготовки данных и отображения результатов. Конкретная методика реализации веб-сервисов существенно зависит от выбранного языка программирования, при этом главным требованием является обеспечение платформенной независимости реализованных решений.

К платформенно-независимым языкам относится язык Java [3], а также языки, реализованные в .NET Framework (например, C#) [4]. По имеющимся возможностям, а также по имеющейся инструментальной поддержке языки Java и C# примерно сопоставимы, однако предпочтение при выборе языка для веб-сервисов следует отдать языку Java, поскольку веб-сервис, реализованный на таком языке, может быть развернут на сервере, где имеется любая Java-виртуальная машина (JVM), что всегда обеспечивается. Вместе с тем для функционирования веб-сервиса на основе C# требуется предварительная установка на сервере среды .NET Framework, что в ряде случаев может ограничить выбор провайдера. С учетом возможности работы веб-сервисов в гетерогенных средах использование языка Java для построения веб-сервисов, естественно, не ограничивает разработчиков в выборе любого языка программирования для построения клиентских приложений.

2 Постановка задачи

Инструментальная поддержка построения веб-ориентированных приложений на Java обеспечивается платформой Web Tools Platform (WTP), которая является верхним уровнем проекта Eclipse [3]. Большинство разработчиков общесистемного программного обеспечения приняло WTP в качестве рабочей среды для создания корпоративных и веб-приложений (BEA WebLogic Workshop, Code Gear JBuilder, My Eclipse Genuitec, IBM Rational Application Developer, JBoss IDE, SAP NetWeaver).

По имеющимся оценкам, проект WTP добился наибольшего успеха в достижении цели обеспечения общего инструментального сопровождения процесса разработки веб-ориентированных приложений. Платформа WTP относится к проектам с открытым кодом и может быть развернута на компьютере вместе с платформой Eclipse. Для инструментальной поддержки разработки веб-сервисов можно также использовать любую из упомянутых выше интегрированных сред. Проект WTP содержит два подпроекта — WST (Web Standard Tools), который поддерживает веб-стандарты, независимые от технологии реализации, и JST (J2EE Standard Tools), поддерживающий технологию разработки Java-приложений J2EE в качестве основной технологии реализации.

В подпроект WST входят инструменты и программные интерфейсы, предназначенные для построения веб-приложений на базе стандартов, не зависящих от технологии реализации. Поскольку веб-приложения выполняются на сервере, который управляет приложением, а не просто обслуживает веб-страницы, то

поддержка базовых функций серверов входит в подпроект WST. Среда WST предоставляет программные интерфейсы, обеспечивающие выполнение серверов в Eclipse при помощи подключаемого модуля, который поддерживает среду выполнения сервера и обычно называется адаптером сервера. При помощи адаптера сервера можно конфигурировать, стартовать, останавливать, перезапускать, отлаживать сервер и публиковать на нем разрабатываемое веб-приложение.

Подпроект WST содержит интернет-подсистему, встроенную в среду Eclipse как компонент текстового редактора и выполняющую функцию собственного веб-браузера. Поэтому при реализации веб-приложений появляется возможность без выхода из среды разработки осуществлять вход в сеть Интернет для ознакомления с реестрами системы UDDI (Universal Description, Discovery, and Integration), обеспечивать мониторинг протокола TCP/IP, анализировать сообщения, посылаемые на заданный порт или получаемые от него. Включенный в состав WST проводник веб-сервисов (Web Service Explorer) позволяет выполнять поиск реестров для веб-сервисов, динамически тестировать удаленные веб-сервисы. Web Service Explorer поддерживает два стандарта распознавания (discovery) веб-сервисов: протокол UDDI работы со сложными реестрами и язык инспектирования веб-сервисов WSIL (Web Service Inspection Language), который имеет формат простого документа XML. Конечным результатом процесса распознавания является для обоих случаев определение местонахождения документа WSDL, описывающего требуемый веб-сервис. Среда WST предоставляет пользователю многофункциональный редактор документов WSDL, который применяется как для построения новых документов WSDL, так и для редактирования существующих. После завершения разработки собственного веб-сервиса проводник Web Service Explorer позволяет зарегистрировать документ WSDL с описанием этого сервиса в реестре UDDI.

Подпроект WST содержит весьма эффективный расширяемый мастер Web Service, связывающий вместе все задачи, возникающие при построении веб-сервисов. Помимо распознавания и публикации веб-сервисов к таким задачам относятся создание, развертывание, тестирование веб-сервисов, а также обеспечение обращения к ним клиентов. В состав WST входит подсистема тестирования, позволяющая проверять разработанный веб-сервис. Инструменты тестирования имеют анализатор сообщений SOAP (Simple Object Access Protocol), а также валидатор документов WSDL, который интегрирован в редактор WSDL.

3 Разработка веб-сервисов по базовым стандартам J2EE

Разработка веб-сервисов согласно базовым стандартам J2EE может быть весьма эффективно реализована в подпроекте JST, который расширяет возможности WST. В JST входят инструменты и программные интерфейсы, которые поддерживают разработку веб-приложений согласно базовым стандартам J2EE. Этот подпроект расширяет серверные инструменты WST возможностью поддержки серверов приложений J2EE, при этом адаптеры серверов могут разрабатываться

с использованием программных интерфейсов Java, а также с помощью обобщенных средств поддержки при помощи XML-сообщений. В состав JST не входят какие-либо среды выполнения серверов, но эта среда содержит адаптеры для многих широко распространенных серверов (Apache Tomcat, Apache Geronimo, JBoss, IBM WebSphere, Object Web Jonas, BEA WebLogic).

Среда JST обеспечивает поддержку разработок проектов J2EE, предоставляя программные интерфейсы доступа к объектам, из которых состоят модули проекта. Наиболее распространенная форма веб-приложений J2EE основана на применении контейнера сервлетов, который содержит как сервлеты, так и JSP-страницы. В JST входит многофункциональный редактор страниц JSP, являющийся расширением редактора HTML и помогающий устанавливать теги HTML и JSP и встраиваемые в JSP-страницу фрагменты Java. В J2EE входит ряд стандартов по веб-сервисам, основными из которых являются JAX-RPC, определяющий связь Java с WSDL и предоставляющий соответствующую модель программирования клиентского приложения, и JSR 109, определяющий развертывание веб-сервиса. Подпроект JST расширяет мастер Web Service, предоставляя ему возможность поддерживать стандарты JAX-RPC и JSR 109.

Подпроект JST поддерживает программный продукт Axis (Apache eXtensible Interaction System), содержащий набор классов, реализующих интерфейсы пакетов Sun Microsystems SAAJ, JAXM, JAX-RPC. Среда Axis (называемая иногда Apache Axis) обеспечивает поддержку построения и последующего развертывания веб-сервисов, а также организацию доступа клиентов. При этом веб-сервисы могут создаваться как нисходящим (Top-Down), так и восходящим (Bottom-Up) методами.

В нисходящем методе сначала создается требуемый документ WSDL, а затем на его основе формируется код построения веб-сервиса. Во втором методе сначала создается Java-класс веб-сервиса, а затем согласно его структуре генерируется WSDL-документ.

Поскольку при использовании нисходящего метода необходимо создание документа WSDL, удовлетворяющего условиям поставленной задачи построения веб-сервиса, то разработчик веб-сервиса должен учитывать особенности базовой структуры спецификации WSDL и допустимых модификаций этой структуры.

Корневым элементом XML-описания документа WSDL является элемент `<definitions>`, содержащий описания пространств имен и имеющий необязательный атрибут `name`. В корневой элемент вложены шесть обязательных и два необязательных элемента. Обязательными элементами являются `<types>`, `<message>`, `<portType>`, `<serviceType>`, `<binding>`, `<service>`, к необязательным относятся `<import>` и `<documentation>`.

Раздел `<types>` определяет на основании языка XSD все используемые веб-сервисом сложные типы данных. Элемент `<types>` — это вложенная XML-схема, и все типы данных, определенные в стандарте XML Schema, являются допустимыми. Если веб-сервис использует простые типы данных, описанные в XSD, то элемент `<types>` в документе WSDL отсутствует.

Раздел `<message>` содержит информацию о сообщениях запроса и ответа, используемых для взаимного обмена данными с веб-сервисом. Здесь описывается каждое послание SOAP (запрос, ответ, пересылка документов) и вкладываются элементы `<part>`, в которых описываются отдельные части послания, являющиеся неделимыми согласно документу WSDL.

Раздел `<portType>` содержит описание интерфейса веб-сервиса, который в WSDL называется *endpoint* (пункт назначения) или *port* (порт). Этот раздел описывается набором веб-услуг, называемых в WSDL операциями и описываемых элементами `<operation>`. Каждая услуга задается действиями, разбитыми на четыре вида: получение послания, отправка ответа, отправка послания с получением ответа, получение послания с отправкой ответа. Получение и отправка послания описываются элементами `<input>` и `<output>`.

Раздел `<serviceType>` перечисляет своими вложенными элементами `<portType>` набор портов, которые связаны с одним веб-сервисом.

Раздел `<binding>` содержит описание конкретного формата пересылки послания: протоколы (SOAP или HTTP), выбранные способы упаковки послания, тип содержимого послания (HTML, XML). С помощью элемента `<binding>` WSDL-документ описывает низкоуровневый протокол связи, который можно использовать с веб-сервисом.

Раздел `<service>` указывает расположение веб-сервиса в виде одного или совокупности портов. Каждый из портов описывается соответствующим вложенным элементом `<port>`, который содержит адрес интерфейса веб-сервиса, заданный согласно правилам, выбранным в элементе `<binding>` способа пересылки.

Раздел `<import>` содержит файл, в котором описана XSD-схема документа WSDL.

Раздел `<documentation>` описывает произвольные комментарии. Этот раздел можно включать в любые элементы описания WSDL-документа.

Разработанный нисходящим методом построения веб-сервисов WSDL-документ должен удовлетворять базовой структуре WSDL и техническому заданию на веб-сервис, включающему в себя описание параметров, передаваемых вызываемому методу, а также возвращаемых методом значений. После создания WSDL-документа возможно сформировать при помощи мастера Web Service скелет Java-класса, который затем дополняется программным кодом, реализующим бизнес-логику веб-сервиса.

Достоинством нисходящего метода построения веб-сервисов следует считать возможность использования произвольных сложных типов данных, которые, естественно, должны быть обязательно описаны в документе WSDL. Серьезным недостатком такого подхода к разработке веб-сервисов является необходимость иметь достаточные навыки проектирования документов с использованием языков XSD и WSDL. Процесс разработки веб-сервисов при нисходящем методе проектирования осложняется также и тем, что существующие редакторы WSDL не выполняют полностью валидацию достоверности созданного WSDL-документа. Поэтому возможные ошибки документа обнаруживаются только при тестирова-

нии сформированного веб-сервиса, что существенно затрудняет, а иногда делает вообще невозможным завершение разработки веб-сервисов нисходящим методом.

Проектирование веб-сервисов восходящим методом полностью устраняет указанную проблему, так как для этого случая WSDL-документ формируется на основании созданного заранее Java-класса, где описаны все передаваемые методу веб-сервиса параметры, а также возвращаемые этим методом значения. Вся имеющаяся в Java-классе информация автоматически преобразуется мастером Web Service в соответствующий WSDL-документ, содержание которого точно соответствует базовой структуре спецификации WSDL и основным характеристикам вызываемого метода веб-сервиса. Продуктивность разработки веб-сервисов резко повышается, и обеспечивается полная достоверность информации, содержащейся в документе WSDL. При этом программист может вообще не рассматривать содержимое созданного мастером Web Service документа WSDL, хотя ознакомление с WSDL-документом, безусловно, полезно, но только для более полного представления о конкретной решаемой задаче построения веб-сервиса.

4 Реализация технологии построения веб-сервисов восходящим методом

Для реализации концепции проектирования веб-сервиса восходящим методом следует создать динамический веб-проект и Java-класс, реализующий веб-сервис и содержащий описание вызываемых методов. Следует заметить, что хорошо спроектированный веб-сервис должен выполнять тщательную валидацию всех входных данных и генерировать информативные исключения при недостоверности этих данных. Процесс развертывания Java-класса в качестве веб-сервиса заключается в его добавлении в конфигурацию SOAP-процессора. При использовании мастера Web Service этот процесс полностью автоматизирован и от программиста не требуется никаких усилий. Достаточно выбрать в проекте созданный Java-класс и выполнить в контекстном меню команду Web Service>Create Web service.

После построения веб-сервиса Java целесообразно проводить его предварительное тестирование. Предварительным тестированием веб-сервисов будем называть процедуру проверки работоспособности вновь созданного веб-сервиса, которую необходимо выполнять еще до построения клиентского приложения. Существует несколько способов проверки результатов процесса построения веб-сервиса до создания клиентских приложений.

Наиболее просто такая задача решается путем использования созданного системой WTP сервлета AxisServlet, который дает возможность убедиться, что Web Service успешно провел развертывание созданного сервиса и построил входящие в него методы. Для вызова сервлета следует выбрать в проекте сервлет AxisServlet и выполнить команду Run as>Run on Server. При этом откроется встроенный в WTP веб-браузер, содержащий имя созданного веб-

сервиса, совпадающее с именем проекта, а также список доступных для вызова методов сервиса. Браузер содержит также ссылку на файл WSDL, при выборе которой откроется созданный WSDL-документ.

Наибольшую информацию о построенном веб-сервисе можно получить, используя входящую в платформу WTP многофункциональную инструментальную среду Web Services Explorer. Поскольку концепция распределенных сервис-ориентированных САПР предполагает создание и постоянное совершенствование клиентских приложений программистами, которые не принимали участия в построении программного обеспечения веб-сервисов, то все разрабатываемые веб-сервисы должны быть самодокументируемыми. Последнее означает, что каждый веб-сервис должен включать в себя информационный метод, например `getInf()`, не имеющий параметров и возвращающий простую строковую переменную, связанную с текстом, где должны содержаться сведения о методах, которые обеспечивают бизнес-логику веб-сервиса, а также описание форматов передаваемых параметров и возвращаемых методами данных [5–8].

Структура информационного метода `getInf()` определяется спецификой распределенной САПР. В частности, для систем автоматизации схемотехнического проектирования может быть предложена следующая унифицированная структура информационного метода `getInf()`:

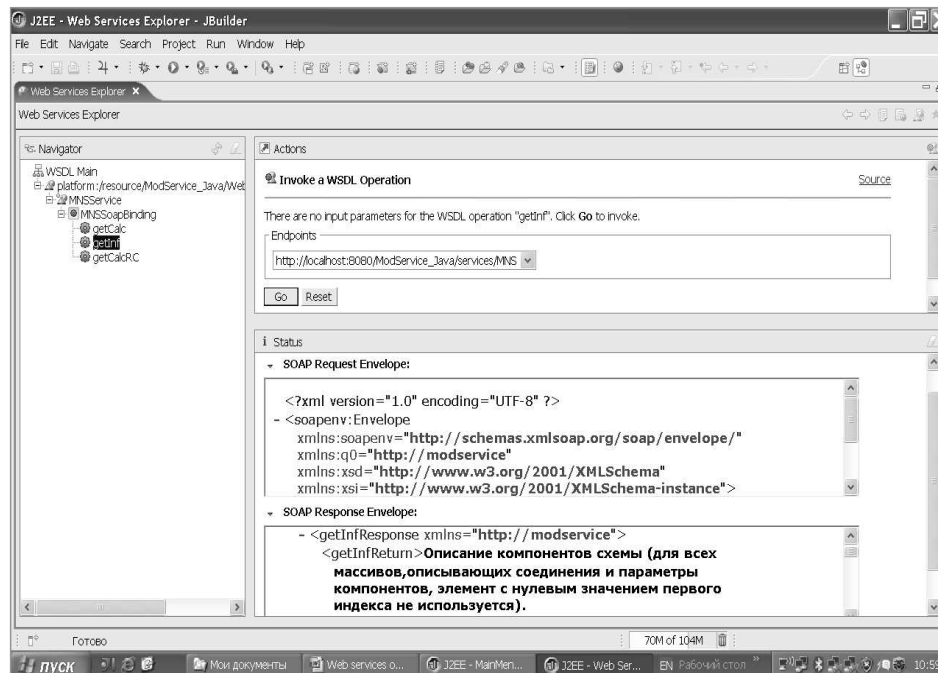
- Описание компонентов схемы (число компонентов данного типа, массив включения, массив значений параметров и порядок формирования этих массивов для каждого типа компонента, используемого в методах веб-сервиса).
- Преобразование массивов описания компонентов схемы к унифицированному виду.
- Синтаксис формирования прокси-объекта и имя рабочей директории для взаимодействия веб-сервиса с клиентскими приложениями.
- Заголовки рабочих методов веб-сервиса.
- Операторы формирования прокси-объекта для взаимодействия веб-сервиса с клиентскими приложениями.
- Описание возвращаемых результатов работы методов веб-сервиса в унифицированной форме.
- Описание преобразования унифицированной формы возвращаемых результатов методов веб-сервиса к виду, требуемому для отображения результатов работы веб-сервиса в клиентских приложениях.

Поскольку основная концепция разработки и дальнейшего развития распределенной системы автоматизированного проектирования предполагает возможность построения клиентских приложений пользователями САПР, которые не участвовали в создании веб-сервисов, то наличие в распоряжении разработчиков клиентских приложений информации, содержащейся в информационном методе `getInf()`, является обязательным условием.

Вызов метода `getInf()` можно осуществить непосредственно из среды разработки клиентского приложения еще до реализации этого приложения путем вызова инструмента *Web Services Explorer*, где по URL-адресу сервиса можно организовать выход к Soap-ответу метода `getInf()`. Конверт этого ответа *Soap Response Envelope* содержит отформатированный текст с полной информацией обо всех основных методах, которые могут быть вызваны из выбранного веб-сервиса, включая форматы всех передаваемых и возвращаемых этими методами данных.

Эта информация обеспечивает возможность согласования интерфейсов клиента и сервера и должна быть учтена программистами, реализующими приложения клиента для соответствующего сервиса. На рисунке приведен пример структуры инструментальной среды *Web Services Explorer*, в которой осуществлен вызов запроса Soap и Soap-ответа для информационного метода `getInf()` веб-сервиса *ModService_Java*, входящего в распределенную систему автоматизации схемотехнического проектирования. Поскольку этот метод не содержит параметров, то конверт *Soap Request Envelope* имеет только заголовок, при этом конверт *Soap Response Envelope* содержит описание возвращаемой строки.

Окончательная проверка работоспособности разрабатываемого веб-сервиса возможна лишь после создания согласованного с ним по интерфейсу клиентского



Soap-запрос и Soap-ответ для информационного метода `getInf()`

приложения и тестирования задачи путем вызова из приложения клиента выбранного метода веб-сервиса. Если необходимо использовать создаваемый веб-сервис в гетерогенных средах, такие приложения должны создаваться и тестироваться для любого типа среды.

5 Заключение

Полученные результаты содержат общую методику построения восходящим методом веб-сервисов Java и соответствующих WSDL-документов на основе технологии JST, а также доставки их на сервер распределенной САПР после проведения автономного тестирования в среде разработки. Внедрение этой методики предоставляет широкие возможности для построения распределенных платформенно-независимых САПР, реализующих взаимодействие клиентских приложений и веб-сервисов в гетерогенных средах.

Литература

1. Ларистов Д. А., Анисимов Д. А. Доступ к Web-ресурсам в САПР систем навигации и управления // Гироскопия и навигация, 2007. № 2. С. 106–107.
2. Гридин В. Н., Анисимов В. И. Методы построения систем автоматизированного проектирования на основе интернет-технологий и компактной обработки разреженных матриц // Информационные технологии в проектировании и производстве, 2009. № 1. С. 3–7.
3. Дей Н., Мандел Л., Райман А. Eclipse: Платформа Web-инструментов / Пер. с англ. — М.: КУДИЦ-Пресс, 2008. 688 с. (Dai N., Mandel L., Ryman A. Eclipse Web Tools Platform. — New York: Addison-Wesley, 2008. 688 p.)
4. Морган С., Райан Б., Хорн Ш., Бломсма М. Разработка распределенных приложений на платформе Microsoft.Net / Пер. с англ. — М.: Русская редакция, 2008. 608 с. (Morgan S., Ryan B., Horn Sh., Blomsma M. .Net Framework 2.0 distributed application development. — New York: Microsoft Press, 2008. 608 p.)
5. Гридин В. Н., Дмитриевич Г. Д., Анисимов Д. А. Построение систем автоматизированного проектирования на основе Web-сервисов // Автоматизация в промышленности, 2011. № 1. С. 9–12.
6. Гридин В. Н., Дмитриевич Г. Д., Анисимов Д. А. Построение систем автоматизированного проектирования на основе Web-технологий // Информационные технологии, 2011. № 5. С. 23–27.
7. Гридин В. Н., Дмитриевич Г. Д., Анисимов Д. А. Построение веб-сервисов систем автоматизации схмотехнического проектирования // Информационные технологии и вычислительные системы, 2012. № 4. С. 79–84.
8. Анисимов Д. А. Методы построения систем автоматизации схмотехнического проектирования на основе веб-сервисов // Известия СПбГЭТУ, 2012. № 10. С. 56–61.

Поступила в редакцию 31.01.14

METHODOLOGY OF CONSTRUCTION OF WEB SERVICES OF DISTRIBUTED PLATFORM-INDEPENDENT COMPUTER-AIDED DESIGN

V. N. Gridin¹, G. D. Dmitrevich¹, and D. A. Anisimov²

¹Design Information Technologies Center, Russian Academy of Sciences, 7a Marshal Biryuzov Str., Odintsovo 143000, Moscow Region, Russian Federation

²St. Petersburg Electrotechnical University “LETI,” 5 Ave. Professor Popov, St. Petersburg 197376, Russian Federation

Abstract: Problems of introduction of Internet technologies into systems of automated design are considered by transition to the distributed systems realizing the tasks of communication and information exchange between applications. Such independently controlled applications are autonomous and can cooperate with each other during performance of the common task. Internet technologies protocols represent a reliable base for linking subsystems and do not demand coordinated use of resources that are located in different nodes of a network which simplifies the processes of construction and operation of distributed systems of automated design. The main requirement for realizing such distributed system is coordination of interfaces that connect separate subsystems. The paper provides a comparative analysis of possible ways of creating the software of distributed platform-independent systems of automated design on the basis of web services technology. The paper presents the description of interaction of the basic protocols used by of web services in distributed systems of automated design.

Keywords: computer-aided design; web technologies; web services; service-oriented architecture; distributed systems

DOI: 10.14357/08696527140113

References

1. Laristov, D. A., and D. A. Anisimov. 2007. Dostup k Web-resursam v SAPR sistem navigatsii i upravleniya. [Access to Web-resources in CAD systems of navigation and management]. *Girokopiya i Navigatsiya* [Gyroscopy and Navigation] 2:106–107.
2. Gridin, V. N., and V. I. Anisimov. 2009. Metody postroeniya sistem avtomatizirovannogo proektirovaniya na osnove internet-tekhnologiy i kompaktnoy obrabotki razrezhennykh matrits. [Methods of construction of systems of the automated designing on the basis of Internet-technologies and compact processing of rarefied matrixes]. *Informatsionnye Tekhnologii v Proektirovanii i Proizvodstve* [Information Technologies in Designing and Manufacture] 1:3–7.
3. Dai, N., L. Mandel, and A. Ryman. 2008. *Eclipse Web Tools Platform*. 1st ed. New York: Addison-Wesley. 688 p.

4. Morgan, S., B. Ryan, Sh. Horn, and M. Blomsma. 2008. *.Net Framework 2.0 distributed application development*. 1st ed. New York: Microsoft Press. 608 p.
5. Gridin, V. N., G. D. Dmitrevich, and D. A. Anisimov. 2011. Postroenie sistem avtomatizirovannogo proektirovaniya na osnove Web-servisov [Construction of systems of the automated designing on the basis of Web-services]. *Avtomatizatsiya v Promyshlennosti* [Automation in Industry] 1:9–12.
6. Gridin, V. N., G. D. Dmitrevich, and D. A. Anisimov. 2011. Postroenie sistem avtomatizirovannogo proektirovaniya na osnove Web-tekhnologiy [Construction of systems of the automated designing on the basis of Web-technologies]. *Informatsionnye Tekhnologii* [Information Technologies] 5:23–27.
7. Gridin, V. N., G. D. Dmitrevich, and D. A. Anisimov. 2012. Postroenie veb-servisov sistem avtomatizatsii skhemotekhnicheskogo proektirovaniya [Construction of web-services of systems of automation circuit designing]. *Informatsionnye Tekhnologii i Vychislitel'nye Sistemy* [Information Technologies and Computing Systems] 4:79–84.
8. Anisimov, D. A. 2012. Metody postroeniya sistem avtomatizatsii skhemotekhnicheskogo proektirovaniya na osnove veb-servisov [Method of construction of systems of automation circuit designing on the basis of web-services]. *Izvestiya SPbGETU* [News of St. Petersburg Electrotechnical University] 10:56–61.

Received January 31, 2014

Contributors

Gridin Vladimir N. (b. 1944) — Doctor of Science in technology, professor, Director, Design Information Technologies Center, Russian Academy of Sciences, 7a Marshal Biryuzov Str., Odintsovo 143000, Moscow Region, Russian Federation; info@ditc.ras.ru

Dmitrevich Gennady D. (b. 1939) — Doctor of Science in technology, professor, principal scientist, Design Information Technologies Center, Russian Academy of Sciences, 7a Marshal Biryuzov Str., Odintsovo 143000, Moscow Region, Russian Federation; info@ditc.ras.ru

Anisimov Denis A. (b. 1987) — Candidate of Science (PhD) in technology, junior scientist, St. Petersburg Electrotechnical University “LETI,” 5 Ave. Professor Popov, St. Petersburg 197376, Russian Federation; anisimovdenis2009@yandex.ru

ЗАДАЧИ ИДЕНТИФИКАЦИИ ИНФОРМАЦИОННЫХ ОБЪЕКТОВ В РАСПРЕДЕЛЕННЫХ МАССИВАХ ДАННЫХ

М. М. Гершкович¹, Т. К. Бирюкова²

Аннотация: Предлагаются подходы к решению задачи идентификации информационных объектов (ИО) в автоматизированных информационных системах (АИС), предназначенных для сбора, хранения и обработки данных, имеющих большое количество узлов и получающих данные из различных источников. Массив данных в рассматриваемых информационных системах, как правило, представляет собой постоянно пополняемый журнал событий. Каждая запись о событии содержит характеристики участника события — ИО и обстоятельства события. Для решения аналитических задач, связанных с ИО, необходимо идентифицировать ИО, т. е. определить множество ИО, представляющих собой с некоторой вероятностью одну и ту же сущность. В работе сформулирована постановка задач идентификации информационных объектов, часто встречающихся на практике при создании больших информационных систем: слияние ИО и кластеризация ИО, т. е. составление совокупностей ИО, «похожих» по некоторому критерию. Отмечено, что с задачей идентификации тесно связана задача поиска связей между ИО, поскольку вероятность идентичности двух ИО повышается, если выявляется связь каждого из них, например, с некоторым третьим ИО. Указаны способы решения этих задач, отмечена специфика идентификации ИО в потоке событий, приведен метод корреляционного поиска для выявления связи между ИО. Приведены методы сравнения имен собственных с учетом возможных искажений в них — фонетических, транскрипционных, а также просто опечаток. Указана целесообразность использования при идентификации физических лиц (ФЛ) данных фамильно-именной группы и в кириллице, и в латинице.

Ключевые слова: идентификация информационных объектов; идентификация объектов; корреляционный поиск; поиск связей; идентичность объектов; слияние информационных объектов; слияние объектов; текстовые атрибуты; искажения данных; фонетические искажения; транскрипционные искажения; перевод из латиницы в кириллицу; Metaphone; расстояние Левенштейна; распределенные информационные системы

DOI: 10.14357/08696527140114

¹Институт проблем информатики Российской академии наук, makmg@mail.ru

²Институт проблем информатики Российской академии наук, yukonta@mail.ru

1 Введение

При создании АИС, предназначенных для сбора, хранения и обработки данных, состоящих из множества территориально-распределенных элементов (узлов) и получающих данные из различных источников, необходимо решить множество задач различного характера [1–3]. Одной из существенных задач, решение которой определяет возможности системы по обработке и анализу данных, является задача идентификации ИО.

Рассмотрим АИС, в которую из источников различного типа поступают данные о событиях, происходящих с некоторыми субъектами — например, с ФЛ, юридическими лицами, транспортными средствами.

Массив данных в рассматриваемых АИС, как правило, представляет собой постоянно пополняемый журнал некоторых происходящих событий. В каждой записи о событии указано время события, субъект события — ИО — и обстоятельства события. Как ИО, так и обстоятельства характеризуются атрибутами, или признаками, представляющими собой или числовые данные, или текст, или изображения, причем набор признаков, характеризующих ИО, может быть разным для источников различного типа.

Для решения аналитических задач, связанных с ИО, крайне важным является идентифицировать ИО (определить множество ИО, представляющих собой с некоторой вероятностью одну и ту же сущность). При этом, с одной стороны, в искомое множество ИО должно быть отобрано обозримое число объектов, иначе результаты будут практически бесполезными. С другой стороны, из-за возможных искажений в записях нельзя чрезмерно строго фильтровать их, иначе можно отсеять нужные объекты. Интересные результаты по проблеме идентификации ИО на основе признаков-изображений (а именно биометрических данных) получены в работе [4]. В настоящей статье основной упор сделан на идентификацию ИО по числовым и особенно по текстовым признакам.

Основная цель идентификации состоит в том, чтобы во множестве событий суметь объединить события, которые происходили с одним объектом с учетом того, что параметры объекта могут быть случайным образом искажены или системным образом изменены. Например, фамилия объекта, представляющего собой ФЛ, может быть случайным образом искажена при вводе в информационную систему либо изменена законным образом. Аналогично, номер паспорта ФЛ может быть искажен при вводе в систему, ФЛ может путешествовать с разными паспортами (с паспортом гражданина РФ, с заграничным паспортом) или получить новый паспорт взамен старого.

В задаче идентификации можно выделить следующие частные задачи, решение которых требуется чаще всего на практике [5, 6]:

- формирование ИО из потока данных о событиях, происходящих с ИО;
- слияние ИО;

- составление «Сложных ИО» — совокупностей ИО, «похожих» по некоторому критерию.

С задачей идентификации тесно связана задача поиска связей между ИО. С одной стороны, для поиска связей между объектами необходимо решить задачу идентификации объектов. С другой стороны, само по себе решение задачи поиска связей позволяет уточнять гипотезы по задаче идентификации.

В следующих разделах рассмотрена постановка задач, относящихся к идентификации ИО, и намечены пути их решения.

2 Постановка задач идентификации информационных объектов

Задача 1. Формирование информационных объектов из потока данных о событиях

Рассмотрим АИС, в которую поступает и сохраняется в базе данных (БД) информация о событиях, происходящих с ИО. События содержат множество атрибутов, среди которых есть данные об ИО — участниках событий. Задача формирования ИО из потока данных о событиях состоит в формировании в БД отдельного множества ИО, которые участвовали в событиях. Так, одно и то же ФЛ может участвовать во многих событиях, и система должна уметь «выделять» данные о ФЛ из данных о событии.

С этой задачей вплотную связаны задачи слияния ИО и составления «Сложных ИО».

Задача 2. Слияние информационных объектов

Слияние ИО — это выявление в БД (и/или в потоке событий) ИО, которые по определенному **критерию слияния** являются одним и тем же ИО, и последующее объединение выявленных ИО (в результате формируется один ИО с атрибутами, возможно, взятыми из всех поступивших ИО, удовлетворяющих критерию слияния).

Задача слияния ИО возникает, например, в иерархических системах с древовидной структурой — когда данные из узлов нижних уровней иерархии поступают в вышестоящие узлы и требуется определять, являются ли ИО, поступившие из разных узлов, фактически одним и тем же ИО.

Задача 3. Кластеризация информационных объектов — составление «Сложного информационного объекта»

«Сложный ИО» — это совокупность ИО, связанных по заданному критерию. Составление «Сложного ИО» заключается в том, чтобы связать «похожие» ИО, которые с заданной вероятностью представляют собой один и тот же ИО (не объединяя их в БД, как в случае слияния, а устанавливая связи между ИО).

Таким образом, появляется возможность анализировать все множество событий, произошедших со «Сложным ИО». Задачу формирования «Сложных ИО» можно рассматривать как задачу кластеризации, т. е. разбиения всего множества ИО в БД на подмножества ИО, связанных между собой.

В задачах 2 и 3 необходимо разработать соответственно **критерии слияния и критерии кластеризации**.

Сложность разработки критериев слияния и кластеризации для конкретной АИС состоит в том, что данные, поступающие от различных внешних систем, могут отличаться по содержанию для ИО одного и того же типа.

В зависимости от конкретных требований по решению задачи идентификации может существенно различаться ее ресурсоемкость. Так, проще всего выполнять идентификацию в БД, окончательно заполненной к моменту поиска. Однако во многих случаях требуется проводить автоматический мониторинг БД при поступлении нового ИО (т. е. при выделении нового ИО из потока входящих событий) и «на ходу» проводить слияние ИО и/или формирование «Сложных ИО», к которым «привязаны» ИО по определенным признакам. При больших объемах данных в БД и при интенсивном входящем потоке данных такие задачи становятся очень ресурсоемкими, и алгоритмы идентификации должны учитывать временные рамки обработки событий. Для упрощения задача может ставиться так, что «Сложные ИО» формируются пользователем (поиск в БД «кандидатов» для «связывания» проводится по запросу пользователя либо пользователь сам «привязывает» ИО к «сложному ИО» — например, из предварительно составленного по запросу списка «похожих» ИО).

Для определения «похожих» ИО целесообразно использовать методы сравнения, учитывающие:

- фонетические и транскрипционные искажения данных;
- написание текстовых атрибутов, характеризующих один и тот же ИО, латиницей и кириллицей (для имен собственных);
- возможные опечатки в несколько букв в текстовых атрибутах ИО.

Под фонетическими искажениями понимаются искажения при внесении в БД текстовых атрибутов ИО, схожих по звучанию. Транскрипционные искажения возникают из-за возможных различий в написании текстовых атрибутов ИО при переводе из латиницы в кириллицу и из кириллицы в латиницу (например, при переводе фамильно-именной группы (ФИО) из кириллицы в латиницу в общегражданских заграничных паспортах граждан РФ ранее использовалась французская транскрипция, а теперь — английская).

Для сравнения текстовых атрибутов с фонетическими и транскрипционными искажениями применяются так называемые «фонетические ключи», строящиеся с учетом особенностей языка (см. подразд. 8.2). Так, схоже звучащие имена и фамилии имеют одинаковый фонетический ключ и при поиске сравниваются не сами текстовые поля, а их фонетические ключи.

Для повышения эффективности поиска ФЛ целесообразно также выполнять перевод в латиницу ФИО, исходно заданных в кириллице, а также перевод в кириллицу ФИО, исходно заданных в латинице (см. подразд. 8.3), и сохранять переведенный вариант. Таким образом обеспечивается возможность поиска по ФИО как в кириллице, так и в латинице.

Для расширения возможностей поиска добавляются методы, учитывающие возможные ошибки в несколько букв вследствие опечаток (см. подразд. 8.4).

Установка различных весов для атрибутов ИО позволяет выполнять поиск данных со значительными искажениями.

Задача 4. Задача поиска связей между объектами

Задачи поиска связей между ИО возникают, когда пользователю БД недостаточно той информации, которая непосредственно содержится в признаках ИО, и он хочет иметь некоторую новую информацию, которая может быть получена на основе имеющейся. Например, если у двух физических лиц A и B одинаковы фамилия и место жительства, отчество A соответствует имени B и при этом A достаточно моложе B , то с большой вероятностью можно утверждать, что между A и B прямая родственная связь, хотя об этом в базе данных сведений нет. Этот элементарный пример помогает понять, сколь многообразны могут быть запросы пользователя и как важна разработка способов поиска связей.

Как отмечалось во введении, установка связей дает возможность уточнить гипотезы при идентификации ИО. Например, если при решении задачи идентификации получилось, что объекты x'_1 и x''_1 являются с некоторой вероятностью одним объектом x_1 , а при решении задачи поиска связей выявлено, что объекты x'_1 и x''_1 связаны с объектом y , то этот факт уточняет гипотезу идентификации x'_1 и x''_1 как одного объекта.

Критерии связи между ИО зависят от предметной области и конкретной постановки задачи.

3 Задача слияния информационных объектов

Рассматривается X — множество ИО, каждый из которых задается набором признаков D . Признаки $d \in D$ неоднородны: часть из них числовые, часть — текстовые (номинальные), могут быть и признаки в виде изображений. Кроме того, набор признаков не одинаков для всех ИО, но одинаков для ИО, происходящих из одного источника. Если число источников K , а набор признаков k -го ($1 \leq k \leq K$) источника есть D_k , то $D = \bigcup_{k=1}^K D_k$, а $D_0 = \bigcap_{k=1}^K D_k$, причем D_0 не пусто (рис. 1). Наконец, некоторые признаки могут быть взаимосвязаны. Так, может существовать функция, преобразующая признак d_i , $d_i \in D_i \setminus D_0$, в признак $d_j \in D_j \setminus D_0$.

Рассмотрим последовательно усложняющиеся задачи идентификации ИО.

Простейшая задача — определение того, являются ли x_1 и x_2 , $x_1 \in X$, $x_2 \in X$, $D(x_1) = D(x_2) = D$ идентичными. Очевидно, что с полной уверенностью утверждать, что x_1 и x_2 идентичны, можно в том и только том случае, если значения δ всех признаков $d \in D$ совпадают:

$$\forall i : d_i \in D \quad d_i(x_1) = d_i(x_2).$$

С другой стороны, если значения всех признаков различны, также с полной уверенностью можно утверждать, что x_1 и x_2 не идентичны.

В случаях же, когда значения признаков частично совпадают или же в какой-то степени близки друг другу, суждение об идентичности x_1 и x_2 можно выносить только с некоторой вероятностью P , $0 \leq P \leq 1$.

По оценке вероятности $P(x_1, x_2)$ можно будет принять решение об идентичности x_1 и x_2 , введя критическое значение $P_{кр}$. Если $P(x_1, x_2) \geq P_{кр}$, то выносится решение об идентичности. Само значение $P_{кр}$ выбирается из соображений, лежащих за пределами решаемой задачи: если больше опасаться ошибок второго рода (принять различные ИО за идентичные), то следует выбирать более мощный критерий, если большую опасность представляют ошибки первого рода (не распознать идентичные ИО), то мощность критерия понижают, снижая при этом и уровень значимости.

Функция $P(x_1, x_2)$ является, по сути, мерой близости x_1 и x_2 , оцениваемой по степени близости каждого из признаков. В простейшем случае это делается по правилу средневзвешенного:

$$P(x_1, x_2) = \sum_{i: d_i \in D} \alpha_i \varphi_i(\delta_i^1, \delta_i^2),$$

где α_i — вес, приписываемый признаку d_i , $\sum_i \alpha_i = 1$; φ_i — функция оценки близости по признаку d_i , область значений каждой функции — $[0, 1]$; δ_i^1 и δ_i^2 — значения признака d_i для x_1 и x_2 .

Очевидно, что функции φ_i для разных признаков могут существенно отличаться друг от друга. В разд. 7 настоящей статьи детально рассматриваются особенности функций φ_i для текстовых признаков.

Конечно, могут быть и более сложные выражения для вероятностной меры $P(x_1, x_2)$, учитывающие взаимосвязи между признаками d_i , оказывающие влияние на веса α_i и функции φ_i . Поясним, что имеется в виду. Для наглядности

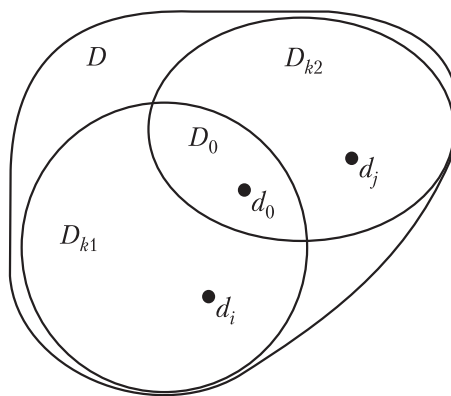


Рис. 1 Пересечение множеств признаков ИО

рассуждений будем проводить их на конкретном примере, когда ИО отражают сведения о ФЛ, а признаками служат «имя», «фамилия», «номер паспорта», «дата рождения» и т. п. Естественно, что такой признак, как «номер паспорта», должен иметь относительно большой вес α , поскольку при совпадении номера паспорта есть очень большие основания считать, что ИО идентичны, даже если в текстовых признаках есть некоторые разночтения. Но, с другой стороны, если остальные признаки совпадают, а номер паспорта совершенно другой, то тоже есть очень большая вероятность идентичности ИО, так как человек мог просто поменять паспорт. И в этом случае следовало бы приписывать признаку «номер паспорта» малый вес. Допустим, что учитывается J ситуаций. Если через α_{ij} обозначить вес i -го признака в j -й ситуации из числа учитываемых, то можно $P(x_1, x_2)$ оценивать по формуле:

$$P(x_1, x_2) = \max_{1 \leq j \leq J} \left(\sum_{i: d_i \in D} \alpha_{ij} \varphi_i(\delta_i^1, \delta_i^2) \right),$$

причем $\forall j \sum_i \alpha_{ij} = 1$.

Еще разумнее было бы пользоваться не единой функцией близости φ_i в различных ситуациях, а ввести набор функций φ_{ij} , но это повлечет за собой рост объема вычислений, что станет препятствием для задач мониторинга в реальном времени.

Особо надо поговорить о виде функций близости φ_i для текстовых признаков. Понятно, что они в первую очередь зависят от количества разночтений между δ_i^1 и δ_i^2 . Но и «качество» разночтений весьма существенно. Интуитивно ясно, что написания «Пугачев» и «Пугачов» ближе друг к другу, чем «Пугачев» и «Пузачев» или «Ругачев», хотя везде разница в одну букву. Для учета этого обстоятельства авторами были введены «фонетические ключи» (см. подразд. 8.1), переводящие исходное написание в последовательность знаков, сравнение которых будет более однородно. Фонетические ключи строятся по правилам, учитывающим фонетические особенности языка. По-разному пишущиеся, но схоже звучащие имена и фамилии имеют одинаковый фонетический ключ, и при поиске учитываются не сами текстовые поля, а их фонетические ключи. Таким образом, меру близости можно вычислять по формуле:

$$P(x_1, x_2) = \max_{1 \leq j \leq J} \left(\sum_{i: d_i \in D} \alpha_{ij} \varphi_i(\psi(\delta_i^1), \psi(\delta_i^2)) \right),$$

где ψ есть функция перевода признака в его фонетический ключ. Для ускорения процесса счета функция ψ применяется однократно при рассмотрении нового ИО и пополняет его признаки.

Усложним исходную задачу. По-прежнему $x_1 \in X$, $x_2 \in X$, но $D(x_1)$ не совпадает с $D(x_2)$, т. е. наборы признаков у x_1 и x_2 различны (рис. 2).

Очевидно, что тогда суммирование надо проводить для тех признаков, которые входят в общую область $D_0 = D(x_1) \cap D(x_2)$. Но выше уже отмечалось, что в остающихся областях $D(x_1) \setminus D_0$ и $D(x_2) \setminus D_0$ могут существовать взаимосвязанные признаки — такие, что существует функция f , переводящая значение признака $d_m^{(1)}(x_1)$ в $d_m^{(2)}(x_2)$, где $d_m^{(1)} \in D(x_1) \setminus D_0$, а $d_m^{(2)} \in D(x_2) \setminus D_0$, а также обратная ей функция f^{-1} , переводящая значение признака $d_m^{(2)}(x_2)$

в $d_m^{(1)}(x_1)$. Речь идет о таких признаках, как «имя кириллицей» и «имя латиницей», причем для ИО, полученных из разных источников, один признак задан, а другой — нет. Но они по определенной процедуре могут друг в друга переводиться. В практических условиях не осуществляется полная однозначность таких преобразований, которая в идеале могла бы быть. Так, при транслитерации имени Ксения по правилам оформления загранпаспортов (приказ МВД РФ № 1047 от 31 декабря 2003 г.) должно получаться Kseniya, но встречаются написания Xenia, Ksenia, Xeniya, причем именно Ksenia рекомендуется приложением № 7 к тому же приказу. Однако использование фонетических ключей позволяет обойти подобную неоднозначность. Таким образом, получаем формулу для вероятностной меры $P(x_1, x_2)$:

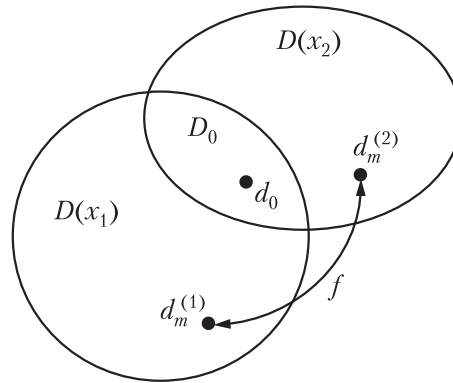


Рис. 2 Пересечение множеств признаков ИО при наличии функциональной связи между признаками

$$P(x_1, x_2) = \max_{1 \leq j \leq J} \left(\sum_{i: d_i \in D_0} \alpha_{ij} \varphi_i (\psi (\delta_i^1), \psi (\delta_i^2)) + \sum_{m: d_m \in D_M} \alpha_{mj} \varphi_m (\psi (\delta_m^1), \psi (\delta_m^2)) \right).$$

Здесь $D_M = \{d_m^{(2)} : \exists f, f(d_m^{(1)}) \Rightarrow d_m^{(2)}, \text{ где } d_m^{(1)} \in D_1 \setminus D_0, d_m^{(2)} \in D_2 \setminus D_0\}$.

Для практической работы алгоритма при принятых процедурах транслитерации f и взятия фонетического ключа ψ необходимо определить значения весов α и оценок близости φ , а также критического значения $P_{кр}$. Это можно сделать только методом экспертных оценок, стремясь подобрать параметры так, чтобы

результаты работы алгоритма были как можно ближе к суждениям человека-эксперта, который взялся бы судить об идентичности x_1 и x_2 по значениям их признаков.

Процесс подбора параметров можно несколько облегчить за счет частичной автоматизации при подготовке материала, делая специальные выборки из общего массива X . Пользуясь ранее отмеченным обстоятельством, что совпадение номера паспорта почти гарантирует идентичность ИО, можно подбирать веса для остальных признаков так, чтобы выводы алгоритма и человека практически совпадали. При работе с такими специальными выборками особо наглядна эффективность (или неэффективность) алгоритма.

Поскольку искажения могут носить не фонетический или транскрипционный характер, а быть вызваны просто опечаткой, следует особым образом рассматривать записи, в которых текстовое поле различается не более чем на определенное число букв, причем более снисходительно трактуются такие частые случаи опечаток, как перестановка букв или замена буквы на другую, находящуюся на соседней клавише.

Можно предусмотреть учет значительных искажений путем приписывания малого веса текстовым атрибутам и большого веса — атрибутам, имеющим численные или дискретные значения.

4 Задача кластеризации информационных объектов

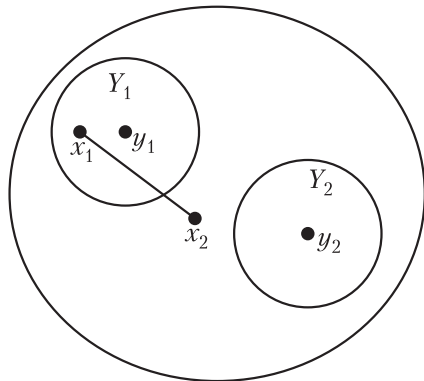


Рис. 3 Кластеризация множества ИО

В задачах кластеризации, в отличие от задач классификации, не указывается заранее требуемое число подмножеств, а также не дается обучающей выборки — число кластеров определяется в результате работы алгоритма. На рис. 3 представлено разбиение множества ИО на кластеры.

Обратим внимание на полную аналогичность циклической задачи внутри алгоритма кластеризации и вышерассмотренной задачи поиска ИО, идентичного новому элементу.

Рассматривается X — множество ИО. Требуется разбить его на непересекающиеся подмножества (кластеры) так, чтобы каждый кластер состоял из объектов, близких по метрике $\rho(x_1, x_2)$, а объекты разных кластеров существенно различались.

В качестве метрики $\rho(x_1, x_2)$ естественно принять дополнение к вероятностной мере идентичности $P(x_1, x_2)$:

$$\rho(x_1, x_2) = 1 - P(x_1, x_2).$$

В задачах кластеризации, в отличие от задач классификации, не указывается заранее требуемое число подмножеств, а также не дается обучающей выборки — число кластеров определяется в результате работы алгоритма.

Действительно, пусть на n -м шаге работы алгоритма кластеризации в рассмотрение берется элемент x_n . Предыдущие $n - 1$ элементов уже разбиты на R кластеров, $R \leq n - 1$.

Введем понятие типичного представителя r -го кластера, $0 \leq r \leq R$. Он представляет собой объединение значений признаков $\delta^{(r)}$, наиболее часто встречающихся среди элементов, составляющих кластер Y_r (или имеющих больше прав представлять данный кластер: так, номер паспорта надо указывать более новый, а не более часто встречающийся):

$$y_r = \bigcup_{d \in D} \delta^{(r)}.$$

После каждого пополнения кластера его типичный представитель пересчитывается.

Производится сравнение x_n со всеми типичными представителями ранее выделенных кластеров и выбирается ближайший кластер:

$$\rho_{\min} = \min_{1 \leq r \leq R} \rho(x_n, y_r); \quad r(n) = \operatorname{argmin}_{1 \leq r \leq R} \rho(x_n, y_r). \quad (1)$$

Если расстояние ρ_{\min} , вычисленное по формуле (1), оказывается достаточно малым ($\rho_{\min} \leq \rho_{\text{кр}}$), элемент x_n включают в кластер с номером $r(n)$, в противном случае он кладет начало новому кластеру Y_{R+1} и служит его типичным представителем.

Критическая величина $\rho_{\text{кр}}$ выбирается из соображений, каких ошибок следует больше опасаться — первого или второго рода. Если нет причин для иного выбора, можно в качестве $\rho_{\text{кр}}$ принять $1 - P_{\text{кр}}$.

5 Корреляционный поиск

Корреляционный поиск — это поиск двух или нескольких объектов, между которыми прослеживается положительная корреляционная связь. Корреляционная связь представляет собой математико-статистическое понятие, и не всегда наличие корреляционной связи свидетельствует о наличии причинно-следственной связи. Тем не менее, в ряде случаев корреляционный поиск может быть полезен именно для предположений о существовании связи по существу.

Например, требуется по одному субъекту события найти иных субъектов, которые чаще других одновременно с ним участвовали в аналогичных событиях. Здесь «одновременность» понимается как расхождение во времени события не более чем на определенный промежуток, а «аналогичность» события — как совпадение определенных атрибутов события.

Коэффициент корреляции в обычном смысле здесь не может быть рассчитан, поскольку рассматривается не случайная величина, а событие, о котором

известно лишь, произошло оно или не произошло. В математической статистике в таких случаях пользуются так называемой «мерой близости». Для ее оценки существует ряд формул, предложенных различными авторами. Мера близости рассчитывается по четырем числам: a — сколько раз произошло и X , и Y ; b — сколько раз произошло X без Y ; c — сколько раз произошло Y без X ; d — не произошло ни X , ни Y .

В рассматриваемом случае число d принципиально не может быть определено, поскольку фиксируются только произошедшие события. Но ряд формул учитывает и такую возможность. Таковы, например, формула Жаккара (Jaccard):

$$\text{JACCARD} = \frac{a}{a + b + c};$$

две формулы Кульчинского:

$$K_1 = \frac{1}{b + c};$$
$$K_2 = \frac{1}{2} \left(\frac{a}{a + b} + \frac{a}{a + c} \right);$$

формула Дайса (Dice):

$$\text{DICE} = \frac{2a}{2a + b + c};$$

вторая формула Сокала и Снита (Sokal & Sneath):

$$\text{SS2} = \frac{a}{a + 2(b + c)};$$

формула Отиаи (Ochiai):

$$\text{OCHIAI} = \sqrt{\frac{a}{a + b} \frac{a}{a + c}}$$

и др. [7].

Вид применяемой формулы и критериальное значение меры близости определяются после решения некоторого числа однотипных задач по экспертным оценкам получаемых результатов. С другой стороны, если сделать критериальное значение плавающим в зависимости от требуемого числа отыскиваемых записей, различие между формулами нивелируется.

6 Фасетный поиск

Фасетный поиск [8] представляет собой пошаговый интерактивный процесс построения формулы запроса.

Каждому объекту массива соответствует его фасетная формула — совокупность признаков, для каждого из которых указаны значения этого признака. Фасетная таблица, сводящая воедино все фасетные формулы, служит базисом для выявления взаимосвязей различных ИО.

Фасетный поиск удобен тем, что на каждом шаге при обращении к конкретному признаку пользователю предоставляется перечень его возможных значений. Выбирая одно или несколько значений из предложенного перечня, пользователь каждый раз получает непустое множество результатов. Последовательно осуществляя управление траекторией поиска, пользователь уточняет формулу запроса, с каждым разом все более приближаясь к искомому объекту.

Фасетный поиск удобен еще и тем, что допускает не только последовательное усечение результирующего множества, но и дополнение его новыми объектами, которые лишь частично отвечают полной формуле запроса, но зато отвечают на ту часть запроса, которая была сделана на последнем шаге. Это позволяет отыскивать объекты, которые ассоциативно связаны с уже включенными в результирующее множество, причем ассоциации могут быть произвольной степени дальности — этот процесс дополнения также идет под контролем пользователя. Таким образом, фасетный поиск может работать и как узко избирательный, и как широкий ассоциативный поиск.

7 Идентификация информационных объектов в потоке событий

Задачу идентификации ИО в потоке событий можно сформулировать следующим образом.

Пусть в текущий момент времени множество ИО X пополняется новым элементом x_{new} . Требуется установить, есть ли во множестве X ИО, идентичные x_{new} . Для решения этой задачи надо пользоваться той же вероятностной мерой $P(x_{\text{new}}, x)$, последовательно подставляя в качестве x все элементы X .

Время расчетов может существенно зависеть от того, требуется ли найти хоть один элемент, идентичный x_{new} , или все такие элементы.

Эта задача решается эффективнее, если предварительно решить задачу кластеризации — все множество X разбить на группы, внутри которых элементы с достаточно большой вероятностью идентичны.

8 Методы сравнения для имен собственных с искажениями

8.1 Исходные положения

Собственные имена при переходе в другую языковую среду часто претерпевают фонетические искажения. Причины этого следующие.

1. Полное и взаимно однозначное соответствие фонетического состава языка и используемого им алфавита — это редкое исключение, а не правило.
2. При этом разные языки, даже использующие одну и ту же графику, для одних и тех же фонем применяют различные способы их передачи. Так, звук «ш» в английском языке передается «sh», во французском «ch», в немецком «sch», в итальянском «sci», в польском «sz» и т. д.
3. Кроме того, в отдельных языках существуют фонемы, в принципе не передаваемые средствами других языков. Так, в испанском существует звук, промежуточный между «б» и «в», в латышском — звук, промежуточный между «к» и «т», и т. д.
4. Случается и обратное явление, когда в данном языке отсутствуют фонемы, широко распространенные в других языках, и, следовательно, в принципе не существует возможности их отображения. Например, в японском языке отсутствует звук «л».
5. В языках, использующих латинский алфавит, принято при передаче иностранных собственных имен сохранять их написание. При этом фактически пренебрегают произношением этих имен.
6. Дополнительным источником искажений является принятая практика отбрасывать диакритические знаки, используемые, например, западнославянскими языками. Так, зачастую чешские буквы «š» (звук «ш») и «č» (звук «ч») переходят соответственно в «s» и «c».
7. В России и во многих других странах, не использующих латинский алфавит, принята, наоборот, транскрипция иностранных имен и названий, когда пытаются в пределах возможного передать произношение, фонетику слова. В тех случаях, когда подлинное произношение неизвестно, пользуются упрощенными правилами транскрипции, отображающими, как это слово должно бы читаться в исходном языке, если оно там читается по правилам.
8. При передаче российских фамилий и названий на латинице используется метод транслитерации — передачи каждой буквы отдельно, по определенным правилам. Методов транслитерации существует много, некоторые закреплены ГОСТом (16976-71, 7.79-2000), но часто используются и ведомственные инструкции. Так, при выдаче загранпаспортов руководствуются «Инструкцией о порядке оформления и выдачи паспортов гражданам Российской Федерации», причем приложение № 6 о бланках паспортов с символикой СССР обязывает придерживаться правил французского написания, как это было принято в Советском Союзе, а приложение № 7 о бланках с российской символикой — правил английского написания.
9. В бывших республиках Советского Союза частично отменена практика передачи фамилий с использованием русского языка в качестве промежуточного. Иногда это ведет к дополнительным искажениям. Так, если раньше в Латвии

окончание «-s», добавляемое к нелатышским фамилиям по чисто грамматическим требованиям латышского языка как показатель именительного падежа, при выдаче загранпаспорта отбрасывалось, то теперь оно сохраняется — по «общеευропейскому» образцу.

С другой стороны, в Азербайджане, перешедшем на латиницу на основе турецкого алфавита (с добавлением трех букв), не придерживаются этой «общеευропейской» практики и при выдаче паспортов заменяют букву «с» на «dzh», букву «i» (без точки) — на «y», букву «ə» — на «a» или «e».

10. Иногда неправомерное использование промежуточного языка приводит к искажениям иного рода. Например, существуют определенные правила перехода от латиницы к кириллице и обратно для языков, не использующих ни ту, ни другую графику, например японского или китайского. Если пытаться прочесть их по-английски, неминуемы ошибки.
11. Для языков, использующих консонантное письмо (еврейское, арабское), характерны вариативные ряды гласных, т. е. одно и то же имя может в зависимости от времени или места произноситься с близкими, но различными гласными: «а» может переходить в «о», «о» в «у», «е» в «и» и т. д.
12. В русском написании восточных имен случаются разночтения одинарного или сдвоенного согласного в положении между гласными: «Хусейн» и «Хуссейн», «Абдурахман» и «Абдуррахман» и т. д.
13. Наконец, еще одной причиной, способной затруднить поиск, является практика неполного следования инструкциям. Так, обычно буква «ё» пишется как «е» и соответственно транслитерируется, мягкий знак часто опускается. Окончание «-ый» нередко записывается как «-у», а не как «-уу».

Из вышеизложенного следует, что поиск с учетом возможных фонетических, транскрипционных и транслитерационных искажений может производиться гораздо качественнее, если известна страна, выдавшая паспорт, а также этническая принадлежность владельца фамилии.

Для более широкой постановки задачи будем считать, что эти сведения отсутствуют. Тогда задача состоит в том, чтобы по данному написанию фамилии (кириллицей или латиницей) предположить, какие ее написания еще возможны:

- (1) если исходная фамилия написана латиницей, прежде всего выясняется, содержатся ли там буквы с диакритическими знаками. Если нет, то, скорее всего, это французское или английское написание (или может таким считаться). Это принимается в качестве основной гипотезы, а в качестве дополнительной — что это упрощенный вариант написания, при котором диакритика была просто отброшена;
- (2) если в слове используются буквы «j», «x» или встречаются сочетания «iou», «ia», «gue», «gui», «guu», «tch», то это французское написание. В противном случае предполагается, что это английское написание.

Для эффективного поиска имен собственных с возможными искажениями, а также с учетом написания имен на разных языках применяются методы построения фонетических ключей, приведенные в подразд. 8.2, и способы перевода из латиницы в кириллицу и из кириллицы в латиницу (см. подразд. 8.3). Метод сравнения строк, учитывающий возможные ошибки в несколько букв вследствие опечаток, приведен в подразд. 8.4.

8.2 Построение фонетических ключей для поиска имен собственных с фонетическими и транскрипционными искажениями

Суть метода построения фонетических ключей состоит в сопоставлении схожим по произношению словам одинаковых кодов, что позволяет осуществлять сравнение таких слов на основе их фонетического сходства. Существует несколько известных алгоритмов построения фонетических ключей в латинице — Soundex, Daitch-Mokotoff Soundex, NYSIIS, Metaphone и др.

Для построения английских фонетических ключей целесообразно применять алгоритм Metaphone (автор — L. Philips, алгоритм создан в 1990 г. [9]), преобразующий исходное слово с учетом правил английского языка, используя заметно более сложные правила, чем в других перечисленных алгоритмах, благодаря чему при преобразовании теряется значительно меньше информации.

Алгоритм основан на разбиении всех букв на группы по звучанию — гласные и согласные, глухие и звонкие. Звонкие согласные преобразуются в соответствующие им парные глухие, объединяются «сливающиеся» при произношении последовательности букв и проводятся некоторые другие действия.

Для построения фонетических ключей русских слов целесообразно использовать алгоритм «Русский Metaphone», предложенный П. Каньковски в 2002 г. [10] и показывающий на практике очень хорошие результаты. Этот алгоритм преобразует исходные слова в соответствии с правилами и нормами русского языка, учитывая фонетическое звучание безударных гласных и возможные «слияния» согласных при произношении. Аналогично алгоритму Metaphone для английских слов все буквы разбиваются на группы по звучанию (гласные и согласные, глухие и звонкие), звонкие согласные преобразуются в соответствующие им глухие, объединяются «сливающиеся» при произношении последовательности букв.

8.3 Перевод из латиницы в кириллицу и из кириллицы в латиницу

Авторами разработаны и реализованы четыре алгоритма перевода имен собственных из одной графики в другую.

Первый алгоритм осуществляет перевод из кириллицы в латиницу в соответствии с принятым в Российской Федерации заполнением бланка паспорта. В основном он следует инструкции МВД, а также учитывает сложившуюся практику заполнения, несколько отступающую от этой инструкции.

Второй алгоритм осуществляет обратный перевод из латиницы в кириллицу на основании тех же правил. Следует заметить, что в инструкции обратный пере-

вод не предусмотрен, и хотя прямой перевод является однозначным, то обратный перевод отнюдь не однозначен. В предлагаемом алгоритме осуществляются наиболее вероятные варианты восстановления кириллического написания исходного имени и фамилии.

Третий алгоритм предназначен для перевода из латиницы в кириллицу для китайских имен. Считается, что китайское имя записано наиболее распространенной системой «пиньинь» и оно переводится в систему Палладия, традиционно используемую для кириллического написания. Послоговый перевод является несколько громоздким, но в принципе затруднения не вызывает, а вот само разбиение на слоги связано со сложностями. Тем не менее оно необходимо, так как одни и те же буквы в начале и конце слога могут обозначать существенно разные фонемы.

Четвертый алгоритм переводит из латиницы в кириллицу японские имена. Предполагается, что японское имя записано по традиционной системе Хэпбёрна и оно переводится в кириллическую систему Поливанова. Авторами предложен оригинальный прием: перевод осуществляется в два этапа — на первом система Хэпбёрна переводится в условную систему на латинской основе, напоминающую ныне не используемую «нихон-сики», для того чтобы кириллизация на втором этапе была полностью единообразной.

8.4 Сравнение строк с ошибками в несколько букв

Вследствие опечаток, а также при внесении текстовых данных (например, ФИО) в БД «со слуха» могут возникать ошибки в несколько букв. При поиске текстовых строк для учета возможности таких ошибок сравнение двух строк целесообразно осуществлять путем вычисления расстояния Левенштейна [11]. Расстояние Левенштейна между двумя строками S_1 и S_2 — это минимальное число операций вставки одного символа, удаления одного символа и замены одного символа на другой, необходимых для преобразования S_1 в S_2 .

В основе алгоритма вычисления расстояния Левенштейна лежит рекуррентная формула:

$$D(i, j) = \min (D(i, j - 1) + 1, D(i - 1, j) + 1, D(i - 1, j - 1) + m(S_1[i], S_2[j])) ,$$

где $S_1[i]$ и $S_2[j]$ — i -й символ строки S_1 и j -й символ строки S_2 соответственно;

$$m(a, b) = \begin{cases} 0, & \text{если } a = b; \\ 1, & \text{если } a \neq b. \end{cases}$$

Представляется разумным (см. положения 11 и 12 в подразд. 8.1) несколько модифицировать рекуррентную формулу. Так, для игнорирования двояких букв необходимо считать

$$D(i, j) = \min (D(i, j - 1) + m(S_2[j - 1], S_2[j]), \\ D(i - 1, j) + m(S_1[i - 1], S_1[i]), D(i - 1, j - 1) + m(S_1[i], S_2[j])) .$$

Кроме того, можно для «родственных» гласных ($a \approx o$, $e \approx и$, $o \approx y$) ввести дробные значения функции m :

$$m(a, b) = \begin{cases} 0, & \text{если } a = b; \\ 0,5, & \text{если } a \approx b; \\ 1 & \text{в остальных случаях.} \end{cases}$$

Для учета опечаток существует модификация расстояния Левенштейна, учитывающая перестановку букв, — расстояние Дамерау–Левенштейна [12]. При этом перестановка букв оценивается не в две редакционные операции, а в одну. Авторы настоящей статьи предложили модифицировать эту функцию еще далее, чтобы учесть такую частую опечатку, как замену одной буквы другой, находящейся на соседней клавише, вводя пониженную цену такой операции: $m(a, b) = 0,5$, если a и b на соседних клавишах.

9 Выводы

1. В работе сформулирована постановка задач идентификации ИО, часто встречающихся на практике при создании больших информационных систем, обрабатывающих данные, поступающие из различных источников.
2. Указаны способы решения задач слияния и кластеризации ИО, отмечена специфика идентификации ИО в потоке событий, приведен метод корреляционного поиска для выявления связи между ИО.
3. Приведены методы сравнения имен собственных с учетом возможных искажений в них — фонетических, транскрипционных, а также просто опечаток. Указана целесообразность использования при идентификации ФЛ данных фамильно-именной группы как в кириллице, так и в латинице и приведены способы перевода слов из кириллицы в латиницу и из латиницы в кириллицу.

Литература

1. Зацаринный А. А., Ионенков Ю. С., Козлов С. В. Некоторые вопросы проектирования информационно-телекоммуникационных систем. — М.: ИПИ РАН, 2010. 218 с.
2. Соколов И. А., Полянский А. В., Киселев Э. В., Сидицын И. Н., Темнов А. И. Проблемы построения информационно-телекоммуникационных систем интегрированного типа. — М.: ИПИ РАН, 2010. 218 с.

3. *Гершкович М. М., Бирюкова Т. К., Сеницын В. И.* Программно-технические решения по организации защищенного информационного обмена в многоуровневых территориально-распределенных системах // Системы и средства защиты информации: Сб. докл. 3-го Межведомственного научно-практич. семинара (Пенза, 7–8 сентября 2011). № 100-дсп. — Пенза: Криптософт, 2012. С. 90–93.
4. *Ушмаев О. С.* Сервисно-ориентированный подход к разработке мультибиометрических технологий // Информатика и её применения, 2008. Т. 2. Вып. 3. С. 41–53.
5. *Гершкович М. М., Бирюкова Т. К., Сеницын В. И.* Проблемы идентификации и распознавания информационных объектов при создании распределенных информационно-телекоммуникационных систем // Оптикоэлектронные приборы и устройства в системах распознавания образов, обработки изображений и символьной информации (Расознавание-2012): Сб. мат-лов X Междунар. конф. — Курск: ЮЗГУ, 2012. С. 24–26.
6. *Гершкович М. М., Бирюкова Т. К., Сеницын В. И.* Основные принципы создания и результаты реализации информационно-телекоммуникационной территориально-распределенной системы в защищенном исполнении // Информационные технологии управления информационными ресурсами двойного применения (V): Препринт / Под ред. акад. И. А. Соколова и проф. И. Н. Сеницына. — М.: ИПИ РАН, 2012. С. 33–46.
7. *Песенко Ю. А.* Принципы и методы количественного анализа в фаунистических исследованиях. — М.: Наука, 1982. 287 с.
8. *Обухова О. Л., Соловьев И. В., Бирюкова Т. К., Гершкович М. М., Чочиа А. П.* Модель фасетного информационного поиска в коллекции научных материалов // Системы и средства информатики. Доп. вып. — М.: Наука, 2009. С. 163–174.
9. *Philips L.* Hanging on the metaphone // Computer Language, 1990. Vol. 7. No. 12. P. 39–43.
10. *Каньковски П.* «Как Ваша фамилия», или Русский MetaPhone // Программист, 2002. № 8. С. 36–39.
11. *Левенштейн В. И.* Двоичные коды с исправлением выпадений, вставок и замещений символов // Докл. АН СССР, 1965. Т. 163. № 4. С. 845–848.
12. *Damerau F. J.* A technique for computer detection and correction of spelling errors // Comm. ACM, 1964. Vol. 7. No. 3. P. 171–176.

Поступила в редакцию 26.02.14

THE TASKS OF IDENTIFICATION OF INFORMATIONAL OBJECTS IN AREA-SPREAD DATA ARRAYS

M. M. Gershkovich and T. K. Birukova

Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str.,
Moscow 119333, Russian Federation

Abstract: An approach for identification of informational objects (IO) in automatic informational systems employed for data collection, storage, and processing is presented. Information systems consist of multiple nodes and acquire

data from multiple sources. In majority of cases, a data array of informational systems is presented as continuously filled event's diary. Each event's record includes characteristics of the event's participant — IO — and of the event's conditions. In order to solve analytical problems related to IO, one should identify IO, i. e., define the array of IOs that are, with certain probability, the same entity. The paper defines typical IO identification tasks for elaboration of large-scale informational systems: IO fusion and IO clustering — forming an aggregate of IOs similar with respect to certain criteria. The identification task is closely connected to the task of identification of links between IOs, as the probability of IO's identity is higher if each IO is associated with another object. The methods for solving these tasks are presented, special features of IO identification in the flow of events are studied, and the correlation search method for detection of associations between IOs is described. The method for comparison of proper names considering probable distortions (phonetic and transcriptional) and misprints is presented. The efficacy of simultaneous Cyrillic and Latin first name–second name blocks application for personal identification is substantiated and the methods for translation from Cyrillic to Latin and *vice versa* are presented.

Keywords: identification of informational objects; identification of objects; correlation search; search for associations; identity of objects; fusion of informational objects; fusion of objects; text attributes; data distortions; phonetic distortions; transcriptional errors; Latin to Cyrillic transcription; Cyrillic to Latin transcription; Metaphone; Levenstein's distance; spread systems; area-spread systems; hierarchical systems; flow of events

DOI: 10.14357/08696527140114

References

1. Zatsarinny, A. A., Y. S. Ionenkov, and S. V. Kozlov. 2010. *Nekotorye voprosy proektirovaniya informatsionno-telekommunikatsionnykh sistem* [Some problems of information-telecommunication systems design]. Ed. A. Zatsarinny. Moscow: IPI RAN. 218 p.
2. Sokolov, I. A., A. V. Polyanskiy, E. V. Kiselev, I. N. Sinitsyn, and A. I. Temnov. 2010. *Problemy postroeniya informatsionno-telekommunikatsionnykh sistem integrirovannogo tipa* [Problems of integrated type information-telecommunication systems design]. Moscow: IPI RAN. 218 p.
3. Gershkovich, M. M., T. K. Biryukova, and V. I. Sinitsyn. 2012. Programmno-tekhnicheskie resheniya po organizatsii zashchishchennogo informatsionnogo obmena v mnogourovnevnykh territorial'no-raspredeleennykh sistemakh [Program-technical solutions for organization of informational exchange within multilevel area distributed systems]. *Sbornik dokladov 3-go Mezhdvdomstvennogo nauchno-prakticheskogo seminara "Sistemy i Sredstva Zashchity Informatsii"* [3rd Interdepartmental Scientific-Technical Seminar 'Systems and Means of Data Protection' Proceedings]. Penza: Izd-vo OOO Nauchno-tekhnicheskoe predpriyatie "Kriptosoft." No. 100-dsp. 90–93.
4. Ushmaev, O. S. 2008. Serviso-orientirovanny podkhod k razrabotke mul'tibiometricheskikh tekhnologiy [Service-oriented approach to multimodal biometrics designing]. *Informatika i ee Primeneniya — Inform. Appl.* 2(3):41–53.

5. Gershkovich, M. M., T. K. Biryukova, and V. I. Sinitsyn. 2012. Problemy identifikatsii i raspoznavaniya informatsionnykh ob"ektov pri sozdanii raspredelennykh informatsionno-telekommunikatsionnykh sistem [Problems of identification and recognition of information's objects in development of area distributed information-telecommunication systems]. *Optiko-elektronnye pribory i ustroystva v sistemakh raspoznavaniya obrazov, obrabotki izobrazheniy i simvol'noy informatsii. Raspoznavanie 1012: Sbornik materialov X Mezhdunarodnoy Nauchno-Tekhnicheskoy Konferentsii* [Optical-Electronic Devices in Image Recognition, Image and Symbol Information Processing Systems, Recognition, 2012: X Scientific-Technical Conference (International) Proceedings]. Kursk: South-West State University. 24–26.
6. Gershkovich, M. M., T. K. Biryukova, and V. I. Sinitsyn. 2012. Printsipy sozdaniya i rezul'taty realizatsii informatsionno-telekommunikatsionnoy territorial'no-raspredelennoy sistemy v zashchishchennom ispolnenii [Principles of design and implementation results of information-telecommunication area distributed systems in secure execution]. *Informatsionnye Tehnologii Upravleniya Informatsionnymi Resursami Dvoynogo Primeneniya (VI)*. [Dual use informational technologies of informational resources management]. Eds. I. A. Sokolov and I. N. Sinitsyn. Moscow: IPI RAN. Preprint. 33–46.
7. Pesenko, Ju. A. 1982. *Printsipy i metody kolichestvennogo analiza v faunisticheskikh issledovaniyakh* [Principles and methods of quantitative analysis in faunal studies]. Moscow: Nauka. 287 p.
8. Obuhova, O. L., I. V. Solovyov, T. K. Biryukova, M. M. Gershkovich, and A. P. Chochia. 2009. Model' fasetnogo informatsionnogo poiska v kolleksii nauchnykh materialov [The model for facet informational search in the collection of the scientific materials]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics*. Additional issue. Moscow: Nauka. 163–174.
9. Philips, L. 1990. Hanging on the metaphone. *Computer Language* 7(12):39–43.
10. Kan'kovski, P. 2002. "Kak Vasha familiya", ili Russkiy MetaPhone ['What is your second name,' or Russian Methaphone]. *Programmist* [Programmer] 8:36–39.
11. Levenshtejn, V. I. 1965. Dvoichnye kody s ispravleniem vypadeniy, vstavok i zameshcheniy simvolov [Binary codes with correction of symbols withdrawal, insertions, and substitutions]. *Dokl. USSR Akad. Sci.* 163(4):845–848.
12. Damerau, F. J. 1964. A technique for computer detection and correction of spelling errors. *Comm. ACM* 7(3):171–176.

Received February 26, 2014

Contributors

Gershkovich Maxim M. (b. 1968) — senior scientist, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; makmg@mail.ru

Biryukova Tatiana K. (b. 1968) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; yukonta@mail.ru

БИБЛИОГРАФИЧЕСКАЯ ИНФОРМАЦИОННО-АНАЛИТИЧЕСКАЯ СИСТЕМА ИПИ РАН

М. Ю. Заикин¹, О. Л. Обухова², И. В. Соловьев³

Аннотация: Дан обзор библиометрических показателей, определяющих результативность публикационной деятельности научных сотрудников. Задачу объединения библиографической информации о публикациях сотрудников ИПИ РАН в едином ресурсе выполняет библиографическая информационно-аналитическая система (БИАС) ИПИ РАН. Источниками заполнения базы данных (БД) БИАС служат реферативные базы научного цитирования Web of Science (WoS), Scopus и Российский индекс научного цитирования (РИНЦ), а также авторитетные тематические интернет-ресурсы, находящиеся в открытом доступе. Программная система БИАС обеспечивает формирование библиометрических показателей научных сотрудников на базе полной библиографии авторов, которая включает не только журнальные публикации, но и монографии, учебники, доклады на конференциях. Организация доступа к информации осуществляется на уровне пользовательского интерфейса посредством формирования запросов по специальным критериям. Также в статье приведены диаграммы сравнения показателей публикационной активности научных сотрудников по версии БИАС ИПИ РАН и РИНЦ.

Ключевые слова: библиометрические показатели; индекс цитирования; индекс Хирша; показатель цитируемости

DOI: 10.14357/08696527140115

1 Введение

Важным компонентом научно-исследовательской деятельности является представление результатов работы в научных статьях. Публикации, вызывающие интерес, цитируются другими авторами. Библиометрический анализ цитирования позволяет определить уровень востребованности и значимости научной работы. Вычисляемые *библиометрические показатели* являются количественной характеристикой продуктивности ученого.

1.1 Библиометрические показатели

Показатель цитируемости ученого отображает общее число распределенных по годам ссылок на публикации, где данный исследователь фигурирует в качестве автора или соавтора.

¹Институт проблем информатики Российской академии наук, maxim@amsd.com

²Институт проблем информатики Российской академии наук, obuhova@amsd.com

³Институт проблем информатики Российской академии наук, soloviev@amsd.com

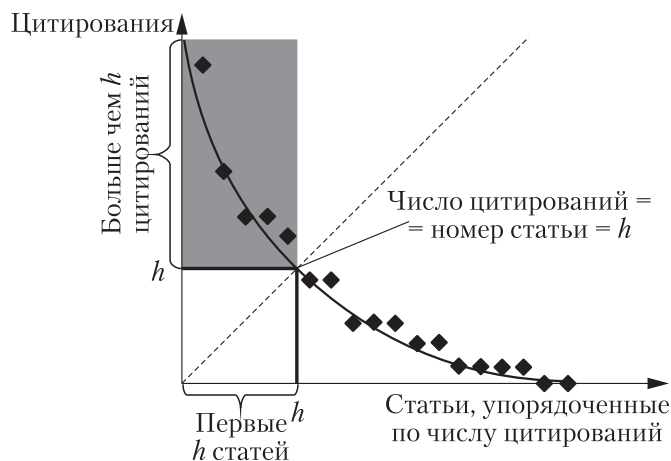


Рис. 1 Вычисление h -индекса

Индекс Хирша (h -индекс), предложенный американским физиком Jorge E. Hirsch [1], основывается на количестве публикаций ученого и числе цитирований этих публикаций. Индекс вычисляется на основе распределения цитирований работ данного исследователя по схеме: «ученый имеет индекс h , если h из его N_p статей цитируются как минимум h раз каждая, в то время как оставшиеся $(N_p - h)$ статей цитируются не более чем h раз каждая».

Графическая иллюстрация вычисления h -индекса приведена на рис. 1.

Импакт-фактор (ИФ, или IF) — численный показатель влиятельности научного журнала, измеряемый путем вычисления среднего числа цитирований на статью за определенный промежуток времени. С 1960-х гг. он ежегодно рассчитывается Институтом научной информации (Institute of Scientific Information, ISI) США для ведущих изданий мира и публикуется в журнале *Journal Citation Report (JCR)* [2].

Расчет импакт-фактора основан на трехлетнем периоде. Например, импакт-фактор журнала в 2013 г. I_{2013} вычислен следующим образом:

$$I_{2013} = \frac{A}{B},$$

где A — число цитирований в течение 2013 г. в журналах, отслеживаемых ISI, статей, опубликованных в данном журнале в 2011–2012 гг.; B — число статей, опубликованных в данном журнале в 2011–2012 гг.

Таким образом, импакт-фактор является мерой, определяющей частоту, с которой цитируется среднестатистическая статья журнала.

1.2 Индексы цитирования

Информационной базой для получения библиометрических показателей служат реферативно-библиографические БД научного цитирования, которые для краткости называются *индексами цитирования*.

Изначально термин «индекс цитирования» ввел Ю. Гарфилд, основатель ISI, внесший серьезный вклад в создание методологии и инструментария библиометрического анализа [3]: «...это упорядоченный список цитированных статей, каждая из которых сопровождается списком цитирующих статей. Цитирующая статья определяется как источник, а цитируемая — как ссылочная. Любая статья-источник может впоследствии стать ссылочной».

В настоящее время понятие «индекса цитирования» используется для обозначения реферативной БД научных публикаций, индексирующей ссылки, указанные в приставных списках этих публикаций, и предоставляющей информацию о количественных показателях, таких как суммарный показатель цитируемости, индекс Хирша. Такая БД позволяет находить как публикации, цитируемые в отдельно взятой статье, так и публикации, цитирующие эту статью. Таким образом, пользователь может проводить эффективный поиск всей библиографии по интересующему его вопросу.

1.3 Библиометрические показатели как оценка эффективности научных исследований

В научном сообществе нет единого мнения по вопросу, правомерно ли использовать показатели цитируемости для определения плодотворности работы ученых. Использование статистики цитирований в качестве основной меры качества исследований не учитывает смысла цитирования, которое само по себе достаточно субъективно [4]. В отчете Международного математического союза, посвященном вопросам статистической оценки труда ученых [5], дано заключение: «Данные цитирований дают лишь ограниченное и неполное представление о качестве научных исследований».

Но, тем не менее, Министерство образования и науки Российской Федерации издало приказ [6] об использовании для оценки деятельности научных коллективов и отдельных ученых библиометрических показателей, которые определяются на основании данных РИНЦ и международных реферативных БД.

2 Международные и отечественные реферативные базы научного цитирования

На мировом рынке научной информации наиболее известны и широко представлены международные индексы цитирования: WoS компании Thomson Reuters (США) и Scopus компании Elsevier (Нидерланды). Оба эти проекта являются коммерческими, доступ к ним осуществляется на платной основе. Кроме того,

значительное число изданий, индексируемых в базах WoS и Scopus, публикует статьи не исходя из критериев качества публикаций, а на платной основе, что существенно снижает доверие к соответствующим индексам, о чем в настоящее время имеется много публикаций в мировой научной печати.

Задача индексирования публикаций российских ученых и учет их цитирования выполняется РИНЦ, БД которого доступна на платформе eLibrary.ru.

2.1 Web of Science

Thomson Reuters WoS [7] — реферативные БД, содержащие каталоги публикаций в научных журналах и ссылок на них. Web of Science входит в состав комплекса сервисов библиометрического анализа ISI Web of Knowledge [8].

К недостаткам WoS как *международного* индекса цитирования относится принцип подбора рецензируемых журналов с явным предпочтением изданиям США. Диаграмма распределения по странам научных журналов, индексируемых WoS, публикуемого в журнале *JCR* компании Thomson Reuters при поддержке ISI Web of Knowledge [9], представлена на рис. 2.

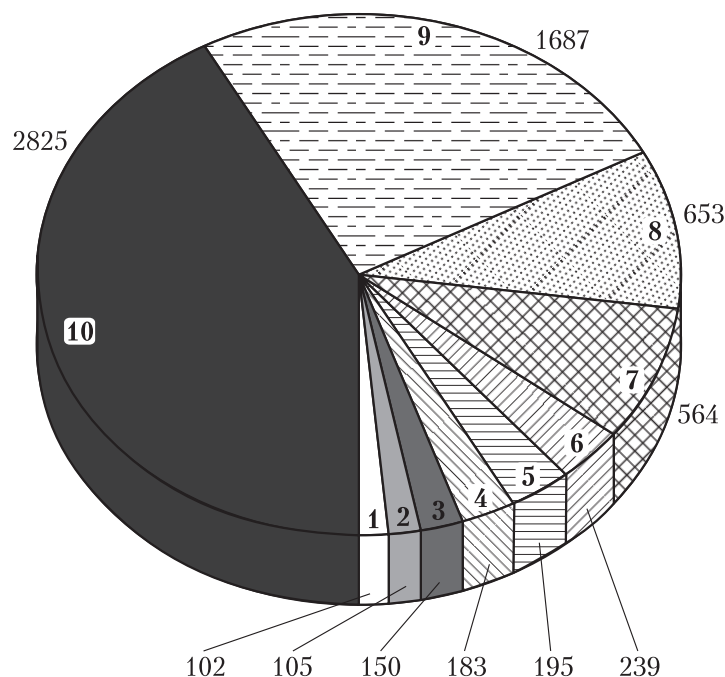


Рис. 2 Диаграмма распределения научных журналов, индексируемых WoS, по странам: 1 — Бразилия; 2 — Индия; 3 — Россия; 4 — Швейцария; 5 — Франция; 6 — Япония; 7 — Германия; 8 — Нидерланды; 9 — Англия; 10 — США

Список отечественных журналов состоит из 150 наименований по всем научным дисциплинам. Для сравнения список рекомендованных Высшей аттестационной комиссией (ВАК) рецензируемых научных журналов для опубликования основных научных результатов диссертаций на соискание ученых степеней доктора и кандидата наук насчитывает более 2000 названий.

Особенно остро вопрос недостаточного представительства в WoS касается публикаций в области информатики и компьютерных технологий. В статье [10] авторы доказали, что **66% публикаций по информатике не учитывается в WoS.**

2.2 Scopus

Scopus — мультидисциплинарная реферативная база, которая включает в себя 50 млн записей (из которых 70% содержат рефераты) из 21 000 наименований изданий, включающих более 20 000 рецензируемых журналов (в том числе 2600 журналов Open Access и около 300 российских журналов на английском языке), 5,5 млн докладов конференций. Это самая крупная в мире реферативная БД, которая обновляется ежедневно [11].

2.3 Российский индекс научного цитирования

Задача индексирования публикаций российских ученых и учет их цитирования в достаточно представительном объеме стоит перед РИНЦ, БД которого представлена на веб-ресурсе eLibrary.ru. РИНЦ находится в стадии становления, но уже сейчас является серьезным инструментом учета и изучения публикационной активности российских ученых, в том числе и в зарубежных изданиях.

Реферативные базы научного цитирования WoS, Scopus и РИНЦ, используемые для подсчета рейтингов как отдельных ученых, так и научных институтов, индексируют только часть работ авторов. Поэтому представляется интересной и важной задачей аккумуляция всех данных о публикационной деятельности научных сотрудников НИИ в едином ресурсе. С этой целью была разработана БИАС ИПИ РАН [12].

3 Структура БИАС ИПИ РАН

3.1 Цели и задачи

Программная система БИАС предназначена для сбора, хранения и предоставления веб-доступа к информации, включающей:

- полные списки научных трудов сотрудников ИПИ РАН;
- списки цитирований публикаций авторов.

Научными трудами считаются журнальные статьи в отечественных и зарубежных изданиях, монографии, учебники и учебные пособия, отчеты по темам НИР, объекты интеллектуальной собственности, диссертации, публикации докладов и тезисов докладов в материалах конференций.

В задачи системы входит формирование библиометрических показателей публикационной активности: показателей цитируемости и индекса Хирша автора. Важной задачей БИАС является формирование статистических отчетов о публикационной деятельности ИПИ РАН.

Программная система состоит из трех компонентов:

- (1) административного модуля, предназначенного для заполнения и редактирования БД публикаций;
- (2) модуля пользовательского интерфейса, обеспечивающего доступ к данным;
- (3) модуля формирования статистических отчетов.

3.2 Программные средства

Программная система БИАС реализована в виде серверного приложения, функционирующего в среде веб-сервера.

Используются следующие программные средства:

- среда разработки веб-приложений: Yii PHP framework (Yiiframework 1.1.10);
- язык разработки: PHP 5.3.13;
- веб-сервер: Apache HTTP Server 2.2.22;
- СУБД: Microsoft SQL Server 2005.

В статье представлена эксплуатационная модель, проходящая опытное тестирование.

3.3 Административный режим

Это привилегированный режим работы системы, вход доступен для зарегистрированных пользователей с соответствующими полномочиями. К привилегированным пользователям относятся, в частности, администраторы и модераторы системы.

В административном режиме происходит занесение и редактирование информации в БД:

- об авторах,
- о публикациях,
- об изданиях,
- об издательствах,
- о типах публикаций.

Форма для заполнения информации о публикациях

С целью наиболее полного представления публикационной деятельности авторов при заполнении БД БИАС используются разнообразные источники, как это представлено на рис. 3.

Поскольку информация о публикациях заносится из источников, имеющих различающиеся форматы библиографической информации, а также неточности и ошибки в библиографических описаниях, формирование точной информации о публикациях требует ручной доработки модераторами системы. В административном режиме используется специальная форма занесения и редактирования записей о публикациях, которая представлена на рис. 4.

Каждая публикация в БД определяется следующим набором свойств:

- название;
- тип публикации;
- издание (совместно с выходными данными публикации);
- год;
- список авторов;
- список библиографических ссылок;



Рис. 3 Источники заполнения БИАС

БИАС ИПИ РАН
Выход (Olga)

[Главная](#) [База](#) [Пользователи](#)

[Главная](#) > [Статьи](#) > [АЛГОРИТМ РАСЧЕТА ПОКАЗАТЕЛЕЙ ПРОИЗВОДИТЕЛЬНОСТИ ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ С ПОВТОРНЫМИ ПЕРЕДАЧАМИ](#)

Редактирование статьи №1868

*Поля помеченные * должны быть заданы.*

название *

тип публикации

издание

издатель

ISBN

год

том

номер

глава

страницы

ссылка в internet

список авторов

список литературы

ключевые слова

аннотация

внешние индексы

[Список статей](#)
[Новая статья](#)
[Просмотр статьи](#)
[Управление](#)

Рис. 4 Форма занесения и редактирования записей о публикациях

– признаки индексации, к которым относятся:

- признак индексации в WoS;
- признак индексации в Scopus;
- признак индексации в РИНЦ на платформе eLibrary;
- признак издания ВАК.

Данные свойства публикаций используются для задания логических поисковых запросов к БД для формирования статистических отчетов, отражающих публикационную активность научных сотрудников ИПИ РАН.

3.4 Концептуальная модель базы данных

Для реализации задач веб-доступа к информации, определяющей публикационную активность авторов, разработана концептуальная модель БД, в которой определены типы объектов и принципы построения связей между ними.

Для описания типов объектов применяются методы фасетной навигации, разработанной авторами [13, 14]. Информационные объекты (ИО) «публикация» представляются фасетной формулой объектов (ФФО) в виде множества совокупностей: ⟨фасетный признак: список значений⟩. Каждый фасетный признак с множеством значений определяет объект «фасет». Объекты одного типа определяют класс. В предлагаемой модели данных определены классы.

SpecializedDataEntity — класс, экземплярами которого являются ИО «публикация». Значения однородных свойств объекта «публикация» сгруппированы в классы $\text{Facet}_1, \dots, \text{Facet}_n$.

Введем следующие обозначения:

1. Множество $X = \{X_j^i | i = 0, \dots, n, j = 1, \dots, n_i\}$, где индекс i определяет номер класса, а j определяет номер экземпляра данного класса, определяет множество всех объектов.
2. X_m^0 — экземпляры класса **SpecializedDataEntity**, $m = 1, \dots, m_0$, где m_0 — количество экземпляров данного класса.
3. X_k^f — экземпляры классов **Facet_f**, $f = 1, \dots, n$, $k = 1, \dots, k_f$, k_f — количество экземпляров соответствующего класса.
4. Одноместный предикат $S(X_m^0, X_k^f)$ обозначает, что объект X_m^0 имеет свойство X_k^f , где $m \subseteq \{1, \dots, m_0\}$; $f \subseteq \{1, \dots, n\}$; $k \subseteq \{1, \dots, k_f\}$.

Поскольку понятия «бинарное отношение» и «признак (свойство)» идентичны в случае, если значением свойства является объект другого класса, то определим итерационную модель построения связей объектов.

Шаг 1. Двухместный предикат $R_{f_1 f_2}^2(X_{i_1}^{f_1}, X_{i_2}^{f_2})$ обозначает бинарное отношение двух объектов $X_{i_1}^{f_1}$ и $X_{i_2}^{f_2}$ классов **Facet_{f₁}** и **Facet_{f₂}**, где $f_1 \neq f_2$,

$i_1 \subseteq \{1, \dots, k_{f_1}\}$, $i_2 \subseteq \{1, \dots, k_{f_2}\}$, k_{f_1} , k_{f_2} — количество объектов соответствующего фасета.

Предикат $R_{f_1 f_2}^2(X_{i_1}^{f_1}, X_{i_2}^{f_2})$ принимает значение «истина» (для краткости записи: = true), если \exists объект X_m^0 , для которого заданы свойства $X_{i_1}^{f_1}$ и $X_{i_2}^{f_2}$, т. е. существуют $S(X_m^0, X_{i_1}^{f_1})$ и $S(X_m^0, X_{i_2}^{f_2})$.

Иными словами, два свойства связаны, если существует такой ИО, у которого задано каждое из этих свойств.

Шаг 2. Трехместный предикат $R_{f_1 f_2 f_3}^3(X_{i_1}^{f_1}, X_{i_2}^{f_2}, X_{i_3}^{f_3})$, где $f_1, f_2, f_3 \subseteq \{1, \dots, n\}$, $i_1 \subseteq \{1, \dots, k_{f_1}\}$, $i_2 \subseteq \{1, \dots, k_{f_2}\}$, $i_3 \subseteq \{1, \dots, k_{f_3}\}$, k_{f_1} , k_{f_2} , k_{f_3} — количество объектов соответствующего фасета, выражающий отношение трех объектов, принимает значение «истина», если $R_{f_1 f_2}^2(X_{i_1}^{f_1}, X_{i_2}^{f_2}) = \text{true}$ и \exists объект $X_{i_3}^{f_3}$ такой, что

$$R_{f_1 f_3}^2(X_{i_1}^{f_1}, X_{i_3}^{f_3}) = \text{true};$$

$$R_{f_2 f_3}^2(X_{i_2}^{f_2}, X_{i_3}^{f_3}) = \text{true}.$$

Шаг n. N -местный предикат $R_{f_1 f_2 \dots f_N}^N(X_{i_1}^{f_1}, X_{i_2}^{f_2}, \dots, X_{i_N}^{f_N})$, где $f_1, f_2, \dots, f_N \subseteq \{1, \dots, n\}$, $i_1 \subseteq \{1, \dots, k_{f_1}\}$, $i_2 \subseteq \{1, \dots, k_{f_2}\}$, \dots , $i_N \subseteq \{1, \dots, k_{f_N}\}$, $k_{f_1}, k_{f_2}, \dots, k_{f_N}$ — количество объектов соответствующего класса, выражающий отношение N объектов, принимает значение истина, если

$$R_{f_1 f_2 \dots f_{N-1}}^{N-1}(X_{i_1}^{f_1}, X_{i_{N-1}}^{f_{N-1}}) = \text{true}$$

и \exists объект $X_{i_N}^{f_N}$ такой, что для $\forall X_{i_j}^{f_j}$, $j \in \{1, \dots, N-1\}$,

$$R_{f_j}^2(X_{i_j}^{f_j}, X_{i_N}^{f_N}) = \text{true}.$$

Построенная модель обеспечивает выполнение следующего алгоритма поиска ИО по заданной непротиворечивой совокупности свойств.

Этап 1. Фиксируем свойство X_1^f .

Выбираем подмножество ИО, у которых задано данное свойство. В каждом из фасетов остаются только те объекты, у которых существует бинарное отношение с X_1^f .

Этап 2. Фиксируем следующее свойство X_2^f .

Подмножество ИО определяется условием того, что у всех объектов задано каждое из свойств X_1^f и X_2^f . Соответственно в каждом из фасетов остаются только те объекты, у которых существует бинарное отношение с X_1^f и с X_2^f .

Число этапов выбора очередного свойства определяется количеством свойств. На каждом этапе получаем подмножество ИО, имеющих заданный набор свойств. Каждый фасет отображает множество допустимых значений свойств сформированного подмножества ИО.

Преимущество данной модели заключается в том, что все связи вычисляются динамически, что позволяет расширять список свойств, не меняя программную реализацию. И второе преимущество заключается в том, что последовательность выбора свойств может быть произвольной. Но при этом на каждом этапе обеспечивается выбор только допустимых свойств.

3.5 Пользовательский интерфейс

Пользовательский интерфейс, представленный на рис. 5, позволяет осуществлять поиск научных трудов по следующим признакам:

The screenshot shows the BIAS IPI RAN search interface. At the top, there is a header with the site name and a 'Выход (Olga)' link. Below the header are navigation tabs: 'Главная', 'База', and 'Пользователи'. The main area is divided into two columns. The left column contains search filters: 'Автор' (with a search box and magnifying glass), 'Год' (with two input boxes), 'Направление' (with a dropdown menu), 'Подразделение' (with a dropdown menu), 'Внешние индексы' (with 'eLibrary' selected), 'AND' (with a dropdown menu), and 'ВАК' (with a dropdown menu). There is a '+ Добавить условие' link and a 'Найти' button. The right column displays search results. At the top right of this column is a 'Название' search box and a 'Найти' button. Below it is an 'Экспорт' button. The results are listed as follows:
1. EXPERIENCE IN CONSTRUCTING ASTROMETRICAL MEASUREMENTS-BASED MODELS OF EARTH POLE LUCTUATIONS 2010
Синицын Игорь Николаевич, Корепанов Эдуард Рудольфович, Семендяев Николай Николаевич, Марков Ю.Г., Синицын Владимир Игоревич
2. SOME IMPROVED METHODS FOR MODELING THE EARTH'S POLAR MOTION 2010
Синицын Игорь Николаевич, Марков Ю.Г., Рыжлова Л.В.
3. PROBABILISTIC MODELS OF THE FLUCTUATIONS IN IRREGULARITY OF THE EARTH'S ROTATION UNDER NONSTATIONARY 2010
Синицын Игорь Николаевич, Марков Ю.Г.
4. РАЗВИТИЕ МЕТОДОВ ПОСТРОЕНИЯ МОДЕЛЕЙ ДВИЖЕНИЯ ПОЛЮСА ЗЕМЛИ 2010
Синицын Игорь Николаевич, Марков Ю.Г., Рыжлова Л.В.
5. ВЕРоятностные модели флуктуаций неравномерности вращения Земли при нестационарных возмущениях 2010
Синицын Игорь Николаевич, Марков Ю.Г.
At the bottom of the results section, there is a 'Страницы' section with links: 'Предыдущая', '1', '2', '3', '4', '5', 'Следующая'.

Рис. 5 Пользовательский интерфейс

- по названию;
- по автору, в том числе среди соавторов;
- по году;
- по направлению научных исследований ИПИ РАН;
- по подразделению;
- по признакам индексации.

3.6 Режим формирования отчетов

Это привилегированный режим работы системы. Содержит форму создания запросов на выдачу статистических отчетов. Список параметров, по которым формируется статистический отчет, определяется динамически в зависимости от поставленной задачи. Вывод осуществляется в excel-файл.

4 Заключение

В статистических отчетах БИАС ИПИ РАН формируются сведения о публикационной деятельности авторов. Каждая публикация сопровождается информацией об индексации в базах научного цитирования WoS, Scopus, РИНЦ. Публикации в изданиях, входящих в перечень ВАК, помечаются соответствующим признаком. Для каждой статьи указывается импакт-фактор журнала по версии JCR для иностранных изданий и по версии РИНЦ для отечественных изданий.

К достоинствам БД БИАС ИПИ РАН нужно отнести тот факт, что в ней собрана полная и достоверная информация о публикациях сотрудников института, включая доклады и тезисы докладов на международных и отечественных конференциях. Поскольку для таких стремительно развивающихся областей знаний, как информатика, конференции — наиболее благоприятная и оперативная среда для представления и обсуждения новых идей, включение библиографической информации о докладах в БД является фактором полноты наполнения БД.

Диаграммы сопоставления результатов заполнения БД РИНЦ и БД публикаций сотрудников ИПИ РАН (БИАС ИПИ РАН) приведены на рис. 6 и 7.

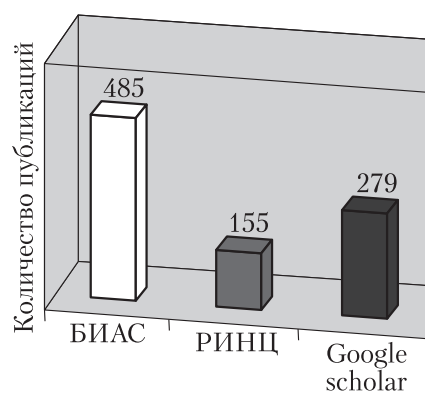


Рис. 6 Диаграммы представленности публикаций сотрудников ИПИ РАН в БИАС, РИНЦ и Google Scholar



Рис. 7 Диаграммы представленности публикаций сотрудников ИПИ РАН в БИАС (1) и РИНЦ (2) по категориям

Модульная структура построения системы БИАС позволяет добавлять новые функциональные блоки в процессе работы системы, не влияя на существующие и отлаженные сервисы. Опытная эксплуатация позволит определить направления развития системы.

Литература

1. *Hirsch J.E.* An index to quantify an individual's scientific research output // *Proceeding of the National Academy of Sciences USA*, 2005. Vol. 102. No. 46. P. 16569–16572.
2. *Journal Citation Report*. http://wokinfo.com/products_tools/analytical/jcr.
3. *Garfield E.* Science Citation Index — a new dimension in indexing // *Science*, 1964. Vol. 144. No. 3619. P. 649–654.
4. *Коголовский М. Р., Паринов С. И.* Классификация и использование семантических связей между информационными объектами в научных электронных библиотеках // *Информатика и её применения*, 2012. Т. 6. Вып. 3. С. 32–42.
5. *Citation statistics: A report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics*. <http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf>.
6. Об утверждении типового положения о комиссии по оценке результативности деятельности научных организаций, выполняющих научно-исследовательские, опытно-конструкторские и технологические работы гражданского назначения, и типовой методики оценки результативности деятельности научных организаций, выполня-

- ющих научно-исследовательские, опытно-конструкторские и технологические работы гражданского назначения: Приказ Министерства образования и науки Российской Федерации № 406 от 14 октября 2009 г. <http://poisk-zakona.ru/35954.html>.
7. Web of Science. http://apps.webofknowledge.com/WOS_GeneralSearch_input.do.
 8. ISI Web of Knowledge. <http://isiwebofknowledge.com>.
 9. Source publication list for Web of Science[®]. http://ip-science.thomsonreuters.com/mjl/publist_sciex.pdf.
 10. *Wainer J., Goldenstein S., Billa C.* Invisible work in standard bibliometric evaluation of computer science // *Comm. ASM*, 2011. Vol. 54. No. 5. P. 141–146.
 11. Scopus. http://www.elsevier.com/_data/assets/pdf_file/0007/148714/scopus_facts_and_figures.pdf.
 12. БИАС ИПИ РАН. <http://ipigate2.portal.amsd.com>.
 13. *Обухова О. Л., Соловьев И. В., Чочиа А. П.* Технология адаптивной фасетной навигации в открытых электронных коллекциях: Свидетельство о регистрации программ № 2008614955, 2008.
 14. *Обухова О. Л., Соловьев И. В., Бирюкова Т. К., Гершкович М. М., Чочиа А. П.* Модель фасетного информационного поиска в коллекции научных материалов // *Системы и средства информатики. Доп. вып.* — М.: Наука, 2009. С. 163–174.

Поступила в редакцию 31.07.13

BIBLIOGRAPHIC INFORMATION-ANALYTICAL SYSTEM OF IPI RAS

M. Yu. Zaikin, O. L. Obuhova, and I. V. Soloviev

Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str.,
Moscow 119333, Russian Federation

Abstract: The article gives an overview of bibliometric indicators that determine effectiveness of publication activity of researchers. The Bibliographic Information-Analytical System (BIAS) of IPI RAS performs the task of combining bibliographic information about publications of IPI RAS staff. The general sources for the BIAS database consist of the bibliographic Science Citation database Web of Science, Scopus, the Russian Science Citation Index, and authoritative themed online resources with open access. The BIAS software system calculates the bibliometric indicators based on a full bibliography of a scientist which includes not only journal publications, but also monographs, textbooks, and conference reports. Access to information is given by forming special queries by means of a user interface. Examples of comparison of the bibliometric indicators of IPI RAS BIAS and the Russian Science Citation Index are given.

Keywords: bibliometric indicators; citation index; h-index; science citation

DOI: 10.14357/08696527140115

References

1. Hirsch, J. E. 2005. An index to quantify an individual's scientific research output. *Proceeding of the National Academy of Science USA* 102(46):16569–16572.
2. Journal Citation Report. Available at: <http://ip-science.thomsonreuters.com/cgi-bin/jrnlst/jlresults.cgi?PC=A> (accessed April 23, 2014).
3. Garfield, E. 1964. Science Citation Index — a new dimension in indexing. *Science* 144(3619):649–654.
4. Kogalovskij, M. R., and S. I. Parinov. 2012. Klassifikatsiya i ispol'zovanie semanticheskikh svyazey mezhdru informatsionnymi ob"ektami v nauchnykh elektronnykh bibliotekakh [Classification and utilization of semantic linkages between information objects in scientific digital libraries]. *Informatika i ee Primeneniya — Inform. Appl.* 6(3):32–42.
5. A report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics "Citation statistics." Available at: <http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf> (accessed April 23, 2014).
6. Prikaz Ministerstva obrazovaniya i nauki Rossiyskoy Federatsii ot 14 oktyabrya 2009 g. N 406 "Ob utverzhdenii tipovogo polozheniya o komissii po otsenke rezul'tativnosti deyatel'nosti nauchnykh organizatsiy, vypolnyayushchikh nauchno-issledovatel'skie, opytно-konstruktorskie i tekhnologicheskie raboty grazhdanskogo naznacheniya, i tipovoy metodiki otsenki rezul'tativnosti deyatel'nosti nauchnykh organizatsiy, vypolnyayushchikh nauchno-issledovatel'skie, opytно konstruktorskie tekhnologicheskie raboty grazhdanskogo naznacheniya." Zaregistrirovan v Minyust Rossii ot 28 yanvarya 2010 g. N 16115. Available at: <http://mon.gov.ru/> (accessed April 23, 2014).
7. Web of Science. Available at: http://apps.webofknowledge.com/WOS_GeneralSearch_input.do (accessed April 23, 2014).
8. ISI Web of Knowledge. Available at: <http://isiwebofknowledge.com> (accessed April 23, 2014).
9. Source Publication List for WoS[®]. Available at: http://ip-science.thomsonreuters.com/mjl/publist_sciex.pdf (accessed April 23, 2014).
10. Wainer, J., C. Billa, and S. Goldenstein. 2011. Invisible work in standard bibliometric evaluation of computer science. *Comm. ACM* 54(5):141–146.
11. Scopus. Available at: http://www.elsevier.com/_data/assets/pdf_file/0007/148714/scopus_facts_and_figures.pdf (accessed April 23, 2014).
12. BIAS IPI RAS. Available at: <http://ipigate2.portal.amsd.com/> (accessed April 23, 2014).
13. Obukhova, O., I. Soloviev, and A. Chochia. 2008. Tekhnologiya adaptivnoy fasetnoy navigatsii v otkrytykh elektronnykh kollektsiyakh [Technology of adaptive facet navigation in open digital collections]. Certificate No. 2008614955.
14. Obukhova O., T. Biryukova, I. Soloviev, M. Gershkovich, and A. Chochia. 2009. Model' fasetnogo informatsionnogo poiska v kollektzii nauchnykh materialov [A model of facet information retrieval within the collection of scientific materials]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics*. Moscow: Nauka. 19(Special Issue):163–174.

Received July 31, 2013

Contributors

Zaikin Maxim Yu. (b. 1973)— scientist, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; maxim@amsd.com

Obuhova Olga L. (b. 1952)— scientist, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; obuhova@amsd.com

Soloviev Ivan V. (b. 1976) — programmer I, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; soloviev@amsd.com

Анисимов Денис Андреевич (р. 1987) — кандидат технических наук, младший научный сотрудник Санкт-Петербургского электротехнического университета

Анкудинов Александр Владимирович (р. 1964) — кандидат технических наук, заместитель начальника управления ОАО «Информационные спутниковые системы» им. академика М. Ф. Решетнёва», Железногорск

Белоусов Василий Владимирович (р. 1977) — кандидат технических наук, заведующий сектором Института проблем информатики Российской академии наук

Бирюкова Татьяна Константиновна (р. 1968) — кандидат физико-математических наук, старший научный сотрудник Института проблем информатики Российской академии наук

Буроменский Николай Григорьевич (р. 1937) — кандидат технических наук, ведущий научный сотрудник Филиала ФГКУ «46 ЦНИИ» Минобороны России, г. Мытищи, Московская область

Воронов Роман Владимирович (р. 1976) — кандидат технических наук, доцент Петрозаводского государственного университета

Вышиванов Максим Александрович (р. 1981) — ведущий программист Владивостокского государственного университета экономики и сервиса

Гаранин Александр Иванович (р. 1951) — кандидат технических наук, старший научный сотрудник Института проблем информатики Российской академии наук

Гершкович Максим Михайлович (р. 1968) — старший научный сотрудник Института проблем информатики Российской академии наук

Гмарь Дмитрий Викторович (р. 1980) — начальник отдела Владивостокского государственного университета экономики и сервиса

Горшенин Андрей Константинович (р. 1986) — кандидат физико-математических наук, старший научный сотрудник ИПИ РАН; доцент Московского государственного технического университета радиотехники, электроники и автоматики (МГТУ МИРЭА)

Гридин Владимир Николаевич (р. 1944) — доктор технических наук, профессор, директор Центра информационных технологий в проектировании Российской академии наук (ЦИТП РАН)

Дмитревич Геннадий Данилович (р. 1939) — доктор технических наук, профессор, главный научный сотрудник Центра информационных технологий в проектировании Российской академии наук (ЦИТП РАН)

Дулин Сергей Константинович (р. 1950) — доктор технических наук, профессор, главный научный сотрудник Открытого акционерного общества «Научно-исследовательский и проектно-конструкторский институт информатизации,

автоматизации и связи на железнодорожном транспорте» (ОАО «НИИАС»); старший научный сотрудник Института проблем информатики Российской академии наук

Заикин Максим Юрьевич (р. 1973) — научный сотрудник Института проблем информатики Российской академии наук

Замковец Сергей Всеволодович (р. 1944) — кандидат технических наук, заведующий лабораторией ИПИ РАН; доцент Московского государственного технического университета радиотехники, электроники и автоматики (МГТУ МИРЭА)

Захаров Виктор Николаевич (р. 1948) — доктор технических наук, доцент, ученый секретарь Института проблем информатики Российской академии наук

Зацаринный Александр Алексеевич (р. 1951) — доктор технических наук, профессор, заместитель директора Института проблем информатики Российской академии наук

Илюшин Геннадий Яковлевич (р. 1947) — кандидат технических наук, заведующий лабораторией ИПИ РАН

Кириков Игорь Александрович (р. 1955) — кандидат технических наук, директор Калининградского филиала Института проблем информатики Российской академии наук

Колесников Александр Васильевич (р. 1948) — доктор технических наук, профессор Балтийского федерального университета им. Иммануила Канта; старший научный сотрудник Калининградского филиала Института проблем информатики Российской академии наук

Корепанов Эдуард Рудольфович (р. 1966) — кандидат технических наук, заведующий сектором Института проблем информатики Российской академии наук

Красненко Сергей Сергеевич (р. 1988) — аспирант Сибирского федерального университета, Красноярск

Крюков Владимир Васильевич (р. 1956) — доктор экономических наук, профессор, проректор Владивостокского государственного университета экономики и сервиса

Легалов Александр Иванович (р. 1956) — доктор технических наук, заведующий кафедрой Института космических и информационных технологий Сибирского федерального университета, Красноярск

Лиманский Валерий Иванович (р. 1952) — кандидат технических наук, старший научный сотрудник Института проблем информатики Российской академии наук

Малахов Дмитрий Валерьевич (р. 1985) — кандидат физико-математических наук, научный сотрудник Института общей физики им. А. М. Прохорова Российской академии наук

Малодушев Сергей Викторович (р. 1985) — ведущий инженер Петрозаводского государственного университета

Нагорняк Константин Сергеевич (р. 1985) — аспирант, руководитель компьютерного центра Владивостокского государственного университета экономики и сервиса

Недорезов Дмитрий Александрович (р. 1984) — аспирант Сибирского федерального университета, Красноярск

Непомнящий Олег Владимирович (р. 1968) — кандидат технических наук, профессор, руководитель научно-исследовательской лаборатории Института космических и информационных технологий Сибирского федерального университета, Красноярск

Обухова Ольга Леонидовна (р. 1952) — научный сотрудник Института проблем информатики Российской академии наук

Рацеев Сергей Михайлович (р. 1979) — кандидат физико-математических наук, доцент Ульяновского государственного университета

Розенберг Игорь Наумович (р. 1965) — доктор технических наук, профессор, первый заместитель директора Открытого акционерного общества «Научно-исследовательский и проектно-конструкторский институт информатизации, автоматизации и связи на железнодорожном транспорте» (ОАО «НИИАС»)

Румовская София Борисовна (р. 1985) — программист 1-й категории Калининградского филиала Института проблем информатики Российской академии наук

Сергеев Игорь Викторович (р. 1965) — кандидат технических наук, заместитель директора Института проблем информатики Российской академии наук

Синицын Владимир Игоревич (р. 1968) — доктор физико-математических наук, доцент, заведующий отделом Института проблем информатики Российской академии наук

Синицын Игорь Николаевич (р. 1940) — доктор технических наук, профессор, заслуженный деятель науки РФ, заведующий отделом Института проблем информатики Российской академии наук

Скворцова Нина Николаевна (р. 1956) — доктор физико-математических наук, доцент; ведущий научный сотрудник им. А. М. Прохорова Российской академии наук; профессор национального исследовательского ядерного университета «МИФИ»

Соловьев Иван Владимирович (р. 1976) — программист 1-й категории Института проблем информатики Российской академии наук

Уманский Владимир Ильич (р. 1954) — доктор технических наук, генеральный директор Закрытого акционерного общества «ИнтехГеоТранс»

Черноусов Андрей Денисович (р. 1987) — аспирант Института общей физики им. А. М. Прохорова Российской академии наук; ведущий инженер Акустического института имени академика Н. Н. Андреева (ОАО «АКИН»)

Шахгельдян Карина Иосифовна (р. 1967) — доктор технических наук, доцент, начальник управления Владивостокского государственного университета экономики и сервиса

Правила подготовки рукописей статей для публикации в журнале «Системы и средства информатики»

Журнал «Системы и средства информатики» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информационных технологий.

Журнал издается на русском языке. По специальному решению редколлегии отдельные статьи могут печататься на английском языке.

Тематика журнала охватывает следующие направления:

- информационно-телекоммуникационные системы и средства их построения;
- архитектура и программное обеспечение вычислительных машин, комплексов и сетей;
- методы и средства защиты информации.

1. В журнале печатаются статьи, содержащие результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях.

Публикация не должна нарушать закон об авторских правах.

Направляя рукопись в редакцию, авторы сохраняют все права собственников данной рукописи и при этом передают учредителям и редколлегии неисключительные права на издание статьи на русском языке (или на языке статьи, если он отличен от русского) и на ее распространение в России и за рубежом. Авторы должны предоставить в редакцию письмо в следующей форме:

Соглашение о передаче права на публикацию:

«Мы, нижеподписавшиеся, авторы рукописи «. . .», передаем учредителям и редколлегии журнала «Системы и средства информатики» неисключительное право опубликовать данную рукопись статьи на русском языке как в печатной, так и в электронной версиях журнала. Мы подтверждаем, что данная публикация не нарушает авторских прав других лиц или организаций.

Подписи авторов: (ф. и. о., дата, адрес)».

Это соглашение может быть предоставлено в бумажном виде или в виде отсканированной копии (с подписями авторов).

Редколлегия вправе запросить у авторов экспертное заключение о возможности публикации представленной статьи в открытой печати.

2. К статье прилагаются данные автора (авторов) (см. п. 8). При наличии нескольких авторов указывается фамилия автора, ответственного за переписку с редакцией.
3. Редакция журнала осуществляет экспертизу присланных статей в соответствии с принятой в журнале процедурой рецензирования.

Возвращение рукописи на доработку не означает ее принятия к печати.

Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.

4. Решение редколлегии о публикации статьи или ее отклонении сообщается авторам. Редколлегия может также направить авторам текст рецензии на их статью. Дискуссия по поводу отклоненных статей не ведется.
5. Редактура статей высылается авторам для просмотра. Замечания к редакции должны быть присланы авторами в кратчайшие сроки.
6. Рукопись предоставляется в электронном виде в форматах MS WORD (.doc или .docx) или L^AT_EX (.tex), дополнительно — в формате .pdf, на дискете, лазерном диске или электронной почтой. Предоставление бумажной рукописи необязательно.

7. При подготовке рукописи в MS Word рекомендуется использовать следующие настройки.

Параметры страницы: формат — А4; ориентация — книжная; поля (см): внутри — 2,5, снаружи — 1,5, сверху и снизу — 2, от края до нижнего колонтитула — 1,3.

Основной текст: стиль — «Обычный», шрифт — Times New Roman, размер — 14 пунктов, абзацный отступ — 0,5 см, 1,5 интервала, выравнивание — по ширине.

Рекомендуемый объем рукописи — не свыше 20 страниц указанного формата.

Сокращения слов, помимо стандартных, не допускаются. Допускается минимальное количество аббревиатур.

Все страницы рукописи нумеруются.

Шаблоны примеров оформления представлены в Интернете:

<http://www.ipiran.ru/publications/collected/template.doc>

8. Статья должна содержать следующую информацию на **русском и английском языках**:

- Название статьи.
- Ф.И.О. авторов, на английском можно только имя и фамилию.
- Место работы, с указанием города и страны и электронного адреса каждого автора.
- Сведения об авторах, в соответствии с форматом, образцы которого представлены на страницах:
http://www.ipiran.ru/journal/collected/2012_22_02_rus/authors.asp и
http://www.ipiran.ru/journal/collected/2012_22_02_eng/authors.asp
- Аннотация (не менее 100 слов на каждом из языков). Аннотация — это краткое резюме работы, которое может публиковаться отдельно. Она является основным источником информации в информационных системах и базах данных. Английская аннотация должна быть оригинальной, может не быть дословным переводом русского текста и должна быть написана хорошим английским языком. В аннотации не должно быть ссылок на литературу и, по возможности, формул.
- Ключевые слова — желательно из принятых в мировой научно-технической литературе тематических тезаурусов. Предложения не могут быть ключевыми словами.

9. Требования к спискам литературы.

Ссылки на литературу в тексте статьи нумеруются (в квадратных скобках) и располагаются в каждом из списков литературы в порядке первых упоминаний.

Списки литературы представляются в двух вариантах:

- (1) **Список литературы к русскоязычной части.** Русские и английские работы — на языке и в алфавите оригинала.
- (2) **References.** Русские работы и работы на других языках — в латинской транслитерации с переводом на английский язык; английские работы и работы на других языках — на языке оригинала.

Необходимо для составления списка “References” пользоваться размещенной на сайте <http://www.translit.ru/> бесплатной программой транслитерации русского текста в латиницу, при этом в закладке «варианты. . . » следует выбрать опцию BNG.

Список литературы “References” приводится полностью отдельным блоком, повторяя все позиции из списка литературы к русскоязычной части, независимо от того, имеются или нет в нем иностранные источники. Если в списке литературы к русскоязычной части есть ссылки на иностранные публикации, набранные латиницей, они полностью повторяются в списке “References”.

Примеры ссылок на различные виды публикаций в списке “References”:

Описание статьи из журнала:

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Rus. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

Описание статьи из электронного журнала:

Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

Описание материалов конференций:

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma “Novye resursoberegayushchie tekhnologii nedropol’zovaniya i povysheniya neftegazootdachi”* [6th Symposium (International) “New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact” Proceedings]. Moscow. 267–272.

Описание книги (монографии, сборники):

Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publs. 352 p.

Описание переводной книги (в списке литературы к русскоязычной части необходимо указать: / Пер. с англ. — после названия книги, а в конце ссылки указать оригинал книги в круглых скобках):

1. В русскоязычной части:

Тимошенко С. П., Янг Д. Х., Уивер У. Колебания в инженерном деле / Пер. с англ. — М.: Машиностроение, 1985. 472 с. (*Timoshenko S. P., Young D. H., Weaver W.* Vibration problems in engineering. — 4th ed. — N.Y.: Wiley, 1974. 521 p.)

2. В англоязычной части:

Timoshenko, S. P., D. H. Young, and W. Weaver. 1974. *Vibration problems in engineering.* 4th ed. N.Y.: Wiley. 521 p.

Описание неопубликованного документа:

Latypov, A. R., M. M. Khasanov, and V. A. Baikov. 2004. Geology and production (NGT GiD). Certificate on official registration of the computer program No. 2004611198. (In Russian, unpubl.)

Описание интернет-ресурса:

Pravila tsitirovaniya istochnikov [Rules for the citing of sources]. Available at: <http://www.scribd.com/doc/1034528/> (accessed February 7, 2011).

Описание диссертации или автореферата диссертации:

Semenov, V. I. 2003. Matematicheskoe modelirovanie plazmy v sisteme kompaktnyy tor [Mathematical modeling of the plasma in the compact torus]. D.Sc. Diss. Moscow. 272 p.

Kozhunova, O. S. 2009. Tekhnologiya razrabotki semanticheskogo slovary informatsonnogo monitoringa [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.

Описание ГОСТа:

GOST 8.586.5-2005. 2007. Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. Moscow: Standardinform Publs. 10 p.

Описание патента:

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

10. Присланные в редакцию материалы авторам не возвращаются.
11. При отправке файлов по электронной почте просим придерживаться следующих правил:
 - указывать в поле subject (тема) название журнала и фамилию автора;
 - использовать attach (присоединение);
 - в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.
12. Журнал «Системы и средства информатики» является некоммерческим изданием. Плата за публикацию не взимается, гонорар авторам не выплачивается.

Адрес редакции журнала «Системы и средства информатики»:

Москва 119333, ул. Вавилова, д. 44, корп. 2, ИПИ РАН
Тел.: +7 (499) 135-86-92 Факс: +7 (495) 930-45-05
e-mail: rust@ipiran.ru (Сейфуль-Мулюков Рустем Бадриевич)
<http://www.ipiran.ru/journal/collected>

Requirements for manuscripts submitted to Journal “Systems and Means of Informatics”

Journal “Systems and Means of Informatics” publishes theoretical, review, and discussion articles on the research and development in the field of information technology.

The journal is published in Russian. By a special decision of the editorial board, some articles can be published in English.

Topics covered include the following areas:

- information and communication systems and tools of their design;
- architecture and software of computational complexes and networks; and
- methods and tools of information protection.

1. The Journal publishes original articles which have not been published before and are not intended for publication in other editions. An article submitted to the Journal must not violate the Copyright law. Sending the manuscript to the Editorial Board, the authors retain all rights of the owners of the manuscript and transfer the nonexclusive rights to publish the article in Russian (or the language of the article, if not Russian) and its distribution in Russia and abroad to the Founders and the Editorial Board. Authors should submit a letter to the Editorial Board in the following form:

Agreement on the transfer of rights to publish:

“We, the undersigned authors of the manuscript “. . . ,” pass to the Founder and the Editorial Board of the Journal “Systems and Means of Informatics” the nonexclusive right to publish the manuscript of the article in Russian (or in English) in both print and electronic versions of the Journal. We affirm that this publication does not violate the Copyright of other persons or organizations.

Author(s) signature(s): (name(s), address(es), date).”

This agreement should be submitted in paper form or in the form of a scanned copy (signed by the authors).

The Editorial Board has the right to request from the authors an official expert conclusion that the submitted article has no secret data prohibited for publication.

2. A submitted article should be attached with **the data on the author(s)** (see item 8). If there are several authors, the contact person should be indicated who is responsible for correspondence with the Editorial Board and other authors about revisions and final approval of the proofs.
3. The Editorial Board of the Journal examines the article according to the established reviewing procedure. If authors receive their article for correction after reviewing, it does not mean that the article is approved to be published. The corrected article should be sent to the Editorial Board for the subsequent review and approval.
4. The decision on the article publication or its rejection is communicated to the authors. The Editorial Board may also send the reviews on the submitted articles to the authors. Any discussion upon the rejected articles is not possible.
5. The edited articles will be sent to the authors for proofread. The comments of the authors to the edited text of the article should be sent to the Editorial Board as soon as possible.
6. The manuscript of the article should be presented electronically in the MS WORD (.doc or .docx) or L^AT_EX (.tex) formats, and additionally in the .pdf format. All documents may be sent by e-mail or provided on a CD or diskette. A hard copy submission is not necessary.

7. The recommended typesetting instructions for manuscript.

Pages parameters: format A4, portrait orientation, document margins (cm): left — 2.5, right — 1.5, above — 2.0, below — 2.0, footer 1.3.

Text: font — Times New Roman, font size — 14, paragraph indent — 0.5, line spacing — 1.5, justified alignment.

The recommended manuscript size: not more than 20 pages of the specified format.

Use only standard abbreviations. Avoid abbreviations in the title and abstract. The full term for which an abbreviation stands should precede its first use in the text unless it is a standard unit of measurement.

All pages of the manuscript should be numbered.

The templates for the manuscript typesetting are presented on site:

<http://www.ipiran.ru/publication/collected/template.doc>

8. Articles should enclose data both in **Russian and English**:

- title;
- author's name and surname;
- affiliation — organization, its address with ZIP code, city, country, and official e-mail address;
- data on authors according to the format (see site):
http://www.ipiran.ru/journal/collected/2012_22_02_rus/authors.asp and
http://www.ipiran.ru/journal/collected/2012_22_02_eng/authors.asp;
- abstract (not less than 100 words) both in Russian and in English. Abstract is a short summary of the article that can be published separately. The abstract is the main source of information on the article and it could be included in leading information systems and data bases. The abstract in English has to be an original text and should not be an exact translation of the Russian one. Good English is required. In abstracts, avoid references and formulae.
- Indexing is performed on the basis of keywords. The use of keywords from the internationally accepted thematic Thesauri is recommended.
Important! Keywords must not be sentences.

9. References. Russian references have to be presented both in English translation and in Latin transliteration.

Please take into account the following examples of Russian references appearance:

Article in journal:

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Rus. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

Journal article in electronic format:

Swaminathan, V., E. Lepkoswka-White, and B.P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

Conference proceedings:

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma "Novye resursoberegayushchie tekhnologii nedropol'zovaniya i povysheniya neftegazootdachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.

Books and other monographs:

Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publs. 352 p.

Dissertation and Thesis:

Kozhunova, O. S. 2009. Tekhnologiya razrabotki semanticheskogo slovaryaya informatsonnogo monitoringa [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.

State standards and patents:

GOST 8.586.5-2005. 2007. Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichstva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. M.: Standardinform Publs. 10 p.

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

References in Latin transcription are presented in the original language.

References in the text are numbered according to the order of their first appearance; the number is placed in square brackets. All items from the reference list should be cited.

10. Manuscripts and additional materials are not returned to Authors by the Editorial Board.
11. Submissions of files by e-mail must include:
 - the journal title and author's name in the "Subject" field;
 - an article and additional materials have to be attached using the "attach" function;
 - an electronic version of the article should contain the file with the text and a separate file with figures.
12. "System and Means of Informatics" journal is not a profit publication. There are no charges for the authors as well as there are no royalties.

Editorial Board address:

IPI RAN, 44, block 2, Vavilov Str., Moscow 119333, Russia

Ph.: +7 (499) 135 86 92, Fax: +7 (495) 930 45 05

e-mail: rust@ipiran.ru (to Prof. Rustem Seyfoul-Mulyukov)

http://www.ipiran.ru/english/journal_systems.asp

SYSTEMS AND MEANS OF INFORMATICS (SISTEMY I SREDSTVA INFORMATIKI)

SCIENTIFIC JOURNAL

Volume 24 No.1 Year 2014

Editor-in-Chief and Chair of Editorial Council
Academician I. A. Sokolov

IN THIS ISSUE:

DISTRIBUTIONS PARAMETRICAL MODELING SOFTWARE FOR INTEGRODIFFERENTIAL STOCHASTIC SYSTEMS <i>I. N. Sinitsyn, I. V. Sergeev, V. I. Sinitsyn, E. R. Korepanov, and V. V. Belousov</i>	4
PARALLELISM IN MICROPROCESSORS <i>A. K. Gorshenin, S. V. Zamkovets, and V. N. Zakharov</i>	46
ON THEORETICALLY PERFECT CIPHERS <i>S. M. Ratseev</i>	61
MUTATION TEST METODOLOGY FOR ONBOARD SPACECRAFTS RADIOELECTRONIC EQUIPMENT GROUND TEST <i>D. A. Nedorezov, A. I. Legalov, O. V. Nepomnjashnij, S. S. Krasnenko, and A. V. Ankudinov</i>	73
DYNAMIC CREATION OF WiFi-SIGNAL LEVEL MAPS FOR LOCATION SYSTEMS <i>R. V. Voronov and S. V. Malodushev</i>	80
AUTOMATED SYSTEM DEVELOPMENT AND CONTROL OF OPERATIONAL TRAIN SCHEDULE DURING "FREE ZONES" <i>S. Dulin, I. Rozenberg, and V. Umanskiy</i>	93
INFORMATION TECHNOLOGY TO RESEARCH THE FINE STRUCTURE OF CHAOTIC PROCESSES IN PLASMA BY THE ANALYSIS OF SPECTRA <i>A. K. Gorshenin</i>	116
ALGORITHM AND PROGRAM COMPLEX FOR ANALYSIS OF THE CHARACTERISTICS OF ACOUSTIC WAVES OF PLASMA ACTUATOR FLUSH-MOUNTED IN THE MODEL PLANE NOZZLE <i>A. D. Chernousov, D. V. Malakhov, and N. N. Skvortsova</i>	128