ISSN 2520-6141

Белорусский государственный технологический университет 1930–2020

ТРУДЫ БГТУ

Научный журнал

Серия 3

ФИЗИКО-МАТЕМАТИЧЕСКИЕ НАУКИ И ИНФОРМАТИКА

№ 1 (230) 2020 год

Рубрики номера:

ФИЗИКО-МАТЕМАТИЧЕСКИЕ НАУКИ
Математика
Теоретическая механика
Физика

ИНФОРМАТИКА И ТЕХНИЧЕСКИЕ НАУКИ
Моделирование процессов и управление в технических системах
Системный анализ и обучающие системы
Обработка и передача информации
Алгоритмизация и программирование

Минск 2020

Учреждение образования
«Белорусский государственный
технологический университет»

ТРУДЫ БГТУ

Научный журнал

Издается с июля 1993 года

Серия 3

ФИЗИКО-МАТЕМАТИЧЕСКИЕ НАУКИ И ИНФОРМАТИКА

№ 1 (230) 2020 год

Выходит два раза в год

Минск 2020

Educational institution
“Belarusian State Technological University”

PROCEEDINGS OF BSTU

Scientific Journal

Published monthly since July 1993

Issue 3

PHYSICS AND MATHEMATICS. INFORMATICS

No. 1 (230) 2020

Published biannually

Minsk 2020

Учредитель – учреждение образования «Белорусский государственный технологический университет»

Главный редактор журнала – Войтов Игорь Витальевич, доктор технических наук, профессор, Республика Беларусь

Адрес редакции: ул. Свердлова, 13а, 220006, г. Минск.
Телефоны: главного редактора журнала – (+375 17) 226-14-32;
главного редактора серии – (+375 17) 399-49-60.
E-mail: root@belstu.by, http://www.belstu.by

Свидетельство о государственной регистрации средств массовой информации
№ 1329 от 23.04.2010, выданное Министерством информации Республики Беларусь.
Журнал включен в «Перечень научных изданий Республики Беларусь для опубликования результатов диссертационных исследований»

Publisher – educational institution "Belarusian State Technological University"

Editor-in-chief – Voitau Ihar Vital'evich, DSc (Engineering), Professor, Republic of Belarus

Contact: 13a, Sverdlova str., 220006, Minsk.
Telephones: editor-in-chief (+375 17) 226-14-32;
managing editor (+375 17) 399-49-60.
E-mail: root@belstu.by, http://www.belstu.by

ФИЗИКО-МАТЕМАТИЧЕСКИЕ НАУКИ

МАТЕМАТИКА

УДК 517.935.2+519.71

В. М. Марченко, И. М. Борковская
Белорусский государственный технологический университет

О СТАБИЛИЗАЦИИ СКАЛЯРНЫХ ГИБРИДНЫХ ДИФФЕРЕНЦИАЛЬНО-РАЗНОСТНЫХ СИСТЕМ

В статье исследуются вопросы стабилизации скалярных гибридных дифференциально-разностных (ГДР) систем в шкалах линейных регуляторов по типу обратной связи. Рассматриваются простейший регулятор, не выводящий систему за пределы заданного класса, и более общий регулятор с интегральными составляющими типа свертки. Представлены необходимые условия стабилизации с помощью указанных регуляторов. Показано, что необходимое условие стабилизации с помощью регулятора с интегральными составляющими типа свертки является одновременно и достаточным. Получены условия стабилизации системы простейшим регулятором. Приведен пример системы, для которой не существует простейшего регулятора, позволяющего ее стабилизировать, но находится регулятор с интегральными элементами. Результаты могут быть применены при синтезе управляющих воздействий в реальных системах управления, описывающихся дифференциально-разностными системами.

Ключевые слова: дифференциально-разностные системы, линейные регуляторы по типу обратной связи, стабилизация.

V. M. Marchenko, I. M. Borkovskaya
Belarusian State Technological University

ON THE STABILIZATION OF SCALAR HYBRID DIFFERENTIAL-DIFFERENCE SYSTEMS

This article explores the problem of scalar hybrid differential-difference systems in the scales of linear regulators according to the type of feedback. The simplest regulator of a given class as well as a general one with integral components of the convolution type are rigorously investigated. The necessary conditions for stabilization with the mentioned regulators are presented. It is shown that the necessary condition for stabilization with the help of a controller with integral components of the convolution type is simultaneously sufficient. The conditions for stabilization of the system by the simplest controller are obtained. An example of the system that could be stabilized by a regulator with integral elements but not by a simplest regulator is provided. The results can be applied in the synthesis of control actions in real control systems described by differential-difference systems.

Key words: differential-difference systems, linear feedback regulators, stabilization.

Введение. При изучении реальных физических процессов наряду с динамическими (дифференциальными) встречаются и алгебраические (функциональные) зависимости. Такие процессы описываются дифференциально-алгебраическими (DAE) системами (отдельные уравнения которых являются дифференциальными, другие – алгебраическими). Эти системы относятся к классу гибридных [1–10]. Следует, однако, признать, что термин «гибридные системы» перегружен [1–16].

Гибридность означает, вообще говоря, неоднородность в природе рассматриваемого процесса или в методах его изучения. Понятие «гибридные системы» относят к системам, описывающим процессы или объекты с существенно различающимися характеристиками, например, содержащие в основной динамике непрерывные и дискретные переменные (сигналы), детерминированные и случайные величины или воздействия и т. д., что, в конечном счете, и определяет характер (природу) гибридных систем.

Имеется много причин для использования гибридных моделей – это, прежде всего, адекватность данных моделей, обоснованное их упрощение, использование цифровых машин (управление с помощью компьютерных программ); гибридные системы возникают при моделировании иерархической структуры реальных систем управления, в частности, при описании динамических, дискретных, стохастических подсистем, комплексных систем и т. д.

Несмотря на бурное развитие теории гибридных систем, предмет изучения этой теории однозначно не обозначен (см. работы [1–16] и ссылки к ним).

Классической в теории регулирования и теории динамических систем является проблема их устойчивости (особенно асимптотической) и стабилизации.

Ниже исследуются вопросы стабилизации гибридных дифференциально-разностных (ГДР) систем в шкалах линейных регуляторов по типу обратной связи.

Основная часть. Рассмотрим стационарную скалярную гибридную дифференциально-разностную систему в симметрической относительно операторов дифференцирования и сдвига форме

$$\dot{x}_1(t) = a_{11}x_1(t) + a_{12}x_2(t) + b_1u(t), \qquad (1)$$

$$x_2(t+h) = a_{21}x_1(t) + a_{22}x_2(t) + b_2u(t), t \geq 0 \qquad (2)$$

с начальными условиями

$$x_1(0) = x_{10}, x_2(\tau) = \psi(\tau), \tau \in [0, h). \qquad (3)$$

Здесь $x_1(t) \in R$, $x_2(kh) \in R$, $u(t) \in R$, $h > 0$, $a_{11}, a_{12}, a_{21}, a_{22}, b_1, b_2$ – действительные числа; $u = u(\cdot)$ – внешнее (кусочно-непрерывное) воздействие – управление; $\psi(\cdot)$ – начальная кусочно-непрерывная функция.

Под решением системы (1), (2) будем понимать абсолютно непрерывную функцию $x_1(\cdot)$ и кусочно-непрерывную функцию $x_2(\cdot)$, которые для всех $t \geq 0$ удовлетворяют уравнению (2) и для почти всех $t \geq 0$ удовлетворяют уравнению (1).

Такое решение начальной задачи (3) для каждого начального значения x_{10} и кусочно-непрерывной функции $\psi(\cdot)$ существует, единственно и может быть найдено методом интегрирования системы (1)–(3) «по шагам».

Присоединим к системе шкалы (классы) линейной обратной связи в виде:

1) простейшего регулятора

$$u(t) = Q_1x_1(t) + Q_2x_2(t), \qquad (4)$$

не выводящего замкнутую систему за пределы рассматриваемого класса;

2) более общего регулятора с интегральными составляющими типа свертки

$$u(t) = Q_1x_1(t) + Q_2x_2(t) + \int_0^t Q_1(s)x_1(t-s)ds + \int_0^t Q_2(s)x_2(t+h-s)ds, \qquad (5)$$

где Q_1 и Q_2 – действительные числа; $Q_1(\cdot)$ и $Q_2(\cdot)$ – кусочно-непрерывные функции с конечным носителем $H > 0$, $Q_1(\cdot) \equiv 0$, $Q_2(\cdot) \equiv 0$ для $t > H$.

Задача. Исследовать задачу стабилизации системы (1), (2) в шкалах (4), (5), т. е. задачу отыскания регуляторов того или иного типа (отыскания чисел Q_1, Q_2, функций $Q_1(\cdot), Q_2(\cdot)$), при которых замкнутая система является устойчивой в том или ином смысле – асимптотически устойчивой, если не оговорено иное.

Рассмотрим невозмущенную ГДР-систему (1), (2), т. е. систему с выключенным управлением $u(t) = 0$ при $t \geq 0$.

Следуя методу Эйлера отыскания решений системы (1) в экспоненциальной форме в теории обыкновенных дифференциальных уравнений

$$x(t) = e^{\lambda t}c_1, y(t) = e^{\lambda t}c_2,$$

получим характеристическое уравнение

$$\det\begin{bmatrix} \lambda I_{n_1} - a_{11} & -a_{12} \\ -a_{21} & e^{\lambda h} - a_{22} \end{bmatrix} = \Delta_1(\lambda) = 0, \lambda \in C,$$

которое назовем основным характеристическим уравнением системы (1), (2); здесь и далее C – поле комплексных чисел. Корни уравнения $\Delta_1(\lambda) = 0$ назовем основными характеристическими значениями этой системы.

Наряду с экспоненциальными решениями невозмущенной системы (1), (2) представляют интерес решения вида

$$x_1(t) \equiv 0,\ x_2(t) = \begin{cases} 0, & t \neq kh, \\ \frac{k}{c\lambda^h}, & t = kh, \end{cases} \quad k = 0,1,2,..., t \geq 0,$$

где $c \neq 0, 0^0 = 1$ (назовем их импульсными решениями), подставляя которые в систему (1), (2), приходим к уравнению

$$\det(\lambda I_{n_2} - a_{22}) = \Delta_2(\lambda) = 0, \lambda \in C,$$

которое назовем присоединенным характеристическим уравнением системы (1), (2), а его корни – присоединенными характеристическими значениями этой системы.

Множество всех характеристических значений (основных и присоединенных) с учетом их кратностей назовем спектром, а решения, его породившие, – спектральными решениями системы (1), (2).

Определения асимптотической и экспоненциальной устойчивости системы (1), (2) понимаются в соответствии с их классическими формулировками для обыкновенных систем.

Определение. Невозмущенную систему (1), (2) будем называть спектрально устойчивой, если все ее спектральные решения являются асимптотически устойчивыми.

Имеет место следующее утверждение.

Утверждение 1. Для спектральной устойчивости невозмущенной системы (1), (2) необходимо и достаточно, чтобы

1) все основные собственные значения имели отрицательные действительные части;

2) все присоединенные собственные значения λ лежали в комплексной плоскости внутри единичного диска: $|\lambda|<1$.

Доказательство утверждения непосредственно вытекает из вида спектральных решений.

Понятие спектральной устойчивости является некоторым ослаблением понятия асимптотической устойчивости, однако во многих случаях решение представляется в виде линейных комбинаций спектральных решений; в таких случаях эти понятия равнозначны.

В исследовании ГДР-систем приходится применять к таким системам преобразование Лапласа. Поэтому возникает необходимость в экспоненциальной оценке роста решений этих систем.

Запишем систему (1), (2) в виде, более удобном для применения преобразования Лапласа. Положим

$$x_2(t)=x_3(t-h),\ t\ge 0. \tag{6}$$

Тогда система запишется в виде ГДР-системы запаздывающего типа

$$\dot{x}_1(t)=a_{11}x_1(t)+a_{12}x_3(t-h)+b_1u(t), \tag{7}$$

$$x_3(t)=a_{21}x_1(t)+a_{22}x_3(t-h)+b_2u(t),\ t\ge 0 \tag{8}$$

с начальными условиями

$$x_1(0)=x_{10},\ x_3(\tau)=\psi(\tau+h),\ \tau\in[-h,0). \tag{9}$$

Регуляторы (4), (5) перепишутся в виде

$$u(t)=Q_1x_1(t)+Q_2x_3(t-h), \tag{10}$$

$$u(t)=Q_1x_1(t)+Q_2x_3(t-h)+$$
$$+\int_0^t Q_1(s)x_1(t-s)ds+\int_0^t Q_2(s)x_3(t-s)ds. \tag{11}$$

Можно показать, что для каждого решения $x_1(\cdot)$, $x_3(\cdot)$ системы (7), (8), порожденного начальными данными (9) и кусочно-непрерывным управлением, имеющим не выше, чем экспоненциальный рост (14), найдутся такие положительные числа L и μ, что

$$\|x_1(t)\|\le Le^{\mu t},\ \|x_3(t)\|\le Le^{\mu t}.$$

Таким образом, имеет место экспоненциальная оценка решений системы (7), (8) (а следовательно, и системы (1), (2)), что позволяет применять к этим системам преобразование Лапласа.

Отметим, что спектры систем (7), (8) и (1), (2) совпадают.

Перейдем к необходимым условиям стабилизируемости.

Теорема 1. Если система (1), (2) является стабилизируемой в шкале (4) (или (5)), то

$$\operatorname{rank}\begin{bmatrix}\lambda-a_{11} & -a_{12}e^{-\lambda h} & b_1\\ -a_{21} & 1-a_{22}e^{-\lambda h} & b_2\end{bmatrix}=2,\ \operatorname{Re}\lambda>0. \tag{12}$$

Доказательство. Предположим противное: система (1), (2) стабилизируема (регулятором (4) или (5)), а условие (12) не выполняется, т. е. существует в общем случае комплексное число $\lambda^*\in C,\ \operatorname{Re}\lambda^*>0$, и числа c_1 и c_2 такие, что

$$[c_1,c_2]\ne 0,$$
$$[c_1,c_2]\begin{bmatrix}\lambda^*-a_{11} & -a_{12} & b_1\\ -a_{21} & e^{\lambda^* h}-a_{22} & b_2\end{bmatrix}=0.$$

Вдоль решений системы (1), (2) имеем

$$0=\int_0^t c_1e^{-\lambda^*\tau}\big(\dot{x}_1(\tau)-a_{11}x_1(\tau)-a_{12}x_2(\tau)-$$
$$-b_1u(\tau)\big)d\tau+\int_0^t c_2e^{-\lambda^*\tau}\big(x_2(\tau+h)-a_{21}x_1(\tau)-a_{22}x_2(\tau)-$$
$$-b_2u(\tau)\big)d\tau=c_1e^{-\lambda^* t}x_1(t)-c_1x_{10}+\int_0^t e^{-\lambda^*\tau}c_1\big(\lambda^*-$$
$$-a_{11}\big)x_1(\tau)d\tau-\int_0^t e^{-\lambda^*\tau}c_1a_{12}x_2(\tau)d\tau-\int_0^t e^{-\lambda^*\tau}c_1b_1\times$$
$$\times u(\tau)d\tau+\int_h^{t+h} e^{-\lambda^* s+\lambda h}c_2x_2(s)ds-\int_0^t e^{-\lambda^*\tau}c_2a_{21}x_1(\tau)d\tau-$$
$$-\int_0^t e^{-\lambda^*\tau}c_2a_{22}x_2(\tau)d\tau-\int_0^t e^{-\lambda^*\tau}c_2b_2u(\tau)d\tau=-c_1x_{10}+$$
$$+c_1e^{-\lambda^* t}x_1(t)+\int_t^{t+h} e^{-\lambda^* s+\lambda^* h}c_2x_2(s)ds-\int_0^h e^{-\lambda^* s+\lambda^* h}c_2\times$$
$$\times x_2(s)ds+\int_0^t e^{-\lambda^*\tau}c_1\big(\lambda^*-a_{11}\big)x_1(\tau)d\tau-\int_0^t e^{-\lambda^*\tau}c_1\times$$
$$\times a_{12}x_2(\tau)d\tau-\int_0^t e^{-\lambda^*\tau}c_1b_1u(\tau)d\tau+\int_0^t e^{-\lambda^* s+\lambda^* h}c_2x_2(s)ds-$$

$$-\int_0^t e^{-\lambda^*\tau}c_2a_{21}x_1(\tau)d\tau-\int_0^t e^{-\lambda^*\tau}c_2a_{22}x_2(\tau)d\tau-\int_0^t e^{-\lambda^*\tau}c_2\times$$

$$\times b_2u(\tau)d\tau-c_1x_{10}+c_1e^{-\lambda^*t}x_1(t)+e^{-\lambda^*(t-h)}\int_0^h e^{-\lambda^*\tau}c_2\times$$

$$\times x_2(t+\tau)d\tau-e^{\lambda^*h}\int_0^h e^{-\lambda^*\tau}c_2\psi(\tau)d\tau+\int_0^t e^{-\lambda^*\tau}c_1\left(\lambda^*-\right.$$

$$\left.-a_{11}\right)x_1(\tau)d\tau-\int_0^t e^{-\lambda^*\tau}c_1a_{12}x_2(\tau)d\tau-\int_0^t e^{-\lambda^*\tau}c_1b_1\times$$

$$\times u(\tau)d\tau+\int_0^t e^{-\lambda^*\tau}c_2\left(e^{\lambda^*h}-a_{22}\right)x_2(\tau)d\tau-$$

$$-\int_0^t e^{-\lambda^*\tau}c_2a_{21}x_1(\tau)d\tau-\int_0^t e^{-\lambda^*\tau}c_2b_2u(\tau)d\tau=c_1e^{-\lambda^*t}x_1(t)+$$

$$+e^{-\lambda^*(t-h)}\int_0^h e^{-\lambda^*\tau}c_2x_2(t+\tau)d\tau-c_1x_{10}-e^{\lambda^*h}\int_0^h e^{-\lambda^*\tau}c_2\times$$

$$\times\psi(\tau)d\tau+\int_0^t e^{-\lambda^*\tau}\left(c_1\left(\lambda^*-a_{11}\right)-c_2a_{21}\right)x_1(\tau)d\tau+$$

$$+\int_0^t e^{-\lambda^*\tau}\left(c_2\left(e^{\lambda^*h}-a_{22}\right)-c_1a_{12}\right)x_2(\tau)d\tau-$$

$$-\int_0^t e^{-\lambda^*\tau}\left(c_1b_1+c_2b_2\right)u(\tau)d\tau=c_1e^{-\lambda^*t}x_1(t)+$$

$$+e^{-\lambda^*(t-h)}\int_0^h e^{-\lambda^*\tau}c_2x_2(t+\tau)d\tau-c_1x_{10}-e^{\lambda^*h}\int_0^h e^{-\lambda^*\tau}c_2\times$$

$$\times\psi(\tau)d\tau+\int_0^t e^{-\lambda^*\tau}[c_1,c_2]\begin{bmatrix}\lambda^*-a_{11} & -a_{12} & b_1\\ -a_{21} & e^{\lambda^*h}-a_{22} & b_2\end{bmatrix}\times$$

$$\times\begin{bmatrix}x_1(\tau)\\ x_2(\tau)\\ u(\tau)\end{bmatrix}d\tau=c_1e^{-\lambda^*t}x_1(t)+$$

$$+e^{-\lambda^*(t-h)}\int_0^h e^{-\lambda^*\tau}c_2x_2(t+\tau)d\tau-$$

$$-c_1x_{10}-e^{\lambda^*h}\int_0^h e^{-\lambda^*\tau}c_2\psi(\tau)d\tau\ \Rightarrow$$

$$c_1e^{-\lambda^*t}x_1(t)+e^{-\lambda^*(t-h)}\int_0^h e^{-\lambda^*\tau}c_2x_2(t+\tau)d\tau=$$

$$=c_1x_{10}-e^{\lambda^*h}\int_0^h e^{-\lambda^*\tau}c_2\psi(\tau)d\tau.$$

Поскольку по условию система (1), (2) стабилизируема, то найдется обратная связь – стабилизирующее управление (4) (или (5)), при котором замкнутая система (1), (2), (4) (или (1), (2), (5)) асимптотически устойчива, стало быть, $\lim\limits_{t\to+\infty}\|x_1(t)\|=0,\quad \lim\limits_{t\to+\infty}\|x_2(t)\|=0.$ Поэтому левая часть в полученном равенстве стремится к нулю при $t\to+\infty$ в случае $\mathrm{Re}\,\lambda^*>0,$ в то время как правую часть можно отграничить от нуля за счет выбора начальных данных. Полученное противоречие доказывает справедливость утверждения теоремы 1.

Теорема 2. Если система является стабилизируемой в шкале (4) (или (5)), то

$$\mathrm{rank}\,[\lambda-a_{22}\quad b_2]=1,\ \forall\lambda\in C,\ |\lambda|<1. \qquad (13)$$

Доказательство. Предположим противное: система (1), (2) стабилизируема (регулятором (4) или (5)), а условие (13) не выполняется, т. е. существует в общем случае комплексное число $\lambda^*\in C, |\lambda^*|\ge 1,$ и число c_2 такое, что

$$c_2\ne 0, c_2\left[\lambda^*-a_{22}\quad b_2\right]=0.$$

Поскольку система (1), (2) стабилизируема, то найдется линейная обратная связь, при которой все решения соответствующей замкнутой системы со временем затухают. В силу уравнения (2) вдоль импульсных решений с учетом вышеуказанного условия имеем

$$c_2x_2(kh+h)=c_2(a_{22}x_2(kh)+b_2u(kh))=$$
$$=\lambda^*c_2x_2(kh)=\ldots=(\lambda^*)^{k+1}c_2\psi(0),$$

стало быть,

$$|c_2x_2(kh+h)|=|\lambda^*|^{k+1}|c_2\psi(0)|.$$

Как и при доказательстве теоремы 1, заключаем, что левая часть последнего равенства при неограниченном возрастании времени стремится к нулю, а правая нет (при подходящем выборе значения $\psi(0)$). Полученное противоречие завершает доказательство теоремы 2.

Будем считать, что необходимые условия (12), (13) стабилизируемости выполнены. Тогда в системе (1), (2) либо

1) $b_2\ne 0$; либо

2) $b_2=0, |a_{22}|<1.$

В первом случае $(b_2\ne 0)$ полагаем

$$u(t)=\frac{1}{b_2}\left(-a_{21}x_1(t)-a_{22}x_3(t-h)+v(t)\right),\ t\ge 0. \qquad (14)$$

Присоединяя к системе (7), (8), (14)

$$\dot{x}_1(t)=a_{11}^*x_1(t)+a_{12}^*x_3(t-h)+b_1^*v(t),$$
$$x_3(t)=v(t), \qquad (15)$$

где $b_1^*=\dfrac{b_1}{b_2}, a_{11}^*=a_{11}-b_1^*a_{21}, a_{12}^*=a_{12}-b_1^*a_{22},$ регулятор (11) и применяя преобразование Лапласа, перейдем в частотную область:

$$(p-a_{11}^*)\breve{x}_1(p)-a_{12}^*e^{-ph}\breve{x}_3(p)-b_1^*\breve{v}(p)=$$

$$=x_{10}+a_{12}^*\int_0^h e^{-pt}\psi(t)dt,\ \breve{x}_3(p)=\breve{v}(p),\ p\in C,$$

здесь

$$\breve{x}_i(p)=\int_0^{+\infty}e^{-pt}x_i(t)dt,\ \breve{v}(p)=\int_0^{+\infty}e^{-pt}v(t)dt,\ i=1,3-$$

Лаплас-образы функций $x_i(t), v(t), t\ge 0, i=1,3.$

Прежде всего заметим, что в случае $a_{11}^*<0$ рассматриваемая система стабилизируема, и стабилизирующий регулятор можно выбрать в виде $v(t)=0, t\ge 0.$ Поэтому далее считаем, что $a_{11}^*>0.$

Записывая регулятор в частотной области и принимая во внимание теорему об образе свертки и теорему Винера – Пэли об образе функции с конечным носителем [17], получаем

$$\breve{v}(p)=Q_1^*(p)\breve{x}_1(p)+Q_2^*(p)\breve{x}_3(p)+Q_2\int_0^h e^{-pt}\psi(t)dt$$

(последнее слагаемое можно опустить, так как при построенном регуляторе (11) оно автоматически учтется в частотной области в Лаплас-образах начальных данных); здесь функции комплексной переменной p

$$Q_1^*(p)=Q_1+\breve{Q}_1(p), Q_2^*(p)=Q_1e^{-ph}+\breve{Q}_2(p),$$
$$\breve{Q}_1(p)=\int_0^{+\infty}e^{-pt}Q_1(t)dt, \breve{Q}_2(p)=\int_0^{+\infty}e^{-pt}Q_2(t)dt$$

являются целыми функциями конечной степени H как Лаплас-образы [17] функций с конечным носителем H.

Рассмотрим характеристическое уравнение полученной замкнутой системы и потребуем, чтобы оно имело желаемый вид:

$$0=\det\begin{bmatrix}p-a_{11}^*-b_1^*Q_1^*(p) & -a_{12}^*e^{-ph}-b_1^*Q_2^*(p)\\ -Q_1^*(p) & 1-Q_2^*(p)\end{bmatrix}=$$

$$=p-a_{11}^*-b_1^*Q_1^*(p)-pQ_2^*(p)+a_{11}^*Q_2^*(p)+b_1^*Q_1^*(p)Q_2^*(p)-$$
$$-a_{12}^*e^{-ph}Q_1^*(p)-b_1^*Q_1^*(p)Q_2^*(p)=p-a_{11}^*-$$
$$-Q_1^*(p)(b_1^*+a_{12}^*e^{-ph})-(p-a_{11}^*)Q_2^*(p)=p+\alpha,\ \alpha>0.$$

Функции должны быть целыми, поэтому нули знаменателя должны быть нулями (той же кратности) числителя, для чего достаточно положить

$$Q_2^*(p)=-\frac{\alpha+a_{11}^*+Q_1^*(p)(b_1^*+a_{12}^*e^{-ph})}{p-a_{11}^*},$$

$$Q_1=-\frac{\alpha+a_{11}^*}{b_1^*+a_{12}^*e^{-a_{11}^*h}},\ \breve{Q}_1(p)\equiv 0,\ p\in C,\qquad(16)$$

что выполнимо, так как в силу условия (12)

$$2=\operatorname{rank}\begin{bmatrix}p-a_{11} & -a_{12}e^{-ph} & b_1\\ -a_{21} & 1-a_{22}e^{-ph} & b_2\end{bmatrix}\underset{b_2\ne 0}{=}$$

$$=\operatorname{rank}\begin{bmatrix}p-a_{11}^* & -a_{12}^*e^{-ph} & b_1^*\\ 0 & 1 & 1\end{bmatrix}=$$

$$=\operatorname{rank}\begin{bmatrix}p-a_{11}^* & -a_{12}^*e^{-ph}-b_1^* & 0\\ 0 & 1 & 1\end{bmatrix},\ \operatorname{Re}\lambda>0\Rightarrow$$
$$b_1^*+a_{12}^*e^{-ph}\ne 0$$

для $p=a_{11}^*, \operatorname{Re}a_{11}^*>0.$

Тогда

$$Q_2^*(p)=-\frac{\alpha+a_{11}^*-(\alpha+a_{11}^*)\dfrac{b_1^*+a_{12}^*e^{-ph}}{b_1^*+a_{12}^*e^{-a_{11}^*h}}}{p-a_{11}^*}=$$

$$=-(\alpha+a_{11}^*)a_{12}^*e^{-a_{11}^*h}\frac{1-e^{-(p-a_{11}^*)h}}{p-a_{11}^*}=$$

$$=\frac{-(\alpha+a_{11}^*)a_{12}^*e^{-a_{11}^*h}}{p-a_{11}^*}+\frac{(\alpha+a_{11}^*)a_{12}^*}{p-a_{11}^*},$$
$$e^{-ph}=\breve{Q}_2(p),\ Q_2=0.$$

Переходя к оригиналам, имеем

$$Q_1(\tau)=\begin{cases}-(\alpha+a_{11}^*)a_{12}^*e^{a_{11}^*(t-h)}, t\in[0,h],\\ 0,\ t>h,\end{cases}\qquad(17)$$

$$v(t)=Q_1x_1(t)+\int_0^t Q_2(s)x_3(t-s)ds, t\ge 0,$$

где Q_1 и $Q_2(\cdot)$ определены в (16) и (17).

Отсюда, возвращаясь к системе (1), (2) и учитывая (14), получаем стабилизирующий регулятор в виде

$$u(t)=\frac{Q_1-a_{21}}{b_2}x_1(t)-\frac{a_{22}}{b_2}x_2(t)+$$
$$+\int_0^t\frac{Q_2(s)}{b_2}x_2(t+h-s)ds, t\ge 0$$

(Q_1 и $Q_2(\cdot)$ найдены в (16) и (17)).

Пусть теперь $b_2=0, |a_{22}|<1.$ Если при этом $b_1=0,$ то внешнее воздействие на систему отсутствует, и смысл задачи стабилизируемости сводится к асимптотической устойчивости первоначальной открытой системы. Поэтому в дальнейшем считаем, что $b_1\ne 0.$ В этом случае необходимое условие (12) выполняется.

Замкнем систему (7), (8) линейной обратной связью вида

$$u(t)=\frac{1}{b_1}\left(-a_{11}x_1(t)-a_{12}x_3(t-h)+v(t)\right),\ t\ge 0,$$

перейдем в частотную область, присоединим к системе регулятор $\breve{v}(p)=Q_1^*(p)\breve{x}_1(p)+Q_2^*(p)\breve{x}_3(p)$, где положим $Q_1^*(p)=-\alpha,\ \alpha>0,\ Q_2^*(p)=0,\ p\in C$.

Характеристическое уравнение полученной замкнутой системы имеет вид

$$0=\det\begin{bmatrix} p-Q_1^*(p) & -Q_2^*(p)\\ -a_{21} & 1-a_{22}e^{-ph}\end{bmatrix}=(p+\alpha)\left(1-a_{22}e^{-ph}\right)$$

и все его корни – основные характеристические значения замкнутой системы – имеют отрицательные действительные части.

Возвращаясь из частотной области в пространство состояний, получаем регулятор $v(t)=-\alpha x_1(t),\ t\ge 0$, что приводит к соответствующей замкнутой системе вида

$$\dot{x}_1(t)=-\alpha x_1(t),\ x_3(t)=a_{21}x_1(t)+a_{21}x_3(t-h),$$

которая в рассматриваемом случае $\alpha>0, |a_{22}|<1$ является асимптотически устойчивой.

Действительно,

$$|x_1(t)|\le|x_{10}|e^{-\alpha t},$$

$$|x_3(t)|\le|a_{21}||x_1(t)|+|a_{22}||x_3(t-h)|\le$$

$$\le|a_{21}|\left(|x_1(t)|+|a_{22}||x_1(t-h)|+\ldots+|a_{22}|^{T_t}|x_1(t-T_th)|\right)+$$

$$+|a_{22}|^{T_t+1}|x_3(t-T_th-h)|\le|a_{21}||x_{10}|e^{-\alpha t}\left(1+|a_{22}|e^{\alpha h}+\right.$$

$$\left.+\ldots+|a_{22}|^{T_t}e^{\alpha T_th}\right)+|a_{22}|^{T_t+1}|\psi(t-T_th)|\le|a_{21}||x_{10}|e^{-\alpha t}\times$$

$$\times\frac{|a_{22}|^{T_t}e^{\alpha T_th}|a_{22}|e^{\alpha h}-1}{|a_{22}|e^{\alpha h}-1}+|a_{22}|^{T_t+1}\sup_{\tau\in[0,h]}|\psi(\tau)|=|a_{22}|^{T_t+1}\times$$

$$\times\left(\frac{|a_{21}||x_{10}|e^{\alpha h}e^{-\alpha(t-T_th)}}{|a_{22}|e^{\alpha h}-1}+\sup_{\tau\in[0,h]}|\psi(\tau)|\right)+\frac{|a_{21}||x_{10}|e^{-\alpha t}}{|a_{22}|e^{\alpha h}-1}\le$$

$$\le\left(\frac{|a_{21}||x_{10}|e^{\alpha h}}{|a_{22}|e^{\alpha h}-1}+\sup_{\tau\in[0,h]}|\psi(\tau)|\right)|a_{22}|^{T_t+1}+\frac{|a_{21}||x_{10}|e^{-\alpha t}}{|a_{22}|e^{\alpha h}-1},$$

откуда видим, что функции $|x_1(t)|, |x_3(t)|$, монотонно затухая $(|x_1(t)|\to 0, |x_3(t)|\to 0)$, стремятся к нулю при $t\to+\infty$, что влечет за собой асимптотическую устойчивость рассматриваемой замкнутой системы и, как следствие, стабилизируемость (в нашем случае $b_2=0, |a_{22}|<1, b_1\ne 0$) системы (1), (2) регулятором вида

$$u(t)=\frac{-a_{11}-\alpha}{b_1}x_1(t)-\frac{a_{12}}{b_1}x_2(t),\ t\ge 0. \qquad (18)$$

Таким образом, доказана теорема 2.

Теорема 3. Необходимые условия (12), (13) стабилизируемости системы (1), (2) в шкале регуляторов (5) являются и достаточными.

В случае $b_2=0, |a_{22}|<1, b_1\ne 0$ регулятор (18), стабилизирующий систему (1), (2), относится к шкале (4). В связи с этим представляет интерес исследование задачи стабилизируемости в этой шкале, что является делом весьма непростым. Как и ранее, рассмотрим систему (7), (8) и соответствующий простейший регулятор

$$u(t)=Q_1x_1(t)+Q_2x_3(t-h),\ t\ge 0. \qquad (19)$$

Замыкая систему этим регулятором, получаем основное характеристическое уравнение

$$0=\det\begin{bmatrix} p-a_{11}-b_1Q_1 & -a_{12}e^{-ph}-b_1Q_2e^{-ph}\\ -a_{21}-b_2Q_1 & 1-a_{22}e^{-ph}-b_2Q_2e^{-ph}\end{bmatrix}=$$

$$=p-a_{11}-b_1Q_1+e^{-ph}(a_{22}a_{11}+a_{22}b_1Q_1+a_{11}b_2Q_2-$$

$$-a_{12}a_{21}-a_{21}b_1Q_2-a_{12}b_2Q_1)-pe^{-ph}(a_{22}+b_2Q_2). \qquad (20)$$

В силу необходимого условия (12) параметры q_1, q_2 надлежит выбирать так, чтобы все корни уравнения (20) имели отрицательные действительные части. В связи с этим представляет интерес следующее утверждение.

Утверждение 2 [2, 3]. Для того, чтобы все корни $(p\in C)$ уравнения

$$p+\alpha_1+\alpha_2e^{-ph}+\alpha_3pe^{-ph}=0,$$

где $\alpha_1, \alpha_2, \alpha_3\in R$, имели отрицательные действительные части, необходимо и достаточно, чтобы либо

1) $|\alpha_3|\le 1, \alpha_1>|\alpha_2|$; либо

2) $|\alpha_3|<1, \alpha_2>|\alpha_1|, h<h^*$,

где $h^*=\sqrt{\dfrac{1-\alpha_3^{\,2}}{\alpha_2^{\,2}-\alpha_1^{\,2}}}\arccos\left(-\dfrac{\alpha_1+\alpha_2\alpha_3}{\alpha_2+\alpha_1\alpha_3}\right)$. (21)

Применим утверждение 2 к уравнению (20) и попытаемся за счет выбора Q_1, Q_2 удовлетворить требованиям, предъявляемым этим утверждением к коэффициентам.

Заметим, условие (13) требует $|\alpha_3|<1$ и в силу вида регулятора (18) неисследованным остается случай $b_2\ne 0$ системы (7), (8). Поэтому сразу можем перейти к рассмотрению характеристического уравнения системы (15), замкнутой регулятором

$$v(t)=Q_1x_1(t)+Q_2x_3(t-h), \qquad (22)$$

$$0=\det\begin{bmatrix} p-a_{11}^*-b_1^*Q_1 & -a_{12}^*e^{-ph}-b_1^*Q_2e^{-ph} \\ -Q_1 & 1-Q_2e^{-ph}\end{bmatrix}=$$

$$=p+(-a_{11}^*-b_1^*Q_1)+(a_{11}^*Q_2-a_{12}^*Q_1)e^{-ph}+(-Q_2)pe^{-ph},$$

$$-a_{11}^*-b_1^*Q_1=\alpha_1,\ a_{11}^*Q_2-a_{12}^*Q_1=\alpha_2,\ -Q_2=\alpha_3.$$

Считаем, что $|Q_2|<1$. Это условие необходимо для асимптотической устойчивости замкнутой системы (15), (22). Далее в соответствии с утверждением 2 потребуем, чтобы

$$-a_{11}^*-b_1^*Q_1>\left|a_{11}^*Q_2-a_{12}^*Q_1\right|, \qquad (23)$$

где $|Q_2|<1$. Очевидно, что в случае $a_{11}^*<0$ неравенство (23) выполняется при $Q_1=Q_2=0$, и стабилизирующий регулятор (22) приобретает вид $v(t)=0$.

Пусть далее $|b_1^*|>|a_{12}^*|$. Тогда при $Q_2=0$, $Q_1=-|Q_1|\operatorname{sgn}b_1^*$ и достаточно больших $|Q_1|$ неравенство (23) будет выполняться, и стабилизирующий регулятор (22) можно выбрать в виде

$$v(t)=-|Q_1|\operatorname{sgn}b_1^*\,x_1(t)$$

при достаточно больших $|Q_1|$.

Действительно, замкнутая этим регулятором система (15) имеет следующий вид:

$$\dot{x}_1(t)=(a_{11}^*-|b_1^*||Q_1|)x_1(t)-a_{12}^*|Q_1|\operatorname{sgn}b_1^*x_1(t-h)x_1(t),$$

$$x_3(t)=-x_1(t).$$

Первое уравнение этой системы – дифференциальное уравнение запаздывающего типа – является асимптотически устойчивым, поскольку все корни его характеристического уравнения (в силу утверждения 2, см. также [4]) имеют отрицательные действительные части, что влечет устойчивость и второго уравнения системы.

Пусть теперь $|b_1^*|<|a_{12}^*|$. Условие 1) утверждения 2 при этом оказывается неэффективным. Учитывая условие 2) утверждения 2, поиск параметров стабилизирующих регуляторов вида (22) сводится в рассматриваемом случае к исследованию задачи математического программирования − максимизации функции

$$h^*=\sqrt{\frac{1-Q_2^2}{\alpha_2^2-\alpha_1^2}}\arccos\left(-\frac{\alpha_1-\alpha_2Q_2}{\alpha_2-\alpha_1Q_2}\right)\to\max$$

при ограничениях

$$\alpha_2=a_{11}^*Q_2-a_{12}^*Q_1>\left|-a_{11}^*-b_1^*Q_1\right|=|\alpha_1|,$$

где $Q_1\in R, |Q_2|<1$.

Рассмотрим некоторые частные случаи. Пусть, например, $a_{11}^*=0$. Тогда, полагая $Q_1=-|Q_1|\operatorname{sgn}a_{12}^*$, $Q_2=0$, приходим к оптимизационной задаче

$$h^*=\frac{1}{|Q_1|\sqrt{|a_{12}^*|^2-|b_1^*|^2}}\arccos\left(\frac{b_1^*\operatorname{sgn}a_{12}^*}{|a_{12}^*|}\right)\to\max$$

при ограничениях $|Q_1|>0$. Очевидно, для каждой системы (15) такая задача разрешима и существует искомый стабилизирующий регулятор вида $v(t)=-|Q_1|\operatorname{sgn}a_{12}^*x_1(t)$ (при достаточно малых $|Q_1|$).

Пусть далее $b_1^*\operatorname{sgn}a_{12}^*>0$. Тогда $b_1^*\operatorname{sgn}a_{12}^*=|b_1^*|$ и при $Q_2=0$, $Q_1=-|Q_1|\operatorname{sgn}b_1^*$ приходим к задаче максимизации функции

$$h^*=$$

$$=\frac{1}{\sqrt{|a_{12}^*|^2|Q_1|^2-\left(a_{11}^*-|b_1^*||Q_1|\right)^2}}\arccos\left(\frac{a_{11}^*-|b_1^*||Q_1|}{|a_{12}^*||Q_1|}\right)$$

при ограничениях $|a_{12}^*||Q_1|>|a_{11}^*-|b_1^*||Q_1||$ или (по раскрытию неравенства) получаем $|Q_1|>q^*$, где $q^*=\dfrac{a_{11}^*}{|a_{12}^*|+|b_1^*|}$. Далее можно найти наибольшее значение (верхнюю точную границу) функции h^*. Вычисляя предел (по правилу Лопиталя) $\lim\limits_{|Q_1|\to q^*}h^*=\left[\dfrac{0}{0}\right]=\dfrac{|a_{12}^*|+|b_1^*|}{a_{11}^*|a_{12}^*|}=h_0^*$, можем сформулировать следующее достаточное условие.

Утверждение 3. Если

$$h<h_0^*,\ a_{11}^*>0,\ |a_{12}^*|>|b_1^*|,\ b_1^*\operatorname{sgn}a_{12}^*>0,$$

то найдется регулятор $v(t)=-|Q_1|\operatorname{sgn}a_{12}^*x_1(t)$, который при некотором $|Q_1|>q^*$ стабилизирует систему (15).

Аналогично можно рассмотреть случай $a_{11}^*>0, |a_{12}^*|>|b_1^*|, b_1^*\operatorname{sgn}a_{12}^*<0$.

В заключение приведем пример системы (15), для которой не существует стабилизирующего регулятора вида (22), однако находится регулятор с интегральными элементами.

Пример. Рассмотрим систему (15) вида

$$\dot{x}_1(t)=x_3(t-h)+v(t),\ x_3(t)=v(t). \qquad (24)$$

В силу утверждения 2 или непосредственно убеждаемся, что регулятор простейшего вида (22), стабилизирующий эту систему, не

существует, однако, замыкая ее регулятором $\breve{v}(p) = Q_1^*(p)\breve{x}_1(p) + Q_2^*(p)\breve{x}_3(p)$ в частотной области, получаем желаемое характеристическое уравнение вида

$$0 = \det\begin{bmatrix} p - Q_1^*(p) & -e^{-ph} - Q_2^*(p) \\ -Q_1^*(p) & 1 - Q_2^*(p) \end{bmatrix} = p - Q_1^*(p) - pQ_2^*(p) - e^{-ph}Q_1^*(p) = p + \alpha,\ \alpha > 0.$$

Отсюда

$$Q_1^*(p) = Q_1 = -\frac{\alpha}{2},\ Q_2^*(p) = \breve{Q}_2(p) = \frac{\alpha + Q_1^*(p) + e^{-ph}Q_1^*(p)}{-p} = \frac{\alpha(1 - e^{-ph})}{-2p}.$$

Возвращаясь к оригиналам, получаем искомый стабилизирующий регулятор вида

$$v(t) = -\frac{\alpha}{2}x_1(t) - \frac{\alpha}{2}\int_0^t q(s)x_3(t-s)ds,\ t \geq 0,$$

где $q(s) = \begin{cases} 1,\ s \in [0,h], \\ 0,\ s > h. \end{cases}$

Рассмотренный пример показывает существенность интегральных элементов при расширении шкалы регуляторов по типу линейной обратной связи.

Заключение. В работе представлены необходимые условия стабилизации с помощью линейных регуляторов типа обратной связи. Показано, что необходимое условие стабилизации с помощью регулятора с интегральными составляющими типа свертки является одновременно и достаточным. Получены условия стабилизации системы простейшим регулятором, не выводящим систему за пределы рассматриваемого класса. Приведен пример системы, для которой не существует простейшего регулятора, позволяющего ее стабилизировать, но находится регулятор с интегральными элементами, решающий эту задачу.

Литература

1. Беллман Р., Кук К. Л. Дифференциально-разностные уравнения. М.: Мир, 1967. 548 с.

2. Марченко В. М., Якименко А. А. К вопросу об устойчивости двумерных дескрипторных систем с запаздывающим аргументом нейтрального типа // Четвертые Богдановские чтения по обыкновенным дифференциальным уравнениям: тез. докл. Междунар. конф., Минск, 7–10 дек. 2005 г. / Ин-т математики НАН Беларуси. Минск, 2005. С. 107–108.

3. Марченко В. М., Луазо Ж.-Ж. Об устойчивости гибридных дифференциально-разностных систем // Дифференциальные уравнения. 2009. Т. 45, № 5. 2009. С. 728–740.

4. Марченко В. М., Якименко А. А. Об устойчивости уравнений с запаздывающим аргументом нейтрального типа // Устойчивость, управление и моделирование динамических систем: материалы Междунар. науч. семинара, Екатеринбург, 15–17 нояб. 2006 г. / УрГУПС. Екатеринбург, 2006. С. 13–14.

5. März R. Solvability of linear differentia algebraic equations with properly stated leading terms // Results in Mathematics 45(2004). Basel: Birkhauser Verlag, 2004. P. 88–95.

6. Кириллова Ф. М., Стрельцов С. В. Необходимые условия оптимальности управлений в гибридных системах // Управляемые системы: сб. ст. Новосибирск, 1975. Вып. 14. С. 24–33.

7. Ахундов А. А. Управляемость линейных гибридных систем // Управляемые системы: сб. ст. Новосибирск, 1975. Вып. 14. С. 4–10.

8. Трофимчук Т. С. Управляемость систем, неразрешенных относительно старшей производной // Управляемые системы: сб. ст. Новосибирск, 1980. Вып. 20. С. 75–82.

9. Марченко В. М., Поддубная О. Н. Представление решений управляемых гибридных систем // Проблемы управления и информатики. 2002. № 6. С. 17–25.

10. Marchenko V. M., Poddubnaya O. N., Zaczkiewicz Z. Hybrid control and observation systems in symmetric form // IEEE conf. «RoMoCo». Poznan, Poland, 2005. P. 137–143.

11. Marchenko V. M., Zaczkiewicz Z. Observability for linear differential-algebraic systems with delay // IEEE conf. «MMAR'2005». Blazejewko, Poland, 2005. P. 299–303.

12. Луазо Ж.-Ж., Марченко В. М. Реализация в шкалах систем с запаздыванием // Доклады РАН. 2002. Т. 383, № 3. С. 305–308.

13. Марченко В. М., Поддубная О. Н. Представление решений и относительная управляемость линейных дифференциально-алгебраических систем со многими запаздываниями // Доклады РАН. 2005. Т. 404, № 4. С. 465–469.

14. De la Sen M. The reachability and observability of hybrid multirate sampling linear systems // Computers Math. Applic. 1996. Vol. 31, no. 1. P. 109–122.

15. Observability of linear hybrid systems / R. Vidal [et al.] // Hybrid systems: Computation and Control. 2003. Vol. 2623 of LNCS. P. 526–539.

16. Gertler J. J., Cruz J. B., Peshkin M. (Eds.) Hybrid systems // Prepr. 13th World Congr. IFAC. 1996. Vol. J. P. 275–311, 473–476.

17. Винер Н., Пэли Р. Преобразование Фурье в комплексной области. М.: Наука, 1964. 269 с.

References

1. Bellman R., Cooke K. L. *Differentsial'no-raznostnyye uravneniya* [Differential-difference equations]. Moscow, Mir Publ., 1967. 548 p.

2. Marchenko V. M., Yakimenka A. A. On the stability of two-dimensional descriptor systems with a delayed argument of neutral type. *Tezisy dokladov Mezhdunarodnoy konferentsii "Chetvertyye Bogdanovskiye chteniya po obyknovennym differentsial'nym uravneniyam"* [Abstracts of the International Conference "Fourth Bogdanov Readings on Ordinary Differential Equations"]. Minsk, 2005, pp. 107–108 (In Russian).

3. Marchenko V. M., Loiseau J.-J. On the stability of hybrid difference-differential systems. *Differentsial'nyye uravneniya* [Differential Equations], 2009, vol. 45, no. 5, pp. 728–740 (In Russian).

4. Marchenko V. M., Yakimenka A. A. On the stability of equations with a delayed argument of neutral type. *Materialy Mezhdunarodnogo nauchnogo seminara "Ustoychivost', upravleniye i modelirovaniye dinamicheskikh system"* [Materials of the International Scientific Workshop "Stability, control and modeling of dynamic systems"]. Yekaterinburg, 2006, pp. 13–14 (In Russian).

5. März R. Solvability of linear differentia algebraic equations with properly stated leading terms. *Results in Mathematics 45(2004)*. Basel, Birkhauser Verlag, 2004. P. 88–95.

6. Kirillova F. M., Strel'tsov S. V. Necessary conditions for optimality of controls in hybrid systems. *Upravlyaemyye sistemy* [Controlled systems]. Novosibirsk, 1975, issue 14, pp. 24–33 (In Russian).

7. Akhundov A. A. Controllability of the linear hybrid systems. *Upravlyaemyye sistemy* [Controlled systems]. Novosibirsk, 1975, issue 14, pp. 4–10 (In Russian).

8. Trofimchuk T. S. Controllability of systems not permitted with respect to the highest derivative. *Upravlyaemyye sistemy* [Controlled systems]. Novosibirsk, 1980, issue 20, pp. 75–82 (In Russian).

9. Marchenko V. M., Poddubnaya O. N. Representation of solutions of controlled hybrid systems. *Problemy upravleniya i informatiki* [Journal of Automation and Information Sciences], 2002, no. 6, pp. 17–25 (In Russian).

10. Marchenko V. M., Poddubnaya O. N., Zaczkiewicz Z. Hybrid control and observation systems in symmetric form. *IEEE conf. "RoMoCo"*. Poznan, 2005, pp. 137–143.

11. Marchenko V. M., Zaczkiewicz Z. Observability for linear differential-algebraic systems with delay. *IEEE conf. "MMAR'2005"*. Blazejewko, 2005, pp. 299–303.

12. Loiseau J.-J., Marchenko V. M. Realization in scales of systems with aftereffect. *Doklady RAN* [Doklady Mathematics], 2002, vol. 383, no. 3, pp. 305–308 (In Russian).

13. Marchenko V. M., Poddubnaya O. N. Representation of solutions and relative controllability of linear differential-algebraic systems with many delays. *Doklady RAN* [Doklady Mathematics], 2005, vol. 404, no. 4, pp. 465–469 (In Russian).

14. De la Sen M. The reachability and observability of hybrid multirate sampling linear systems. *Computers Math. Applic.*, 1996, vol. 31, no. 1, pp. 109–122.

15. Vidal R., Chiuso A., Soato S., Sastry S. Observability of linear hybrid systems. *Hybrid systems: Computation and Control*, 2003, vol. 2623 of LNCS, pp. 526–539.

16. Gertler J. J., Cruz J. B., Peshkin M. (Eds.) Hybrid systems. *Prepr. 13th World Congr. IFAC*, 1996, vol. J, pp. 275–311, 473–476.

17. Wiener N., Paley R. *Preobrazovaniye Fur'ye v kompleksnoy oblasti* [Fourier transform in the complex domain]. Moscow, Nauka Publ., 1964. 269 p.

Информация об авторах

Марченко Владимир Матвеевич − доктор физико-математических наук, профессор. E-mail: vladimir.marchenko@gmail.com

Борковская Инна Мечиславовна − кандидат физико-математических наук, доцент кафедры высшей математики. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: borkovskaia@gmail.com

Information about the authors

Marchenko Vladimir Matveevich − DSc (Physics and Mathematics), Professor. E-mail: vladimir.marchenko@gmail.com

Borkovskaya Inna Mechislavovna − PhD (Physics and Mathematics), Assistant Professor, the Department of Higher Mathematics. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: borkovskaia@gmail.com

Поступила после доработки 11.11.2019

УДК 517.977

А. А. Якименко
Белорусский государственный технологический университет

НЕОБХОДИМОЕ УСЛОВИЕ МОДАЛЬНОЙ УПРАВЛЯЕМОСТИ ЧЕТЫРЕХМЕРНОЙ СИСТЕМОЙ НЕЙТРАЛЬНОГО ТИПА

В публикации получено необходимое условие модальной управляемости четырехмерной стационарной динамической системой с запаздывающим аргументом нейтрального типа с одним запаздыванием для одного класса регуляторов по типу обратной связи. Дано определение задачи модального управления для исследуемой системы. Изучен специальный класс регуляторов по типу обратной связи, с помощью которых может быть решена задача модального управления. Эти регуляторы могут содержать как дифференциально-разностную, так и интегральную части. Показано, что в случае разрешимости задачи модального управления для рассматриваемой системы один из важнейших параметров интегральной части таких регуляторов может быть найден из решения специально подобранного алгебраического уравнения.

Ключевые слова: системы нейтрального типа, модальное управление, дифференциально-разностные регуляторы, обратная связь, запаздывание.

A. A. Yakimenka
Belarusian State Technological University

NECESSARY CONDITION OF MODAL CONTROLLABILITY FOR FOUR-DIMENSIONAL NEUTRAL TYPE SYSTEM

The publication obtained a new necessary condition for modal controllability of a four-dimensional stationary dynamic system with a delayed argument of a neutral type with one delay for one class of feedback regulators. The definition of the modal control problem for the system under study is given. A special class of feedback regulators is considered, with the help of which the modal control problem can be solved. These regulators can contain both differential-difference and integral parts. It is shown that in the case of solvability of the modal control problem for the system under consideration, one of the most important parameters of the integral part of such regulators can be found from the solution of one algebraic equation.

Key words: neutral type systems, modal control, differential-difference regulators, feedback control, delay.

Введение. Задача модального управления является одной из основных задач теории управления. Такая задача хорошо изучена для систем без запаздывания. Для систем с запаздывающим аргументом нейтрального типа [1–10] решение задачи модального управления значительно сложнее. Это обусловлено тем, что пространство состояний таких систем, как правило, бесконечномерно. Задача модального управления требует нахождения регуляторов, ее решающих.

Такие регуляторы можно искать в различных классах. В статье предложены регуляторы, реализация которых достаточна проста. Показано, что если задача модального управления разрешима, то вид этих регуляторов зависит от решения алгебраического уравнения, построенного с помощью коэффициентов рассматриваемой системы.

Основная часть. Рассмотрим четырехмерную линейную стационарную систему с запаздывающим аргументом нейтрального типа с одним входом:

$$\begin{bmatrix}\dot{x}_1(t)\\ \dot{x}_2(t)\\ \dot{x}_3(t)\\ \dot{x}_4(t)\end{bmatrix}=\begin{bmatrix}a_{111} & a_{112} & a_{113} & a_{114}\\ a_{121} & a_{122} & a_{123} & a_{124}\\ a_{131} & a_{132} & a_{133} & a_{134}\\ a_{141} & a_{142} & a_{143} & a_{144}\end{bmatrix}\begin{bmatrix}x_1(t)\\ x_2(t)\\ x_3(t)\\ x_4(t)\end{bmatrix}+$$
$$+\begin{bmatrix}a_{211} & a_{212} & a_{213} & a_{214}\\ a_{221} & a_{222} & a_{223} & a_{224}\\ a_{231} & a_{232} & a_{233} & a_{234}\\ a_{241} & a_{242} & a_{243} & a_{244}\end{bmatrix}\begin{bmatrix}x_1(t-h)\\ x_2(t-h)\\ x_3(t-h)\\ x_4(t-h)\end{bmatrix}+$$
$$+\begin{bmatrix}a_{311} & a_{312} & a_{313} & a_{314}\\ a_{321} & a_{322} & a_{323} & a_{324}\\ a_{331} & a_{332} & a_{333} & a_{334}\\ a_{341} & a_{342} & a_{343} & a_{344}\end{bmatrix}\begin{bmatrix}\dot{x}_1(t-h)\\ \dot{x}_2(t-h)\\ \dot{x}_3(t-h)\\ \dot{x}_4(t-h)\end{bmatrix}+\begin{bmatrix}0\\ 0\\ 0\\ 1\end{bmatrix}u(t), \quad (1)$$

где $h > 0$ – постоянное запаздывание.

Характеристический квазиполином системы (1) имеет вид

$$\sum_{i=0}^{4}\sum_{j=0}^{4}\alpha_{ij}\lambda^i e^{-j\lambda h}, \quad (2)$$

где $\alpha_{ij} \in \mathbb{R}$ – числа, зависящие от коэффициентов системы (1), $\alpha_{40} = 1$. Задача модального управления состоит в том, чтобы для любых наперед заданных чисел β_{ij}, $i = 1, 2, 3, 4, j = 1, 2, 3, 4$, $\beta_{30} = 1$ найти такой линейный регулятор, при котором система (1), замкнутая этим регулятором, имеет характеристический квазиполином

$$\sum_{i=0}^{4}\sum_{j=0}^{4}\beta_{ij}\lambda^i e^{-j\lambda h}. \qquad (3)$$

Регулятор будем искать в форме

$$u(t) = q'_{00}x(t) + \sum_{i=0}^{L}\sum_{j=1}^{N} q'_{ij}x^{(i)}(t-jh) + \int_{-h}^{0} q'(s)x(t+s)\,ds, \qquad (4)$$

где $q_{ij} \in \mathbb{R}^4$, штрих $(\cdot)'$ означает транспонирование, $L, N \in \mathbb{N}$,

$$x(t) = \begin{bmatrix} x_1(t) & x_2(t) & x_3(t) & x_4(t)\end{bmatrix}',$$

$$x^{(i)}(\cdot) \equiv \frac{d^i x(\cdot)}{dt^i},\ x^{(0)}(\cdot) \equiv x(\cdot).$$

В частотной области регулятор (4) примет вид

$$U(\lambda) = q'_{00} + \sum_{i=0}^{L}\sum_{j=1}^{M} q'_{ij}\lambda^i e^{-j\lambda h} + G'(\lambda). \qquad (5)$$

Если ядро интегральной части регулятора $G'(\lambda) \equiv 0$, то регулятор (5) примет наиболее простой дифференциально-разностный вид. Однако задача модального управления при этом решается лишь в исключительных случаях. Пусть теперь $G'(\lambda) \not\equiv 0$ и имеет вид

$$G'(\lambda) = [g_1 \quad g_2 \quad g_3 \quad g_4],$$

где

$$g_i = \sum_{j=1}^{S}\beta_{ij}\frac{\alpha_j(\lambda, e^{-\lambda h})}{(\lambda-\xi)^j}, \qquad (6)$$

$i = \overline{1, 4}$, $\xi \in \mathbb{C}$; $\alpha_j(\lambda, e^{-\lambda h})$, $j = \overline{1, S}$ выбираются так, чтобы функции $\dfrac{\alpha_j(\lambda, e^{-\lambda h})}{(\lambda-\xi)^j}$ были целыми и удовлетворяли условиям теоремы Винера – Пэли, например, $\alpha_1(\lambda, e^{-\lambda h}) = e^{-\xi h} - e^{-\lambda h}$. Пусть $S = 1$. Тогда характеристический квазиполином замкнутой этим регулятором системы (1) имеет следующий вид:

$$\begin{vmatrix} a_{11}(\cdot)-\lambda & a_{12}(\cdot) & a_{13}(\cdot) & a_{14}(\cdot) \\ a_{21}(\cdot) & a_{22}(\cdot)-\lambda & a_{23}(\cdot) & a_{24}(\cdot) \\ a_{31}(\cdot) & a_{32}(\cdot) & a_{33}(\cdot)-\lambda & a_{34}(\cdot) \\ \tilde{a}_{41}(\cdot)+\beta_1(\cdot) & \tilde{a}_{42}(\cdot)+\beta_2(\cdot) & \tilde{a}_{43}(\cdot)+\beta_3(\cdot) & \tilde{a}_{44}(\cdot)-\lambda+\beta_4(\cdot) \end{vmatrix} =$$

$$= \begin{vmatrix} a_{11}(\cdot)-\lambda & a_{12}(\cdot) & a_{13}(\cdot) & a_{14}(\cdot) \\ a_{21}(\cdot) & a_{22}(\cdot)-\lambda & a_{23}(\cdot) & a_{24}(\cdot) \\ a_{31}(\cdot) & a_{32}(\cdot) & a_{33}(\cdot)-\lambda & a_{34}(\cdot) \\ \tilde{a}_{41}(\cdot) & \tilde{a}_{42}(\cdot) & \tilde{a}_{43}(\cdot) & \tilde{a}_{44}(\cdot)-\lambda \end{vmatrix} +$$

$$+ \begin{vmatrix} a_{11}(\cdot)-\lambda & a_{12}(\cdot) & a_{13}(\cdot) & a_{14}(\cdot) \\ a_{21}(\cdot) & a_{22}(\cdot)-\lambda & a_{23}(\cdot) & a_{24}(\cdot) \\ a_{31}(\cdot) & a_{32}(\cdot) & a_{33}(\cdot)-\lambda & a_{34}(\cdot) \\ \beta_1(\cdot) & \beta_2(\cdot) & \beta_3(\cdot) & \beta_4(\cdot) \end{vmatrix}, \qquad (7)$$

где $a_{ij}(\cdot) = a_{1ij} + a_{2ij}e^{-\lambda h} + a_{3ij}\lambda e^{-\lambda h}$, $i = \overline{1,3}$, $j = \overline{1, 4}$, $\tilde{a}_{4j} = q_j(\lambda, e^{-\lambda h}) + a_{14j} + a_{24j}e^{-\lambda h} + a_{34j}\lambda e^{-\lambda h}$, a_{ijk}, $i = \overline{1, 3}$, $j = \overline{1, 4}$, $k = \overline{1, 4}$ определены в (1), $q_j(\lambda, e^{-\lambda h})$ – квазиполиномы, определяемые дифференциально-разностной частью регулятора (5), $\beta_j(\cdot) = \beta_{j1}\dfrac{\alpha_1(\lambda, e^{-\lambda h})}{(\lambda-\xi)}$, $j = \overline{1, 4}$; β_{j1} определены в (6).

Второе слагаемое в правой части (7) перепишем в виде

$$\begin{vmatrix} a_{11}(\cdot)-\lambda & a_{12}(\cdot) & a_{13}(\cdot) & a_{14}(\cdot) \\ a_{21}(\cdot) & a_{22}(\cdot)-\lambda & a_{23}(\cdot) & a_{24}(\cdot) \\ a_{31}(\cdot) & a_{32}(\cdot) & a_{33}(\cdot)-\lambda & a_{34}(\cdot) \\ \beta_{11} & \beta_{21} & \beta_{31} & \beta_{41} \end{vmatrix} \frac{e^{-\xi h} - e^{-\lambda h}}{\lambda-\xi}. \qquad (8)$$

Обозначим $m = e^{-\lambda h}$. Представим элементы матрицы, определитель которой записан в (8) в следующем виде:

$$a_{ij}(\cdot) = a_{1ij} + a_{2ij}m + a_{3ij}\lambda m =$$
$$= a_{1ij} + a_{2ij}m + a_{3ij}\xi m + a_{3ij}(\lambda-\xi)m,\ i \neq j,$$
$$a_{ii}(\cdot) = a_{1ii} + a_{2ii}m + a_{3ii}\lambda m - \lambda =$$
$$= a_{1ii} + a_{2ii}m + a_{3ii}\xi m - \xi + a_{3ii}(\lambda-\xi)m - (\lambda-\xi),$$
$$i = \overline{1, 3},\ j = \overline{1, 4}.$$

Тогда, выделяя слагаемые, которые содержат множители $(\lambda - \xi)$ и используя следующее свойство определителей: определитель, в каждом

элементе строки которого есть сумма двух слагаемых, равен сумме двух определителей, (8) перепишется в виде

$$c\left(\lambda, e^{-\lambda h}, \xi\right) + \frac{e^{-\xi h} - e^{-\lambda h}}{\lambda - \xi} \times$$

$$\times \left(\begin{vmatrix} a_{111}-\xi & a_{112} & a_{113} & a_{114} \\ a_{121} & a_{122}-\xi & a_{123} & a_{124} \\ a_{131} & a_{132} & a_{133}-\xi & a_{134} \\ \beta_{11} & \beta_{21} & \beta_{31} & \beta_{41} \end{vmatrix} + \right.$$

$$+m \left(\begin{vmatrix} a_{111}-\xi & a_{112} & a_{113} & a_{114} \\ a_{121} & a_{122}-\xi & a_{123} & a_{124} \\ a_{231}+a_{331}\xi & a_{232}+a_{332}\xi & a_{233}+a_{333}\xi & a_{234}+a_{334}\xi \\ \beta_{11} & \beta_{21} & \beta_{31} & \beta_{41} \end{vmatrix} + \right.$$

$$+ \begin{vmatrix} a_{111}-\xi & a_{112} & a_{113} & a_{114} \\ a_{221}+a_{321}\xi & a_{222}+a_{322}\xi & a_{223}+a_{323}\xi & a_{224}+a_{324}\xi \\ a_{131} & a_{132} & a_{133}-\xi & a_{134} \\ \beta_{11} & \beta_{21} & \beta_{31} & \beta_{41} \end{vmatrix} +$$

$$\left. + \begin{vmatrix} a_{211}+a_{311}\xi & a_{212}+a_{312}\xi & a_{213}+a_{313}\xi & a_{214}+a_{314}\xi \\ a_{121} & a_{122}-\xi & a_{123} & a_{124} \\ a_{131} & a_{132} & a_{133}-\xi & a_{134} \\ \beta_{11} & \beta_{21} & \beta_{31} & \beta_{41} \end{vmatrix} \right) +$$

$$+m^2 \left(\begin{vmatrix} a_{111}-\xi & a_{112} & a_{113} & a_{114} \\ a_{221}+a_{321}\xi & a_{222}+a_{322}\xi & a_{223}+a_{323}\xi & a_{224}+a_{324}\xi \\ a_{231}+a_{331}\xi & a_{232}+a_{332}\xi & a_{233}+a_{333}\xi & a_{234}+a_{334}\xi \\ \beta_{11} & \beta_{21} & \beta_{31} & \beta_{41} \end{vmatrix} + \right.$$

$$+ \begin{vmatrix} a_{211}+a_{311}\xi & a_{212}+a_{312}\xi & a_{213}+a_{313}\xi & a_{214}+a_{314}\xi \\ a_{121} & a_{122}-\xi & a_{123} & a_{124} \\ a_{231}+a_{331}\xi & a_{232}+a_{332}\xi & a_{233}+a_{333}\xi & a_{234}+a_{334}\xi \\ \beta_{11} & \beta_{21} & \beta_{31} & \beta_{41} \end{vmatrix} +$$

$$\left. + \begin{vmatrix} a_{211}+a_{311}\xi & a_{212}+a_{312}\xi & a_{213}+a_{313}\xi & a_{214}+a_{314}\xi \\ a_{221}+a_{321}\xi & a_{222}+a_{322}\xi & a_{223}+a_{323}\xi & a_{224}+a_{324}\xi \\ a_{131} & a_{132} & a_{133}-\xi & a_{134} \\ \beta_{11} & \beta_{21} & \beta_{31} & \beta_{41} \end{vmatrix} \right) +$$

$$+m^3 \begin{vmatrix} a_{211}+a_{311}\xi & a_{212}+a_{312}\xi & a_{213}+a_{313}\xi & a_{214}+a_{314}\xi \\ a_{221}+a_{321}\xi & a_{222}+a_{322}\xi & a_{223}+a_{323}\xi & a_{224}+a_{324}\xi \\ a_{231}+a_{331}\xi & a_{232}+a_{332}\xi & a_{233}+a_{333}\xi & a_{234}+a_{334}\xi \\ \beta_{11} & \beta_{21} & \beta_{31} & \beta_{41} \end{vmatrix}, \quad (9)$$

где $c\left(\lambda, e^{-\lambda h}, \xi\right)$ – некоторый квазиполином, однозначно определяемый соотношением (8).

Из того, что характеристический квазиполином (3) замкнутой системы не содержит слагаемые, умножаемые на функцию $\frac{e^{-\xi h} - e^{-\lambda h}}{\lambda - \xi}$, следует, что второе слагаемое в (9) должно быть тождественно равно нулю. Для этого нужно, чтобы многочлен относительно m в скобках правой части (9) был бы нулевым. Так как из условия $G'(\lambda) \not\equiv 0$ вытекает, что $\beta_{11}^2 + \beta_{21}^2 + \beta_{31}^2 + \beta_{41}^2 \neq 0$, то отсюда с учетом (9) следует, что система линейных алгебраических уравнений

$$\begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} \\ \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} \end{bmatrix} \begin{bmatrix} \beta_{11} \\ \beta_{21} \\ \beta_{31} \\ \beta_{41} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad (10)$$

где

$$\gamma_{11} = \begin{vmatrix} a_{112} & a_{113} & a_{114} \\ a_{122}-\xi & a_{123} & a_{124} \\ a_{132} & a_{133}-\xi & a_{134} \end{vmatrix},$$

$$\gamma_{12} = \begin{vmatrix} a_{111}-\xi & a_{113} & a_{114} \\ a_{121} & a_{123} & a_{124} \\ a_{131} & a_{133}-\xi & a_{134} \end{vmatrix},$$

$$\gamma_{13} = \begin{vmatrix} a_{111}-\xi & a_{112} & a_{114} \\ a_{121} & a_{122}-\xi & a_{124} \\ a_{131} & a_{132} & a_{134} \end{vmatrix},$$

$$\gamma_{14} = \begin{vmatrix} a_{111}-\xi & a_{112} & a_{113} \\ a_{121} & a_{122}-\xi & a_{123} \\ a_{131} & a_{132} & a_{133}-\xi \end{vmatrix},$$

$$\gamma_{21} = \begin{vmatrix} a_{112} & a_{113} & a_{114} \\ a_{122}-\xi & a_{123} & a_{124} \\ a_{232}+a_{332}\xi & a_{233}+a_{333}\xi & a_{234}+a_{334}\xi \end{vmatrix} +$$

$$+ \begin{vmatrix} a_{112} & a_{113} & a_{114} \\ a_{222}+a_{322}\xi & a_{223}+a_{323}\xi & a_{224}+a_{324}\xi \\ a_{132} & a_{133}-\xi & a_{134} \end{vmatrix} +$$

$$+ \begin{vmatrix} a_{212}+a_{312}\xi & a_{213}+a_{313}\xi & a_{214}+a_{314}\xi \\ a_{122}-\xi & a_{123} & a_{124} \\ a_{132} & a_{133}-\xi & a_{134} \end{vmatrix},$$

$$\gamma_{22} = \begin{vmatrix} a_{111}-\xi & a_{113} & a_{114} \\ a_{121} & a_{123} & a_{124} \\ a_{231}+a_{331}\xi & a_{233}+a_{333}\xi & a_{234}+a_{334}\xi \end{vmatrix} +$$

$$+\begin{vmatrix} a_{111}-\xi & a_{113} & a_{114} \\ a_{221}+a_{321}\xi & a_{223}+a_{323}\xi & a_{224}+a_{324}\xi \\ a_{131} & a_{133}-\xi & a_{134} \end{vmatrix}+$$

$$+\begin{vmatrix} a_{211}+a_{311}\xi & a_{213}+a_{313}\xi & a_{214}+a_{314}\xi \\ a_{121} & a_{123} & a_{124} \\ a_{131} & a_{133}-\xi & a_{134} \end{vmatrix},$$

$$\gamma_{23}=\begin{vmatrix} a_{111}-\xi & a_{112} & a_{114} \\ a_{121} & a_{122}-\xi & a_{124} \\ a_{231}+a_{331}\xi & a_{232}+a_{332}\xi & a_{234}+a_{334}\xi \end{vmatrix}+$$

$$+\begin{vmatrix} a_{111}-\xi & a_{112} & a_{114} \\ a_{221}+a_{321}\xi & a_{222}+a_{322}\xi & a_{224}+a_{324}\xi \\ a_{131} & a_{132} & a_{134} \end{vmatrix}+$$

$$+\begin{vmatrix} a_{211}+a_{311}\xi & a_{212}+a_{312}\xi & a_{214}+a_{314}\xi \\ a_{121} & a_{122}-\xi & a_{124} \\ a_{131} & a_{132} & a_{134} \end{vmatrix},$$

$$\gamma_{24}=\begin{vmatrix} a_{111}-\xi & a_{112} & a_{113} \\ a_{121} & a_{122}-\xi & a_{123} \\ a_{231}+a_{331}\xi & a_{232}+a_{332}\xi & a_{233}+a_{333}\xi \end{vmatrix}+$$

$$+\begin{vmatrix} a_{111}-\xi & a_{112} & a_{113} \\ a_{221}+a_{321}\xi & a_{222}+a_{322}\xi & a_{223}+a_{323}\xi \\ a_{131} & a_{132} & a_{133}-\xi \end{vmatrix}+$$

$$+\begin{vmatrix} a_{211}+a_{311}\xi & a_{212}+a_{312}\xi & a_{213}+a_{313}\xi \\ a_{121} & a_{122}-\xi & a_{123} \\ a_{131} & a_{132} & a_{133}-\xi \end{vmatrix},$$

$$\gamma_{31}=\begin{vmatrix} a_{112} & a_{113} & a_{114} \\ a_{222}+a_{322}\xi & a_{223}+a_{323}\xi & a_{224}+a_{324}\xi \\ a_{232}+a_{332}\xi & a_{233}+a_{333}\xi & a_{234}+a_{334}\xi \end{vmatrix}+$$

$$+\begin{vmatrix} a_{212}+a_{312}\xi & a_{213}+a_{313}\xi & a_{214}+a_{314}\xi \\ a_{122}-\xi & a_{123} & a_{124} \\ a_{232}+a_{332}\xi & a_{233}+a_{333}\xi & a_{234}+a_{334}\xi \end{vmatrix}+$$

$$+\begin{vmatrix} a_{212}+a_{312}\xi & a_{213}+a_{313}\xi & a_{214}+a_{314}\xi \\ a_{222}+a_{322}\xi & a_{223}+a_{323}\xi & a_{223}+a_{323}\xi \\ a_{132} & a_{133}-\xi & a_{134} \end{vmatrix},$$

$$\gamma_{32}=\begin{vmatrix} a_{111}-\xi & a_{113} & a_{114} \\ a_{221}+a_{321}\xi & a_{223}+a_{323}\xi & a_{224}+a_{324}\xi \\ a_{231}+a_{331}\xi & a_{233}+a_{333}\xi & a_{234}+a_{334}\xi \end{vmatrix}+$$

$$+\begin{vmatrix} a_{211}+a_{311}\xi & a_{213}+a_{313}\xi & a_{214}+a_{314}\xi \\ a_{121} & a_{123} & a_{124} \\ a_{231}+a_{331}\xi & a_{233}+a_{333}\xi & a_{234}+a_{334}\xi \end{vmatrix}+$$

$$+\begin{vmatrix} a_{211}+a_{311}\xi & a_{213}+a_{313}\xi & a_{214}+a_{314}\xi \\ a_{221}+a_{321}\xi & a_{223}+a_{323}\xi & a_{224}+a_{324}\xi \\ a_{131} & a_{133}-\xi & a_{134} \end{vmatrix},$$

$$\gamma_{33}=\begin{vmatrix} a_{111}-\xi & a_{112} & a_{114} \\ a_{221}+a_{321}\xi & a_{222}+a_{322}\xi & a_{224}+a_{324}\xi \\ a_{231}+a_{331}\xi & a_{232}+a_{332}\xi & a_{234}+a_{334}\xi \end{vmatrix}+$$

$$+\begin{vmatrix} a_{211}+a_{311}\xi & a_{212}+a_{312}\xi & a_{214}+a_{314}\xi \\ a_{121} & a_{122}-\xi & a_{124} \\ a_{231}+a_{331}\xi & a_{232}+a_{332}\xi & a_{234}+a_{334}\xi \end{vmatrix}+$$

$$+\begin{vmatrix} a_{211}+a_{311}\xi & a_{212}+a_{312}\xi & a_{214}+a_{314}\xi \\ a_{221}+a_{321}\xi & a_{222}+a_{322}\xi & a_{224}+a_{324}\xi \\ a_{131} & a_{132} & a_{134} \end{vmatrix},$$

$$\gamma_{34}=\begin{vmatrix} a_{111}-\xi & a_{112} & a_{113} \\ a_{221}+a_{321}\xi & a_{222}+a_{322}\xi & a_{223}+a_{323}\xi \\ a_{231}+a_{331}\xi & a_{232}+a_{332}\xi & a_{233}+a_{333}\xi \end{vmatrix}+$$

$$+\begin{vmatrix} a_{211}+a_{311}\xi & a_{212}+a_{312}\xi & a_{213}+a_{313}\xi \\ a_{121} & a_{122}-\xi & a_{123} \\ a_{231}+a_{331}\xi & a_{232}+a_{332}\xi & a_{233}+a_{333}\xi \end{vmatrix}+$$

$$+\begin{vmatrix} a_{211}+a_{311}\xi & a_{212}+a_{312}\xi & a_{213}+a_{313}\xi \\ a_{221}+a_{321}\xi & a_{222}+a_{322}\xi & a_{223}+a_{323}\xi \\ a_{131} & a_{132} & a_{133}-\xi \end{vmatrix},$$

$$\gamma_{41}=\begin{vmatrix} a_{212}+a_{312}\xi & a_{213}+a_{313}\xi & a_{214}+a_{314}\xi \\ a_{222}+a_{322}\xi & a_{223}+a_{323}\xi & a_{224}+a_{324}\xi \\ a_{232}+a_{332}\xi & a_{233}+a_{333}\xi & a_{234}+a_{334}\xi \end{vmatrix},$$

$$\gamma_{42}=\begin{vmatrix} a_{211}+a_{311}\xi & a_{213}+a_{313}\xi & a_{214}+a_{314}\xi \\ a_{221}+a_{321}\xi & a_{223}+a_{323}\xi & a_{224}+a_{324}\xi \\ a_{231}+a_{331}\xi & a_{233}+a_{333}\xi & a_{234}+a_{334}\xi \end{vmatrix},$$

$$\gamma_{43}=\begin{vmatrix} a_{211}+a_{311}\xi & a_{212}+a_{312}\xi & a_{214}+a_{314}\xi \\ a_{221}+a_{321}\xi & a_{222}+a_{322}\xi & a_{224}+a_{324}\xi \\ a_{231}+a_{331}\xi & a_{232}+a_{332}\xi & a_{234}+a_{334}\xi \end{vmatrix},$$

$$\gamma_{44}=\begin{vmatrix} a_{211}+a_{311}\xi & a_{212}+a_{312}\xi & a_{213}+a_{313}\xi \\ a_{221}+a_{321}\xi & a_{222}+a_{322}\xi & a_{223}+a_{323}\xi \\ a_{231}+a_{331}\xi & a_{232}+a_{332}\xi & a_{233}+a_{333}\xi \end{vmatrix},$$

имеет нетривиальное решение. Отсюда определитель матрицы системы (10) должен быть равен нулю. Очевидно, что определитель матрицы системы (10) представляет собой многочлен

относительно переменной ξ степени не выше, чем 12.

Таким образом, доказана следующая теорема.

Теорема. Пусть система (1) модально управляема регулятором (4) (или (5)), причем в (5) $G'(\lambda) \not\equiv 0$ имеет вид (6). Тогда ξ будут корнями многочлена степени не выше 12, который является определителем матрицы системы (10).

Заключение. Полученная теорема дает одно необходимое условие модальной управляемости системы (1) в классе регуляторов (4) (или (5)). Вопрос о том, достаточно ли регуляторов (4) (или (5)) для решения задачи модальной управляемости остается открытым. Доказано [6–9], что для аналогичной системы второго порядка такие регуляторы решают задачу модального управления.

Условие теоремы позволяет свести бесконечномерную вариационную задачу нахождения регуляторов к конечномерной задаче нахождения коэффициентов регулятора (4), что существенно упрощает решение задачи модального управления.

Литература

1. Марченко В. М. О проблеме модального управления в линейных системах с запаздыванием // Доклады Академии наук БССР. 1978. № 5. С. 401–404.

2. Salamon D. Control and Observation of Neutral Systems. London: Pitman Press, 1984. 362 p.

3. Wonham W. M. On pole assignment in multi-input controllable systems // IEEE Trans. Automat. Contr. 1967. Vol. AC-12, no. 6. P. 660–665.

4. Кириллова Ф. М., Марченко В. М. Функциональные преобразования и некоторые канонические формы в линейных системах с запаздывающим аргументом. Минск, 1978. 28 с. (Препринт / Акад. наук БССР, № 7 (39)).

5. Spong M. W. A semistate approach to feedback stabilization of neutral delay systems // Circuits Systems Signal Process. 1986. Vol. 5, no. 1. P. 69–84.

6. Якименко А. А. Модальное управление одной запаздывающей системой // Труды БГТУ. 2013. № 6: Физ.-мат. науки и информатика. С. 3–7.

7. Якименко А. А. Модальное управление одной системой нейтрального типа // Труды БГТУ. 2016. № 6: Физ.-мат. науки и информатика. С. 18–21.

8. Якименко А. А. Модальное управление одной системой нейтрального типа в общециклическом случае // Труды БГТУ. Сер. 3, Физ.-мат. науки и информатика. 2017. № 2. С. 25–27.

9. Якименко А. А. Модальное управление одной системой нейтрального типа в общециклическом случае при кратных корнях // Труды БГТУ. Сер. 3, Физ.-мат. науки и информатика. 2018. № 1. С. 5–8.

10. Якименко А. А. Достаточное условие модальной управляемости для систем нейтрального типа с соизмеримыми запаздываниями // Труды БГТУ. Сер. 3, Физ.-мат. науки и информатика. 2019. № 2. С. 17–21.

References

1. Marchenko V. M. On problem of modal control in linear systems with delay. *Doklady Akademii nauk BSSR* [Reports of the BSSR Academy of Science], 1978, no. 5, pp. 401–404 (In Russian).

2. Salamon D. Control and Observation of Neutral Systems. London, Pitman Press, 1984. 362 p.

3. Wonham W. M. On pole assignment in multi-input controllable systems. *IEEE Trans. Automat. Contr.*, 1967, vol. AC-12, no. 6, pp. 660–665.

4. Kirillova F. M., Marchenko V. M. *Funktsional'nyye preobrazovaniya i nekotoryye kanonicheskiye formy v lineynykh sistemakh s zapazdyvayushchim argumentom* [Functional transforms and some canonical forms for linear retarded systems]. Minsk, 1978. 28 p.

5. Spong M. W. A semistate approach to feedback stabilization of neutral delay systems. *Circuits Systems Signal Process*, 1986, vol. 5, no. 1, pp. 69–84.

6. Yakimenka A. A. Modal control for one delayed system. *Trudy BGTU* [Proceedings of BSTU], 2013, no. 6: Physics and Mathematics. Informatics, pp. 3–7 (In Russian).

7. Yakimenka A. A. Modal control for one neutral type system. *Trudy BGTU* [Proceedings of BSTU], 2016, no. 6: Physics and Mathematics. Informatics, pp. 18–21 (In Russian).

8. Yakimenka A. A. Modal control for one neutral type system in general cyclic case. *Trudy BGTU* [Proceedings of BSTU], series 3, Physics and Mathematics. Informatics, 2017, no. 2, pp. 25–27 (In Russian).

9. Yakimenka A. A. Modal control for one neutral type system in general cyclic case with double roots. *Trudy BGTU* [Proceedings of BSTU], series 3, Physics and Mathematics. Informatics, 2018, no. 1, pp. 5–8 (In Russian).

10. Yakimenka A. A. Sufficient condition of modal controllability for neutral type systems with commensurate delays. *Trudy BGTU* [Proceedings of BSTU], series 3, Physics and Mathematics. Informatics, 2019, no. 2, pp. 17–21 (In Russian).

Информация об авторе

Якименко Андрей Александрович – кандидат физико-математических наук, доцент, доцент кафедры высшей математики. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: yakimenko@belstu.by

Information about the author

Yakimenka Andrei Aliaksandravich – PhD (Physics and Mathematics), Associate Professor, Assistant Professor, the Department of Higher Mathematics. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: yakimenko@belstu.by

Поступила после доработки 19.11.2019

УДК 512.815.6

Н. П. Можей
Белорусский государственный университет информатики и радиоэлектроники

ЛИНЕЙНЫЕ АЛГЕБРЫ ЛИ, СОСТОЯЩИЕ ИЗ НИЛЬПОТЕНТНЫХ ЭНДОМОРФИЗМОВ

Исследование линейных групп Ли сопряжено, с одной стороны, с более общей задачей изучения произвольных линейных групп, с другой стороны, линейные группы Ли тесно связаны с алгебраическими группами. Цель работы – описание с точностью до сопряженности подалгебр алгебры Ли $\mathfrak{gl}(4,\mathbb{C})$, состоящих из нильпотентных эндоморфизмов. Решение этой задачи является первым шагом в классификации всех подалгебр алгебры Ли $\mathfrak{gl}(4,\mathbb{C})$. Определены основные понятия: линейная алгебра Ли, разделяющая алгебра Ли, разделяющая оболочка, линейный нильрадикал, подалгебра Мальцева, разложение Мальцева, ступень нильпотентности. Приведен алгоритм классификации линейных алгебр ступени нильпотентности n по алгебре ступени $n-1$, а также показано, что решение задачи классификации подалгебр алгебры Ли $\mathfrak{gl}(4,\mathbb{C})$ сводится к классификации линейных алгебр Ли, состоящих из нильпотентных эндоморфизмов, классификации максимальных разделяющих алгебр Ли с каждым линейным нильрадикалом, классификации немаксимальных разделяющих алгебр Ли и классификации неразделяющих линейных алгебр Ли с каждой разделяющей оболочкой. Рассмотрено в явном виде описание линейных алгебр Ли на четырехмерном пространстве, состоящих из нильпотентных эндоморфизмов. Алгоритмы, приведенные в работе, могут быть компьютеризованы и использованы для решения аналогичных задач в больших размерностях.

Ключевые слова: нильпотентный эндоморфизм, линейная группа Ли, алгебра Ли, разделяющая алгебра Ли.

N. P. Mozhey
Belarusian State University of Informatics and Radioelectronics

LINEAR LIE ALGEBRAS CONSISTING OF NILPOTENT ENDOMORPHISMS

The study of linear Lie groups is connected, on the one hand, with the more general problem of studying arbitrary linear groups, on the other hand, linear Lie groups are closely connected with algebraic groups. The purpose of the paper is to describe, up to conjugacy, subalgebras of the Lie algebra $\mathfrak{gl}(4,\mathbb{C})$, consisting of nilpotent endomorphisms. The solution to this problem is the first step in the classification of all subalgebras of the Lie algebra $\mathfrak{gl}(4,\mathbb{C})$. The basic concepts are defined: linear Lie algebra, dividing Lie algebra, dividing cover, linear nilradical, Maltsev's subalgebra, Maltsev's decomposition, nilpotency class. An algorithm for the classification of linear algebras of the nilpotency class n by the algebra of the degree $n-1$ is obtained, and it is also shown that the solution to the classification of subalgebras of the Lie algebra $\mathfrak{gl}(4,\mathbb{C})$ reduces to the classification of linear Lie algebras consisting of nilpotent endomorphisms, the classification of maximal dividing Lie algebras with each linear nilradical, and the classification of non-maximal dividing Lie algebras; and classifications of non-dividing linear Lie algebras with each dividing cover. An explicit description is given of linear Lie algebras on a four-dimensional space consisting of nilpotent endomorphisms. The algorithms described in the work can be computerized and used to solve similar problems in large dimensions.

Key words: nilpotent endomorphism, linear Lie group, Lie algebra, dividing Lie algebra.

Введение. Исследование линейных групп Ли сопряжено, с одной стороны, с более общей задачей изучения произвольных линейных групп (не обязательно групп Ли и над произвольными полями), о таких группах см., например, [1, 2]. С другой стороны, линейные группы Ли тесно связаны с алгебраическими группами над полями $\mathbb{R}$ и $\mathbb{C}$. А именно, любая алгебраическая, т. е. замкнутая в топологии Зарисского, линейная группа является линейной алгебраической группой. Алгебраические линейные группы составляют наиболее изученный класс линейных групп.

Целью данной работы является описание классификации подалгебр алгебр Ли $\mathfrak{gl}(4,P)$, где $P=\mathbb{R}$ или $\mathbb{C}$, состоящих из нильпотентных эндоморфизмов, с точностью до сопряженности.

Основная часть. Пусть V – фиксированное конечномерное векторное пространство над полем нулевой характеристики. Напомним, что

подалгебры алгебры Ли $\mathfrak{gl}(V)$ называются *линейными алгебрами Ли*. Будем говорить, что линейные алгебры Ли $\mathfrak{g}_1$ и $\mathfrak{g}_2$ сопряжены, если найдется такой элемент $\varphi \in GL(V)$, что $\varphi \cdot \mathfrak{g}_1 \cdot \varphi^{-1} = \mathfrak{g}_2$. Основой описываемой методики является теория разделяющих алгебр Ли, развитая в [3].

Любой эндоморфизм x пространства V единственным образом может быть представлен в виде суммы коммутирующих полупростого s и нильпотентного n эндоморфизмов пространства V. При этом s называется полупростой компонентой эндоморфизма x, а n – его нильпотентной компонентой. Разложение $x = s + n$ называется разложением Жордана эндоморфизма x. Линейная алгебра Ли называется *разделяющей*, если она содержит полупростую и нильпотентную компоненты каждого своего элемента [3]. *Разделяющей оболочкой* линейной алгебры Ли $\mathfrak{g}$ называется минимальная разделяющая линейная алгебра Ли, содержащая $\mathfrak{g}$, она обозначается $e(\mathfrak{g})$ [3].

Пусть $\mathfrak{g}$ – разделяющая линейная алгебра Ли и $\mathfrak{n} = \mathfrak{n}_V(\mathfrak{g})$ – наибольший идеал нильпотентности тождественного представления алгебры Ли $\mathfrak{g}$ в V, который мы будем называть *линейным нильрадикалом* линейной алгебры Ли $\mathfrak{g}$. Известно, что существует такая подалгебра $\mathfrak{m} \subset \mathfrak{g}$, редуктивная в $\mathfrak{gl}(V)$, что $\mathfrak{g}$ является прямой суммой подпространств $\mathfrak{m}$ и $\mathfrak{n}$ [3]. Любая подалгебра $\mathfrak{m}$ с указанными свойствами называется *подалгеброй Мальцева* разделяющей алгебры Ли $\mathfrak{g}$. Разложение $\mathfrak{g} = \mathfrak{m} \oplus \mathfrak{n}$ называется *разложением Мальцева* разделяющей алгебры Ли $\mathfrak{g}$. Заметим, что подалгебра Мальцева $\mathfrak{m}$ является прямой суммой подалгебры Леви $\mathfrak{s} = [\mathfrak{m}, \mathfrak{m}]$ алгебры Ли $\mathfrak{g}$ и центра $\mathfrak{t}$ алгебры Ли $\mathfrak{m}$, а $\mathfrak{t}$ – максимальным элементом множества коммутативных подалгебр радикала $\mathfrak{r}$ алгебры Ли $\mathfrak{g}$, состоящих из полупростых эндоморфизмов. Кроме того, идеал $\mathfrak{n}$ совпадает с множеством нильпотентных эндоморфизмов, лежащих в радикале $\mathfrak{r}$, и $\mathfrak{r} = \mathfrak{t} \oplus \mathfrak{n}$.

Пусть далее $\mathfrak{n}$ – линейная алгебра Ли, состоящая из нильпотентных эндоморфизмов. Множество разделяющих алгебр Ли с линейным нильрадикалом $\mathfrak{n}$ обладает *максимальным элементом* $\overline{\mathfrak{g}}$, причем все максимальные элементы этого множества сопряжены алгебре Ли $\overline{\mathfrak{g}}$. Пусть $\overline{\mathfrak{g}} = \overline{\mathfrak{m}} \oplus \mathfrak{n}$ – разложение Мальцева разделяющей алгебры Ли $\overline{\mathfrak{g}}$. Любая разделяющая алгебра Ли с линейным нильрадикалом $\mathfrak{n}$ сопряжена некоторой разделяющей алгебре Ли $\mathfrak{g}$ с разложением Мальцева $\mathfrak{g} = \mathfrak{m} \oplus \mathfrak{n}$, у которого $\mathfrak{m} \subset \overline{\mathfrak{m}}$.

Таким образом, решение задачи классификации подалгебр алгебр Ли $\mathfrak{gl}(4, P)$ разбивается на следующие подзадачи:

1) классификация линейных алгебр Ли, состоящих из нильпотентных эндоморфизмов;

2) классификация максимальных разделяющих алгебр Ли с каждым линейным нильрадикалом из пункта 1;

3) классификация немаксимальных разделяющих алгебр Ли с каждым линейным нильрадикалом из пункта 1 и с учетом результатов пункта 2;

4) классификация неразделяющих линейных алгебр Ли с каждой разделяющей оболочкой из пунктов 2 и 3.

Заметим также, что (за исключением пункта 1) остальная часть работы может быть без изменений перенесена на случай подалгебр произвольной редуктивной алгебры Ли.

Остановимся подробнее на классификации линейных алгебр Ли, состоящих из нильпотентных эндоморфизмов, что и является целью настоящей работы.

Пусть $\mathfrak{g}$ – подалгебра Ли $\mathfrak{gl}(V)$, состоящая из нильпотентных элементов. Определим убывающую фильтрацию векторного пространства V следующим образом: $V_1 = V$, $V_{k+1} = \mathfrak{g}(V_k)$ при $k \geq 1$. Из теоремы Энгеля следует, что существует такое натуральное число n, что $V_n \neq \{0\}$, а $V_{n+1} = \{0\}$. Назовем это число n *ступенью нильпотентности подалгебры* $\mathfrak{g} \subset \mathfrak{gl}(V)$.

Пусть $\mathfrak{g}$ – подалгебра алгебры Ли $\mathfrak{gl}(V)$ ступени n. Тогда $\mathfrak{g}(V_n) = \{0\}$. Положим $W = V / V_n$. Пусть $\pi : \mathfrak{g} \to \mathfrak{gl}(W)$ – гомоморфизм алгебр Ли вида: $\pi(x) : v + V_n \mapsto x(v) + V_n$. Положим $\mathfrak{a} = ker\pi$, $\mathfrak{h} = \pi(\mathfrak{g})$. Тогда очевидно, что $\mathfrak{h}$ – подалгебра в $\mathfrak{gl}(W)$, состоящая из нильпотентных элементов и имеющая ступень нильпотентности $n-1$.

Алгебра Ли $\mathfrak{a}$ является коммутативным идеалом в $\mathfrak{g}$ и может быть естественным образом отождествлена с подпространством в $\mathcal{L}(W, V_n) : y(v + V_n) = y(v)$ для всех $y \in \mathfrak{a}, v \in V$.

Определим действие алгебры Ли $\mathfrak{h}$ на $\mathcal{L}(W, V_n)$ следующим образом:

$$x.\alpha = -\alpha \circ x, \text{ где } x \in \mathfrak{h}, \ \alpha \in \mathcal{L}(W, V_n).$$

Тогда $x.y = [\pi^{-1}(x), y]$ для $x \in \mathfrak{h}$, $y \in \mathfrak{a}$. Следовательно, $\mathfrak{a}$ является подмодулем $\mathfrak{h}$-модуля $\mathcal{L}(W, V_n)$. Далее алгебра Ли $\mathfrak{g}$ естественным образом определяет элемент $[\omega] \in H^1(\mathfrak{h}, \mathcal{L}(W, V_n) / \mathfrak{a})$. Действительно, пусть $s : W \to V$ – произвольное сечение канонической проекции $p : V \to W = V / V_n$. Отождествим векторное пространство V с $V_n \times W$ при помощи изоморфизма $I_s : x \mapsto (x - s \circ p(x), p(x))$. Определим отображение $\overline{\omega}_s : \mathfrak{h} \to \mathcal{L}(W, V_n)$ следующим образом:

$$\overline{\omega}_s(x) : w \mapsto s(x(w)) - \pi^{-1}(x).s(w).$$

Тогда, как нетрудно проверить, $\overline{\omega}_s \in Z^1(\mathfrak{h}, \mathcal{L}(W, V_n))$. Посредством факторизации

возникает элемент $\omega_s \in Z^1(\mathfrak{h}, \mathcal{L}(W, V_n)/\mathfrak{a})$. При этом, если s' – другое сечение, то $\omega_s - \omega_{s'} \in B^1(\mathfrak{h}, \mathcal{L}(W, V_n)/\mathfrak{a})$, и поэтому класс $[\omega_s]$ элемента ω_s не зависит от выбора сечения.

Пусть $\mathfrak{g}$ – подалгебра ступени нильпотентности n, $\mathfrak{h} = \pi(\mathfrak{g})$, $\mathfrak{a} = ker\pi$ и $[\omega]$ – соответствующий подалгебре $\mathfrak{g}$ элемент пространства $H^1(\mathfrak{h}, \mathcal{L}(W, V_n)/\mathfrak{a})$. Тогда имеет место равенство

$$(\omega(\mathfrak{h}) + \mathfrak{a}).W_{n-1} = V_n.$$

Действительно, как и ранее, отождествим V с $V_n \times W$ при помощи некоторого сечения $s: W \to V$. Поскольку $\mathfrak{g}(V_n) = 0$ и $\mathfrak{h}(W_{n-1}) = 0$, то

$$V_n = \mathfrak{g}^{n-1}(V) = \mathfrak{g}^{n-1}(\{0\} \times W) =$$

$$= (\omega(\mathfrak{h}) + \mathfrak{a}).W_{n-1} \times \mathfrak{h}^{n-1}(W) = (\omega(\mathfrak{h}) + \mathfrak{a}).W_{n-1} \times \{0\}.$$

Обратно, пусть $\mathfrak{h}$ – подалгебра в $\mathfrak{gl}(W)$ ступени нильпотентности $n-1$, U – некоторый тривиальный $\mathfrak{h}$-модуль, $\mathfrak{a}$ – подмодуль $\mathfrak{h}$-модуля $\mathcal{L}(W, U)$, и элемент $[\omega] \in H^1(\mathfrak{h}, \mathcal{L}(W, U)/\mathfrak{a})$ таков, что

$$(\omega(\mathfrak{h}) + \mathfrak{a}).W_{n-1} = U.$$

Тогда в векторном пространстве $V = U \times W$ может быть определена структура линейной алгебры Ли:

$$\mathfrak{g} = \{(u, w) \mapsto ((y + \omega(x))(w), x(w)) \mid x \in \mathfrak{h}, y \in \mathfrak{a}\},$$

превращающая $\mathfrak{g}$ в подалгебру ступени нильпотентности n, такую, что $\mathfrak{g}(V_{n-1}) = U$ и $\mathfrak{h} = \pi(\mathfrak{g})$.

Таким образом, показано, что *существует взаимно-однозначное соответствие между подалгебрами ступени нильпотентности n и тройками* $(\mathfrak{h}, \mathfrak{a}, [\omega])$, *где* $\mathfrak{h}$ *– подалгебра ступени нильпотентности* $n-1$, $\mathfrak{a}$ *– подмодуль* $\mathfrak{h}$*-модуля* $\mathcal{L}(W, U)$ *и* $[\omega] \in H^1(\mathfrak{h}, \mathcal{L}(W, U)/\mathfrak{a})$, $(\omega(\mathfrak{h}) + \mathfrak{a}).W_{n-1} = U$.

Для подалгебры $\mathfrak{h} \subset \mathfrak{gl}(W)$ и подмодуля $\mathfrak{a} \subset \mathcal{L}(W, U)$ определим следующие подгруппы:

$$A(\mathfrak{h}) = \{\varphi \in GL(W) \mid \varphi \cdot \mathfrak{h} \cdot \varphi^{-1} = \mathfrak{h}\},$$

$$A(\mathfrak{h}, \mathfrak{a}) = \{(\psi, \varphi) \in GL(U) \times GL(W) \mid (\psi, \varphi).\mathfrak{a} =$$

$$= \mathfrak{a}, \varphi \in A(\mathfrak{h}), \psi \in GL(U)\},$$

где $(\psi, \varphi).y = \psi y \varphi^{-1}$ для $y \in \mathfrak{a}$.

Пусть $\mathfrak{g}_1$ и $\mathfrak{g}_2$ – две алгебры Ли, соответствующие тройкам $(\mathfrak{h}, \mathfrak{a}, [\omega_1])$ и $(\mathfrak{h}, \mathfrak{a}, [\omega_2])$. Тогда для того, чтобы алгебра Ли $\mathfrak{g}_1$ была сопряжена алгебре Ли $\mathfrak{g}_2$, необходимо и достаточно, чтобы элементы $[\omega_1]$ и $[\omega_2]$ были сопряжены относительно действия группы $A(\mathfrak{h}, \mathfrak{a})$ в пространстве $H^1(\mathfrak{h}, \mathcal{L}(W, U)/\mathfrak{a})$.

Действительно, пусть подалгебры $\mathfrak{g}_1$ и $\mathfrak{g}_2$ сопряжены при помощи элемента $h \in GL(V)$, т. е. $h \cdot \mathfrak{g}_1 \cdot h^{-1} = \mathfrak{g}_2$. Тогда изоморфизм h может быть записан в виде

$$h: (u, \omega) \mapsto (\psi(u) + \alpha(\omega), \varphi(\omega)),$$

где $\psi \in GL(U)$, $\varphi \in GL(W)$, $\alpha \in \mathcal{L}(W, U)$. Следовательно, $(\psi, \varphi) \in A(\mathfrak{h}, \mathfrak{a})$ и

$$\psi \cdot \omega_1(x) \cdot \varphi^{-1} + \psi \cdot \mathfrak{a} \cdot \varphi^{-1} + \alpha(x) \cdot \varphi^{-1} = \omega_2(\varphi x \varphi^{-1}) + \mathfrak{a}$$

для всех $x \in \mathfrak{h}$.

Поскольку $\alpha(x) = x.(-\alpha) \in B^1(\mathfrak{h}, \mathcal{L}(W, U))$, то элементы $[\omega_1]$ и $[\omega_2]$ сопряжены относительно группы $A(\mathfrak{h}, \mathfrak{a})$. Обратное утверждение очевидно.

Таким образом, получаем алгоритм классификации линейных алгебр ступени нильпотентности n по алгебре $\mathfrak{h} \subset \mathfrak{gl}(W)$ ступени $n-1$ и векторному пространству U.

1. Определяем группу $A(\mathfrak{h})$ и ее действие в пространстве $\mathcal{L}(W, U)$, где U – тривиальный $\mathfrak{h}$-модуль.

2. С точностью до группы $A(\mathfrak{h})$ производим классификацию подмодулей $\mathfrak{h}$-модуля $\mathcal{L}(W, U)$.

3. Для всякого подмодуля $\mathfrak{a}$ из предыдущего пункта классифицируем элементы $[\omega] \in H^1(\mathfrak{h}, \mathcal{L}(W, U)/\mathfrak{a})$, такие, что $(\omega(\mathfrak{h}) + \mathfrak{a}).W_{n-1} = U$.

4. Для каждой тройки $(\mathfrak{h}, \mathfrak{a}, [\omega])$ строим соответствующую подалгебру $\mathfrak{g} \subset \mathfrak{gl}(V)$, где $V = U \times W$.

Тогда всякая подалгебра $\mathfrak{g}$ ступени нильпотентности n, такая, что $\mathfrak{g}(V) = U$ и $\pi(\mathfrak{g}) = \mathfrak{h}$, сопряжена одной и только одной из построенных алгебр.

Получим классификацию подалгебр в $\mathfrak{gl}(4, \mathbb{C})$, состоящих из нильпотентных элементов. Решение этой задачи является первым шагом в алгоритме классификации всех подалгебр алгебры Ли $\mathfrak{gl}(4, \mathbb{C})$, описанном ранее. При классификации подалгебр, состоящих из нильпотентных элементов, мы используем методику, приведенную выше.

Теорема. *Любая подалгебра в* $\mathfrak{gl}(4, \mathbb{C})$, *состоящая из нильпотентных элементов, сопряжена одной и только одной из следующих подалгебр:*

$$\begin{pmatrix} 0 & x & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & x & 0 \\ 0 & 0 & 0 & y \\ 0 & 0 & 0 & x \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\begin{pmatrix} 0 & x & 0 & 0 \\ 0 & 0 & x & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & x & 0 & y \\ 0 & 0 & x & 0 \\ 0 & 0 & 0 & x \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\begin{pmatrix} 0 & 0 & x & 0 \\ 0 & 0 & 0 & x \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & x & y & 0 \\ 0 & 0 & x & y \\ 0 & 0 & 0 & x \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\begin{pmatrix}0&x&0&0\\0&0&x&0\\0&0&0&x\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&x&y&0\\0&0&0&y\\0&0&0&x\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&0&x&0\\0&0&y&z\\0&0&0&x\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&y&z&u\\0&0&x&0\\0&0&0&0\\0&0&0&0\end{pmatrix},$$

$$\begin{pmatrix}0&x&0&0\\0&0&0&0\\0&0&0&y\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&0&0&x\\0&0&z&y\\0&0&0&z\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&x&y&z\\0&0&x&0\\0&0&0&x\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&0&y&z\\0&0&x&u\\0&0&0&0\\0&0&0&0\end{pmatrix},$$

$$\begin{pmatrix}0&0&x&y\\0&0&0&x\\0&0&0&0\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&0&0&x\\0&0&0&y\\0&0&0&z\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&x&0&y\\0&0&x&z\\0&0&0&x\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&0&0&y\\0&0&z&u\\0&0&0&x\\0&0&0&0\end{pmatrix},$$

$$\begin{pmatrix}0&0&0&x\\0&0&0&y\\0&0&0&0\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&x&y&z\\0&0&x&0\\0&0&0&0\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&\alpha x&y&z\\0&0&x&y\\0&0&0&x\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&y&x&z\\0&0&0&u\\0&0&0&x\\0&0&0&0\end{pmatrix},$$

$$\begin{pmatrix}0&0&x&y\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&x&y&z\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&-x&y&z\\0&0&0&y\\0&0&0&x\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&y&z&u\\0&0&0&x\\0&0&0&y\\0&0&0&0\end{pmatrix},$$

$$\begin{pmatrix}0&x&y&0\\0&0&0&0\\0&0&0&y\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&0&x&y\\0&0&0&z\\0&0&0&0\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&(1+\alpha)x&(1-\alpha)y&z\\0&0&0&y\\0&0&0&x\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&-y&z&x\\0&0&u&z\\0&0&0&y\\0&0&0&0\end{pmatrix},$$

$$\begin{pmatrix}0&0&x&y\\0&0&0&0\\0&0&0&x\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&0&x&y\\0&0&y&z\\0&0&0&0\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&-x&x+y&z\\0&0&0&y\\0&0&0&x\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&y-z&u&x\\0&0&y&u\\0&0&0&z\\0&0&0&0\end{pmatrix},$$

$$\begin{pmatrix}0&x&y&z\\0&0&x&y\\0&0&0&0\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&x&0&y\\0&0&x&z\\0&0&0&0\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&y&z&x\\0&0&y&u\\0&0&0&y\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&y&u&v\\0&0&0&x\\0&0&0&z\\0&0&0&0\end{pmatrix},$$

$$\begin{pmatrix}0&x&y&0\\0&0&z&0\\0&0&0&0\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&0&y&z\\0&0&x&u\\0&0&0&x\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&y&z&u\\0&0&x&0\\0&0&0&x\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&y&u&v\\0&0&x&z\\0&0&0&0\\0&0&0&0\end{pmatrix},$$

$$\begin{pmatrix}0&x&y&z\\0&0&z&0\\0&0&0&0\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&y&z&u\\0&0&y&x\\0&0&0&0\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&y&z&u\\0&0&y&x\\0&0&0&z\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&y&v&x\\0&0&u&z\\0&0&0&y\\0&0&0&0\end{pmatrix},$$

$$\begin{pmatrix}0&y&0&z\\0&0&y&u\\0&0&0&x\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&y&u&v\\0&0&y&x\\0&0&0&z\\0&0&0&0\end{pmatrix},$$

$$\begin{pmatrix}0&y&u&x\\0&0&z&y\\0&0&0&z\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&y+u&v&x\\0&0&u&z\\0&0&0&y\\0&0&0&0\end{pmatrix},$$

$$\begin{pmatrix}0&y-u&z+u&x\\0&0&y&z\\0&0&0&u\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&y&v&x\\0&0&u&z\\0&0&0&u\\0&0&0&0\end{pmatrix},$$

$$\begin{pmatrix}0&-y&y+u&x\\0&0&z&u\\0&0&0&y\\0&0&0&0\end{pmatrix},\quad \begin{pmatrix}0&y&v&w\\0&0&x&z\\0&0&0&u\\0&0&0&0\end{pmatrix},$$

$$\begin{pmatrix}0&0&y&u\\0&0&v&x\\0&0&0&z\\0&0&0&0\end{pmatrix},$$ нулевая.

Действительно, приведем пример классификации подалгебр из нильпотентных элементов в $\mathfrak{gl}(4,\mathbb{C})$ по заданной подалгебре $\mathfrak{h}$ в $\mathfrak{gl}(3,\mathbb{C})$. Пусть

$$\mathfrak{h}=\{x\cdot e_1+y\cdot e_2\}=\left\{\begin{pmatrix}0&0&x\\0&0&y\\0&0&0\end{pmatrix}\middle| x,y\in\mathbb{C}\right\} -$$

подалгебра ступени нильпотентности 2, $\{u_1,u_2,u_3,u_4\}$ – стандартный базис в $\mathbb{C}^4$ и $U=\langle u_1\rangle$, $W=\langle u_2,u_3,u_4\rangle$. Тогда в базисе $\{u_2^*\otimes u_1, u_3^*\otimes u_1, u_4^*\otimes u_1\}$ $\mathfrak{h}$-модуль $\mathcal{L}(W,U)$ имеет вид

$$\mathfrak{h}=\left\{\begin{pmatrix}0&0&0\\0&0&0\\x&y&0\end{pmatrix}\middle| x,y\in\mathbb{C}\right\}.$$

Всякий его подмодуль сопряжен одному из следующих: $\{0\}$, $\langle u_4^*\otimes u_1\rangle$, $\langle u_3^*\otimes u_1, u_4^*\otimes u_1\rangle$, $\mathcal{L}(W,U)$.

1. Пусть $\mathfrak{a}=\{0\}$. Векторы

$$v_1:\{e_1\mapsto u_2^*\otimes u_1\},$$
$$v_2:\{e_1\mapsto u_3^*\otimes u_1, e_2\mapsto u_2^*\otimes u_1\},$$
$$v_3:\{e_2\mapsto u_3^*\otimes u_1\},$$
$$v_4:\{e_1\mapsto u_4^*\otimes u_1\},$$
$$v_5:\{e_2\mapsto u_4^*\otimes u_1\}$$

образуют базис пространства $Z^1(\mathfrak{h},\boxplus W,U))$ и при этом $B^1(\mathfrak{h},\mathcal{L}(W,U))=\langle v_4,v_5\rangle$. Отождествим трехмерное пространство $H^1(\mathfrak{h},\mathcal{L}(W,U))$ с множеством симметрических матриц второго порядка. Тогда действие группы $A(\mathfrak{h},\mathfrak{a})$ примет вид

$$(a,A).B=a\cdot{}^tA^{-1}\cdot B\cdot A^{-1},$$

где $a\in\mathbb{C}^*$, $A\in GL(2,\mathbb{C})$. Условие $\omega(\mathfrak{h}).W_2=U$ равносильно $B\neq 0$. С точностью до указанного действия матрица B принимает одну из следующих форм:

$$\left\{\begin{pmatrix}1&0\\0&0\end{pmatrix},\begin{pmatrix}1&0\\0&1\end{pmatrix}\right\}.$$

Соответствующие подалгебры имеют вид

$$\left\{\begin{pmatrix}0&x&0&0\\0&0&0&x\\0&0&0&y\\0&0&0&0\end{pmatrix},\begin{pmatrix}0&x&y&0\\0&0&0&x\\0&0&0&y\\0&0&0&0\end{pmatrix}\right\}, x,y\in\mathbb{C}.$$

2. Пусть $\mathfrak{a}=\langle u_4^*\otimes u_1\rangle$. Тогда $\mathfrak{h}$-модуль $\mathcal{L}(W,U)/\mathfrak{a}$ тривиален и четырехмерное пространство $H^1(\mathfrak{h},\boxplus W,U)/\mathfrak{a})$ можно отождествить с множеством матриц второго порядка. При этом классы когомологий, соответствующие матрицам B и B', сопряжены относительно группы $A(\mathfrak{h},\mathfrak{a})$, если $B'=a\cdot{}^tA^{-1}\cdot B\cdot A^{-1}$ для некоторых $a\in\mathbb{C}^*$, $A\in GL(2,\mathbb{C})$. Классы эквивалентности матриц имеют вид

$$\left\{\begin{pmatrix}0&1\\-1&0\end{pmatrix},\begin{pmatrix}1&0\\0&0\end{pmatrix},\begin{pmatrix}1&1\\-1&0\end{pmatrix},\begin{pmatrix}1&\alpha\\-\alpha&1\end{pmatrix},\alpha\sim-\alpha\right\}.$$

Им соответствуют следующие подалгебры:

$$\begin{pmatrix}0&y&-x&z\\0&0&0&x\\0&0&0&y\\0&0&0&0\end{pmatrix},\begin{pmatrix}0&x&0&z\\0&0&0&x\\0&0&0&y\\0&0&0&0\end{pmatrix},$$

$$\begin{pmatrix}0&x+y&-x&z\\0&0&0&x\\0&0&0&y\\0&0&0&0\end{pmatrix},$$

$$\begin{pmatrix}0&x+\alpha y&-\alpha x+y&z\\0&0&0&x\\0&0&0&y\\0&0&0&0\end{pmatrix},\alpha\sim-\alpha,x,y,z\in\mathbb{C}.$$

3. Пусть $\mathfrak{a}=\langle u_3^*\otimes u_1, u_4^*\otimes u_1\rangle$. Тогда $\mathfrak{h}$-модуль $\mathcal{L}(W,U)/\mathfrak{a}$ является тривиальным. Отож-

дествим пространство $H^1(\mathfrak{h}, \mathcal{L}(W,U)/\mathfrak{a})$ с $\mathbb{C}^2$. При этом действие группы $A(\mathfrak{h}, \mathfrak{a})$ запишется в виде

$$(a, A).v = a \cdot {}^t A^{-1} \cdot v,$$

где $v \in \mathbb{C}^2$, $a \in \mathbb{C}^*$ и A – невырожденная верхнетреугольная матрица второго порядка. При заданном действии вектор v эквивалентен одному из следующих:

$$\left\{\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right\}.$$

Соответствующие подалгебры имеют вид

$$\left\{\begin{pmatrix}0&0&z&u\\0&0&0&x\\0&0&0&y\\0&0&0&0\end{pmatrix}, \begin{pmatrix}0&x&z&u\\0&0&0&x\\0&0&0&y\\0&0&0&0\end{pmatrix}, \begin{pmatrix}0&y&z&u\\0&0&0&x\\0&0&0&y\\0&0&0&0\end{pmatrix}\right\},$$

$$x, y, z, u \in \mathbb{C}.$$

4. Пусть $\mathfrak{a} = \mathcal{L}(W, U)$. Тогда подмодулю $\mathfrak{a}$ соответствует единственная подалгебра:

$$\left\{\left.\begin{pmatrix}0&z&u&w\\0&0&0&x\\0&0&0&y\\0&0&0&0\end{pmatrix}\right| x, y, z, u, v, w \in \mathbb{C}\right\}.$$

Аналогичным образом могут быть рассмотрены все оставшиеся случаи.

Заключение. Описаны с точностью до сопряженности подалгебры алгебры Ли $\mathfrak{gl}(4, \mathbb{C})$, состоящие из нильпотентных эндоморфизмов. Решение этой задачи является первым шагом в классификации всех подалгебр алгебры Ли $\mathfrak{gl}(4, \mathbb{C})$.

Алгоритмы классификации, описанные в работе, могут быть компьютеризованы и использованы для решения аналогичных задач в больших размерностях.

Литература

1. Мерзляков Ю. И. Рациональные группы. М.: Наука, 1987. 464 с.
2. Супруненко Д. А. Группы матриц. М.: Наука, 1979. 351 с.
3. Бурбаки Н. Группы и алгебры Ли. М.: Мир, 1972–1978. Гл. I–VIII.

References

1. Merzlyakov Yu. I. *Ratsional'nyye gruppy* [Rational groups]. Moscow, Nauka Publ., 1987. 464 p.
2. Suprunenko D. A. *Gruppy matrits* [Matrix groups]. Moscow, Nauka Publ., 1979. 351 p.
3. Burbaki N. *Gruppy i algebry Li* [Groups and Lie algebras]. Moscow, Mir Publ., 1972–1978.

Информация об авторе

Можей Наталья Павловна − кандидат физико-математических наук, доцент, доцент кафедры программного обеспечения информационных технологий. Белорусский государственный университет информатики и радиоэлектроники (220013, г. Минск, ул. П. Бровки, 6, Республика Беларусь). E-mail: mozheynatalya@mail.ru

Information about the author

Mozhey Natalya Pavlovna − PhD (Physics and Mathematics), Associate Professor, Assistant Professor, the Department of Software for Information Technologies. Belarusian State University of Informatics and Radioelectronics (6, P. Brovki str., 220013, Minsk, Republic of Belarus). E-mail: mozheynatalya@mail.ru

Поступила после доработки 15.11.2019

ТЕОРЕТИЧЕСКАЯ МЕХАНИКА

УДК 531.19

Р. Н. Ласовский, Д. В. Гапанюк, Я. Г. Грода
Белорусский государственный технологический университет

МОДЕЛИРОВАНИЕ ТРЕХМЕРНОГО ТВЕРДОТЕЛЬНОГО ЭЛЕКТРОЛИТА СО СЛЯБ ГЕОМЕТРИЕЙ

Рассматривается трехмерная модель твердотельного электролита со сляб геометрией, содержащей зерно и межзеренную прослойку, описываемую слоями, которые характеризуются дополнительными межузловыми энергетическими барьерами. Взаимодействие между подвижными ионами принимается состоящим из дальнодействующего электростатического отталкивания и короткодействующего притяжения. Система находится между двумя противоположно заряженными электродами. Выполняется моделирование описанной системы по методу Монте-Карло. При этом кулоновская энергия определяется суммированием по методу Эвальда с добавлением дополнительных «вакуумных областей» с обеих сторон сляба. Исследуется профиль концентрации подвижных ионов при различных условиях. Сопоставление результатов с аналитическими вычислениями показывает их количественное расхождение в пределах 15%.

Ключевые слова: твердотельный электролит, межзеренная граница, электрод, метод Монте-Карло, суммирование Эвальда, электропроводность, профиль концентрации.

R. N. Lasovsky, D. V. Gapanjuk, Ya. G. Groda
Belarusian State Technological University

MODELING OF THREE-DIMENSIONAL SOLID ELECTROLYTE WITH A SLAB GEOMETRY

A three-dimensional model of a solid-state electrolyte with a slab geometry containing grain and intergranular interlayer is considered. The interlayer is described by layers characterized by additional interstitial energy barriers. The interaction between mobile ions consists of long-range electrostatic repulsion and short-range attraction. The system is located between two oppositely charged electrodes. The system is simulated by the Monte Carlo method. The Coulomb energy is determined by Ewald summation with the addition “vacuum regions” on both sides of the slab. The concentration profile of mobile ions was investigated under various conditions. A comparison of the results with analytical calculations shows their quantitative discrepancy of not more than 15%.

Key words: solid-state electrolyte, grain boundary, electrode, Monte Carlo method, Ewald summation, electrical conductivity, concentration profile.

Введение. В настоящее время в электрохимических системах активно используются жидкие растворы ионных солей или полимерные ионообменные мембраны, что сопряжено с опасностью появления утечек и воспламенения. Переход к электрохимическим элементам с твердотельными электролитами может обеспечить повышение прочности, долговечности, экологичности и безопасности источников энергии, расширить диапазон рабочих температур [1–4].

Твердые электролиты являются предметом интенсивной научной деятельности ввиду широких перспектив их промышленного применения. На их основе, например, изготавливаются аккумуляторные батареи [2], топливные элементы [4], суперконденсаторы [5], запоминающие устройства и т. д. Во многих случаях твердые электролиты представляют собой керамику или поликристаллиты, что требует разработки специфических методов их получения и экспериментальных исследований. Теоретические исследования таких материалов также сталкиваются с серьезными осложнениями вследствие необходимости учета дальнодействующего кулоновского взаимодействия, наличия межфазных границ и приэлектродных областей.

В работе [6] рассматривалась трехмерная решеточная система, в которой учитывалось только кулоновское взаимодействие. В данной статье также принимается во внимание близкодействующее ван-дер-ваальсовое притяжение.

Модель твердотельного электролита со сляб геометрией. Компьютерное моделирование реальных трехмерных кулоновских систем обычно выполняется на относительно небольшом числе частиц в пределах $100 < N < 10\,000$ [7, 8]. Размер системы ограничен скоростью компьютерного выполнения программы. Моделируемая система занимает весь предоставляемый ей объем. Поэтому, если в начале расчета задать координаты частиц системы в некоторой конечной области, через определенное время они способны разлететься на большие расстояния. Чтобы моделировать поведение системы при заданной плотности или давлении, необходимо поместить эти частицы в непроницаемый ящик. При этом граничные условия будут нарушать однородность системы, и для исследования макроскопических свойств потребуется значительно увеличить ее размеры, что приведет к существенному увеличению необходимых вычислительных мощностей.

Проблема поверхностных эффектов может быть преодолена путем реализации периодических граничных условий. Ящик реплицируется периодическим образом в пространстве, образуя бесконечную последовательность в трех измерениях.

В работе рассматривается трехмерная решеточная модель керамического ионного проводника размером $L_x \times L_y \times L_z$ (моделируемая система принималось размером 30×10×10 узлов, образующих простую кубическую решетку), которая находится между двумя электродами. При этом, кроме дальнодействующего электростатического отталкивания между ионами, в модели учитывается ван-дер-ваальсовое притяжение ближайших соседей. Ионы выполняют термоактивированные прыжки в случайно выбранные вакантные узлы. Электронейтральность системы обеспечивается наличием неподвижных ионов противоположного заряда. Электроды моделируются перпендикулярными к оси x стенками, по узлам которых случайным образом разбросаны противоположно заряженные частицы со средней концентрацией c_w.

Ввиду наличия электродов возникает проблема моделирования системы со сляб геометрией, когда в двух направлениях, параллельных электродам, можно воспользоваться периодическими граничными условиями, а в третьем – система конечна. При такой геометрии для возможности использования метода Эвальда для суммирования дальнодействующих кулоновских взаимодействий с периодичностью в трехмерном пространстве необходимо добавлять «вакуумные области» с обеих сторон сляба внутри ячейки моделирования и корректировать вклады фиктивных копий системы в направлении, перпендикулярном ограничивающим плоскостям сляба [9, 10]. Эффективные численные алгоритмы возможны, если коррекция производится лишь по дипольному взаимодействию, что обусловливает необходимость использования достаточно больших «вакуумных областей». Проблема заключается в корректном выборе размера «вакуумной области» и пределов суммирования в обратном пространстве. Размер «вакуумной области» при моделировании принимался равным 20 параметрам решетки.

Для воспроизводства эффекта повышенного сопротивления межзеренной границы будем моделировать ее слоем из двух плоскостей, перпендикулярных оси x, с отличающимися межузловыми энергетическими барьерами $U_{\mathrm{bar}i}$ (рис. 1), где индекс i нумерует узлы решетки.

При моделировании системы случайным образом выбирается ион, находящийся в узле j. Направление его возможного перехода определяется случайным образом в один из свободных узлов i. Интенсивности перескоков ионов пропорциональны величине:

$$w_{jl} = \exp\left(-\frac{(U_{\mathrm{bar}i} - U_{\mathrm{bar}j}) + J(z_j - z_i) + \Delta U_{\mathrm{Coul}}}{k_{\mathrm{B}} T} \right), \quad (1)$$

где J – энергия ван-дер-ваальсового притяжения ближайших соседей; z_j, z_i – число ближайших соседей j-го и i-го узла соответственно; k_{B} – постоянная Больцмана; T – температура.

Моделировалась одна траектория. Первые 500 Монте-Карло шагов (МКШ) отводились на установление стационарного состояния, последующее усреднение производилось по 10^4 МКШ.

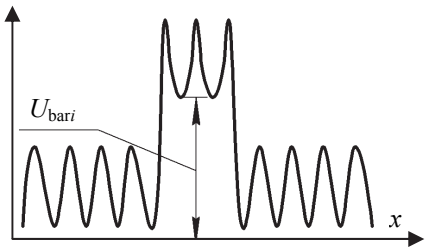


Рис. 1. Модель энергетических барьеров вдоль оси, перпендикулярной плоскостям электродов

Разность кулоновских энергий иона до и после прыжка определяется выражением

$$\Delta U_{\mathrm{Coul}} = U_{\mathrm{Coul}}^{\mathrm{end}} - U_{\mathrm{Coul}}^{\mathrm{start}}. \quad (2)$$

Величину энергии кулоновских взаимодействий для системы со сляб геометрией с использованием суммирования Эвальда можно записать в виде [10]:

$$U_{\text{Coul}} = \frac{1}{4\pi\varepsilon\varepsilon_0}\left[\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(q_i q_j \frac{\operatorname{erfc}\left(\alpha\left|\mathbf{r}_{ij}\right|\right)}{\left|\mathbf{r}_{ij}\right|} + \right.\right.$$

$$+\frac{1}{\pi V}\sum_{\mathbf{k}\neq 0} q_i q_j \frac{4\pi^2}{\alpha^2}\exp\left(-\frac{k^2}{\alpha^2}\right)\cos\left(\mathbf{k}\cdot\mathbf{r}_{ij}\right)\Bigg) +$$

$$\left. + \frac{2\pi}{V}\left|\sum_{i=1}^{N} q_i \mathbf{r}_i\right|^2\right], \qquad (3)$$

где q_i – заряд i-го иона; $\alpha = 5/L$ – параметр Эвальда; $\mathbf{r}_{ij}$ – радиус-вектор частицы j относительно частицы i; $V = L_x L_y L_z$ – объем системы; $\mathbf{k} = 2\pi(\gamma'_x / L_x, \gamma'_y / L_y, \gamma'_z / L_z)$ – вектор обратной решетки; $\gamma'_x, \gamma'_y, \gamma'_z$ – целые числа.

При моделировании обрезание при суммировании в прямом пространстве производилось по половине длины вдоль каждого направления, а также принималось $\gamma'_{\max y} = \gamma'_{\max z} = 4$, $\gamma'_{\max x} = 12$, что соответствует размеру «вакуумных областей», равному двум размерам основной ячейки моделирования в направлении оси x, и учету 334 k-векторов в обратном пространстве. При этом достигается точность около 1%, и дальнейшее увеличение области суммирования мало влияет на точность моделирования.

Результаты моделирования. С целью сопоставления результатов с аналитическими вычислениями [11] моделирование было выполнено для двух средних концентраций подвижных ионов $\rho = 0{,}1$ и 0,03 для системы с диэлектрической проницаемостью $\varepsilon = 41{,}8$ и параметром решетки $a = 0{,}4$ нм при температуре $T = 1000$ К.

На рис. 2 и 3 показан профиль концентрации подвижных ионов при средней концентрации $\rho = 0{,}1$, отсутствии/наличии притяжения ближайших соседей и отсутствии межзеренной границы.

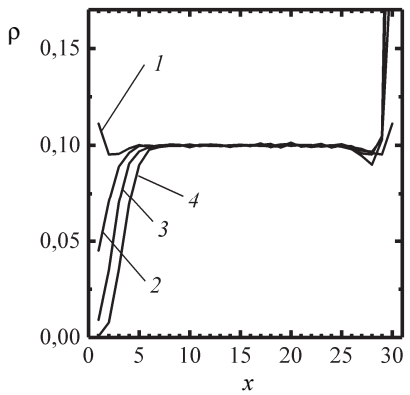


Рис. 2. Профиль концентрации подвижных ионов при $\rho = 0{,}1$, отсутствии притяжения ближайших соседей и отсутствии межзеренной границы: *1* – $c_w = 0$; *2* – $c_w = 0{,}1$; *3* – $c_w = 0{,}2$; *4* – $c_w = 0{,}3$

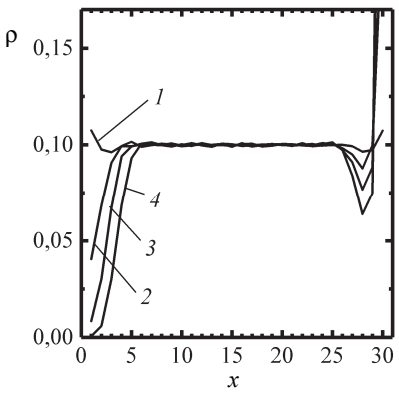


Рис. 3. Профиль концентрации подвижных ионов при $\rho = 0{,}1$, наличии притяжения ближайших соседей и отсутствии межзеренной границы: *1* – $c_w = 0$; *2* – $c_w = 0{,}1$; *3* – $c_w = 0{,}2$; *4* – $c_w = 0{,}3$

На левой положительно заряженной стенке ширина слоя пониженной концентрации растет с увеличением заряда на стенке. В то же время ионы накапливаются на противоположной стене. Слой с пониженной концентрацией появляется из-за кулоновского отталкивания от левой стенки и удовлетворения условию электронейтральности. Кроме того, учет короткодействующего притяжения между ионами приводит к более глубокому провалу концентрации вблизи правой стенки.

Такая же картина наблюдается при $\rho = 0{,}03$ (рис. 4 и 5). Необходимо лишь отметить увеличение ширины слоя пониженной концентрации вследствие малой объемной концентрации подвижных ионов, что требует большего объема приэлектродной области системы для обеспечения условия электронейтральности.

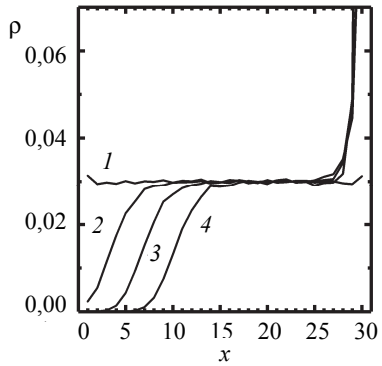


Рис. 4. Профиль концентрации подвижных ионов при $\rho = 0{,}03$, отсутствии притяжения ближайших соседей и отсутствии межзеренной границы: *1* – $c_w = 0$; *2* – $c_w = 0{,}1$; *3* – $c_w = 0{,}2$; *4* – $c_w = 0{,}3$

На рис. 6–9 показан профиль концентрации подвижных ионов при средней концентрации $\rho = 0{,}1$ и 0,03, отсутствии/наличии притяжения ближайших соседей и наличии межзеренной границы.

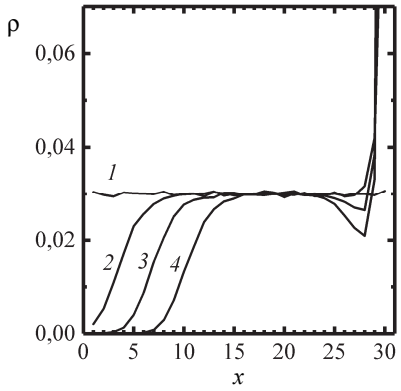


Рис. 5. Профиль концентрации подвижных ионов при ρ = 0,03, наличии притяжения ближайших соседей и отсутствии межзеренной границы: *1* – $c_w = 0$; *2* – $c_w = 0{,}1$; *3* – $c_w = 0{,}2$; *4* – $c_w = 0{,}3$

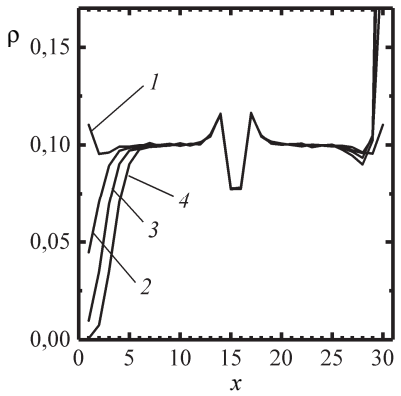


Рис. 6. Профиль концентрации подвижных ионов при ρ = 0,1, отсутствии притяжения ближайших соседей и наличии межзеренной границы: *1* – $c_w = 0$; *2* – $c_w = 0{,}1$; *3* – $c_w = 0{,}2$; *4* – $c_w = 0{,}3$

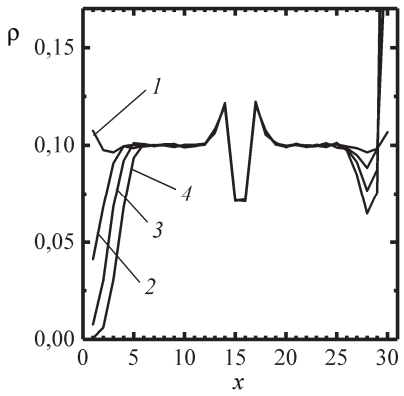


Рис. 7. Профиль концентрации подвижных ионов при ρ = 0,1, наличии притяжения ближайших соседей и наличии межзеренной границы: *1* – $c_w = 0$; *2* – $c_w = 0{,}1$; *3* – $c_w = 0{,}2$; *4* – $c_w = 0{,}3$

Выбранный размер системы оказался достаточным, чтобы поведение распределения заряда вблизи межзеренной границы практически не влияло на явления вблизи стенок и наоборот.

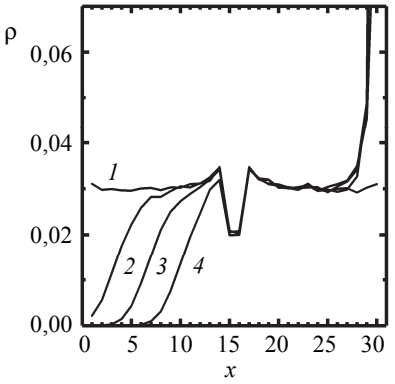


Рис. 8. Профиль концентрации подвижных ионов при ρ = 0,03, отсутствии притяжения ближайших соседей и наличии межзеренной границы: *1* – $c_w = 0$; *2* – $c_w = 0{,}1$; *3* – $c_w = 0{,}2$; *4* – $c_w = 0{,}3$

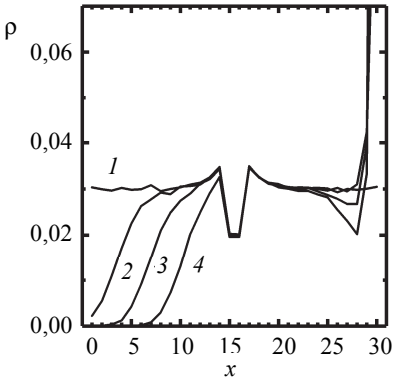


Рис. 9. Профиль концентрации подвижных ионов при ρ = 0,03, наличии притяжения ближайших соседей и наличии межзеренной границы: *1* – $c_w = 0$; *2* – $c_w = 0{,}1$; *3* – $c_w = 0{,}2$; *4* – $c_w = 0{,}3$

Сопоставление результатов с аналитическими вычислениями [11] показывает полное качественное совпадение результатов и количественное расхождение не более 15%, что может быть обусловлено приближениями, принятыми при аналитических вычислениях.

Заключение. Рассмотрена трехмерная модель твердотельного электролита со сляб геометрией, содержащей зерно и межзеренную прослойку, описываемую слоями, которые характеризуются дополнительными межузловыми энергетическими барьерами.

Отмечено увеличение слоя с пониженной концентрацией вблизи положительно заряженной стенки с увеличением заряда на ней. Этот эффект становится более заметным с уменьшением средней концентрации подвижных ионов.

Показано, что наличие межзеренной границы практически не влияет на распределение концентрации вблизи стенок и наоборот.

Публикация содержит результаты исследований, выполненных при грантовой поддержке научной программы Евросоюза HORIZON-2020 (проект AMD-734276-CONIN) и Министерства образования Республики Беларусь.

Литература

1. Solid Oxide Fuel Cells: Materials Properties and Performance / J. Zhang [et al.] // CRC Press. 2016. 298 p.
2. High-Energy All-Solid-State Lithium Batteries with Ultralong Cycle Life / X. Yao [et al.] // Nano Lett. 2016. Vol. 16, no. 11. P. 7148–7154.
3. Ласовский Р. Н., Бокун Г. С., Вихренко В. С. Диаграммное приближение для неравновесных и неоднородных состояний решеточных систем // Труды БГТУ. Сер. VI, Физ.-мат. науки и информатика. 2010. Вып. XVII. С. 59–62.
4. Nanostructurization caused by first order phase transitions in systems with hopping dynamics / G. S. Bokun [et al.] // Solid State Ionics. 2013. Vol. 251. P. 51–54.
5. Unusual properties of a model of an intergrain boundary in solid oxide ceramic electrolytes / G. S. Bokun [et al.] // Solid State Ionics. 2017. Vol. 302. P. 25–29.
6. Ласовский Р. Н., Гапанюк Д. В., Пацаган Т. Н. Моделирование трехмерного керамического электролита с межзеренной границей // Труды БГТУ. Сер. 3, Физ.-мат. науки и информатика. 2019. № 1. С. 15–19.
7. Allen M. P., Tildesley D. J. Computer simulation of liquids. New York: Clarendon Press, 1989. 385 p.
8. Frenkel D., Smit B. Understanding Molecular Simulation. San Diego: Academic Press, 2002. 638 p.
9. Yeh C., Berkowitz M. L. Ewald summation for systems with slab geometry // J. Chem. Phys. 1999. Vol. 111. P. 3155–3162.
10. Santos A., Girotto M., Levin Y. Simulations of Coulomb systems with slab geometry using an efficient 3D Ewald summation method // J. Chem. Phys. 2016. Vol. 144. Art. 144103.
11. The effect of short-range interaction and correlations on the charge and electric field distribution in a model solid electrolyte / T. Patsahan [et al.] // Solid State Ionics. 2019. Vol. 335. P. 156–163.

References

1. Zhang J., Hui R., Wilkinson D., Li X. Solid Oxide Fuel Cells: Materials Properties and Performance. *CRC Press*, 2016. 298 p.
2. Yao X., Liu D., Wang C., Long P., Peng G., Hu Y., Li H., Chen L., Xu X. High-Energy All-Solid-State Lithium Batteries with Ultralong Cycle Life. *Nano Lett.*, 2016, vol. 16, no. 11, pp. 7148–7154.
3. Lasovsky R. N., Bokun G. S., Vikhrenko V. S. Diagram approximation for nonequilibrium and inhomogeneous states of lattice systems. *Trudy BGTU* [Proceedings of BSTU], series VI, Physics and Mathematics. Informatics, 2010, issue XVIII, pp. 59–62 (In Russian).
4. Bokun G. S., Lasovsky R. N., Vikhrenko V. S. Nanostructurization caused by first order phase transitions in systems with hopping dynamics. *Solid State Ionics*, 2013, vol. 251, pp. 51–54.
5. Bokun G. S., Groda Y. G., Lasovsky R. N., Vikhrenko V. S. Unusual properties of a model of an intergrain boundary in solid oxide ceramic electrolytes. *Solid State Ionics*, 2017, vol. 302, pp. 25–29.
6. Lasovsky R. N., Gapanjuk D. V., Patsahan T. N. Modeling of a three-dimensional ceramic electrolyte with an intergranular border. *Trudy BGTU* [Proceedings of BSTU], series 3, Physics and Mathematics. Informatics, 2019, no. 1, pp. 15–19 (In Russian).
7. Allen M. P., Tildesley D. J. Computer simulation of liquids. New York, Clarendon Press, 1989. 385 p.
8. Frenkel D., Smit B. Understanding Molecular Simulation. San Diego, Academic Press, 2002. 638 p.
9. Yeh C., Berkowitz M. L. Ewald summation for systems with slab geometry. *J. Chem. Phys.*, 1999, vol. 111, pp. 3155–3162.
10. Santos A., Girotto M., Levin Y. Simulations of Coulomb systems with slab geometry using an efficient 3D Ewald summation method. *J. Chem. Phys.*, 2016, vol. 144, art. 144103.
11. Patsahan T., Bokun G., di Caprio D., Holovko M., Vikhrenko V. The effect of short-range interaction and correlations on the charge and electric field distribution in a model solid electrolyte. *Solid State Ionics*, 2019, vol. 335, pp. 156–163.

Информация об авторах

Ласовский Руслан Николаевич − кандидат физико-математических наук, доцент кафедры механики и конструирования. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: lasovsky@tut.by

Гапанюк Дмитрий Владимирович − кандидат физико-математических наук, доцент кафедры механики и конструирования. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: gapdm@mail.ru

Гродa Ярослав Геннадьевич – кандидат физико-математических наук, доцент, доцент кафедры механики и конструирования. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: groda@belstu.by

Information about the authors

Lasovsky Ruslan Nikolaevich – PhD (Physics and Mathematics), Assistant Professor, the Department of Mechanics and Engineering. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: lasovsky@tut.by

Gapanjuk Dmitry Vladimirovich − PhD (Physics and Mathematics), Assistant Professor, the Department of Mechanics and Engineering. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: gapdm@mail.ru

Groda Yaroslav Gennad'yevich − PhD (Physics and Mathematics), Associate Professor, Assistant Professor, the Department of Mechanics and Engineering. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: groda@belstu.by

Поступила после доработки 23.11.2019

ФИЗИКА

УДК 535.37+541.65+543.4

Н. Н. Крук[1], Д. В. Кленицкий[1], В. Маес[2]
[1]Белорусский государственный технологический университет
[2]Хассельтский университет (Бельгия)

ИССЛЕДОВАНИЕ СТРУКТУРНЫХ ФАКТОРОВ, ОПРЕДЕЛЯЮЩИХ ОСНОВНОСТЬ АЛКИЛИРОВАННЫХ ПРОИЗВОДНЫХ СВОБОДНОГО ОСНОВАНИЯ КОРРОЛА

С использованием методов квантово-химических расчетов определена равновесная молекулярная конформация длинноволнового NH таутомера семейства алкилированных производных свободного основания коррола. Установлено, что величина двугранного угла φ_D между пирролениновым кольцом, которое протонируется, и средней плоскостью макроцикла коррелирует с длиной связи C_1C_{19} в дипиррольном фрагменте. Длина связи C_1C_{19} зависит от архитектуры периферического замещения, которая определяет силу и локализацию стерических взаимодействий на периферии макроцикла. Если в положениях C_2 и C_{18} пиррольных колец дипиррольного фрагмента нет заместителей, то длина связи C_1C_{19} минимальная, а угол наклона пирролениного кольца φ_D максимальный. Основность таких производных будет наибольшей. Напротив, введение заместителей в дипиррольный фрагмент приводит к увеличению длины связи C_1C_{19} и уменьшению величины угла φ_D. В результате основность макроцикла снижается. На основании установленной взаимосвязи структура – свойство можно предложить молекулярные структуры с заданной основностью макроцикла.

Ключевые слова: коррол, неплоскостные искажения, периферическое замещение, основность.

M. M. Kruk[1], D. V. Klenitsky[1], W. Maes[2]
[1]Belarusian State Technological University
[2]Hasselt University (Belgium)

STUDY OF STRUCTURAL FACTORS DETERMINATIVE FOR BASICITY OF THE ALKYLATED DERIVATIVES OF THE FREE BASE CORROLES

Equilibrium molecular conformation of the long wavelength NH tautomer for the family of alkylated derivatives of the free base corrole has been determined by using quantum-chemical methods. It was found that dihedral angle φ_D between the pyrrolenine ring undergoing protonation and macrocycle mean plane correlates with the C_1C_{19} bond length in the dipyrrole fragment. C_1C_{19} bond length depends on the architecture of peripheral substitution which determines the strength and localization of the steric hindrances on the macrocycle periphery. When there are no substituents in the C_2 and C_{18} positions of the pyrrole rings belonging to the dipyrrole fragment, the C_1C_{19} bond length would be of minimum, and the tilting angle φ_D reaches the maximum value. Basicity of such derivatives is the highest. On the contrary, introduction of substituents into dipyrrole fragment leads to increase in the C_1C_{19} bond length and decrease in tilting angle φ_D. As a result the macrocycle basicity decreases. The molecular structures with given basicity can be proposed based on the established structure – property relationship.

Key words: corrole, nonplanar distortions, peripheral substitution, basicity.

Введение. Тетрапиррольные соединения проявляют как основные, так и кислотные свойства, которые проявляются в присоединении либо диссоциации протона(ов) в соответствующих условиях. Основные свойства исследовались для большого количества производных порфиринов, различающихся по архитектуре периферического замещения [1, 2]. Было установлено, что протонирование сопровождается значительными неплоскостными искажениями тетрапиррольного макроцикла, результатом которых является экспонирование протонов пиррольных колец и неподеленной электронной пары атомов азота пирролениновых колец

в растворитель. Такие структурные изменения приводят к значительному росту основности и кислотности макроцикла вследствие благоприятных условий для межмолекулярных взаимодействий.

В молекулах корролов отсутствует один из атомов углерода в мезоположении макроцикла и два соседних пиррольных фрагмента соединены C_a–C_a связью [3]. Ароматическая стабилизация такого сокращенного макроцикла оказывается возможной благодаря увеличению числа пиррольных колец за счет сокращения числа пирролениновых колец. В результате молекула свободного основания коррола содержит в ядре три протона, а не два, как молекулы свободных оснований порфиринов. Однако сопряженная электронная система макроцикла свободного основания коррола также содержит 26 π-электронов, как и в порфириновом макроцикле. Сокращение размеров ядра макроцикла и наличие дополнительного протона в ядре способствует тому, что, даже при отсутствии стерических взаимодействий периферических заместителей, формируется непланарная конформация тетрапиррольного макроцикла корролов.

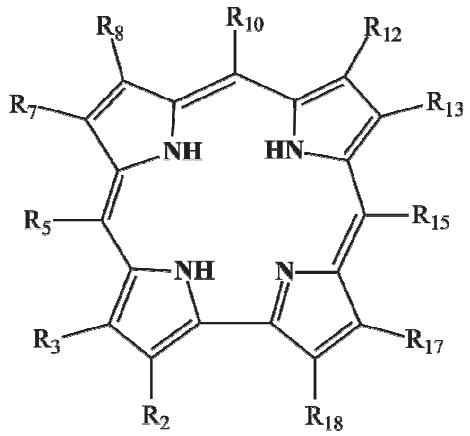
Как было показано, молекулярная конформация и величина неплоскостных искажений макроцикла коррола будет зависеть от архитектуры замещения и типа периферических заместителей [3–5]. Асимметричный характер макроцикла корролов приводит к тому, что углы наклона относительно средней плоскости макроцикла будут различными для всех пиррольных фрагментов. Ранее нами предложено, что данные отличия обусловливают тот факт, что основность всех пиррольных колец будет различной [6]. Методами квантово-химических расчетов установлено, что на величину основности тетрапиррольного макроцикла также влияет характер электростатического потенциала молекулы, однако структурные факторы являются определяющими [7].

Степень неплоскостных искажений макроцикла семейства алкилзамещенных корролов изучалась нами в предыдущих работах, в которых выполнен анализ молекулярной конформации длинноволновых NH таутомеров 18 соединений [2, 3]. Показано, что они могут быть разделены на четыре группы, различающиеся по величине неплоскостных искажений макроцикла, охарактеризованной с использованием параметра Δ23, а критерием для дифференциации является характер локализации стерических взаимодействий, индуцированных присоединением метильных групп. При этом было отмечено, что ряд структурных элементов, характеризующих молекулярную конформацию макроцикла, изменяется параллельно изменениям параметра Δ23, но в то же время конформационные изменения отдельных фрагментов определяются локальным взаимодействием соседних метильных групп. Показано, что длину связи C_1C_{19} в дипиррольном фрагменте можно связать с размерами сопряженной π-системы макроцикла коррола, и она, по-видимому, отражает баланс между непланарными и планарными искажениями молекулярной структуры макроцикла.

Определяющий характер молекулярной конформации, и, в первую очередь, тип и величина неплоскостных искажений макроцикла, в формировании его основности предоставляет возможность оценки основности макроцикла исходя из анализа молекулярной структуры. В настоящей работе исследованы длинноволновые NH таутомеры 18 метилированных производных свободных оснований корролов и выявлены локальные характеристики молекулярной конформации макроцикла, которые определяют его основность. Проведен анализ данных характеристик для исследуемых соединений и определена архитектура периферического замещения макроцикла, необходимая для наибольшей и наименьшей основности макроцикла.

Основная часть. Молекулярная структура исследованных длинноволновых NH таутомеров алкилированных производных свободного основания коррола приведена на рис. 1. Исследуемые соединения различаются архитектурой периферического замещения, которая определяет локализацию стерических взаимодействий на периферии тетрапиррольного макроцикла. Усиление стерических взаимодействий и их распространение на весь макроцикл приводит к существенному росту величины степени неплоскостных искажений макроцикла Δ23 (таблица).

С точки зрения определения основности макроцикла представляет интерес не только величина степени неплоскостных искажений макроцикла Δ23 в целом, но и локальные конформационные характеристики, определяющие положение пирроленинового фрагмента, который протонируется. Атом азота пирроленинового кольца может либо экспонироваться в растворитель, либо экранироваться близлежащими фрагментами макроцикла от межмолекулярных взаимодействий. В первом случае основность молекулы будет возрастать, а во втором уменьшаться. В таблице приведены рассчитанные величины двугранных углов φ_D между плоскостью пирроленинового кольца D и средней плоскостью макроцикла 7C. Анализ полученных значений указывает на отсутствие однозначной корреляции величины угла φ_D и величины Δ23 для всего массива данных.



1 – $R_2 = R_3 = R_5 = R_7 = R_8 = R_{10} = R_{12} = R_{13} = R_{15} = R_{17} = R_{18} = H$;
2 – $R_3 = R_8 = R_{13} = R_{18} = CH_3$, $R_2 = R_5 = R_7 = R_{10} = R_{12} = R_{15} = R_{17} = H$;
3 – $R_2 = R_3 = R_5 = R_7 = R_8 = R_{10} = R_{12} = R_{13} = R_{15} = R_{17} = R_{18} = CH_3$;
4 – $R_2 = R_3 = R_7 = R_8 = R_{12} = R_{13} = R_{17} = R_{18} = CH_3$, $R_5 = R_{10} = R_{15} = H$;
5 – $R_5 = R_{10} = R_{15} = CH_3$, $R_2 = R_3 = R_7 = R_8 = R_{12} = R_{13} = R_{17} = R_{18} = H$;
6 – $R_2 = R_7 = R_{12} = R_{17} = CH_3$, $R_3 = R_5 = R_8 = R_{10} = R_{13} = R_{15} = R_{18} = H$;
7 – $R_2 = R_3 = R_{17} = R_{18} = CH_3$, $R_5 = R_7 = R_8 = R_{10} = R_{12} = R_{13} = R_{15} = H$;
8 – $R_2 = R_3 = R_5 = R_{15} = R_{17} = R_{18} = CH_3$, $R_7 = R_8 = R_{10} = R_{12} = R_{13} = H$;
9 – $R_2 = R_3 = R_5 = R_7 = R_{13} = R_{15} = R_{17} = R_{18} = CH_3$, $R_8 = R_{10} = R_{12} = H$;
10 – $R_2 = R_3 = R_5 = R_7 = R_8 = R_{12} = R_{13} = R_{15} = R_{17} = R_{18} = CH_3$, $R_{10} = H$;
11 – $R_2 = R_3 = R_5 = R_7 = R_{13} = R_{15} = R_{17} = R_{18} = CH_3$, $R_8 = R_{10} = R_{12} = H$;
12 – $R_2 = R_3 = R_7 = R_8 = R_{12} = R_{13} = R_{17} = R_{18} = CH_3$, $R_5 = R_{10} = R_{15} = H$;
13 – $R_5 = R_7 = R_8 = R_{10} = R_{12} = R_{13} = R_{15} = CH_3$, $R_2 = R_3 = R_{17} = R_{18} = H$;
14 – $R_3 = R_5 = R_7 = R_8 = R_{10} = R_{12} = R_{13} = R_{15} = R_{17} = CH_3$, $R_2 = R_{18} = H$;
15 – $R_2 = R_5 = R_7 = R_8 = R_{10} = R_{12} = R_{13} = R_{15} = R_{18} = CH_3$, $R_3 = R_{17} = H$;
16 – $R_2 = R_3 = R_{10} = R_{17} = R_{18} = CH_3$, $R_5 = R_7 = R_8 = R_{12} = R_{13} = R_{15} = H$;
17 – $R_2 = R_3 = R_5 = R_{10} = R_{15} = R_{17} = R_{18} = CH_3$, $R_7 = R_8 = R_{12} = R_{13} = H$;
18 – $R_2 = R_5 = R_{10} = R_{15} = R_{18} = CH_3$, $R_3 = R_7 = R_8 = R_{12} = R_{13} = R_{17} = H$

Рис. 1. Структура и нумерация исследованных соединений. Заместители обозначены в соответствии с нумерацией атомов в макроцикле (согласно номенклатуре IUPAC)

Однако можно заметить, что исследованные соединения можно разделить на две группы (рис. 2). Для первой из них (соединения **1**, **2**, **5**, **6**, **13**, **14**) величина двугранного угла φ_D варьирует от 7 до 15° и обнаруживает линейную зависимость от величины параметра Δ23 с коэффициентом парной корреляции 0,67. Величина угла φ_D для всех остальных соединений лежит в интервале 0,5–3,5° и не зависит от степени неплоскостных искажений макроцикла в целом (коэффициент парной корреляции 0,12).

Анализ архитектуры периферического замещения соединений **1**, **2**, **5**, **6**, **13**, **14** свидетельствует о том, что у них отсутствуют стерические взаимодействия между пиррольными кольцами в дипиррольном фрагменте из-за того, что в положениях C_2 и C_{18} либо нет метильных заместителей вообще (**1**, **5**, **13**, **14**), либо есть только один (**2**, **6**). В данных структурах отсутствует отталкивание пиррольных колец, на что указывает малая длина связи C_1C_{19}, которая составляет 141,9–142,4 пм. Следовательно, размеры ядра тетрапиррольного макроцикла уменьшаются, что приводит к необходимости минимизирования стерических взаимодействий протонов в ядре за счет формирования неплоскостных конформеров с большими углами наклона пиррольных (пирролениновых) колец относительно средней плоскости макроцикла.

Ключевые структурные параметры исследованных соединений

№ п/п	Δ23, пм* [4]	φ_D, угл. град.	Длина связи C_1C_{19}, пм
1	28,1	12,0	142,3
2	28,5	12,3	142,4
3	41,3	2,5	143,1
4	27,2	2,0	143,3
5	27,9	10,4	142,3
6	29,4	15,4	141,9
7	26,7	1,6	143,3
8	31,0	1,5	143,4
9	32,0	0,7	143,3
10	32,3	0,7	143,3
11	32,6	0,7	143,2
12	32,0	0,5	142,9
13	37,7	9,1	142,0
14	37,8	7,2	142,1
15	38,0	1,2	142,8
16	26,8	2,3	143,2
17	30,4	1,7	143,3
18	29,3	3,5	143,0

*Величина среднеквадратичного отклонения атомов от средней плоскости тетрапиррольного макроцикла Δ23 рассчитывалась по формуле

$$\Delta 23 = \sqrt{\frac{1}{23}\sum_{i=1}^{23}\Delta z_i^2},$$

где Δz_i – отклонение i-го атома макроцикла от средней плоскости макроцикла 7С [3].

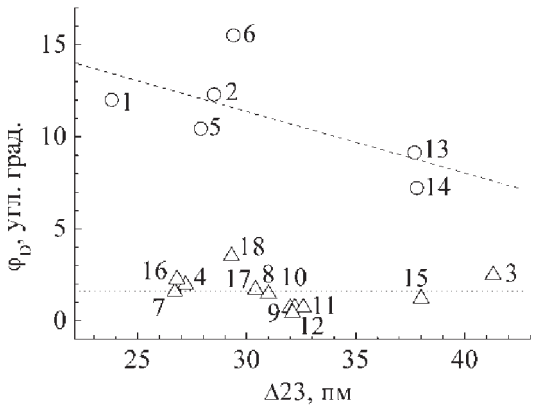


Рис. 2. Зависимость величины двугранного угла φ_D между протонирующимся пирролениновым кольцом и средней плоскостью макроцикла от параметра Δ23 для соединений с длиной связи C_1C_{19}: ○ – до 142,4 пм; △ – более 142,4 пм

Во всех остальных соединениях существует отталкивание метильных групп, присоединенных в положениях C_2 и C_{18}, в результате чего длина связи C_1C_{19} существенно увеличивается (таблица). Вследствие этого увеличиваются размеры макроциклического ядра, что приводит к заметному снижению стерических напряжений в макроцикле при относительно небольших углах наклона пиррольных колец. Равновесная молекулярная конформация формируется главным образом за счет изменения углов и длин связей в плоскости макроцикла, а не за счет формирования конформации с большими отклонениями от планарного строения.

Следует отметить, что в этой группе выделяются соединения **12** и **15**, у которых длина связи C_1C_{19} имеет несколько меньшее значение, чем у остальных, несмотря на наличие метильных заместителей в положениях C_2 и C_{18}. По нашему мнению, это обусловлено возможностью частичного минимизирования стерических взаимодействий не за счет «расталкивания» пиррольных колец в дипиррольном фрагменте, а за счет их поворота навстречу друг другу. Данная конформационная релаксация возможна, так как в этих соединениях не замещены либо два ближайших C_m-положения макроцикла (**12**), либо положения C_2 и C_{18} (**15**). У соединения **18** положения C_2 и C_{18} также не замещены, однако длина связи C_1C_{19} обнаруживает тенденцию к увеличению. Это можно объяснить тем, что при отсутствии метильных заместителей в положениях C_7, C_8, C_{12} и C_{13} конформационная релаксация посредством увеличения длины связи C_1C_{19} в дипиррольном фрагменте требует меньших энергетических затрат, чем поворот пиррольных колец.

Противоположная картина наблюдается у соединений **13** и **14**, для которых стерически напряженный домен включает все три C_m-положения макроцикла, как и у соединения **15**. Однако в этом случае отсутствие стерического взаимодействия заместителей в положениях C_2 и C_{18} приводит к уменьшению длины связи C_1C_{19}. В результате, как было отмечено выше, формируются конформеры с неплоскостными искажениями макроцикла. Но величина двугранного угла φ_D оказывается несколько меньше, чем у соединений **1**, **2**, **5**, **6** (таблица). Таким образом, результаты проведенного анализа подтверждают существование взаимосвязи между длиной связи C_1C_{19} и величиной двугранного угла φ_D. График зависимости величины угла φ_D от длины связи C_1C_{19} представлен на рис. 3.

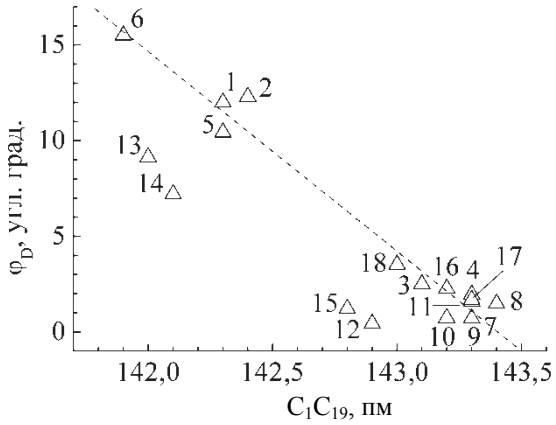


Рис. 3. Зависимость величины двугранного угла φ_D между протонирующимся пирролениновым кольцом и средней плоскостью макроцикла от длины связи C_1C_{19}. Линия показывает результат линейной регрессии без учета соединений **12–15**

Анализ графика действительно указывает на наличие линейной зависимости между двумя величинами. У соединений с меньшей длиной связи C_1C_{19} величина угла φ_D существенно возрастает. Коэффициент парной корреляции, полученный в результате линейной регрессии всего массива данных, равен 0,89. Однако следует отметить, что описанные выше особенности конформационной релаксации в соединениях **12–15** позволяют выделить их в отдельную подгруппу. Тогда как для остальных соединений коэффициент парной корреляции при аппроксимации зависимости линейной функцией оказывается равным 0,98. При этом надо подчеркнуть, что в подгруппе соединений **12–15** угол наклона линейной зависимости оказывается таким же. Это говорит о том, что функциональная взаимосвязь двух величин не изменяется, а описанные выше другие каналы конформационной релаксации в соединениях **12–15** вносят некоторый дополнительный

вклад в постоянный член линейной зависимости. Этот вклад имеет отрицательный знак и приводит к уменьшению величины угла φ_D.

Заключение. Методами квантовой химии рассчитана молекулярная структура макроцикла длинноволновых NH таутомеров семейства метилзамещенных производных свободного основания коррола с различной архитектурой периферического замещения. Установлено, что величина двугранного угла φ_D между протонирующимся пирролениновым кольцом и средней плоскостью макроцикла 7С зависит от длины связи C_1C_{19}, которая, в свою очередь, определяется архитектурой периферического замещения. Установление корреляции позволяет предложить молекулярные структуры, которые будут существенно различаться основностью макроцикла.

Производные свободных оснований корролов с заместителями в квадрантах, не включающих дипиррольный фрагмент, должны иметь высокую основность в результате формирования конформера с большим углом наклона пирроленинового кольца относительно средней плоскости макроцикла. При этом атом азота пирроленинового кольца экспонируется в раствор и доступен для межмолекулярных взаимодействий. Напротив, если заместители размещены в C_b-положениях пиррольных колец дипиррольного фрагмента и соседних с ним C_m-положениях макроцикла, то соединения должны иметь низкую основность из-за того, что неподеленная электронная пара азота пирроленинового кольца лежит практически в плоскости тетрапиррольного макроцикла и экранируется от межмолекулярных взаимодействий в растворе. Замещенные в трех C_m-положениях макроцикла производные обладают большей основностью по сравнению с C_b-замещенными производными. При этом основность C_m-замещенных производных будет существенно зависеть от типа заместителя, поскольку степень стерических взаимодействий с макроциклом (например, из-за наличия или отсутствия объемных групп в ортоположениях арильных заместителей) способна влиять на длину связи C_1C_{19} в дипиррольном фрагменте.

Литература

1. Андрианов В. Г., Малкова О. В., Березин Д. Б. Кислотно-основные свойства порфиринов // Успехи химии порфиринов. В 5 т. Т. 3 / под ред. О. А. Голубчикова. СПб., 2001. С. 107–129.

2. Kruk M. M., Starukhin A. S., Maes W. Influence of macrocycle protonation on the photophysical properties of porphyrins // Macroheterocycles. 2011. Vol. 4, no. 2. P. 69–79.

3. Kruk M. M., Klenitsky D. V., Maes W. Molecular structure and conformation of free base corroles // Macroheterocycles. 2019. Vol. 12, no. 1. P. 58–67.

4. Крук Н. Н., Кленицкий Д. В., Маес В. Квантово-химическое исследование молекулярной структуры алкилированных корролов // Труды БГТУ. Сер. 3, Физ.-мат. науки и информатика. 2018. № 1. С. 36–42.

5. Влияние молекулярной структуры на энергию нижних возбужденных электронных синглетных и триплетных состояний свободных оснований корролов / Н. Н. Крук [и др.] // Труды БГТУ. Сер. 3, Физ.-мат. науки и информатика. 2019. № 1. С. 20–26.

6. Corrole NH Tautomers: Spectral Features and Individual Protonation / Yu. B. Ivanova [et al.] // Journal of Physical Chemistry, A. 2012. Vol. 116, no. 44. P. 10683–10694.

7. Quantum chemical insights into the dependence of porphyrin basicity on the *meso*-aryl substituents: thermodynamics, buckling, reaction sites and molecular flexibility / M. Presselt [et al.] // Phys. Chem. Chem. Phys. 2015. Vol. 17, no. 21. P. 14096–14106.

References

1. Andrianov V. G., Malkova O. V., Berezin D. B. Acid-base properties of porphyrins. *Uspekhi khimii porfirinov* [Advances in Porphyrin Chemistry]. Ed. by O. A. Golubchikov, St. Petersburg, 2001, vol. 3, pp. 107–129 (In Russian).

2. Kruk M. M., Starukhin A. S., Maes W. Influence of macrocycle protonation on the photophysical properties of porphyrins. *Macroheterocycles*, 2011, vol. 4, no. 2, pp. 69–79.

3. Kruk M. M., Klenitsky D. V., Maes W. Molecular structure and conformation of free base corroles. *Macroheterocycles*, 2019, vol. 12, no. 1, pp. 58–67.

4. Kruk M. M., Klenitsky D. V., Maes W. Quantum-chemical study of the molecular structure of alkylated corroles. *Trudy BGTU* [Proceedings of BSTU], series 3, Physics and Mathematics. Informatics, 2018, no. 1, pp. 36–42 (In Russian).

5. Kruk M. M., Klenitsky D. V., Gladkov L. L., Maes W. Influence of the molecular structure on the energy of lowest excited electronic singlet and triplet states of the free base corroles. *Trudy BGTU* [Proceedings of BSTU], series 3, Physics and Mathematics. Informatics, 2019, no. 1, pp. 20–26 (In Russian).

6. Ivanova Yu. B., Savva V. A., Mamardashvili N. Zh., Starukhin A. S., Ngo T. H., Dehaen W., Maes W., Kruk M. M. Corrole NH Tautomers: Spectral Features and Individual Protonation. *Journal of Physical Chemistry*, *A*, 2012, vol. 116, no. 44, pp. 10683–10694.

7. Presselt M., Dehaen W., Maes W., Klamt A., Martinez T. J., Beenken W. J. D., Kruk M. M. Quantum chemical insights into the dependence of porphyrin basicity on the *meso*-aryl substituents: thermodynamics, buckling, reaction sites and molecular flexibility. *Phys. Chem. Chem. Phys*., 2015, vol. 17, no. 21, pp. 14096–14106.

Информация об авторах

Крук Николай Николаевич − доктор физико-математических наук, заведующий кафедрой физики. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: m.kruk@belstu.by

Кленицкий Дмитрий Викентьевич − кандидат физико-математических наук, доцент, доцент кафедры физики. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: klen@belstu.by

Маес Воутер − кандидат химических наук, профессор. Хассельтский университет (г. Дипенбек, B-3590, Бельгия). E-mail: wouter.maes@uhasselt.be

Information about the authors

Kruk Mikalai Mikalaevich − DSc (Physics and Mathematics), Head of the Department of Physics. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: m.kruk@belstu.by

Klenitsky Dmitry Vikentievich − PhD (Physics and Mathematics), Associate Professor, Assistant Professor, the Department of Physics. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: klen@belstu.by

Maes Wouter − PhD (Chemistry), Professor. Hasselt University (B-3590, Diepenbeek, Belgium). E-mail: wouter.maes@uhasselt.be

Поступила после доработки 28.11.2019

УДК 535.543.1

Л. В. Танин, А. И. Горчарук, П. В. Моисеенко, В. А. Танин
ЗАО «Голографическая индустрия»

СОЗДАНИЕ НОВОГО ПОКОЛЕНИЯ КОМБИНИРОВАННЫХ ГОЛОГРАФИЧЕСКИХ ЗАЩИТНЫХ ЭЛЕМЕНТОВ НА ОСНОВЕ ПОЛИМЕРИЗОВАННЫХ ЖИДКИХ КРИСТАЛЛОВ

Основой современных методов защиты продукции, товаров, документов и ценных бумаг являются голографические защитные технологии, развитие которых свидетельствует о том, что они совершенствуются, появляются новые методы записи голограмм, новое оборудование и материалы, повышается их технологичность, упрощается идентификация и проверка подлинности. В современной голографии комбинированные изображения применяются очень широко. При создании комбинированных изображений возникают различные оптические эффекты, такие как муар, параллакс, изменение цвета и др., которые в комбинации между собой, а также с другими изображениями (микротекст, скрытые изображения, последовательная нумерация, маркировка, кодирование, химические индикаторы) позволяют использовать их как для защиты документов, так и для получения оригинального художественного эффекта.

В статье рассматриваются комбинированные защитные элементы на основе рельефно-фазовой голограммы с нанесенным полимерным слоем-носителем, содержащим скрытое изображение, видимое в поляризованном свете. Данный защитный элемент получил название кристаллограмма. В процессе разработки кристаллограммы был освоен синтез мономеров и приготовление анизотропной поляризуемой композиции, получен слой полимеризуемых жидких кристаллов (ПЖК) с контрастной визуализацией скрытого изображения. Отработана технология совмещения рельефно-фазовой голограммы с нанесенным полимерным слоем-носителем с последующим блокированием слоя ПЖК защитными лаковыми слоями.

Ключевые слова: голографические защитные технологии, методы записи голограмм, комбинированные изображения, кристаллограмма, полимеризуемые жидкие кристаллы.

L. V. Tanin, A. I. Harcharuk, P. V. Moiseenko, V. A. Tanin
CJSC “Holography industry”

CREATION OF A NEW GENERATION OF COMBINED HOLOGRAPHIC SECURITY ELEMENTS BASED ON POLYMERIZED LIQUID CRYSTALS

The basis of modern methods of protecting products, goods, documents and securities are holographic protective technologies, the development of which indicates that they are being improved, new recording holograms methods, new equipment and materials are appearing, their manufacturability is being improved, identification and authentication are being simplified. In modern holography, combined images are used very widely. When creating combined images, various optical effects arise, such as moire, parallax, color change, etc., which are combined with each other, as well as with other images (microtext, hidden images, sequential numbering, marking, coding, chemical indicators), they can be used both to protect documents and to obtain an original artistic effect.

The article discusses combined protective elements based on a relief-phase hologram with a deposited polymer carrier layer containing a latent image visible in polarized light. This protective element is called crystallogram. In the process of developing a crystallogram, was the synthesis of monomers and the preparation of an anisotropic polarizable composition was mastered, a layer of polymerizable liquid crystals (PLC) was obtained with contrast visualization of a latent image. The technology for combining the relief-phase hologram with the applied polymer carrier layer with the subsequent blocking of the polarizable liquid crystals (PLC) layer with protective varnish layers has been developed.

Key words: holographic protective technologies, recording holograms methods, combined images, crystallogram, polarizable liquid crystals.

Введение. В области защитных технологий число методов получения надежных средств защиты с каждым годом растет. Однако одновременно с этим совершенствуются и способы подделки, что требует создания более совершенных элементов защиты. Следует также отметить, что не существует оптимального защитного средства, которого было бы невозможно воспроизвести. Лишь в комбинации с различными методами это средство может служить одним из критериев оригинальности защищаемого объекта.

Среди защитных элементов с переменными оптическими свойствами, наиболее трудно воспроизводимыми, выделяются элементы, содержащие скрытые изображения, видимые только в поляризованном свете.

Основная часть. Как правило, скрытые поляризованные изображения получаются на поверхности или в объеме полимерного слоя-носителя скрытого изображения в результате формирования оптической анизотропии в локальной области данного полимерного слоя, а именно изменения величины двулучепреломления [1].

Локальное изменение величины двулучепреломления может быть достигнуто либо вариацией направления оптической анизотропии прозрачного материала в определенных его участках, либо модулированием толщины прозрачного анизотропного материала в определенных участках слоя.

Эти эффекты получаются путем механического, химического, фотофизического или термомеханического воздействий. Во всех этих случаях скрытое поляризованное изображение формируется в предварительно подготовленном специальным образом полимерном слое либо на поверхности твердой подложки [2].

Известные методы изготовления скрытого изображения и защитного элемента на его основе предполагают наличие сплошного, как правило, полимерного слоя, несущего скрытое поляризованное изображение, на обратную сторону которого наносится клей, после чего производится вырубка и получается конечный продукт – самоклеющиеся этикетки и метки.

Среди применяемых на сегодняшний день методов изготовления комбинированных защитных элементов с переменными оптическими свойствами, эквивалентными представленной в статье технологии изготовления оптических защитных элементов, выделяются следующие методы.

Известно применение элементов, содержащих скрытое изображение, изготовленное с помощью локального механического удаления твердого анизотропного полимерного материала с поверхности жесткой подложки. Однако данная технология приводит к формированию контрастного поляризованного изображения с ясно видимым невооруженным глазом контуром изображения.

Другой метод заключается в том, что на подложку наносят отражающий слой. Впечатывают скрытое изображение. Осуществляют структурирование участка слоя, несущего скрытое изображение на предварительно заданную глубину с обеспечением оптической анизотропии в данном структурированном участке слоя, которая обусловливает невидимость изображения при наблюдении его невооруженным глазом и его четкую контрастную видимость при просмотре в поляризованном свете. После этого по всей поверхности защитного элемента наносят тонкий прозрачный защитный слой. При этом коэффициент преломления защитного слоя, по существу, совпадает с коэффициентом преломления участка слоя, несущего скрытое изображение. Однако такой способ имеет существенный недостаток, а именно: отражающий слой, согласно данному изобретению, является непрозрачным для электромагнитных лучей видимого диапазона. В этом случае при нанесении данного защитного элемента на документ от визуального просмотра скрыта та часть документа, где находится защитный элемент, что снижает защитные функции элемента, так как на этом месте нельзя расположить и, соответственно, защитить оригинальную подпись, например, владельца документа или разместить символы, выполненные люминесцентными красками, которые визуализируются при наличии, например, ультрафиолетового излучения.

Широкое применение на сегодняшний день получило средство защиты под названием юниграмма. Данный защитный элемент имеет в качестве основы голограмму с присущими ей защитными элементами и специальный слой, интегрированный с голограммой и содержащий латентное изображение, визуализируемое с помощью поляроида [3].

В юниграмме для визуализации скрытой латентной информации используют видоизменный полярископ (рис. 1).

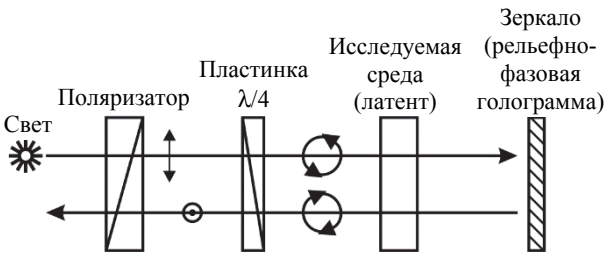


Рис. 1. Оптическая схема идентификации скрытого латентного изображения юниграммы

Полярископ состоит из поляризатора, после прохождения которого свет становится линейно поляризованным, и фазовой пластинки λ/4, которая преобразует линейно поляризованный свет в свет, поляризованный по кругу (циркулярная поляризация с правым или левым направлением вращения светового вектора, в данном случае это непринципиально). За фазовой пластинкой устанавливают зеркало, которое отражает свет обратно, направляя его через

фазовую пластинку с другой стороны. При отражении от зеркала меняется направление вращения светового вектора на противоположное (с правого на левое или, наоборот, с левого на правое). В результате после повторного прохождения через пластинку λ/4 получается линейно поляризованный свет. Вследствие чего перед поляризатором свет приобретает ортогональное направление колебаний светового вектора относительно исходного. Это направление колебаний поляризатор не пропускает, и мы наблюдаем темное поле. Если вместо зеркала поместить голографическую фольгу (которая играет роль зеркала) с нанесенной на ее поверхность полимерной многослойной пленкой со скрытым поляризационным изображением между фазовой пластиной, то в местах, где среда изотропна, поле останется темным. В местах, где среда анизотропна, свет пройдет через поляризатор, поле просветлится и можно увидеть локализацию этих участков с анизотропией [4].

Визуализация скрытого изображения юниграммы производится простым в обращении и недорогим прибором, представляющим собой поляризатор, совмещенный с фазовой пластинкой λ/4 [5]. Следует отметить, что визуализируемая информация при идентификации скрытых изображений не зависит от угла поворота идентификатора относительно юниграммы. Также длительный срок нахождения в качестве средства защиты и широкое распространение знаний технологии изготовления юниграммы ставят задачи перед производителями защитной продукции в необходимости замены анизотропной среды.

Кристаллограмма является следующим поколением многокомпонентных комбинированных защитных оптических элементов на базе рельефно-фазовой голограммы и полимерных слоев со скрытым поляризационным изображением [6]. Уникальность данного защитного средства обусловлена тем, что в процессе визуализации закодированной информации при вращении идентификатора скрытых изображений наблюдается смена одних элементов на другие, а также меняется цветовая гамма элементов скрытого поляризационного изображения. Формирование скрытого поляризационного изображения осуществляется в слое жидкокристаллического материала, который наносится на защитную голограмму.

Рассмотрим формирование цветных скрытых изображений, которые имеют место в кристаллограмме. Эти области можно визуализировать при помощи поляризатора и анализатора. Классическая схема полярископа показана на рис. 2.

Свет, проходя через поляризатор, становится линейно поляризованным. Анализатор, пропускающий свет только с одним направлением колебаний светового вектора, устанавливают так, чтобы свет не выходил, т. е. скрещивают оси поляризатора и анализатора.

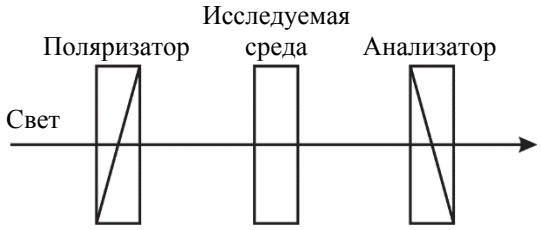


Рис. 2. Классическая схема полярископа

В рабочее пространство между поляризатором и анализатором вводят слой полимеризованных жидких кристаллов (ПЖК). Параллельный пучок естественного света, направленный на поляризатор, превращаясь в линейно поляризованный, падает на ПЖК перпендикулярно его поверхности.

При нормальном падении пучка лучей на одноосный кристалл, оптическая ось в котором параллельна преломляющей поверхности, возникают два луча **е** и **о**. Эти лучи (обыкновенный и необыкновенный) будут распространяться в одном направлении, но с разными скоростями.

Лучи (обыкновенный и необыкновенный), созданные линейно поляризованным светом, являются когерентными, а пройдя аннализатор будут иметь колебания векторов E_o и E_e в одной плоскости [2].

Различие в скоростях обыкновенного и необыкновенного лучей внутри ПЖК приводит к возникновению некоторой разности фаз, а следовательно, к оптической разности хода между двумя когерентными лучами. Таким образом, вышедшие из анализатора два луча удовлетворяют всем условиям, необходимым для осуществления интерференции.

Когда оптическая разность хода обыкновенного и необыкновенного лучей равна целому числу волн, то выходящий из анализатора свет будет максимальной интенсивности. Минимумы интенсивности будут наблюдаться, если оптическая разность хода двух лучей будет равна нечетному числу полуволн. Происходит гашение, но не полное. Следовательно, и исходящий из анализатора свет будет меньшей интенсивности.

Если на пластинку направлять не монохроматический, а белый свет, то благодаря частичному гашению некоторых участков спектра прошедший свет уже будет не белым, а окрашенным.

На рис. 3 представлен общий вид структуры кристаллограммы в поперечном сечении.

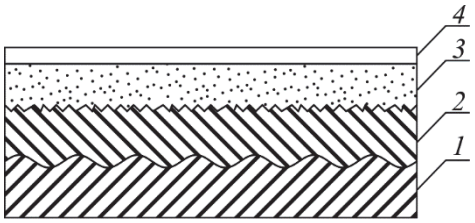


Рис. 3. Структура кристаллограммы:
1 – подложка; *2* – полимерная композиция;
3 – композиция, содержащая мономеры жидких кристаллов; *4* – защитный слой

На подложку *1* с отражающим слоем, которая может быть представлена в виде голографической фольги с нанесенным оригинальным рисунком, или позолотной фольги, или металлизированной бумаги наносят полимерную композицию *2*, затем производят ее сушку и получают на подложке полимерный слой. Далее осуществляют структурирование методом тиснения всей поверхности полимерного слоя дифракционными рельефными структурами в виде символов (изображений) и пробельных мест, направление ориентации которых постоянно изменяется. На структурированную поверхность наносят композицию, содержащую мономеры жидких кристаллов *3*. Под действием теплового воздействия происходит ориентирование и выстраивание жидкокристалических мономеров вдоль направления штрихов дифракционных структур, при этом меняются оптические свойства слоя *3* – возникает двулучепреломление.

Производят УФ-полимеризацию нанесенной композиции, что стабилизирует свойства слоя *3*, и наносят защитный прозрачный полимерный слой *4* на всю поверхность слоя, содержащего символы (изображения) и пробельные места из заполимеризованных жидких кристаллов. Если участок, созданный посредством такой дифракционной структуры, рассматривают сквозь линейный поляризатор и вращают плоскость пропускания такого поляризатора, то визуализируются защитные признаки в виде символов, цифр или какой-либо другой закодированной информации, которые генерируются на основании изменения направления поляризации зон с скрытыми изображениями.

Заключение. Предложенный в работе способ получения комбинированных защитных оптических элементов на базе рельефно-фазовой голограммы под названием кристаллограмма существенно увеличивает защищенность документа, усложняет его подделку, изменение и несанкционированный доступ к защищаемой информации. Освоено серийное производство данного изделия. Особую сложность составляли такие этапы производства, как: синтез мономеров и приготовление анизотропной поляризуемой композиции, получение слоя ПЖК с контрастной визуализацией скрытого изображения, блокирование слоя ПЖК защитными лаковыми слоями, организация эффективного контроля на всех этапах изготовления и использования ПЖК.

Литература

1. Калитеевский Н. И. Волновая оптика. М.: Высшая школа, 1995. 463 с.

2. Кольер Р., Беркхарт К., Лин Л. Оптическая голография. М.: Мир, 1973. 698 с.

3. Голографические методы записи и средства идентификации комбинированных объемных и плоских изображений / Л. В. Танин [и др.] // Голография в России и за рубежом. Наука и практика: тез. докл. III Междунар. науч.-практ. конф., Москва, 26–28 окт. 2006 г. М., 2006. С. 19–20.

4. Идентификационная метка: пат. 5765 Респ. Беларусь. № 20080089; заявл. 11.02.2008; опубл. 30.12.2009. Бюл. № 6. С. 224.

5. Устройство для идентификации голографических марок: пат. 307 Респ. Беларусь. № и20000162; заявл. 20.10.2000; опубл. 02.04.2001. Бюл. № 3. С. 147.

6. Защитная метка: пат. 9286 Респ. Беларусь. № и20120980; заявл. 11.09.2012; опубл. 30.06.2013. Бюл. № 3. С. 188.

References

1. Kaliteevskiy N. I. *Volnovaya optika* [Wave optics]. Moscow, Vysshaya shkola Publ., 1995. 463 p.

2. Kol’yer R., Berkkhart K., Lin L. *Opticheskaya golografiya* [Optical holography]. Moscow, Mir Publ., 1973. 698 p.

3. Tanin L. V., Korolenko A. A., Moiseenko P. V., Vitkevich L. I. Holographic recording methods and means of identification of combined volume and flat images. *Tezisy dokladov III Mezhdunarodnoy nauchno-prakticheskoy konferentsii “Golografiya v Rossii i za rubezhom. Nauka i praktika”* [Abstract for the III International scientific and practical conference “Holography in Russia and abroad. Science and practice”]. Moscow, 2006, pp. 19–20 (In Russian).

4. Tanin L. V., Moiseenko P. V., Manikalo V. V., Boborenko A. G., Lushchikov M. N., Gorcharuk A. I., Korolenko A. A., Tolstik A. L., Vasilenok G. D., Kazak N. S., Kabanov V. V., Belyy V. N.,

Smirnov A. G., Kislukhin S. V., Korochkin L. S., Gorelenko A. Ya., Nikolaychik O. K., Makarevich N. E., Shevtsov V. A. *Identifikatsionnaya metka* [Identification label]. Patent BY, no. 5765, 2009.

5. Tanin L. V., Rubanov A. S., Erokhovets V. K., Moiseenko P. V., Ryzhechkin S. A., Manikalo V. V., Burskiy V. A. *Ustroystvo dlya identifikatsii golograficheskikh marok* [Device for identification of holographic marks]. Patent BY, no. 307, 2001.

6. Tanin L. V., Boboreko A. G., Moiseenko P. V., Kabanov V. V., Altshuler V. D., Shangin S. V., Kislukhin S. V., Gorelenko A. Ya., Shevtsov V. A., Nikolaychik O. K., Makarevich N. E., Burskiy V. A., Ginak S. N., Rak A. V. *Zashchitnaya metka* [Security label]. Patent BY, no. 9286, 2013.

Информация об авторах

Танин Леонид Викторович – доктор физико-математических наук, академик Международной инженерной академии, председатель Совета директоров – главный советник. ЗАО «Голографическая индустрия» (220012, г. Минск, пер. Калинина, 12, Республика Беларусь). E-mail: leonidtanin@gmail.com

Горчарук Андрей Иванович – начальник отдела матриц. ЗАО «Голографическая индустрия» (220012, г. Минск, пер. Калинина, 12, Республика Беларусь). E-mail: gaiholin@gmail.com

Моисеенко Петр Васильевич – кандидат технических наук, заместитель директора по науке и инновационной деятельности. ЗАО «Голографическая индустрия» (220012, г. Минск, пер. Калинина, 12, Республика Беларусь). E-mail: moi@holography.by

Танин Вячеслав Андреевич – аспирант, заместитель директора по коммерческим вопросам. ЗАО «Голографическая индустрия» (220012, г. Минск, пер. Калинина, 12, Республика Беларусь). E-mail: tanin@holography.by

Information about the authors

Tanin Leonid Viktorovich – DSc (Physics and Mathematics), Academician of the International Academy of Engineering, Chairman of the Board of Directors – Chief Adviser. CJSC “Holography industry” (12, Kalinina Lane, 220012, Minsk, Republic of Belarus). E-mail: leonidtanin@gmail.com

Harcharuk Andrey Ivanovich – Head of Matrix Department. CJSC “Holography industry” (12, Kalinina Lane, 220012, Minsk, Republic of Belarus). E-mail: gaiholin@gmail.com

Moiseenko Petr Vasil’yevich – PhD (Engineering), Deputy Director for Science and Innovation. CJSC “Holography industry” (12, Kalinina Lane, 220012, Minsk, Republic of Belarus). E-mail: moi@holography.by

Tanin Vyacheslav Andreevich – PhD student, Deputy Director for Commercial Affairs. CJSC “Holography industry” (12, Kalinina Lane, 220012, Minsk, Republic of Belarus). E-mail: tanin@holography.by

Поступила после доработки 08.10.2019

ИНФОРМАТИКА И ТЕХНИЧЕСКИЕ НАУКИ

МОДЕЛИРОВАНИЕ ПРОЦЕССОВ И УПРАВЛЕНИЕ В ТЕХНИЧЕСКИХ СИСТЕМАХ

УДК 621.3.29

М. Блащак[1], П. П. Урбанович[2]
[1]Люблинский католический университет Иоанна Павла II (Польша)
[2]Белорусский государственный технологический университет

АТАКИ НА МНОГОПОЛЬЗОВАТЕЛЬСКИЕ КОМПЬЮТЕРНЫЕ ИГРЫ И НЕКОТОРЫЕ МЕТОДЫ ЗАЩИТЫ ОТ НИХ

Приведен анализ некоторых уязвимостей многопользовательских компьютерных игр и атак на сервер такой игры (на примере игры «Project I.G.I. 2: Covert Strike»). Часто в основе взаимодействия поставщиков и пользователей рассматриваемой услуги лежит модель free-to-play. Каждому участнику игры нужно соперничать с другими виртуальными игроками. Одна из самых распространенных целей атак злоумышленников – кража паролей и других учетных данных, необходимых для доступа к аккаунту игроков. В статье проанализированы механизмы и программные особенности реализации некоторых видов атак, основанных на ошибках в кодах программных платформ («движков») игр. Анализ сетевого трафика между сервером и клиентским приложением показал, что большинство атак можно блокировать, перехватывая и анализируя сетевой трафик. Одной из лучших систем для обеспечения безопасности сервера считается Linux, поскольку он имеет очень высокоэффективный брандмауэр, функционал которого является ключевым аспектом в решении проблемы нейтрализации атак на серверы многопользовательских компьютерных игр. В статье также описано авторское приложение, предназначенное для нейтрализации некоторых атак на сервер анализируемой игры.

Ключевые слова: многопользовательская компьютерная игра, атака, переполнение буфера, безопасность, программная платформа, сервер, клиент.

M. Błaszczak[1], P. P. Urbanovich[2]
[1]The John Paul II Catholic University of Lublin (Poland)
[2]Belarusian State Technological University

ATTACKS ON MULTIPLAYER COMPUTER GAMES AND SOME METHODS OF PROTECTION AGAINST THEM

The analysis of some vulnerabilities of multiplayer computer games and attacks on the server of such a game is given (on example the game “Project I.G.I. 2: Covert Strike”). Often the free-to-play model is the basis of interaction between suppliers and users of the analyzed service. Each participant in the game needs to rival not only with other virtual players. One of the most common targets of attackers is theft of passwords and other credentials necessary to access the player’s accounts. The article analyzes the mechanisms and software features of the implementation of some types of attacks based on errors in the codes of software platforms (“engines”) of the games. An analysis of network traffic between the server and the client application showed that most attacks can be blocked by intercepting and analyzing network traffic. One of the best systems for providing of server security is Linux, because it has a very high-performance firewall, the functionality of which is a key aspect in solving the problem of neutralizing attacks on multiplayer computer game servers. The article also describes an authoring software application intended for neutralization of some attacks on the server of the analyzed game.

Key words: multiplayer computer game, attack, buffer overflow, security, software platform, server, client.

Введение. Современные компьютерные игры (КИ) – огромная индустрия с денежным оборотом, сопоставимым с нефтяным бизнесом. Особой популярностью пользуются мультиплеерные (многопользовательские) игры (МПИ, англ. Mass Multiplayer Online Game, MMOG) [1]. МПИ – сетевая компьютерная игра, в которой большое количество игроков взаимодействуют друг с другом в виртуальном мире. Указанная популярность во многом связана с тем, что в основе взаимодействия пользователей лежит модель free-to-play – игра доступна бесплатно, а прибыль идет от продажи игровых предметов, ускоряющих получение опыта, и различной декоративной экипировки.

В индустрии КИ появились разнообразные способы монетизации. Разработчики МПИ создают виртуальные пространства, функционирующие на основе собственной экономической системы. Деньги этой системы привлекают не только инвесторов, но и злоумышленников. Число вредоносных программ, «ворующих» игровые предметы и «угоняющих» аккаунты пользователей, растет быстрыми темпами. Особенно уязвимы мобильные приложения, так как многие из них требуют ввода игроком данных банковской карты [2]. Любая отрасль, которая оперирует персональными данными, как правило, становится объектом атак со стороны тех, кто хочет получить эти данные. Игровая индустрия – не исключение.

Типичная онлайновая игра разделена на серверную часть и игровой клиент, устанавливаемый на компьютерах или мобильных устройствах игроков (пользователей).

Программная платформа КИ обеспечивает техническую базу, на основе которой реализуются такие функции, как рендеринг графики, имитация физических процессов, искусственный интеллект, управляющий поведением игровых персонажей, сеть, управление памятью и т. д. Учитывая очень высокую сложность этих программных платформ, невозможно ожидать отсутствия в них багов. И они действительно есть всегда. Эти недостатки сказываются на работе самих игр, а не аппаратных платформ, на которых они реализованы и функционируют, – локальных компьютерах, серверах или мобильных устройствах.

В [3] кратко проанализированы некоторые уязвимости сервера игры «Project I.G.I. 2: Covert Strike» [4].

В настоящей статье будут рассмотрены более подробно особенности некоторых атак на сервер МПИ (на примере игры [4]), а также меры по обеспечению его безопасности.

Основная часть. Существует много типов атак с использованием ошибок в программных кодах игр [5]. Одна из таких атак – «Format string attack» [6] – заключается в неправильной передаче параметров в функцию *printf*.

Атака на основе форматирования последовательности знаков. Атакующий обычно использует директиву *%n*, которая записывает количество символов, сохраненных данной функцией под область памяти, указанную в следующем аргументе, как в примере на рис. 1.

```
0000   2f 25 6e 25 6e 25 6e                                  /%n%n%n
```

Рис. 1. Пакет, используемый в атаке, форматирующей строки

Эту атаку очень легко осуществить. Самый простой способ – ввести комбинации символов *%n%n* в игровом чате. Это закрывает приложение на сервере. Не только чат подвержен такой ошибке.

Ввод указанной комбинации символов в любом месте приводит к той же реакции приложения. Приведенная строка кода воспринимается как команда сервера. Эта комбинация символов также может быть отправлена в пакете, содержащем имя игрока. На рис. 2 приведен фрагмент такого пакета.

```
0080   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0090   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
00a0   00 00 00 00 00 00 00 00 00 00 00 00 25 6e 25 6e   ............%n%n
00b0   25 6e 25 6e 25 6e 00 6c 6d 6e 6f 70 71 72 73 74   %n%n%n.lmnopqrst
00c0   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
00d0   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
```

Рис. 2. Атакующая строка в пакете с именем игрока

Другое место, где можно использовать атаку, – это чат игры. В качестве содержимого сообщения отправляется указанная строка. На рис. 3 показан фрагмент такого пакета.

```
0000   de ad be ef 0c 00 00 00 00 00 00 47 5a 01 18   ...........GZ..
0010   03 ab ab ab 01 00 00 00 ff ff ff ff 54 41 48 43   ............TAHC
0020   4c a4 00 00 4a 6f 6e 65 73 3a 20 25 6e 25 6e 0a   L...Jones: %n%n.
0030   00 5f 00 00 00 00 80 3f 00 00 00 00 00 00 00 00   ._.....?........
0040   00 00 00 00 00 00 80 3f 00 00 00 00 00 00 00 00   .......?........
```

Рис. 3. Фрагмент пакета с сообщением, содержащим атакующую строку

Для этого типа атаки была создана защита в виде патчей. Однако такая защита приводит к возможности реализации иных видов атак, не менее опасных. Например, проблема смены так называемой игровой карты. Если карта на сервере поменялась, игроки могут встретить трудности с идентификацией.

Атака с переполнением буфера имени игрока. В области информационной безопасности

корпоративных ресурсов очень остро стоит проблема атак на сеть путем переполнения буфера [6]. Основная особенность такой атаки состоит в следующем: если атакующий сможет «подсунуть» компьютеру некоторые инструкции в виде кода, компьютер выполнит эти инструкции. Это является основой для нападения, связанного с переполнением буфера. С формальной стороны переполнение буфера возникает, когда компьютерная программа записывает данные («подсунутые» инструкции) за пределами пространства, выделенного в памяти буфера.

Это приводит к перезаписи других данных в памяти, которые могут потребоваться для правильного функционирования приложения. Одним из способов использования этой атаки является переполнение буфера для имени игрока. Приложение имеет ограничение имени в 19 символов. Можно войти в игру, введя соответствующие параметры в командной строке. Если вводится строка символов длиной больше 64 в параметре *name*, игрок войдет в систему с именем лишь из 64 символов. Чтобы воспользоваться ошибкой переполнения буфера, нужно подготовить сетевые пакеты, отвечающие за присоединение к игре. Для этого следует отправить пакет с именем, например, длиной 66 символов (рис. 4).

Рис. 4. Фрагмент пакета переполнения буфера имени игрока

Довольно хорошо с подобной проблемой можно справиться, используя межсетевые экраны для анализа сетевого трафика [7].

Атака с переполнением буфера ключа. Это атака, аналогичная предыдущей, но здесь используется ключ к игре. При вступлении в игру каждый игрок отправляет свой ключ (CD-KEY) в пакете, который верифицируется. Пример пакета показан на рис. 5.

Содержимое пакета включает в себя зашифрованный ключ игрока. Это содержимое можно разделить на две части. Первая, отмеченная подчеркиванием, никогда не меняется. Ее размер составляет 32 байта. Остальная часть изменяется в каждом отправленном пакете.

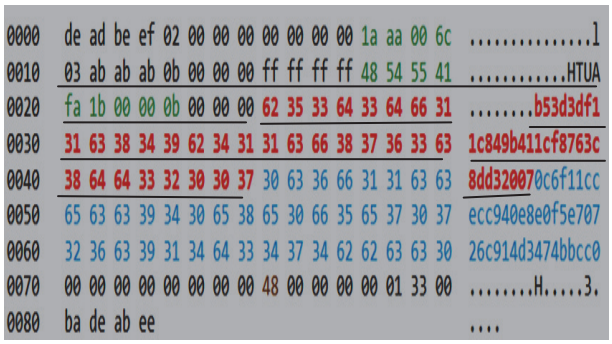


Рис. 5. Пример пакета авторизации с уникальным CD-ключом

Если отсылается пакет с более длинным ключом на сервер, серверное приложение перестанет работать. Защита сервера от переполнения буфера имени игрока также является защитой от атаки с переполнением буфера ключа, поскольку этому пакету предшествует подключение к игре.

Атака по запросу о ключе. Механизм проверки CD-ключа имеет и иное слабое место. Чтобы проверить, используется ли иными игроками ключ, представленный данным игроком, приложение использует онлайн-валидацию. Ключ отправляется на серверы *Gamespy*, и затем отправитель получает ответ о результатах валидации. Все сообщения шифруются с помощью функции XOR, используя для шифрования строку *gamespy*. В указанном ответе встречаются запросы, разделенные символом *backslash*. При чтении запросов появляется ошибка. Если игрок отправит на сервер сообщение с одним символом *backslash*, приложение закроется. Проблема здесь кроется в плохо запрограммированном анализаторе запросов. Это можно проиллюстрировать следующим фрагментом кода:

```
int size = strchr(buff + 1, '\\') - buff;
  if(size > 32) return;
  strncpy(querybuff, buff + 1, size);
```

Переменная *buff* содержит запрос. В нем отыскивается знак «\». Затем проверяется условие, и извлеченный текст помещается в переменную *querybuff*. Заметна ошибка в этом программном коде. Значение, возвращаемое функцией *strchr*, не проверяется, поэтому, если функция не находит косую черту и возвращает «0», функция *strncpy* выдаст исключение, потому что значение переменной *size* будет отрицательным.

Решением описанной проблемы является патч, выпущенный Luigi Auriemm. Тип *signed* изменен на *unsigned*, поэтому значение не может быть отрицательным [8].

Атака на основе модификации карты. О карте мы вспоминали в анализе атаки на основе форматирования последовательности знаков.

Редактор карт доступен каждому. Некоторые модификации могут закрыть программу. Файлы карты на сервере должны совпадать с файлами карты у игрока. Карта содержит различные объекты: здание, стена, лестница и др. Каждый объект имеет свой идентификационный номер. Это важно для интерактивных объектов, таких как двери, кнопки, лестницы. Когда игрок использует один из объектов, он отправляет на сервер пакет с идентификационным номером объекта. Затем сервер отсылает пакеты с информацией о деятельности игрока другим участникам игры. Благодаря этому каждый может увидеть эффект от использования объекта, например открытие двери. Проблема возникает, когда игрок (или злоумышленник) использует объект с идентификационным номером, которого сервер «не знает». В этом случае программа закрывается.

Карта может состоять из ограниченного количества объектов. В анализируемой игре можно создать максимум 4096 объектов. Это облегчает защиту от этой атаки.

Если у всех будет одинаковый файл, который использует максимальное количество объектов, игра не закроется. Другой способ – создать базу данных всех используемых идентификационных номеров объектов и проверять, содержит ли пакет идентификатор из этой базы данных.

Кроме рассмотренных, существуют и иные виды атак на серверные и клиентские приложения данной и других компьютерных игр. Часто решением возникающих проблем занимаются не только разработчики игр, но и сами игроки. В последнем случае появляются специализированные программные средства.

Специализированные программные средства для защиты сервера игры. В доступных источниках содержится мало информации о программных продуктах, предназначенных для решения указанных задач.

Вероятно, одним из первых было многооконное приложение Project1. Оно имеет много функций, облегчающих работу администратора (например, отправка команд на сервер, написание общих сообщений игрокам, механизмы предупреждения обмана игроков). Project1 предоставляет много важной информации о сервере: количество игроков, текущая карта, время игры, список игроков каждой команды, статистика игроков, IP-адреса, разговоры в чате.

Приложение Mautorun основано на анализе сетевых пакетов. Приобрело большую популярность среди администраторов серверов.

Одной из программ, предотвращающих атаки на сервер, является AutoBan. Его основная функция заключается в обнаружении атак с переполнением буфера. Это обнаружение основано на анализе разницы во времени, соответствующего отправлению пакетов присоединения к игре. Как правило, атакующая сторона отправляет пакеты в течение 1 с. Пример соответствующих линеек кода (в действительности – двух) выглядит так:

```
[13:20:01] Server info sent to
192.168.1.1:26014
[13:20:01] NETWORKPACKET_TYPE_
CLIENTCONNECT [192.168.1.1:26015]
```

Можно заметить, что обе строки были созданы на протяжении 1 с. Это означает, что были отправлены вредоносные пакеты. Программа AutoBan извлекает IP-адрес и блокирует его. Если сервер работает быстро, последний пакет, который должен закрыть сервер, будет нейтрализован.

Второй способ обнаружить атаку – это проверить порты. Если хакер отправляет каждый пакет из другого сокета, то это приводит к изменению порта. Если две линии указывают на разные порты, это – вероятно, атака.

Авторское приложение для защиты сервера МПИ. Для нейтрализации описанных выше атак на сервер анализируемой игры нами разработано специальное приложение.

Вся система состоит из трех модулей (рис. 6).

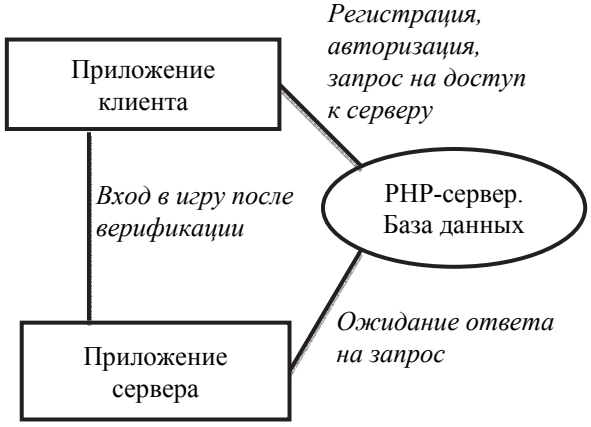


Рис. 6. Общая схема взаимодействия модулей системы

Клиентская программа используется для взаимодействия с пользователем, отвечает за регистрацию и вход в систему, настройку учетной записи и присоединение к игре. Второй модуль – это PHP-сервер с базой данных на платформе MySQL. Его задача – проанализиро-

вать данные, полученные из клиентской программы, проверить их корректность и возвратить необходимую информацию игроку. Информация о пользователях хранится в базе данных. Последний модуль представляет собой серверное приложение, которое было интегрировано в существующую программу управления сервером. Модуль предназначен для проверки того, запрашивает ли пользователь доступ к серверу, а также для контроля исключений в брандмауэре. Добавляя исключения в брандмауэр, пользователь получает доступ к серверу. На рис. 7 представлен алгоритм регистрации пользователя.

Клиентская часть написана на C# с использованием технологии .NET. Основным преимуществом этой технологии является доступ ко многим библиотекам, содержащим готовые решения анализируемой проблемы.

Серверное приложение создано на Java с использованием технологии Maven и библиотеки *jnetpcap*, предназначенной для анализа сетевого трафика. Коммуникационный сервер создан с помощью технологии управления базами данных MySQL и языка PHP.

К особенностям разработанного приложения можно отнести следующее.

После получения списка ожидающих пользователей программа вызывает команду для добавления инструкции в брандмауэр сервера:

```
iptables -A INPUT -s 192.168.0.1 -p udp -
-dport 26001 -j ACCEPT
```

Такое правило принимает UDP-пакеты, поступающие на порт 26001 с IP-адреса 192.168.0.1.

Уникальный ключ CD-KEY должен быть сохранен в системном реестре.

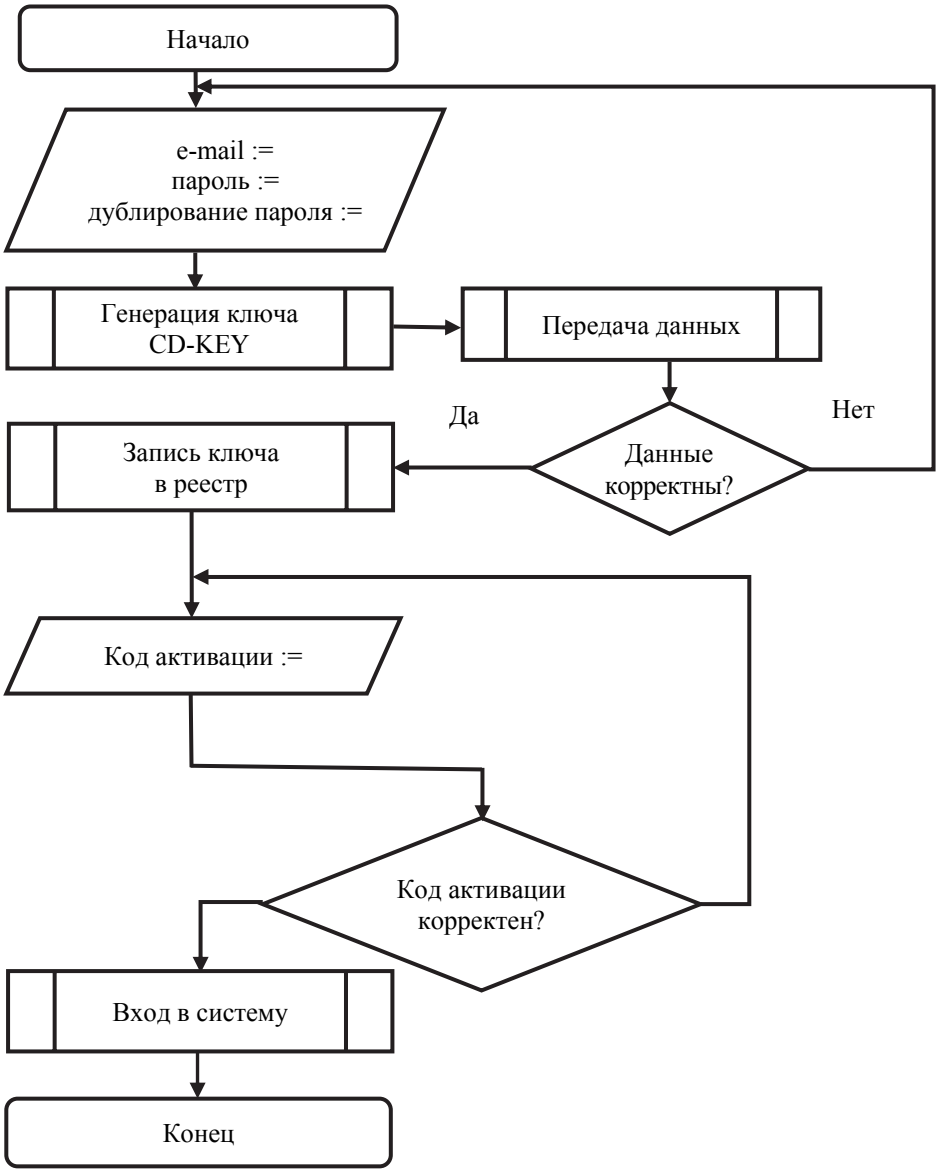


Рис. 7. Алгоритм регистрации пользователя для авторского приложения

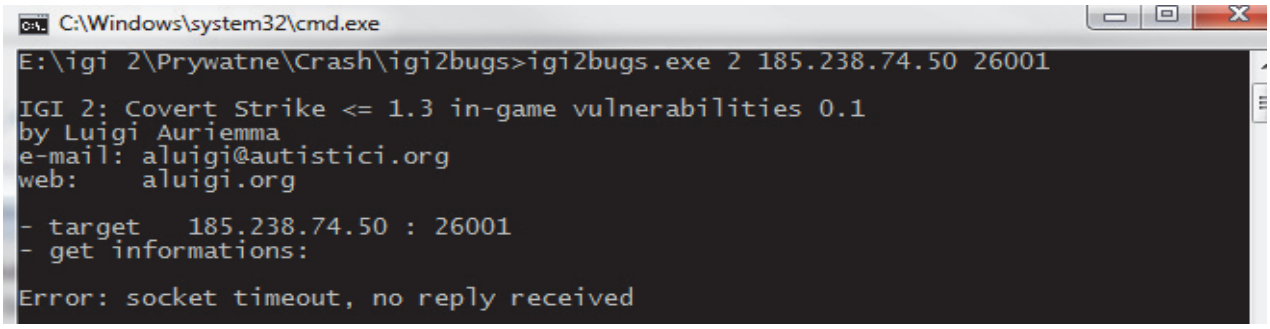


Рис. 8. Иллюстрация реализации атаки без регистрации пользователя

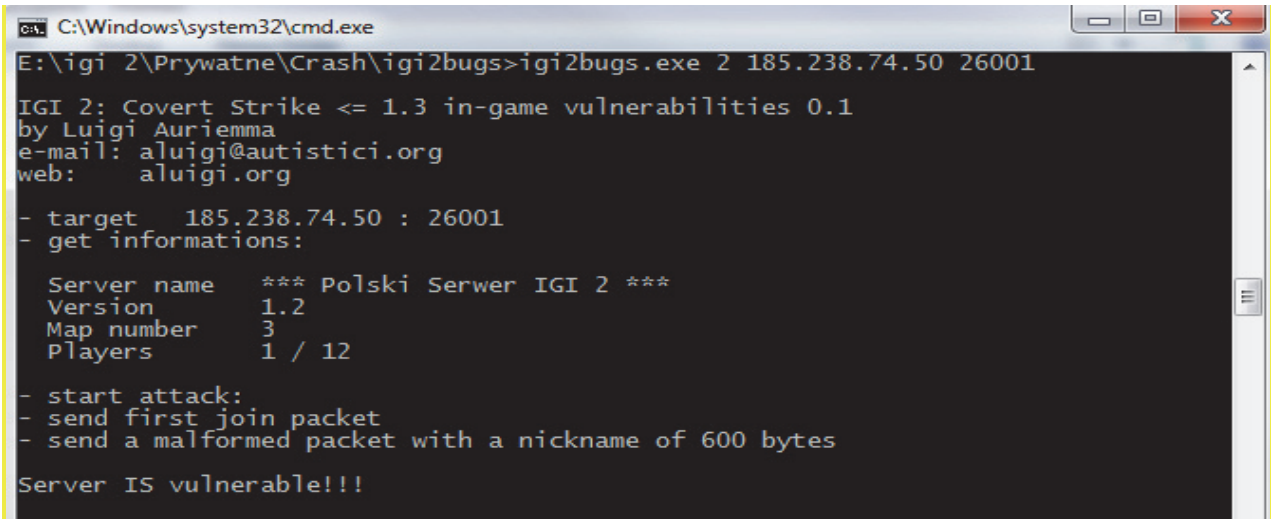


Рис. 9. Иллюстрация реализации атаки зарегистрированного пользователя

Программа на сервере проверяет статус игроков каждые 5 с, отправляя запрос в базу данных. Доступ к серверу получается путем добавления необходимой инструкции в брандмауэр. Если попытаться получить доступ к серверу без запроса доступа, то соответствующие пакеты будут проигнорированы, потому что весь входящий трафик будет заблокирован. Попытка войти на сервер без использования клиентской программы не приведет к получению ответа от сервера.

Попытка атаковать сервер с помощью хорошо известной программы *igi2bugs* также потерпит неудачу, как это показано на скриншоте (рис. 8), поскольку сервер отклоняет весь трафик. Это делает систему невосприимчивой к атакам незарегистрированных пользователей.

Если пользователь зарегистрируется, он все равно не сможет атаковать, поскольку он не отправил запрос на доступ к серверу. Только после нажатия кнопки «Присоединиться к игре» («Join game») клиентская программа отправляет такой запрос. К сожалению, система не в состоянии заблокировать попытку атаки зарегистрированного игрока. Такую ситуацию иллюстрирует рис. 9.

В большинстве компьютерных игр последнего поколения практически нет уязвимостей, свойственных описанной игре. Но нет ни одного приложения, абсолютно защищенного перед атаками. Невозможно также защитить многопользовательскую компьютерную игру на 100%. Однако можно снизить ее уязвимости на основе анализа предыдущих событий.

Заключение. Анализ сетевого трафика между сервером и клиентским приложением (игроком) показал, что большинство атак на сервер многопользовательской компьютерной игры можно заблокировать.

При выборе системы, на которой будет запускаться игровой сервер, необходимо учитывать время отклика при добавлении необходимых инструкций в брандмауэр. Система Linux позволяет фильтровать сетевой трафик, используя *iptables* [8]. Брандмауэр Windows медленнее реагирует на добавление новых инструкций. Также доступны улучшенные серверные приложения, которые блокируют некоторые виды атак.

Наш практический опыт показал достаточно высокую эффективность использования масштабирования базовой программной платформы компьютерной игры «Project I.G.I. 2: Covert Strike» для защиты игрового сервера от злоумышленников.

Литература

1. Ciesielka P., Urbanovich P. P. Security of applications for computer games [Электронный ресурс] // Информационные технологии: материалы 83-й науч.-техн. конф. проф.-препод. состава, науч. сотр. и аспирантов, Минск, 4–15 февр. 2019 г. / Белорус. гос. технол. ун-т. Минск, 2019. С. 26–28. URL: https://www.belstu.by/Portals/0/userfiles/37/09-tezisi-PPS-IT-2019_2.pdf (дата обращения: 20.09.2019).

2. Видеоигры и информационная безопасность: как не проиграть [Электронный ресурс]: [сайт]. [2019]. URL: https://www.securitylab.ru/blog/company/falcongaze/338191.php (дата обращения: 20.09.2019).

3. Błaszczak M., Urbanovich P. P. Server security of the multiplayer game «PROJECT I.G.I. 2: COVERT STRIKE» [Электронный ресурс] // Информационные технологии: материалы 83-й науч.-техн. конф. проф.-препод. состава, науч. сотр. и аспирантов, Минск, 4–15 февр. 2019 г. / Белорус. гос. технол. ун-т. Минск, 2019. С. 117–119. URL: https://www.belstu.by/Portals/0/userfiles/37/09-tezisi-PPS-IT-2019_2.pdf (дата обращения: 20.09.2019).

4. Логинов А. Краткие обзоры. IGI 2: Covert Strike [Электронный ресурс]: [сайт]. [2019]. URL: https://www.igromania.ru/article/7721/Kratkie_obzory_IGI_2_Covert_Strike.html (дата обращения: 21.04.2019).

5. Урбанович П. П. Защита информации методами криптографии, стеганографии и обфускации. Минск: БГТУ, 2016. 220 с.

6. Howard M., LeBlanc D. Writing Secure Code, Redmond. Washington: Microsoft Press, 2003. 768 p.

7. Урбанович П. П., Романенко Д. М., Кабак Е. В. Компьютерные сети. Минск: БГТУ, 2011. 400 с.

8. Luigi Auriemma. Gamespy SDK used for online cd-keys validation in third party code [Электронный ресурс]: [сайт]. [2019]. URL: http://aluigi.altervista.org/adv/gshboom-adv.txt (дата обращения: 30.09.2019).

References

1. Ciesielka P., Urbanovich P. P. [Security of applications for computer games]. *Informatsionnyye tekhnologii*, Minsk, 2019, pp. 26–28 (In Russian). Available at: https://www.belstu.by/Portals/0/userfiles/37/09-tezisi-PPS-IT-2019_2.pdf (accessed 20.09.2019).

2. *Videoigry i informatsionnaya bezopasnost'*: *kak ne proigrat'* [Video games and information security: how not to lose]. Available at: https://www.securitylab.ru/blog/company/falcongaze/338191.php (accessed 20.09.2019).

3. Błaszczak M., Urbanovich P. P. [Server security of the multiplayer game "PROJECT I.G.I. 2: COVERT STRIKE"]. *Informatsionnyye tekhnologii*, Minsk, 2019, pp. 26–28 (In Russian). Available at: https://www.belstu.by/Portals/0/userfiles/37/09-tezisi-PPS-IT-2019_2.pdf (accessed 20.09.2019).

4. Loginov A. *Kratkiye obzory. IGI 2: Covert Strike* [Brief reviews. IGI 2: Covert Strike]. Available at: https://www.igromania.ru/article/7721/Kratkie_obzory_IGI_2_Covert_Strike.html (accessed 21.04.2019).

5. Urbanovich P. P. *Zashchita informatsii metodami kriptografii, steganografii i obfuskatsii* [Information protection using cryptography, steganography and obfuscation methods]. Minsk, BGTU Publ., 2016. 220 p.

6. Howard M., LeBlanc D. Writing Secure Code, Redmond. Washington, Microsoft Press, 2003. 768 p.

7. Urbanovich P. P., Romanenko D. M., Kabak E. V. *Komp'yuternyye seti* [Computer networks]. Minsk, BGTU Publ., 2011. 400 p.

8. Luigi Auriemma. Gamespy SDK used for online cd-keys validation in third party code. Available at: http://aluigi.altervista.org/adv/gshboom-adv.txt (accessed 30.09.2019).

Информация об авторах

Блащак Матэуш – магистрант. Люблинский католический университет Иоанна Павла II (20-950, г. Люблин, Аллеи Рацлавицке, 14, Польша). E-mail: mateuszblaszczakb@gmail.com

Урбанович Павел Павлович – доктор технических наук, профессор, профессор кафедры информационных систем и технологий. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: p.urbanovich@belstu.by, uppkul@kul.lublin.pl

Information about the authors

Błaszczak Mateusz – Master's degree student. The John Paul II Catholic University of Lublin (14, Aleje Racławickie, 20-950, Lublin, Poland). E-mail: mateuszblaszczakb@gmail.com

Urbanovich Pavel Pavlovich – DSc (Engineering), Professor, Professor, the Department of Information Systems and Technology. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: p.urbanovich@belstu.by, uppkul@kul.lublin.pl

Поступила после доработки 25.11.2019

УДК 621.391

И. О. Оробей[1], Д. А. Гринюк[1], Н. М. Олиферович[1], М. Ф. Лукашевич[2], М. А. Анкуда[1]
[1]Белорусский государственный технологический университет
[2]НПП «Герда»

ФИЛЬТР С АДАПТАЦИЕЙ ПО ВЕРОЯТНОСТНОМУ КРИТЕРИЮ

Проведен анализ основных недостатков существующих методов адаптивной фильтрации. Предложен вариант построения адаптивного цифрового фильтра с бесконечной импульсной характеристикой, который характеризуется минимальным временем запаздывания при нестационарном исследуемом процессе за счет установления сигнала на основе показателя нестационарности по критерию серий. Нестационарность вероятностных характеристик потока данных с выхода аналогово-цифрового преобразователя приводит к появлению тренда, т. е. к отсутствию статистической независимости. Поскольку данные могут иметь разные функции распределения, то их исследования удобно проводить на основе свободных от распределений непараметрических методов, например, с помощью критерия серий или инверсий, причем первый вариант является более предпочтительным, поскольку не требует хранения всей выборки значений.

Ключевые слова: адаптивные фильтры, аналогово-цифровой преобразователь, конечная импульсная характеристика, критерий серии, экспоненциальный фильтр.

I. O. Orobei[1], D. A. Hryniuk[1], N. M. Oliferovich[1], M. F. Lukashevich[2], M. A. Ankuda[1]
[1]Belarusian State Technological University
[2]SPE “Gerda”

FILTER WITH ADAPTATION BY PROBABILITY CRITERIA

The analysis of the main disadvantages of the existing methods of adaptive filtering is carried out. A variant of constructing an adaptive digital filter with an infinite pulse characteristic is proposed, which is characterized by a minimum delay time in the non-stationary process under study by establishing a signal based on the non-stationary index according to the series criterion. The non-stationarity of the probabilistic characteristics of the data stream from the output of the analog-to-digital converter leads to the appearance of a trend, that is, to the lack of statistical independence. Since the data can have different distribution functions, it is convenient to study them on the basis of nonparametric methods free of distributions, for example, using the criterion of series or inversions, the first option being more preferable because it does not require storage of the entire sample of values.

Key words: adaptive filters, analog-to-digital converter, finite impulse response, series criterion, exponential filter.

Введение. Для обработки в режиме реального времени слабых сигналов с нестационарным характером шумов и помех в измерительной технике [1], в системах исследования конденсированных сред методами ядерного магнитного резонанса [2], а также при измерении электрохимических характеристик [3] можно использовать цифровые адаптивные фильтры (АФ), обеспечивающие увеличение количества усредняемых значений при стационарном процессе либо снижение при появлении нестационарности. Применение АФ актуально при использовании для обработки сигнала контроллеров с быстродействующим аналогово-цифровым преобразователем (АЦП). В существующих методах цифровой адаптивной фильтрации применяют фильтры скользящего среднего с изменяемыми весовыми коэффициентами, имеющие конечную импульсную характеристику (КИХ). Весовые коэффициенты выбирают по вектору ошибки из перестраиваемой матрицы, рассчитывают на основе минимизации ошибок градиентным методом [4] или определяют методом наименьших квадратов [5].

Для реализации алгоритмов оценивания и расчета весовых коэффициентов необходима высокая вычислительная мощность, определяемая размерностью системы уравнений. Фильтр с КИХ требует большого объема памяти для хранения усредняемых значений и матрицы весовых коэффициентов и имеет ограниченное возрастание точности с течением времени даже при соблюдении модельных и реальных значений сигнала и шума. Неидентичность реальных данных и теоретических значений, полученных на основе модели, может привести к неустойчивой работе АФ. При наличии нескольких локальных экстремумов в функциях ошибок выбор весовых коэффициентов по градиентным методам приводит к экстремуму, который не

допускает получение минимального отклонения. Статистические методы определения коэффициентов также не обеспечивают устойчивость работы фильтра в случае нестационарности вероятностных характеристик процесса.

Целью работы являлась разработка альтернативного метода цифровой фильтрации, который характеризуется минимальным временем запаздывания при нестационарном исследуемом процессе и обладает бесконечной импульсной характеристикой.

Основная часть. Сущность предложенного цифрового АФ основывается на использовании критерия серий для оценки статистической независимости или тренда данных с АЦП. Нестационарность вероятностных характеристик потока данных приводит к появлению тренда, т. е. к отсутствию статистической независимости. Поскольку данные с АЦП могут иметь разные функции распределения, то их исследования удобно проводить на основе свободных от распределений непараметрических методов, например, с помощью критерия серий или инверсий, причем первый вариант является более предпочтительным, поскольку не требует хранения всей выборки значений [6, 7].

Серией называется последовательность однотипных наблюдений, перед и после которой следуют наблюдения противоположного типа или таковые вообще отсутствуют [8]. Для последовательности N наблюдений случайной величины каждое наблюдение y_i $(i = 1, 2, \ldots, N)$ можно отнести к одному из двух классов (+) или (–). При выполнении условия $y_i \geq Y_{\text{ср}}$, где $Y_{\text{ср}}$ – среднее значение или медиана последовательности y_i, наблюдение можно отнести к классу (+) (с ошибкой $e \geq 0$); в противном случае наблюдение относится к классу (–) ($e < 0$). Наблюдения с $e = 0$ можно отбрасывать или относить к классу предыдущего наблюдения. Число серий r в последовательности позволяет выяснить, являются ли отдельные результаты статистически независимыми наблюдениями одной случайной величины. Если последовательность N наблюдений состоит из независимых исходов случайной величины, т. е. вероятность отдельных исходов (+) или (–) не меняется, то число серий r является случайной величиной, распределенной по нормальному закону, со средним значением и дисперсией [7]:

$$\mu = \frac{2N^+N^-}{N} + 1; \quad \sigma^2 = \frac{2N^+N^-\left(2N^+N^- - N\right)}{N^2\left(N-1\right)}, \quad (1)$$

где N^+, N^- – число исходов, относящихся к классам (+) и (–) соответственно.

При статистической независимости $N^+ = N^- = N/2$, что позволяет преобразовать среднее значение и дисперсию к виду [8]

$$\mu = \frac{N}{2} + 1; \quad \sigma^2 = \frac{N^2 - 2N}{4\left(N-1\right)}. \quad (2)$$

Для малых N среднее число серий имеет вид

$$\mu = \frac{2N^+N^-}{N} + 0{,}5; \quad \mu = \frac{N}{2} + 0{,}5. \quad (3)$$

После определения μ, σ^2 и r задается уровень значимости и сравнивается экспериментальное число серий r с границами принятия гипотезы статистической независимости, определяемыми относительно μ по уровню значимости. Если r окажется вне этой области, то гипотезу статистической независимости отвергают с принятым уровнем значимости, в противном случае процесс считают статистически независимым. В разработанном АФ использовано определение уровня значимости, соответствующего принятию гипотезы статистической независимости по μ, σ^2 и r, т. е. уровня значимости, соответствующего границам $[\mu - r;\ \mu + r]$. Через уровень значимости или связанные с ним величины можно определить вероятность статистической независимости данных.

Функциональная схема АФ приведена на рис. 1. В фильтре применен непосредственный подсчет серий r в блоке *8*, причем наблюдение с $e = 0$ получает знак ошибки предыдущего наблюдения (блок *7*). Вместо непосредственного подсчета r можно находить N^+ и N^- за N наблюдений в реальном процессе с последующим расчетом r по формуле (1). Переменные на рис. 1 имеют следующее назначение: z – целая переменная состояния фильтра; c – целая постоянная наращивания z; γ – вероятность статистической независимости отсчетов за период числа наблюдений; i – переменная числа наблюдений; h – шаг дискретизации АЦП, k-й шаг дискретизации соответствует последнему цифровому отсчету; $e(kh)$, $e((k-1)h)$ – ошибки между цифровым отсчетом АЦП соответственно на k-м и на $(k-1)$-м шагах дискретизации и выходом фильтра; $s(kh)$, $s((k-1)h)$ – знаки ошибки соответственно на k-м и на $(k-1)$-м шагах дискретизации; N^+, N^- – соответственно число положительных (с положительной ошибкой) и отрицательных исходов за N наблюдений при их статистической независимости; σ – среднеквадратичное отклонение числа серий r за N наблюдений при их статистической независимости (σ^2 – дисперсия); $y(kh)$ – цифровой отсчет АЦП на последнем k-м шаге; $Y_{\text{ср}}(kh)$ – среднее значение отсчетов АЦП за k шагов; ceil – операция округления до ближайшего большего целого; α – весовой коэффициент фильтра.

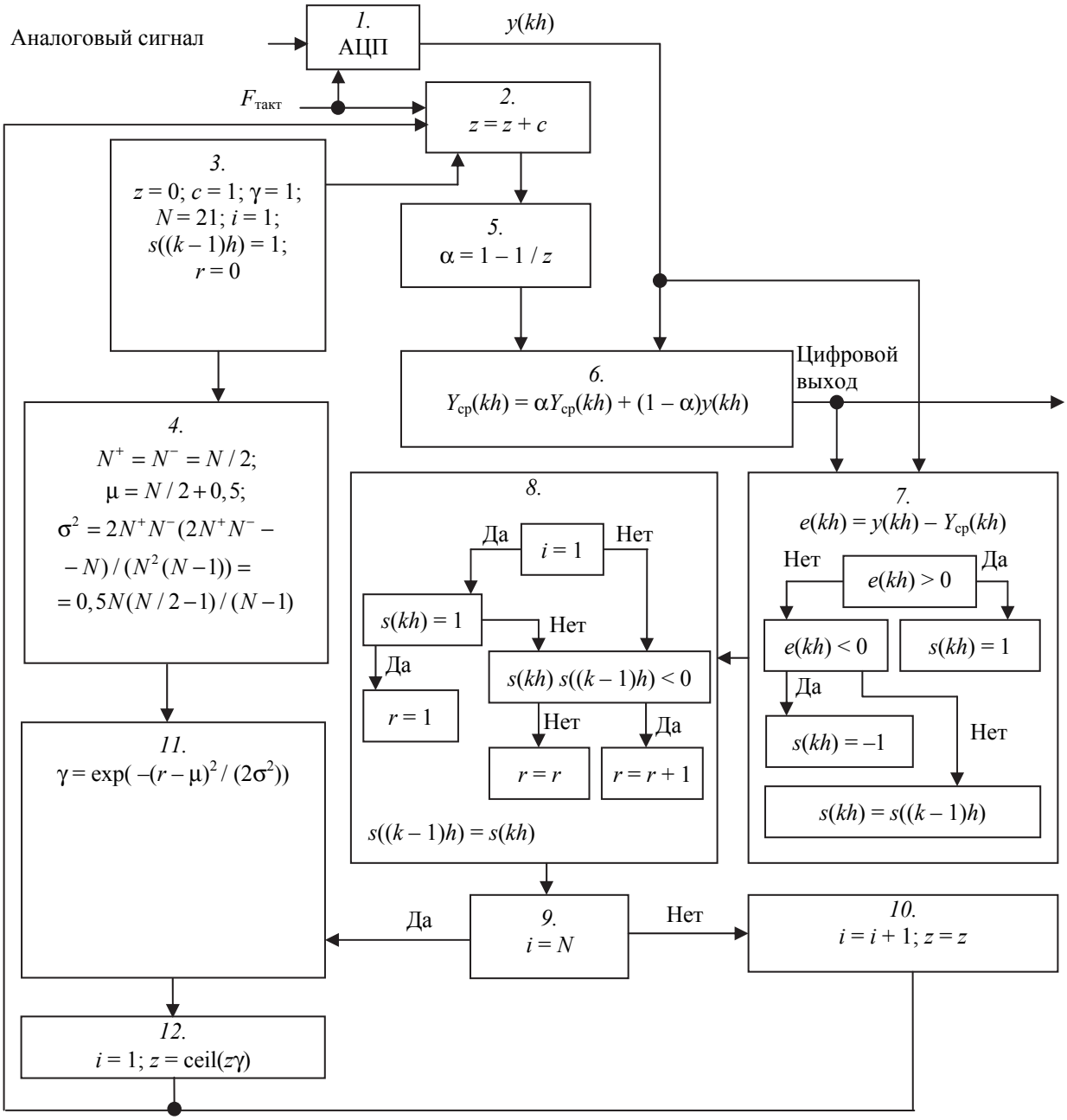


Рис. 1. Адаптивный фильтр:
1 – АЦП; *2* – наращивание переменной состояния; *3* – инициализация начальных установок; *4* – расчет среднего и дисперсии для *N* наблюдений при статистической независимости отсчетов; *5* – вычисление весового коэффициента; *6* – авторегрессивная фильтрация скользящего среднего; *7* – определение ошибки и ее знака; *8* – подсчет числа серий; *9* – проверка конца наблюдений; *10* – наращивание переменной наблюдений при сохранении переменной состояния; *11* – определение вероятности статистической независимости; *12* – сброс переменной наблюдений с изменением переменной состояния

Операцию авторегрессивной фильтрации выполняет экспоненциальный фильтр первого порядка, работа которого определяется уравнением [9]

$$Y_{\text{ср}}(kh) = \alpha\big((k-1)h\big) + (1-\alpha)\,y(kh). \qquad (4)$$

На рис. 1 в блоке *6* уравнение (4) модифицировано, это связано с экономией памяти за счет среднего значения на предыдущем шаге. Ее возможность обусловлена тем, что $Y_{\text{ср}}(kh)$ фактически становится $Y_{\text{ср}}((k-1)h)$ с приходом на блок *6* нового цифрового отсчета с АЦП $y(kh)$. Замена фильтра первого порядка на фильтр второго порядка с ненулевым значением коэффициента перед последним отсчетом приводит к увеличению времени вычислений, не влияя при этом на характеристики АФ.

Сокращение памяти для хранения отдельных значений усредняемой величины достигается за счет фильтра с ненулевым значением коэффициента только перед последним значением усредняемой величины (блок *6* на рис. 1) [9]. Весовые коэффициенты перед остальными отдельными отсчетами равны нулю, что исключает необходимость хранения всех цифровых отсчетов АЦП, кроме последнего. Пошаговое возрастание точности для последовательности статистически независимых данных со стационарными вероятностными характеристиками обеспечивается благодаря наращиванию постоянной состояния фильтра в блоке *2*. При этом снижается весовой коэффициент перед последним цифровым отсчетом и увеличивается вклад в последующую величину среднего значения предыдущего среднего (блок *5*). Увеличение переменной состояния производится в цикле для каждого значения цифрового отсчета, причем в пределах одной последовательности N наблюдений отсчеты АЦП принимаются статистически независимыми. Для каждого наблюдения анализируется знак ошибки (блок *7*) и подсчитывается текущее число серий знаков ошибок (блок *8*). В качестве текущего среднего значения при расчете ошибки используется выходной сигнал фильтра. В каждом цикле проверяется условие достижения N наблюдений по значению переменной наблюдений i (блок *9*). Если набор последовательности из N наблюдений не закончен, то производится наращивание i (блок *10*) с последующим переходом к наращиванию переменной состояния (к блоку *2*). По достижению N переходят к определению вероятности статистической независимости γ в законченной последовательности (блок *11*), которая производится сопоставлением экспериментального числа серий r, определенного в блоке *8*, с расчетными характеристиками для N наблюдений при их статистической независимости, получаемыми в блоке *4* после инициализации начальных установок. Вероятность статистической независимости γ умножают на переменную состояния фильтра, округляют полученную величину до ближайшего большего целого значения, которое присваивают переменной состояния фильтра, и сбрасывают переменную наблюдений (блок *12*), после чего переходят к новому набору последовательности N наблюдений.

Вероятность статистической независимости процесса в блоке *11* определяется как отношение плотности вероятности, соответствующей экспериментальному r, к плотности вероятности, соответствующей среднему числу серий μ, рассчитываемому по (3), при дисперсии σ^2, найденной по (1), (2). Такое определение γ требует меньше вычислительных средств, чем критерии с непосредственным определением уровня значимости. Статистическая независимость последовательности данных нарушается при появлении нестационарности случайного процесса, т. е. при непостоянстве вероятностных характеристик (среднего значения, медианы и т. д.) случайного процесса, поэтому предлагаемый алгоритм адаптируется к случайным нестационарным процессам в случае, когда спектральные составляющие дрейфа вероятностных характеристик имеют период, превышающий время выборки N наблюдений.

Для определения γ в блоке *11* может служить кусочно-линейная аппроксимация вероятностного критерия по отношению плотностей вероятности (рис. 2).

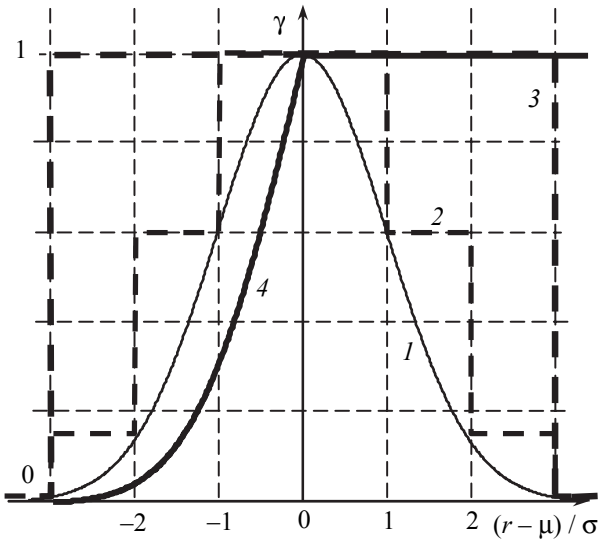


Рис. 2. Определение статистической независимости γ для блока *11*:
линии *1–3* – по отношению плотностей вероятности (*1* – по формуле (5) при $K_1 = 1$;
2 – кусочно-линейная аппроксимация по диапазонам отклонений от среднего значения σ, 2σ и 3σ;
3 – кусочно-линейная аппроксимация по диапазону отклонений от среднего значения 3σ);
линия *4* – по функции распределения серий

Для оптимизации вычислительных ресурсов лучше использовать линию *3* на рис. 2, считая, что $\gamma \approx 0{,}01$ при отклонении, большем или равном 3σ. Вероятность статистической независимости также можно находить в виде

$$\gamma = \exp\left(-\frac{(r-\mu)^2}{K_1\sigma^2}\right), \qquad (5)$$

где K_1 – коэффициент сжатия-растяжения вероятности статистической независимости, $0 < K_1 < \infty$.

При $K_1 > 2$ снижается скорость изменения весовых коэффициентов фильтра, т. е. замедляется процесс адаптации, но снижается чувствительность к спектральным составляющим трендов, имеющим период, сопоставимый с длительностью последовательности наблюдений.

Для определения γ можно использовать функцию распределения случайной величины r, например вероятность попадания в интервал, не включающий $[\mu - r; \mu + r]$, т. е. вероятность отклонения, превышающего либо равного отклонению экспериментального числа серий от среднего, рассчитываемого в предположении статистической независимости:

$$\gamma = 1 - \frac{1}{\sigma\sqrt{2\pi}} \int \exp\left(-\frac{x-\mu}{2\sigma^2} \right) dx. \qquad (6)$$

В некоторых случаях стационарность вероятностных характеристик случайного процесса сохраняется и при отсутствии статистической независимости данных, что ограничивает область применимости стандартного критерия серий в АФ [10, 11]. Применимость АФ в таких случаях обеспечивается предположением, что серии распределены по закону, отличающемуся от нормального до некоторого значения R только постоянным множителем. При получении $r > R$ вероятностные характеристики процесса можно считать стационарными, несмотря на отсутствие статистической независимости данных при нормальном распределении серий. Тогда для числа серий можно ввести распределение:

$$\begin{cases} f(x) = \dfrac{K_2}{\sigma\sqrt{2\pi}} \exp\left(-\dfrac{(x-\mu)^2}{2\sigma^2} \right), & x \le R, \\ f(x) = 0, & x > R, \end{cases} \qquad (7)$$

где K_2 – коэффициент, определяемый из условия нормировки

$$\int_{-\infty}^{+} f(x)dx = 1. \qquad (8)$$

С учетом (8) распределение (7) преобразовывается к виду

$$\begin{cases} f(x) = \dfrac{1}{\int_{-\infty}^{R} g(x)dx} g(x), & x \le R, \\ f(x) = 0, & x > R, \end{cases} \qquad (9)$$

где

$$g(x) = \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right).$$

Для распределения, описываемого (9), в качестве γ можно использовать вероятность попадания в интервал $[-\infty; r]$, т. е. функцию распределения $F(x)$

$$\gamma = F(x) = \frac{1}{\int_{-\infty}^{R} g(x)dx} \int_{-\infty}^{r} g(x)dx. \qquad (10)$$

На рис. 2 показан вариант исполнения операции определения γ по функции распределения серий в соответствии с (10) при $R = \mu$. Формулу (10) можно модифицировать, заменив нижние пределы в интегралах на единицу с учетом того, что экспериментальное число серий $r \ge 1$, т. е. вероятность статистической независимости данных в последовательности из N наблюдений определяется как нормированная вероятность попадания числа серий в интервал от единицы до значения экспериментального числа серий r.

Ограниченное время переходного процесса, реализующего операцию в блоке *6*, обеспечивается за счет снижения весового коэффициента перед предыдущим значением (значениями) среднего при увеличении коэффициента перед последним отсчетом АЦП, что происходит при изменении переменной состояния по вероятности статистической независимости.

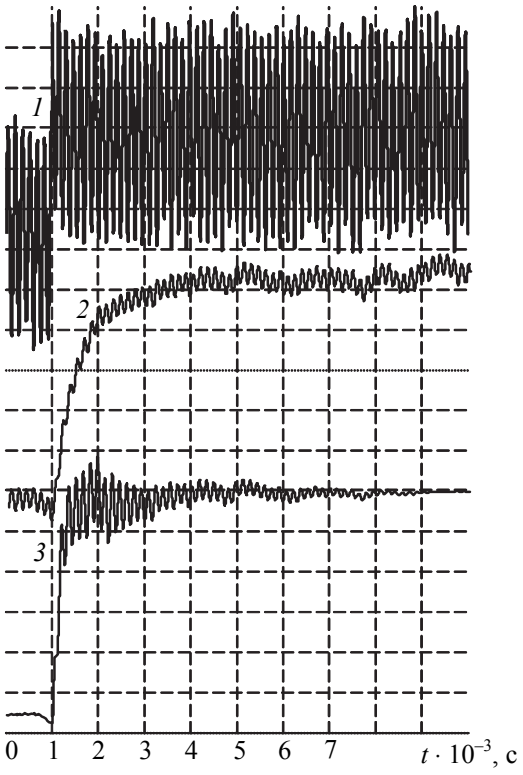


Рис. 3. Временные диаграммы цифровых сигналов:
1 – цифровой сигнал на выходе АЦП без фильтра;
2 – сигнал на выходе оптимизированного фильтра с неизменными характеристиками;
3 – сигнал на выходе адаптивного фильтра

При близкой к нулю вероятности коэффициент перед предыдущим значением среднего также приближается к нулю, а коэффициент перед последним цифровым отсчетом АЦП – к единице, т. е. в качестве среднего значения фильтра используется последний цифровой отсчет. Статистическая независимость последовательности данных нарушается при появлении нестационарности случайного процесса, т. е. при непостоянстве вероятностных характеристик (среднего значения, медианы и т. д.) случайного процесса, поэтому алгоритм адаптируется к случайным нестационарным процессам в случае, когда составляющие дрейфа вероятностных характеристик имеют период, превышающий время выборки N наблюдений.

Заключение. Таким образом, проведенный анализ метода адаптивной фильтрации на основе критерия серий показал, что при появлении нестационарности запаздывание, соответствующее установлению нового значения, не превышает 2–3 тактов. После чего фильтр переходит в новое состояние с наращиванием времени усреднения. Временные диаграммы, характеризующие работу АФ (рис. 3, кривая *3*), показывают, что наряду с высокой скоростью установления в динамическом режиме при скачке входных данных разработанный фильтр обеспечивает возрастающее с течением времени ослабление случайных погрешностей в статическом режиме, что позволяет обеспечить снижение динамических и случайных погрешностей.

Литература

1. Ростами Х. Р. Высокочувствительный холловский магнитометр // Приборы и техника эксперимента. 2016. № 2. С. 112–116.

2. О возможности регистрации спектров ядерного магнитного резонанса жидких сред в слабых полях в экспресс-режиме / В. В. Давыдов [и др.] // Журнал технической физики. 2018. Т. 88, вып. 12. С. 1885–1889.

3. Астафьев Е. А., Манжос Р. А. Прибор с широким динамическим диапазоном для измерения электрохимических шумов // Приборы и техника эксперимента. 2018. № 1. С. 149–150.

4. Зубарев Ю. Б., Витязев В. В., Дворкович В. П. Цифровая обработка сигналов – информатика реального времени // Цифровая обработка сигналов. 1999. № 1. С. 5–17.

5. Витязев В. В. Банки фильтров в системах широкополосной передачи данных // Цифровая обработка сигналов. 2016. № 2. С. 44–52.

6. Сухорукова И. Г., Оробей И. О., Гринюк Д. А. Эффективность работы адаптации фильтра на критерии серий // Труды БГТУ. 2011. № 6: Физ.-мат. науки и информатика. С. 107–111.

7. Сухорукова И. Г., Гринюк Д. А., Оробей И. О. Адаптация критерия серий к применению в управлении технологическими процессами // Труды БГТУ. 2014. № 6: Физ.-мат. науки и информатика. С. 92–95.

8. Бендат Дж., Пирсол А. Прикладной анализ случайных данных. М.: Мир, 1989. 545 с.

9. Олссон Г., Пиани Дж. Цифровые системы автоматизации и управления. СПб.: Невский Диалект, 2001. 557 с.

10. Сухорукова И. Г., Гринюк Д. А., Оробей И. О. Влияние условий фильтрации и сглаживания в информационных каналах на критерий серий // Труды БГТУ. 2016. № 6: Физ.-мат. науки и информатика. С. 117–121.

11. Oliferovich N., Hryniuk D., Orobei I. Measuring the speed of capillary soaking with adaptation regarding coordinates // Open Conference of Electrical, Electronic and Information Sciences (eStream). Vilnius, 2015. P. 1–4. DOI: 10.1109/eStream.2015.7119495.

References

1. Rostami Kh. R. Highly sensitive Hall magnetometer. *Pribory i tekhnika eksperimenta* [Instruments and experimental technique], 2016, no. 2, pp. 112–116 (In Russian).

2. Davydov V. V., Myazin N. S., Dudkin V. I., Grebenikova N. M. On the possibility of recording nuclear magnetic resonance spectra of liquid media in weak fields in express mode. *Zhurnal tekhnicheskoy fiziki* [Journal of Technical Physics], 2018, vol. 88, issue 12, pp. 1885–1889 (In Russian).

3. Astaf'ev E. A., Manzhos R. A. Wide dynamic range instrument for measuring electrochemical noise. *Pribory i tekhnika eksperimenta* [Instruments and experimental technique], 2018, no. 1, pp. 149–150 (In Russian).

4. Zubarev Yu. B., Vityazev V. V., Dvorkovich V. P. Digital signal processing – real-time computer science. *Tsifrovaya obrabotka signalov* [Digital signal processing], 1999, no. 1, pp. 5–17 (In Russian).

5. Vityazev V. V. Filter banks in broadband data transmission systems. *Tsifrovaya obrabotka signalov* [Digital signal processing], 2016, no. 2, pp. 44–52 (In Russian).

6. Sukhorukova I. G., Hryniuk D. A., Orobei I. O. Efficiency of filter adaptation on series criteria. *Trudy BGTU* [Proceedings of BSTU], 2011, no. 6: Physics and Mathematics. Informatics, pp. 107–111 (In Russian).

7. Sukhorukova I. G., Hryniuk D. A., Orobei I. O. Adaptation of the series criterion for use in process control. *Trudy BGTU* [Proceedings of BSTU], 2014, no. 6: Physics and Mathematics. Informatics, pp. 92–95 (In Russian).

8. Bendat Dzh., Pirsol A. *Prikladnoy analiz sluchaynykh dannykh* [Applied Random-Data Analysis]. Moscow, Mir Publ., 1989. 545 p.

9. Olsson G., Piani Dzh. *Tsifrovyye sistemy avtomatizatsii i upravleniya* [Digital automation and control systems]. St. Petersburg, Nevskiy Dialekt Publ., 2001. 557 p.

10. Sukhorukova I. G., Hryniuk D. A., Orobei I. O. Influence of conditions of filtering and smoothing in information channels runs tests. *Trudy BGTU* [Proceedings of BSTU], 2016, no. 6: Physics and Mathematics. Informatics, pp. 117–121 (In Russian).

11. Oliferovich N., Hryniuk D., Orobei I. Measuring the speed of capillary soaking with adaptation regarding coordinates. *Open Conference of Electrical, Electronic and Information Sciences* (*eStream*). Vilnius, 2015, pp. 1–4. DOI: 10.1109/eStream.2015.7119495.

Информация об авторах

Оробей Игорь Олегович – кандидат технических наук, доцент, доцент кафедры автоматизации производственных процессов и электротехники. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: orobei@tut.by

Гринюк Дмитрий Анатольевич – кандидат технических наук, доцент, доцент кафедры автоматизации производственных процессов и электротехники. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: hryniuk@tut.by

Олиферович Надежда Михайловна – ассистент кафедры автоматизации производственных процессов и электротехники. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: oliferovich@belstu.by

Лукашевич Максим Францевич – главный инженер. НПП «Герда» (220125, г. Минск, ул. Шафарнянская, 11, Республика Беларусь). E-mail: mluk_75@mail.ru

Анкуда Максим Анатольевич – ассистент кафедры автоматизации производственных процессов и электротехники. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: max.ankuda@gmail.com

Information about the authors

Orobei Igor' Olegovich – PhD (Engineering), Associate Professor, Assistant Professor, the Department of Automation of Production Processes and Electrical Engineering. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: orobei@tut.by

Hryniuk Dzmitry Anatol'yevich – PhD (Engineering), Associate Professor, Assistant Professor, the Department of Automation of Production Processes and Electrical Engineering. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: hryniuk@tut.by

Oliferovich Nadezhda Mikhaylovna – assistant lecturer, the Department of Automation of Production Processes and Electrical Engineering. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: oliferovich@belstu.by

Lukashevich Maksim Frantsevich – Chief Engineer. SPE "Gerda" (11, Shafarnyanskaya str., 220125, Minsk, Republic of Belarus). E-mail: mluk_75@mail.ru

Ankuda Maksim Anatol'yevich – assistant lecturer, the Department of Automation of Production Processes and Electrical Engineering. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: max.ankuda@gmail.com

Поступила после доработки 19.12.2019

УДК 681.3.06

А. А. Сущеня, Е. А. Блинова
Белорусский государственный технологический университет

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ СТЕГАНОГРАФИЧЕСКОЙ СИСТЕМЫ С ИСПОЛЬЗОВАНИЕМ СТЕГАНОГРАФИЧЕСКОГО КОНТЕЙНЕРА В ВИДЕ ЭЛЕКТРОННОЙ КНИГИ ФОРМАТА EPUB

Описывается модель стеганографической системы, основанная на использовании предлагаемого метода, в котором в качестве контейнера применяется файл-архив электронного формата EPUB. При помощи указанного метода предлагается скрывать тайное сообщение в файлах различного формата, что обеспечит повышение стеганографической стойкости системы. Новизна рассматриваемого метода заключается в размещении тайного сообщения, а также его контрольной суммы в трех видах контейнеров: XHTML-файлах, в которых находится текстовое содержание электронной книги, JPG-изображении, которое представляет собой обложку электронного издания, и CSS-файлах, в которых описаны правила визуального отображения книги. Излагаются возможности использования стеганографического метода в электронных изданиях для осаждения цифровых водяных знаков с целью защиты документов-контейнеров от несанкционированного копирования и распространения. Представлено разработанное программное средство «EPUB Modifier», демонстрирующее работу описанной стеганографической модели. Показан пользовательский интерфейс приложения, технология разработки, а также основные структурные элементы его архитектуры. Для грамотного построения архитектуры был использован паттерн проектирования «Chain of responsibility». Программное средство работает с электронным документом формата EPUB, изменяя его основные семантические части для создания стеганографического контейнера.

Ключевые слова: стеганография, формат EPUB, XHTML, CSS, LSB.

A. A. Sushchenia, E. A. Blinova
Belarusian State Technological University

MATHEMATICAL DESCRIPTION OF A STEGANOGRAPHIC SYSTEM FOR EMBEDDING INFORMATION IN THE EPUB FORMAT CONTAINER

The article describes a model of the steganographic system based on the proposed method, in which a file-archive of an EPUB electronic format is used as a container. Using this method, it is proposed to hide the secret message in files of different formats, which increases the steganographic stability of the system. The novelty of the proposed method is to place the secret message, as well as its checksum in three types of containers: XHTML-files, which contain the text content of the e-book, JPG-image, which is the cover of the electronic edition and CSS-files, which describe the rules of visual display of the book. The possibilities of using the steganographic method in electronic publications for the deposition of digital watermarks in order to protect container documents from unauthorized copying and distribution are explored. The developed software “EPUB Modifier” demonstrating the operation of the described steganographic model is presented. The user interface of the application, the development technology, as well as the main structural elements of its architecture are described. The “Chain of responsibility” design pattern was used for competent architecture construction. The software works with an electronic document of the EPUB format, changing its basic semantic parts to create a steganographic container.

Key words: steganography, format EPUB, XHTML, CSS, LSB.

Введение. Цифровой контент приобретает все большую популярность. Электронные книги не являются исключением. Сегодня намного быстрее и удобнее можно приобрести электронный вариант издания через Интернет, при этом не выходя из дома. Однако наряду с удобством также возрастает риск несанкционированного распространения и копирования цифровых экземпляров книг. Проблема защиты информации в процессе передачи от одного абонента к другому является актуальной для человечества на протяжении длительного периода времени. На сегодняшний день найдено большое количество способов, которые позволяют тайно осуществить процесс обмена информацией для обеих сторон. Способы скрытной передачи информации изучает стеганография [1].

Общей чертой таких способов является то, что скрываемое сообщение встраивается в не привлекающий внимание объект, который открыто отправляется адресату.

Секретность системы защиты передаваемых сообщений должна содержаться в ключе – фрагменте информации, предварительно разделенном между адресатами. Ключом выступает алгоритм внедрения информации в объект. Объект, в который внедрена информация, называется стеганографическим контейнером [2].

В качестве контейнера может выступать электронная книга одного из самых популярных форматов на сегодняшний день – EPUB, который по своей сути представляет собой набор XHTML-, JPG- и CSS-файлов [3].

Предлагаемая модель стеганографической системы основана на использовании метода, комбинирующего три разных типа контейнеров для улучшения стеганографической стойкости.

Основная часть. Объектом исследования в данной работе являются стеганографические методы защиты прав интеллектуальной собственности на электронные книги. Предметом – модели стеганографических процессов.

Предлагаемая модель строится на основе следующих обозначений и положений:

пусть **M** – это конечное множество сообщений, которые могут быть тайно размещены в контейнере: $\mathbf{M} = \{M_1, M_2, \ldots, M_n\}$. В предлагаемом методе **M** подразделяется на M_O – само внедряемое сообщение и M_H – контрольная сумма M_O, вычисленная на основе алгоритма MD5;

C – это конечное множество всех допустимых контейнеров (файлов-контейнеров или документов-контейнеров): $\mathbf{C} = \{C_1, C_2, \ldots, C_p\}$, причем $p > n$;

K – множество всех ключей, под которыми в общем случае понимаются методы или алгоритмы осаждения сообщения в контейнер или иные операции по предварительному преобразованию осаждаемого сообщения либо выбору элементов контейнера для такого осаждения: $\mathbf{K} = \{K_1, K_2, \ldots, K_z\}$ [4].

Произвольное тайное сообщение M_i можно скрыть в контейнере C_j при использовании ключа K_m: $M_i \in \mathbf{M}$, $i = 1, 2, \ldots, n$; $C_j \in \mathbf{C}$, $j = 1, 2, \ldots, p$; $K_m \in \mathbf{K}$, $m = 1, 2, \ldots, z$. Результатом такого типа преобразований будет заполненный контейнер (или стеганосообщение) S_q, относящийся к множеству заполненных контейнеров или стеганосообщений **S**: $\mathbf{S} = \{S_1, S_2, \ldots, S_r\}$, $q = 1, 2, \ldots, r$.

Функцию **F**, определенную на $\mathbf{M} \times \mathbf{C} \times \mathbf{K}$ со значениями в **S**, будем отождествлять с осаждением или встраиванием сообщения M_i из множества **M** в контейнер C_j из множества **C** на основе ключа из множества **K**, предусматривающего использование соответствующего алгоритма осаждения:

$$\mathbf{F}: \mathbf{M} \times \mathbf{C} \times \mathbf{K} \longrightarrow \mathbf{S}. \qquad (1)$$

Соотношение (1) формально описывает процедуру осаждения тайного сообщения M_i в контейнер C_j на основе выбранного ключа K_m.

Функцию $\mathbf{F}^{-1}$, определенную на $\mathbf{S} \times \mathbf{K}$ со значениями в **M**, будем отождествлять с извлечением тайного сообщения $M_i \in \mathbf{M}$ из стеганосообщения $S_q \in \mathbf{S}$:

$$\mathbf{F}^{-1}: \mathbf{S} \times \mathbf{K} \longrightarrow \mathbf{M}, \mathbf{C}. \qquad (2)$$

Таким образом, выражение (2) определяет обратное по отношению к (1) отображение, которое каждому элементу S_q множества **S** и фиксированному элементу множества **K** ставит в соответствие элемент M_i множества **M** и элемент C_j множества **C**.

Соотношение (2) формально описывает процедуру извлечения сообщения из контейнера на основе того же выбранного метода [1]. При извлечении сообщения для подтверждения того, что оно не было модифицировано, вычисляется его контрольная сумма и сравнивается с M_H.

С учетом всех вышеуказанных обозначений опишем стеганографическую систему, в которой в качестве контейнера используется электронный документ формата EPUB.

В составе электронной книги выделим три контейнера, в которые производится внедрение информации:

$$\mathbf{C} = \{C_{JPG}, C_{CSS}, C_{XHTML}\}, \qquad (3)$$

где C_{JPG} – обложка книги, изображение, представленное в формате JPG; C_{CSS} – файл каскадных таблиц стиля, хранящий конфигурацию отображения книги; C_{XHTML} – набор XHTML-файлов, количество которых зависит от глав в книге, а также несущий основную текстовую информацию.

Для выполнения процедуры внедрения информации в соответствии с контейнерами (3) используются следующие стеганографические ключи:

$$\mathbf{K} = \{K_{LSB}, K_{CSS}, K_Q\}, \qquad (4)$$

где K_{LSB} – ключ, представляющий собой осаждение информации стеганографическим методом LSB (Least Significant Bits) [1]. Суть метода замены наименее значащего бита заключается в сокрытии информации путем изменения последних битов изображения, кодирующих цвет, на биты скрываемого сообщения. Разница между пустым и заполненным контейнерами должна быть не ощутима для органов восприятия человека. Рассмотрим этот метод на примере 24-битного растрового RGB-изображения.

Приведенный на рис. 1 пример показывает, как сообщение «100011» может быть скрыто в двух пикселях 24-битного изображения.

Пустой контейнер:	11111010	10101111
	10000101	10110110
	11000101	10010101
Осаждаемое сообщение:	100 011	
Заполненный контейнер:	1111101**1**	1010111**0**
	1000010**0**	1011011**1**
	1100011**0**	1001010**1**

Рис. 1. Пример осаждения двоичного сообщения с использованием метода LSB

Каждый пиксель кодируется тремя байтами, каждый байт определяет интенсивность красного (Red), зеленого (Green) и синего (Blue) цвета. Совокупность интенсивностей цвета в каждом из трех каналов определяет оттенок пикселя. Изменяя наименее значащий бит, меняется значение байта на единицу.

Таким образом, контейнер C_{JPG} является основным из используемых и содержит непосредственно само осаждаемое сообщение M_O.

K_{CSS} – ключ, реализующий осаждение сообщения в файл каскадных таблиц стилей. Для визуального оформления XHTML-разметки предназначена технология CSS. CSS (англ. Cascading Style Sheets – каскадные таблицы стилей) – технология описания внешнего вида документа, оформленного языком разметки. Если XHTML предоставляет информацию о структуре электронной книги, то таблицы стилей сообщают, как она должна выглядеть. Стиль – это совокупность правил, применяемых к элементу гипертекста и определяющих способ его отображения. Стиль включает все типы элементов дизайна: шрифт, фон, цвета ссылок, поля и расположение объектов. Таблица стилей – это совокупность стилей, применимых к гипертекстовому документу. Каскадирование – это порядок присвоения различных стилей.

При создании CSS-файлов существует практика хранения изображений в формате base64.

Для реализации данного подхода необходимо закодировать изображение в формате base64. Далее положить получившуюся строку в CSS-файл, заменяя «ТИП» на MIME-тип изображения – JPEG/PNG/GIF или BMP и «КОД» на нужную строку в base64 (листинг 1).

```
.some_class {
    background-image: url("КОД");
}
```

Листинг 1. Пример использования кодировки base64 для задания фона в стиле

При помощи ключа K_{CSS} происходит внедрение обложки книги в файл со стилями [5]. Предварительно в обложку внедряется сообщение с использованием стеганографического алгоритма LSB. Контейнер C_{CSS} предназначен для дублирования основного сообщения, что в свою очередь повышает стеганографическую стойкость.

K_Q – ключ, реализующий осаждение сообщения в главы книги с использованием метода замены кавычек в файлах XHTML. XHTML – это основанный на XML язык разметки гипертекста, максимально приближенный к стандартам HTML. XHTML отличается от HTML строгостью написания кода. Если HTML позволяет писать практически любые конструкции и браузер их корректно распознает, то с появлением XHTML это стало невозможным. XHTML требует строгого соблюдения всех правил, предъявляемых W3C.

Известно, что интерпретатор XHTML-документа не придает значения, какой тип кавычек используется при его создании. Следовательно, если заменить какую-нибудь пару кавычек в валидном XHTML-документе, например с двойной на одинарную, то семантический смысл документа не изменится. Используя эту технику, в XHTML можно осадить бинарную последовательность.

При встраивании последовательности бит условимся, что единице будет соответствовать двойная кавычка, а нулю – одинарная. Начиная с первой пары кавычек в документе, будем ставить ей в соответствие бит встраиваемого сообщения и, при необходимости, изменять тип кавычки на противоположный. (Например, первая пара кавычек в документе двойная, а первый бит осаждаемой последовательности нулевой, следовательно, необходимо тип кавычек заменить на одинарный.) После того, как место в одном XHTML-файле закончилось, следует перейти к следующей главе. Данную процедуру необходимо выполнять до тех пор, пока сообщение не закончится, либо же когда глав в книге больше не останется [5, 6]. Таким образом, контейнер C_{XHTML} используется в качестве дополнительного и хранит хэш основного сообщения M_H.

С учетом всех описанных элементов стеганографической системы функция встраивания сообщения **F** будет выглядеть следующим образом:

$$\mathbf{F}\colon \mathbf{M}\ \{M_O, M_H\} \times \mathbf{C}\ \{C_{JPG}, C_{CSS}, C_{XHTML}\} \times \\ \times\ \mathbf{K}\ \{K_{LSB}, K_{CSS}, K_Q\} \longrightarrow \mathbf{S}.$$

Функция извлечения $\mathbf{F}^{-1}$ будет следующей:

$$\mathbf{F}^{-1}\colon \mathbf{S} \times \mathbf{K}\ \{K_{LSB}, K_{CSS}, K_Q\} \longrightarrow \mathbf{M}\ \{M_O, M_H\}, \\ \mathbf{C}\ \{C_{JPG}, C_{CSS}, C_{XHTML}\}.$$

Таким образом, стеганографическая модель при помощи нового метода была адаптирована для осуществления процедуры внедрения информации в цифровой файл формата EPUB.

Для демонстрации описанной стеганографической системы осаждения метки в электронный документ формата EPUB создано программное средство «EPUB Modifier». В качестве технологии для создания приложения была выбрана Windows Forms, которая позволяет разработать приложение с полнофункциональным графическим интерфейсом, простое в развертывании и обновлении, способное работать при наличии или отсутствии подключения к Интернету и использующее более безопасный доступ к ресурсам на локальном компьютере по сравнению с традиционными приложениями Windows. В Windows Forms форма – это визуальная поверхность, на которой выводится информация для пользователя [7]. Приложение Windows Forms строится путем помещения элементов управления на форму и написания кода для реагирования на действия пользователя, такие как щелчки мыши или нажатия клавиш. Элемент управления – это отдельный элемент пользовательского интерфейса, предназначенный для отображения или ввода данных.

Выбранная технология обладает достаточным набором инструментов для осуществления процедуры осаждения/извлечения информации в стеганографический контейнер формата EPUB [7].

В процессе внедрения сообщения M_i контейнер **C** проходит через несколько этапов обработки, число которых совпадает с количеством используемых в модели ключей **K**. Для грамотного построения процедуры осаждения был применен шаблон проектирования «Chain of responsibility». Цепочка обязанностей («Chain of responsibility») – поведенческий шаблон проектирования, благодаря которому удается избежать «жесткой» привязки отправителя запроса к получателю, позволяя тем самым нескольким объектам обработать запрос. Все возможные обработчики запроса образуют цепочку, а сам запрос перемещается по этой цепочке, пока один из ее объектов не обработает запрос. Каждый объект при получении запроса выбирает, либо обработать запрос, либо передать выполнение запроса следующему по цепочке. Данный шаблон применяется в следующих случаях: наличие более одного объекта, который может обработать определенный запрос; необходимость передачи запроса на выполнение одному из нескольких объектов, точно не определяя, какому именно; необходимость задания объектов динамически. Исходя из описания шаблона «Chain of responsibility», можно сказать, что его использование в разработке приложения является целесообразным. UML-диаграмма паттерна «Chain of responsibility» представлена на рис. 2.

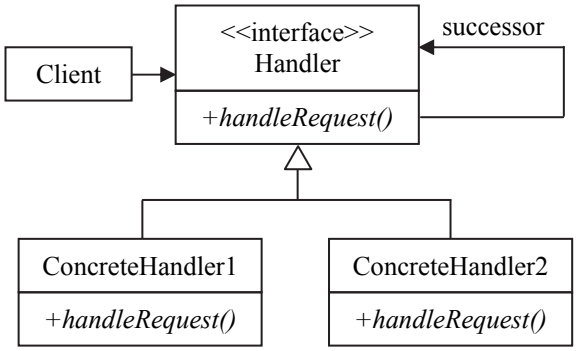


Рис. 2. UML-диаграмма паттерна «Chain of responsibility»

Handler определяет интерфейс для обработки запроса, а также может определять ссылку на следующий обработчик запроса. *ConcreteHandler1* и *ConcreteHandler2* – конкретные обработчики, которые реализуют функционал для обработки запроса. *Client* отправляет запрос объекту *Handler*.

В приложении «EPUB Modifier» описанный паттерн реализован на основе двух классов *Embedder* и *Extracter*. В классе *Embedder* содержатся свойства для хранения следующего звена цепочки, обрабатываемой книги, внедряемого сообщения и для директории, в которой находится файл-контейнер. Основным методом является *EmbedMessage*, который выполняет переопределенный метод *Embed* класса потомка, а также запускает выполнение следующего звена цепочки, если оно есть. Для добавления класса в цепочку осаждения сообщения необходимо реализовать класс *Embedder*, а также переопределить абстрактный метод *Embed*. Далее следует разместить объект класса в свойстве *Next* предыдущего звена цепочки.

Реализация родительского класса *Embedder* представлена в листинге 2.

```
abstract class Embedder{
    public Embedder Next;
    public EpubBook Book;
    public Dictionary<string, string> files
        = new Dictionary<string, string>();
    public string ContentDirectoryPath;
    public string embedMessage;
    3 references
    public Embedder(EpubBook book, string message){
        Book = book;
        ContentDirectoryPath = book.Schema.ContentDirectoryPath;
        embedMessage = message;}
    2 references
    public void EmbedMessage() {
        Embed();
        if (Next != null) Next.EmbedMessage();}
    4 references
    public abstract void Embed();
}
```

Листинг 2. Программный код класса *Embedder*

Использование цепочки обязанностей дает следующие преимущества: ослабление связан-

ности между объектами (отправителю и получателю запроса ничего неизвестно друг о друге; клиенту неизвестна цепочка объектов, какие именно объекты составляют ее, как запрос в ней передается); в цепочку можно добавлять новые типы объектов, которые реализуют общий интерфейс; расположение последовательности объектов-обработчиков в цепочке в зависимости от их приоритета.

Заключение. Рассмотренная в данной статье модель основана на использовании комбинации стеганографических методов. В список методов входят LSB, метод закодированного изображения в каскадных таблицах стилей, а также метод замены кавычек в файлах разметки. Применение перечисленного сочетания методов позволяет увеличить стеганографическую стойкость системы и добавить возможность проверки осажденного сообщения в цифровом файле формата EPUB.

Представленная модель стеганографической системы позволяет проводить процедуру внедрения сообщения в электронные книги формата EPUB с учетом особенностей содержимого файла-архива. Система на основе рассмотренной модели может быть применена для внесения цифрового водяного знака в электронные книги с целью защиты авторского права на интеллектуальную собственность и подтверждения целостности документа, а также для размещения различных скрытых стеганографических меток в каждую копию электронной книги, для выявления канала несанкционированного копирования и распространения.

В сравнении с системой, в которой в качестве контейнера используется только метод замены кавычек и закодированного изображения в каскадных таблицах стилей, описанная система обладает более надежной стеганографической стойкостью за счет применения дополнительного контейнера для дублирования сообщения, а также отдельного контейнера для хранения контрольной суммы исходного сообщения.

Представлено программное средство, позволяющее производить процедуру осаждения/извлечения сообщения в стеганографический контейнер формата EPUB.

Литература

1. Урбанович П. П. Защита информации методами криптографии, стеганографии и обфускации. Минск: БГТУ, 2016. 220 с.

2. Конахович Г. Ф., Пузыренко А. Ю. Компьютерная стеганография. Теория и практика. Киев: МК-Пресс, 2006. 288 с.

3. Сущеня А. А. Программное средство стеганографического преобразования текстов-контейнеров на основе языка разметки XML // 69-я науч.-техн. конф. учащихся, студентов и магистрантов. Минск, 2–13 апр. 2018 г. / Белорус. гос. технол. ун-т. Минск, 2018. С. 81–84.

4. Шутько Н. П., Романенко Д. М., Урбанович П. П. Математическая модель системы текстовой стеганографии на основе модификации пространственных и цветовых параметров символов текста // Труды БГТУ. 2015. № 6: Физ.-мат. науки и информатика. С. 152–156.

5. Сущеня А. А. Стеганографический метод внедрения текстовой информации в контейнер формата EPUB // 70-я науч.-техн. конф. учащихся, студентов и магистрантов. Минск, 15–20 апр. 2019 г. / Белорус. гос. технол. ун-т. Минск, 2019. С. 247–251.

6. Шутько Н. П. Особенности и формальное описание процесса осаждения секретной информации в текстовые документы на основе стеганографии // Труды БГТУ. 2014. № 6: Физ.-мат. науки и информатика. С. 121–124.

7. Docs.microsoft.com: сайт. URL: https://docs.microsoft.com/ru-ru/dotnet/framework/winforms/windows-forms-overview (дата обращения: 02.11.2019).

References

1. Urbanovich P. P. *Zashchita informatsii metodami kriptografii, steganografii i obfuskatsii* [The protection of information based on the methods by cryptography, steganography and obfuscation]. Minsk, BGTU Publ., 2016. 220 p.

2. Konakhovich G. F., Puzyrenko A. Yu. *Komp'yuternaya steganografiya. Teoriya i praktika* [Computer steganography. Theory and practice]. Kyiv, MK-Press Publ., 2006. 288 p.

3. Sushchenia A. A. Software tool for steganographic transformation of container texts based on XML markup language. *69-ya nauchno-tekhnicheskaya konferentsiya uchashchikhsya, studentov i magistrantov* [69th Scientific and technical conference of students, students and undergraduates]. Minsk, 2018, pp. 81–84 (In Russian).

4. Shutko N. P., Romanenko D. M., Urbanovich P. P. Mathematical model of textual shorthand system based on modification of spatial and color parameters of text symbols. *Trudy BGTU* [Proceedings of BSTU], 2015, no. 6: Physics and Mathematics. Informatics, pp. 152–156 (In Russian).

5. Sushchenia A. A. Steganographic method of introduction of textual information into the container of EPUB format. *70-ya nauchno-tekhnicheskaya konferentsiya uchashchikhsya, studentov i magistrantov* [70th Scientific and technical conference of students, students and undergraduates]. Minsk, 2019, pp. 247–251 (In Russian).

6. Shutko N. P. Features and formal description of the process of deposition of secret information in text documents based on shorthand. *Trudy BGTU* [Proceedings of BSTU], 2014, no. 6: Physics and Mathematics. Informatics, pp. 121–124 (In Russian).

7. Docs.microsoft.com. Available at: https://docs.microsoft.com/ru-ru/dotnet/framework/winforms/windows-forms-overview (accessed 02.11.2019).

Информация об авторах

Сущеня Артем Александрович – магистрант. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: asuschenya@gmail.com

Блинова Евгения Александровна – старший преподаватель кафедры информационных систем и технологий. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: evgenia.blinova@belstu.by

Information about the authors

Sushchenia Artsiom Aleksandrovich – Master's degree student. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: asuschenya@gmail.com

Blinova Evgeniya Aleksandrovna – Senior Lecturer, the Department of Information Systems and Technology. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: evgenia.blinova@belstu.by

Поступила после доработки 13.11.2019

СИСТЕМНЫЙ АНАЛИЗ И ОБУЧАЮЩИЕ СИСТЕМЫ

УДК 512.8:681

О. В. Герман[1], Ю. О. Герман[2], М. В. Кузнецов[2]
[1]Белорусский государственный технологический университет
[2]Белорусский государственный университет информатики и радиоэлектроники

ПОДХОД К ВЫБОРУ УПРАВЛЕНИЯ В СИСТЕМЕ КЛАСТЕРОВ

Рассматривается техника управления сложной системой на основе кластеров с «серой областью». «Серая область» определяет множество состояний системы, которые не удается однозначно отнести к тому или иному четкому классу. Первая задача, которая решается в настоящей работе, связана с определением «серой области». Для этой цели формулируем задачу линейного программирования, вводя в неравенства, ограничивающие область допустимых значений, дополнительные переменные, моделирующие размытость «границ» кластеров. Когда «серая область» определена, возникает вторая задача: принятия управления для каждого состояния, попадающего в «серую область». Для решения этой задачи рассматриваем подход, связанный с построением кластеров в «серой области». Наблюдаемый вектор переменных состояния «прогоняется» по кластерам. Используется байесовская функция штрафа за неправильное отнесение вектора состояния к кластерам. Этот подход более точен при значительном размере кластеров, когда их статистические характеристики становятся достаточно устойчивыми. В статье предлагается вариант замены вероятностей принадлежности к кластерам нечеткой мерой принадлежности и демонстрируется техника ее вычисления. Описанный в статье метод можно применить при выборе управления в производственных, финансовых и иных организациях, базируясь только на обучающих таблицах наблюдений за определенный период времени в прошлом.

Ключевые слова: система управления, кластер, функция штрафа, байесовская стратегия распознавания.

O. V. German[1], Yu. O. German[2], M. V. Kuznetsov[2]
[1]Belarusian State Technological University
[2]Belarusian State University of Informatics and Radioelectronics

AN APPROACH TO CONTROL DEFINITION IN THE SYSTEM OF CLUSTERS

An approach to control definition in complex system on the basis of clusters with “gray area” is represented. “Gray area” defines a set of the system states which cannot be assigned to strict clusters in determinate way. The first problem solved in the paper is connected to a definition of the “gray area”. For this purpose, we state a linear programming problem by putting into inequalities which define the set of valid solutions new variables modeling fuzziness of the clusters boundaries. After the “gray area” is defined, the second problem arises connected to control definition within the boundaries of the “gray area”. To solve it, an approach is considered connected to building clusters in “gray area” and usage of penalty function for incorrect classifying. The observed state vector of the regulated system traverses the clusters. For each cluster the Bayesian technique of penalty evaluation for incorrect assigning the state vector to it is realized. This technique provides more exact results in the case of clusters with big sizes when their statistical properties become stable. It is proposed some way of substituition of the probabilities of the stets vectors belonging to clusters by the adequately computed fuzzy measures with the corresponding technique outlined. The general approach described in the paper may be used for control definition in manufacturing, financial and other organizations with the help of learning tables containing only data of the system behavior for some time period in the past.

Key words: control system, cluster, penalty function, Bayesian recognition strategy.

Введение. В статье излагается техника принятия управленческих решений, использующая обучающие таблицы, на которых определяется множество кластеров состояний. При этом трудности связаны с так называемой «серой областью», в которой перемешаны объекты из

разных кластеров. Выделяем в общем случае четыре кластера состояний: *N* (нормальный кластер, определяющий множество штатных состояний), *F* (финальный кластер проигрышных состояний), *PS* (кластер отклоняющихся от нормы состояний, которые идентифицируются как позитивные, более близкие к *N*) и *NG* (кластер отклоняющихся от нормы состояний, которые идентифицируются как негативные, более близкие к *F*). В качестве примера можно взять любую производственную структуру или финансовую организацию, состояния которой можно связать с критериями пятифакторной модели Альтмана [1]:

$$z = 1{,}2x_1 + 1{,}4x_2 + 3{,}3x_3 + 0{,}6x_4 + x_5, \qquad (1)$$

где x_1 – отношение размера оборотного капитала к стоимости активов; x_2 – отношение величины чистой прибыли к стоимости активов; x_3 – отношение размера прибыли до налогообложения к стоимости активов; x_4 – отношение рыночной стоимости акций предприятия к суммарной величине обязательств; x_5 – отношение объема выручки от продаж к стоимости активов.

При $z > 2{,}9$ попадаем в кластер *N*, при $z < 1{,}8$ – в кластер *F*. В остальных случаях попадаем в объединенный кластер $PS \cup NG$ («серую область»). Нас будет интересовать принятие управляющих воздействий в «серой области».

В литературе [2–4] представлены различные подходы к нечеткой кластеризации. Здесь мы пытаемся объединить две различные стратегии: механизм классифицирующих деревьев и байесовскую стратегию минимизации функции штрафа за неверную классификацию [5–7].

Основная часть. Отправной позицией нашего подхода является таблица наблюдений за динамикой сложной системы (например, финансовой, производственной, транспортной и т. п.). Система описывается вектором параметров $V(t)$, состояния векторов изменяются в результате реализуемых управлений $U(V, t)$, так что поведение системы описывается достаточно сложной траекторией. Для учета динамики можно воспользоваться наблюдениями по разным фирмам за достаточно длительный период времени. Такие данные можно описывать многомерными матрицами $V[t, n, Z_t]$, где $t = 0, 1, 2, ..., T$ задает моменты наблюдения (дискретные величины), *n* определяет значения наблюдаемых критериев. В качестве исходных данных будем использовать обучающую табл. 1 (знак «?» означает неизвестное значение).

В табл. 1 указаны только два управляющих воздействия: Y_1 и Y_2 (соответственно для классов *A* и *B*).

Таблица 1

Исходные данные

№ п/п	x_1	x_2	Класс	Управление
1	1	6	*A*	$Y_1[1{,}0]$
2	2	3	*A*	$Y_1[0{,}8]$
3	2	5	*A*	$Y_1[1{,}0]$
4	3	2	*A*	$Y_1[?]$
5	4	4	*A*	$Y_1[?]$
6	4	5	*A*	$Y_1[0{,}6]$
7	3	1	*B*	$Y_2[1{,}0]$
8	3	6	*B*	$Y_2[?]$
9	3	3	*B*	$Y_2[?]$
10	3	4	*B*	$Y_2[0{,}8]$
11	4	1	*B*	$Y_2[1{,}0]$
12	5	2	*B*	$Y_2[1{,}0]$
13	4	3	*B*	$Y_2[?]$
14	5	6	*B*	$Y_2[0{,}6]$

Ищем дискриминаторную функцию в виде $f_1(x_1, x_2) = a_0 + a_1x_1 + a_2x_2$. Базируясь на идеях Zimmermann [4], составим задачу линейного нечеткого программирования по данным табл. 1:

$$\begin{gathered}
\lambda_1 + \lambda_2 + ... + \lambda_{10} \to \min, \\
a_0 + a_1 + 6a_2 \ge 0, \\
a_0 + 2a_1 + 3a_2 \ge 0 - \lambda_1, \\
a_0 + 2a_1 + 5a_2 \ge 0, \\
a_0 + 3a_1 + 2a_2 \ge 0 - \lambda_2, \\
a_0 + 4a_1 + 4a_2 \ge 0 - \lambda_3, \\
a_0 + 4a_1 + 5a_2 \ge 0 - \lambda_4, \\
a_0 + 3a_1 + a_2 \le -0{,}1, \\
a_0 + 3a_1 + 6a_2 \le -0{,}1 + \lambda_5, \\
a_0 + 3a_1 + 3a_2 \le -0{,}1 + \lambda_6, \\
a_0 + 3a_1 + 4a_2 \le -0{,}1 + \lambda_7, \\
a_0 + 4a_1 + a_2 \le -0{,}1, \\
a_0 + 5a_1 + 2a_2 \le -0{,}1, \\
a_0 + 4a_1 + 3a_2 \le -0{,}1 + \lambda_8, \\
a_0 + 5a_1 + 6a_2 \le -0{,}1 + \lambda_9, \\
\lambda_1, \lambda_2, ..., \lambda_9 \ge 0.
\end{gathered} \qquad (2)$$

Неравенства получаем так: подставляем значения x_1, x_2 из каждой строки таблицы в $a_0 + a_1x_1 + a_2x_2$. Если объект из класса *A*, то правая часть неравенства ≥ 0. Если объект из класса *B*, то правая часть неравенства ≤ –0,1 (число –0,1 взято условно как достаточно малая величина). Строгий выбор ограничивается точностью вычислений на компьютере). В тех строках, где значение управления точно не 1,0, вводим компенсирующую неотрицательную величину λ_i. Решим систему, например, в Excel. Получим следующие значения:

$$\begin{gathered}
a_0 = -0{,}167;\ a_1 = 0;\ a_2 = 0{,}033;\ \lambda_1 = 0{,}067; \\
\lambda_2 = 0{,}1;\ \lambda_3 = 0{,}033;\ \lambda_4 = 0;\ \lambda_5 = 0{,}3; \\
\lambda_6 = 0{,}033;\ \lambda_7 = 0{,}067;\ \lambda_8 = 0{,}033; \\
\lambda_9 = 0{,}133.
\end{gathered}$$

Видим, что «серая область» есть, и нам следует ее конкретизировать. При этом оказалось, что точка (4, 5) ($\lambda_4 = 0$) строго принадлежит классу A (не является «серой»). Такие точки следует пересмотреть на предмет их строгой принадлежности к соответствующему классу. Мы поступим именно таким образом. Также может получиться, что система окажется несовместной. Это связано с неправильным определением строгого включения точек (с мерой, равной 1,0) в соответствующий класс. В табл. 1 такие точки определены под номерами 1, 3, 7, 11, 12. Практически выяснить, какие точки с мерой, равной 1,0, должны быть отнесены к «серой области», можно также по схеме Zimmermann, т. е. составить задачу типа (2) для этих точек, введя в каждом неравенстве компенсирующую переменную и устремив их сумму в целевой функции к минимуму. Итак, полагаем, что «серая область» определена. В этой области уже нетрудно найти кластеры, объединяющие объекты из одного и того же класса. В примере таких классов два. Для выбора одного из этих кластеров в «серой области» воспользуемся оценкой величины штрафа за неправильную классификацию по Байесу. Разумеется, можно исходить из того, что каждый кластер описывается как множество реализаций многомерной случайной величины. Тогда можно говорить о задаче отыскания вероятности принадлежности к кластеру. Будем использовать функцию штрафа за неправильную классификацию:

$$F_i = \sum C_{ik} P(k \mid V), \qquad (3)$$

$$F_i = \sum C_{ik} P_k P(V \mid k). \qquad (4)$$

Здесь F_i – это штраф за отнесение объекта V к кластеру i; P_k – вероятность выбора кластера k; $P(V \mid k)$ – условная вероятность появления объекта V в кластере k. На практике полагают, что $C_{ik} = 1$, если $i <> k$, и $C_{ik} = 0$ в противном случае. Для определения вероятностей P_k мы должны располагать временной серией измерений векторов состояний системы. Будем использовать табл. 2 наблюдений о состоянии системы за некоторый период времени.

Таблица 2

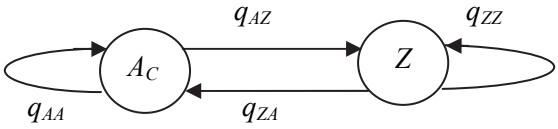
Исходные данные и временная серия наблюдений

№ п/п	x_1	x_2	Класс	Управление
1	1	6	A_C	Y_1[1,0]
2	2	3	A_C	Y_1[0,8]
3	2	5	A_C	Y_1[1,0]
4	3	2	Z	Y_1[0,73]
5	4	4	A_C	Y_1[0,9]

Окончание табл. 2

№ п/п	x_1	x_2	Класс	Управление
6	4	5	Z	Y_1[0,6]
7	3	1	Z	Y_2[1,0]
8	3	6	Z	Y_2[0,3]
9	3	3	Z	Y_2[0,9]
10	3	4	Z	Y_2[0,8]
11	4	1	Z	Y_2[1,0]
12	5	2	Z	Y_2[1,0]
13	4	3	Z	Y_2[0,9]
14	5	6	Z	Y_2[0,6]

Здесь выделены два класса: A_C, где применяется управление Y_1 с мерой, определенной на подынтервале $B(0{,}77;\ 1{,}0]$, и класс Z, который относится ко всем остальным случаям. Номера наблюдений соответствуют временной упорядоченности векторов состояний. Предполагается, что интервал между каждой парой соседних наблюдений один и тот же. Обратимся к формуле (4). Для оценки P_k воспользуемся иллюстрацией системы переходов между классами, показанной на рисунке.



Переходы между классами

Обозначим стационарную вероятность состояния A_C через P_{AC}, а стационарную вероятность состояния Z через P_Z. Имеем систему уравнений:

$$\begin{aligned} P_{AC} + P_Z &= 1, \\ P_{AC} &= P_{AC}\, q_{AA} + P_Z\, q_{ZA}, \\ P_Z &= P_Z q_{ZZ} + P_{AC}\, q_{AZ}. \end{aligned} \qquad (5)$$

Последнее уравнение в (5) избыточно. Вероятности q_{AA}, q_{AZ}, q_{ZZ}, q_{ZA} отыскиваем непосредственно из табл. 2. Имеем:

$$q_{AA} = 0{,}5,\ q_{AZ} = 0{,}5,\ q_{ZZ} \approx 0{,}9,\ q_{ZA} \approx 0{,}1.$$

Отсюда

$$P_{AC} \approx 0{,}17,\ P_Z \approx 0{,}83.$$

Иначе обстоит дело с $P(V \mid k)$. Вместо $P(V \mid k)$ будем использовать нечеткие величины принадлежности объектов к кластерам. Для вычисления этих величин снова обратимся к табл. 2. Для каждого вектора определен класс и соответствующее управление.

Пусть наблюдаем некоторый объект r с координатами x_{1r}, x_{2r} = <2, 6>. Определим евклидово расстояние этого вектора до кластеров A_C, Z. Обозначим эти расстояния соответственно $d_{r,A}$ и $d_{r,Z}$. Найдем в каждом классе типичного

представителя (его координаты вычисляются как средние значения по кластеру). Так, в кластере A_C типичный представитель имеет координаты <2,25; 4,25>, а в кластере Z – <3,7; 3,3>. Обозначим нечеткую меру принадлежности объекта r к кластеру A_C через $\mu(r \mid A_C)$. Тогда

$$\mu(r \mid Z) = 1 - \mu(r \mid A_C). \quad (6)$$

Согласно нашему допущению,

$$\mu(r \mid Z) / \mu(r \mid A_C) = d_{r,A} / d_{r,Z}. \quad (7)$$

Евклидово расстояние находим обычным способом, т. е. как корень из суммы квадратов покоординатных разностей. Имеем:

$$d_{r,A} = 1{,}76, \; d_{r,Z} = 3{,}19. \quad (8)$$

Из (6)–(8) сразу находим:

$$\mu(r \mid A_C) = 0{,}64, \; \mu(r \mid Z) = 0{,}36. \quad (9)$$

Теперь на основании (4) получаем:

$$F_{AC} = 0{,}29, \; F_Z = 0{,}108.$$

Таким образом, наблюдаемый вектор следует отнести к классу Z, поскольку в этом случае штраф будет меньшим.

Теперь необходимо решить вопрос о выборе управления в классе Z. Из табл. 2 следует, что нужно отыскать нечеткое управление из двух кандидатов: Y_1 и Y_2. В классе Z состояния не упорядочены в общем случае по равноотстоящим интервалам наблюдений. Класс Z разбивается на два кластера: I, II. В первом принимается управление Y_1, во втором – Y_2. Будем использовать формулу (4). Вероятности кластеров пропорциональны числу их появлений в табл. 2. Имеем $P(\mathrm{I}) = 0{,}2$, $P(\mathrm{II}) = 0{,}8$. Как и выше, вероятность принадлежности наблюдаемого вектора к кластеру оценим на основе евклидова расстояния до кластеров. Типичный представитель кластера I имеет значение <3,5; 3,5>, а типичный представитель кластера II – <3,75; 3,25>. На основании условия

$$\mu(r \mid \mathrm{I}) / \mu(r \mid \mathrm{II}) = d_{r,\mathrm{II}} / d_{r,\mathrm{I}}$$

находим

$$\mu(r \mid \mathrm{I}) = 0{,}48, \; \mu(r \mid \mathrm{II}) = 0{,}52.$$

Из (4) получаем значения функции штрафа:

$$F_{\mathrm{I}} = 0{,}42, \; F_{\mathrm{II}} = 0{,}096.$$

Наименьший штраф для второго кластера определяет управление Y_2.

Наконец, остается выполнить последний шаг – отыскать нечеткую меру для этого управления. Опишем два способа решения этой проблемы. Согласно первому способу, воспользуемся общей схемой метода k-средних. Примем, например, $k = 2$ и найдем два наиболее близких вектора к входному вектору r. Это векторы v = <3, 6> и w = <3, 4>. Оценим степень близости к ним вектора r, как делали выше. Имеем:

$$\mu(r \mid v) = 0{,}7, \; \mu(r \mid w) = 0{,}3.$$

Взвешенная по этим оценкам мера управления Y_2 на векторе r составит:

$$(0{,}7 \cdot 0{,}3 + 0{,}3 \cdot 0{,}8) / (0{,}7 + 0{,}3) = 0{,}45.$$

Итак, для вектора состояния r = <2, 6> следует применить управление Y_2 с мерой 0,45 (содержательно: ослабив величину этого воздействия почти в 2 раза).

Второй способ базируется на технике интерполирования многомерных данных. Проиллюстрируем его суть на примере. Для этих целей мы поступим следующим образом. Пусть дана табл. 3 со значением управления y на известных состояниях <x_1, x_2>.

Таблица 3

Исходные данные для многомерного интерполирования

№ п/п	x_1	x_2	y
1	1	1	2
2	1	2	5
3	1	3	10
4	2	1	5
5	2	2	8
6	2	3	13
7	3	2	13
8	3	3	18

Нормируем данные в столбцах табл. 3. С этой целью заменим значения в столбцах, используя формулу

$$x_i^* = (x_i - \min_{x1}) / (\max_{xi} - \min_{xi}).$$

Здесь применены максимальное и минимальное значения в соответствующем столбце ($\max_{xi}$, $\min_{xi}$). Результат представлен в табл. 4.

Таблица 4

Нормированные значения данных

№ п/п	x_1^*	x_2^*	y
1	0	0	2
2	0	0,5	5
3	0	1	10
4	0,5	0	5
5	0,5	0,5	8
6	0,5	1	13
7	1	0,5	13
8	1	1	18

Переходим от многомерных данных к скалярам, например, используя формулу

$$z = x_1^2 + 2x_2^2 + x_1 + 3x_2.$$

Преобразованные данные сведем в табл. 5.

Таблица 5

Преобразованные данные для интерполяции

№ п/п	z	y
1	0	2
2	2	5
3	5	10
4	0,75	5
5	2,75	8
6	5,75	13
7	4	13
8	7	18

Разместим значения z по неубыванию (табл. 6).

Таблица 6

Упорядоченные данные

№ п/п	z	y
1	0	2
2	0,75	5
3	2	5
4	2,75	8
5	4	13
6	5	10
7	5,75	13
8	7	18

Теперь мы можем рассчитать коэффициенты интерполяционного полинома и найти значение этого полинома в промежуточных точках интервала. Выбор управления в любой промежуточной точке интервала реализуется прямым вычислением на основе интерполяционного многочлена. Воспользуемся языком Python и реализуем следующий скрипт:

```
import numpy as np
from scipy import interpolate
from scipy.interpolate import interp1d
import matplotlib.pyplot as plt
x = np.array([0, 0.75, 2, 2.75, 4, 5, 5.75, 7])
y = np.array([2,5,5,8,13,10,13,18])
print np.interp(0.31, x, y)
```

Рассчитаем значение интерполяционной функции для $x_1 = 1{,}5$, $x_2 = 1$ (значения взяты в качестве иллюстрации). Этим значениям соответствует $z = 0{,}31$. Значение функции интерполяции равно 3,24. Оно и определяет управляющее воздействие на данном входе.

Заключение. Используя подход Zimmermann, адаптированный в этой работе для нахождения объектов «серой области», всегда можно получить ответ на вопрос, лежит ли предъявляемый объект в «серой области» или нет. При положительном ответе на этот вопрос следует выполнить процедуру классификации на кластерах «серой области». В одних случаях множество кластеров определено заранее (например, известен тип реализуемого в каждом кластере управления, которое однозначно соответствует кластеру). В других случаях необходимо сформировать кластерную структуру «серой области» (используя, например, метод k-средних). Для определения того, к какому кластеру принадлежит заданный многомерный объект, использован подход, в основе которого лежит функция штрафа за неправильную классификацию объекта. В этой работе нами показано, как данную функцию вычислять, базируясь не на вероятностях, а на устанавливаемых нечетких мерах принадлежности входного объекта к соответствующему кластеру. В сравнении с нейронными сетями [8] описанный подход не требует длительного обучения и не критичен к размерности обучающей таблицы. Заметим, что вероятностный подход также чувствителен к объему используемых опытных данных и, кроме того, требует установления закона и параметров распределения характеристик многомерных объектов с учетом их взаимосвязи. Предложенный в статье подход устраняет отмеченные трудности, объединяя байесовский метод минимизации риска неправильной классификации и нечеткую меру принадлежности объектов к кластерам, устанавливаемую на основе евклидовой метрики.

Литература

1. Altman I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy // Journal of Finance. 1968. P. 189–209.

2. Mamdani E. H., Assilian S. An experiment in linguistic synthesis with a fuzzy logic controller // Int. J. Man Mach. Stud. 1975. No. 7. P. 1–13.

3. Fuzzy Logic Toolbox. User's Guide // The MathWorks Inc. 1995. P. 962.

4. Fuller R., Zimmermann H.-J. Fuzzy reasoning for solving fuzzy mathematical programming problems // Fuzzy Sets and Systems. 1993. No. 60. P. 121–133.

5. Jin L., Xin F., Xu Y. A method of multi-attribute decision making under uncertainty using evidental reasoning and prospect theory // International Journal of Computational Intelligence Systems. 2015. Vol. 8. P. 48–62.

6. Classification and regression trees / L. Breiman [et al.]. Wadsworth: Belmont CA, 1984. 368 p.

7. An application of ID3 Decision Tree Algorithm in land capability classification / N. Kumar [et al.] // Agropedology. 2012. Vol. 22, no. 1. P. 35–42.

8. Горбань А. Н. Обучение нейронных сетей. М.: Параграф, 1990. 160 с.

References

1. Altman I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance*, 1968, pp. 189–209.

2. Mamdani E. H., Assilian S. An experiment in linguistic synthesis with a fuzzy logic controller. *Int. J. Man Mach. Stud.*, 1975, no. 7, pp. 1–13.

3. Fuzzy Logic Toolbox. User's Guide. *The MathWorks Inc*., 1995, p. 962.

4. Fuller R., Zimmermann H.-J. Fuzzy reasoning for solving fuzzy mathematical programming problems. *Fuzzy Sets and Systems*, 1993, no. 60, pp. 121–133.

5. Jin L., Xin F., Xu Y. A method of multi-attribute decision making under uncertainty using evidental reasoning and prospect theory. *International Journal of Computational Intelligence Systems*, 2015, vol. 8, pp. 48–62.

6. Breiman L., Freidman J. H., Olshen R. A., Stone C. J. Classification and regression trees. Wadsworth, Belmont CA, 1984. 368 p.

7. Kumar N., Obi Reddy G., Chatterji S., Sapkar D. An application of ID3 Decision Tree Algorithm in land capability classification. *Agropedology*, 2012, vol. 22, no. 1, pp. 35–42.

8. Gorban A. N. *Obucheniye neyronnykh setey* [Neuronets Learning]. Moscow, Paragraf Publ., 1990. 160 p.

Информация об авторах

Герман Олег Витольдович – кандидат технических наук, доцент кафедры информационных систем и технологий. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: ovgerman@tut.by

Герман Юлия Олеговна – кандидат технических наук, доцент кафедры программного обеспечения информационных технологий. Белорусский государственный университет информатики и радиоэлектроники (220600, г. Минск, ул. П. Бровки, 6, Республика Беларусь). E-mail: juliagerman@tut.by

Кузнецов Михаил Владимирович – аспирант. Белорусский государственный университет информатики и радиоэлектроники (220600, г. Минск, ул. П. Бровки, 6, Республика Беларусь). E-mail: mishaky@mail.ru

Information about the authors

German Oleg Vitol'dovich – PhD (Engineering), Assistant Professor, the Department of Information Systems and Technology. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: ovgerman@tut.by

German Yulia Olegovna – PhD (Engineering), Assistant Professor, the Department of Software of Information Technology Systems. Belarusian State University of Informatics and Radioelectronics (6, P. Brovki str., 220600, Minsk, Republic of Belarus). E-mail: juliagerman@tut.by

Kuznetsov Michail Vladimirovich – PhD student. Belarusian State University of Informatics and Radioelectronics (6, P. Brovki str., 220600, Minsk, Republic of Belarus). E-mail: mishaky@mail.ru

Поступила после доработки 16.12.2019

УДК 316.776

G. Jaber, N. V. Patsei, F. Rahal
Belarusian State Technological University

SEMANTIC INFORMATION-CENTRIC NETWORKING NAMING SCHEMA

The article describes a new semantic-base naming schemer. This proposal takes into consideration the problem of data communication types that traverse the ICN. The legacy proposals in ICN have weaknesses in dealing with some type of communication. In order to deal with this problem, a three-dimension addressing scheme was presented. It includes Geographical, Semantic, and Publisher ID addresses. The article discusses the process of forming a Semantic address on the basis of Network Universal Language with the construction of a semantic graph. We used the IPv6 extension header to define a new routing scheme that can work with a three-dimension address. In conclusion, the routing scheme and tables are briefly described. As a result, the proposed scheme will evolve the interests of Subscribers to a higher abstract level and will reduce the name resolution brokers and delays in some cases.

Key words: routing, information-centric networks, semantic, geographical, address, publisher, subscriber request, IPv6.

Introduction. Information-Centric Networking (ICN) or its other names including Data-Oriented Networking, Content-Based Networking or Content-Centric Networking/Named Data Networking, is a substitute paradigm for the present architecture of the Internet that focuses on naming data for its model of communication [1]. There are some problems in the present architecture of internet for which the ICN is able to find resolutions. The problems include ineffective use of resources, Distributed Denial of Service (DDoS) attacks, lack of security, and problems in the fields of mobility, scalability, routing protocol as well as economic problems [2].

The routing protocol defines the manner of communication between network routers. This protocol sends required information to routers and enables them to select possible routes between two existing nodes in the network. On the other hand, routing algorithms are responsible to make decision about the appropriate selection of the route [3]. Each router is equipped with the knowledge of specific networks with direct connections to it. The routing protocol distributes this information to adjacent neighbors in the first place, and to the whole network in the second place. That is how the routers gain knowledge about the topology of the network [4]. The routing approach can be considered the heart of any ICN architecture, in this regard, each ICN routing protocol tries to find one or more copies of the distributed information within the network [5]. There are different routing protocols offered in different ICN architectures, from which name resolution and data routing are the most common protocols.

There are two roles defined for routers in the ICN architectures at the time of a request for a particular Named Data Object (NDO). The first task of the routers is finding a node that has a copy of the required piece of information, and forwarding a request the node. The second task is finding a route from the node to the user who had asked for the information piece. A method of doing these two tasks is called *name resolution*. This method includes finding one or more lower-layer locators for the name of NDO. These locators are able to call back the requested NDO. The other way to do the routing tasks is called *name-based routing*. In this method, the request for the NDO is directly routed to the node that has a copy of the content (based on the NDO's name). The name resolution phase in the name-based routing is removed [6, 7]. Fig. 1 displays the types of routing in the popular ICN architectures.

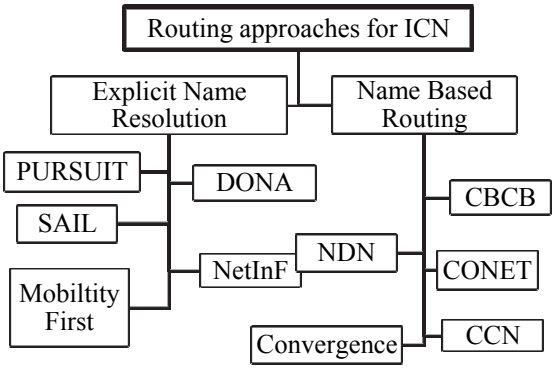


Fig. 1. Types of routing in different ICN architectures

Challenges in the name resolution and routing process are the following: ensured delivery and detection of the nearest copy of required content, scalability (it includes the development of an information model and a naming framework which support efficient information dissemination with improved security properties; it also includes the development of a world-wide scalable name resolution mechanism for a new namespace), excessive current on routing tables (if an overflow takes place, the router rejects request packets, the user experiences a low transmission rate, and the whole network will crash as a result), single point for failure (this problem happens when a great number of published and registered NDOs in the Name Resolution System (NRS) go unavailable), security and filtering.

Most of the proposed techniques in ICN are not suitable to deal with all data transmission types between Publishers/Subscribers. Another problem that resides in the proposed ICN schemas (fig. 1) is the limitation to deal with knowledge searching.

To solve these problems the types of data communication were examined and there was suggested the classification of data transmission into four types based on the number of subscriptions and frequency of data object use [8]. There are four scenarios for working with data, conditionally named A, B, C and D:

– type A: one subscriber – one use (voice call);

– type B: one subscriber – reusable (cloud storage);

– type C: several subscribers – one use (video streaming);

– type D: multiple subscribers – reusable (YouTube).

Besides, we also classified subscribers' requests into types. Subscribers' Requests may be of four types: *R1* – requesting any data content from a specific Publisher; *R2* – requesting specific data content from a specific Publisher; *R3* – requesting specific data content from a specific Publisher; *R4* – requesting information with any data content from any Publisher.

Theoretical base. Due to the high mobility of terminals in nowadays networks, the Publisher and Subscriber should hold a *Dynamic Address* that may be changed according to their geography in the network. In addition, the name should represent the content (an intuitive address) to serve Subscribers' requesting information (*R4* type) and should be unique to serve *R3* Subscribers. Thus, three dimensions for a naming scheme are proposed in a model of ICN network called Semantic Information-Centric Networking (SINC).

SINC naming scheme is based on the principle that the user (Publisher/Subscriber) should label the data with at least one dimension. The three dimensions (3D-address) are: Publisher ID, Semantic name, and Geographical ID.

Geographical address. A geographical address is a 128-bit unique address assigned by local host itself to route data towards a particular known location in the network based on the hierarchal structure. The IP address is an application on the geographical address that routes data from a source to a destination in a very flexible and fast way. This address is used here since it will facilitate routing towards the Publisher and the Subscriber taking into consideration the mobility of the Subscriber or the Publisher.

When the Subscriber moves from one sub network to another, his geographical address (IPv6) should change based on his new sub network (location/geography), so in proposed scheme we suggest to use EUI64 addressing technique to all mobile users (Publishers/Subscribers). This address allocation technique will allow each user in the network to have a unique address suffix due to the fact that the last part of EUI64 address is based on the MAC address of the user interface.

Considering a user interface with the following MAC address: 20-68-9D-94-77-1E moving to a subnet 2000**::**/64 will automatically assign the following IPv6 address: 2000**::**22:68:9D:FF:FE:94:77:1E/64.

Suppose that the user changes his sub network, it could be easily reached by his EUI64 suffix. A suggestion to reach this user is through the packet broadcast to all the nearest sub networks by changing the subnet address prefix part and fixing its EUI suffix address thus fixing the suffix and changing the prefix (table 1). This process will ensure the roaming of the users (Publishers or Subscribers) between subnets even in case of high mobility.

Table 1

Geographical Address Structure

Prefix: Subnet (mobile)	Suffix: Mac Address (fixed)
2000**::**/64	20-68-9D-94-77-1E
2000**::**22:68:9D:FF:FE:94:77:1E/64	

Publisher ID address. It is a set of addresses, built on the root or main unique address which is assigned by a central authority (Assigned Names and Numbers) ICANN. ICANN authorizes domain name registrars, through which domain names may be registered and reassigned. Publisher ID address is a 128-bit hierarchal address. This address is flat human friendly address that is readable by human (Domain Name Space). Each content within the Publisher can be addressed with other sub address that is assigned locally by the Publisher itself.

Let's take "BELSTU" as an example. It is a Publisher, that has a global unique 128-bit address assigned from ICANN. "BELSTU" will give each content (faculties and departments) it publishes a 128-bit sub address. This address is important to be used as in *R1* type request. Another example where this address shows high significance is the necessity to verify the publisher's ID. In case of *R1* and *R3* Subscriber's request (e.g. voice call, video call), a central agent (e.g. WhatsApp sever) should have a public address and manage the data transmission between two Subscribers.

Semantic address. A semantic address is formed on the basis of Universal Networking Language (UNL). UNL is a declarative formal language specifically designed to represent semantic data extracted from natural language texts [9]. The pivot paradigm is used: the representation of an utterance in the UNL interlingua is a hypergraph where normal nodes bear UWs (Universal Words) with semantic attributes (@a), and arcs bear semantic relations (R) as it shown on fig. 2 [10, 11].

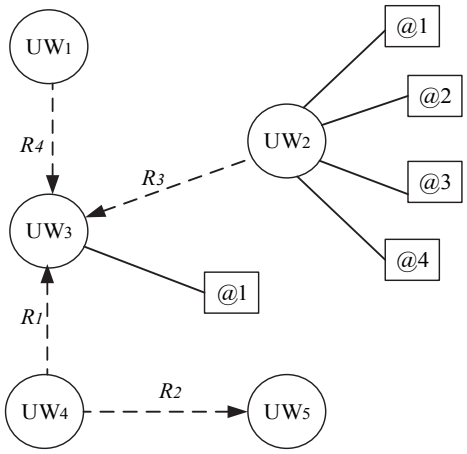


Fig. 2. UNL structure

The term "Universal Word" represents simple or compound concepts. There are three types of UWs: basic, restricted and extra UWs. A UW can have an UW ID. It is used to refer to some information and there are thirty-six UW-IDs (numbers from 0 to 9 and letters from A-Z).

Compound UWs represents a set of binary relations that are grouped together to express a concept. A sentence itself is considered a compound UW. For the graph fig. 2 the UNL representation is the following:

```
R1 (ID4: UW4: ———, ID3: UW3: @1 ———)
R2 (ID4: UW4: ———, ID5: UW5: ———)
R3 (ID2: UW2: @1@2@3@4, ID3: UW3: @1 ———)
R4 (ID1: UW1: ———, ID3: UW3: @1)
```

The sentence [BSTU held a conference] could be presented with UW as follows:

```
agt(held(icl>do), BSTU(icl>organization))
obj(held(icl>do), conference(icl>event))
```

Relations are binary connecting two UWs. There are forty labels that represent the relations between UWs in binary relation. They can be *ontological* (such as "icl" and "iof", referred to above), *logical* (such as "and" and "or"), and *thematic* (such as "agt" = agent, "ins" = instrument, "tim" = time, "plc" = place, etc.).

Attributes of UWs are used to describe subjectivity of sentences. Attributes represent information that cannot be conveyed by UWs and relations. UNL attributes shows view, aspect, time of event, etc. Normally, they represent information concerning *time* ("@past", "@future", etc.), *reference* ("@def", "@indef", etc.), *modality* ("@can", "@must", etc.), *focus* ("@topic", "@focus", etc.), and so on. There are 58 attributes in UNL [12].

For example:

```
agt(held(icl>do).@entry.@past, BSTU
(icl>organization))
obj(held(icl>do).@entry.@past,
conference(icl>event).@indef)
```

The attribute @entry denotes the main predicate of the sentence, @past – the present tense, and @indef – a non-specific class.

In proposed name scheme for SINC, UNL is adapted to create Semantic addresses.

In SICN name we assign for R (relation) 6 bits. 12 bits for the weight of the relation between two Universal Words (fig. 3). In SICN scheme we assign for each UW 31 bits, 6 bits for UW-ID and in each UW up to three attributes for each – 6 bits.

As you can see on fig. 3 for every relation we have 128 bits. So, a semantic address is a set of relations and descriptions of a semantic graph.

SICN Header Format. In accordance with added addresses, the IPv6 header structure has been redesigned (fig. 4). We left the fixed part of the header unchanged: *Version* (4 bits) indicates version of Internet Protocol; *Traffic Class* (8 bits) indicates class or priority of IPv6 packet, it helps routers to handle the traffic based on priority of the packet; *Flow Label* (20 bits) is used by source to label the packets belonging to the same flow; *Payload Length* (16 bits) – indicates total size of the payload which tells routers about the amount of information a particular packet contains in its payload; *Next Header* (8 bits) – indicates type of extension header; *Hop Limit* (8 bits) is same as TTL in IPv4 packets and indicates the maximum number of intermediate nodes IPv6 packet is allowed to travel; *Source Address* (128 bits) – an address of the original source of the packet; *Destination Address* (128 bits) – field indicates the IPv6 address of the final destination [13].

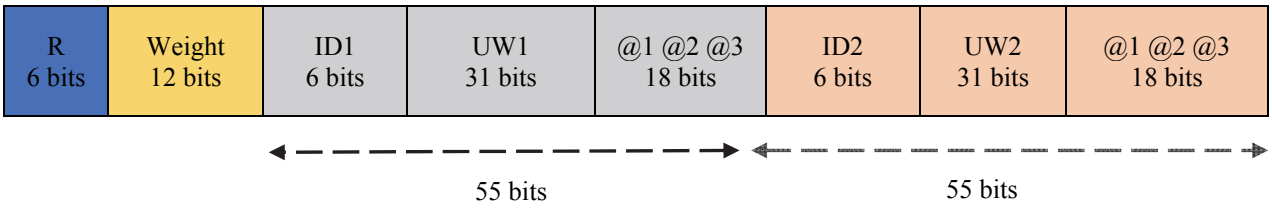


Fig. 3. Structure of UNL based semantic relation

Then we will use the *Extension Headers* for storage of Metadata and a three-dimension-naming scheme (3D-address): Geographical, Publisher ID and Semantic address.

Metadata Addressing fields are used for the address classification. 128 bit field divided into 12 parts of 10 bits each and with a remaining part of 8 bits as shown in fig. 4. Thus, each 10 bits part of the 12 parts in this field is classified into two sub parts where the mask part has 7 bits and the logic relation part has 3 bits.

Version (4 bits)	*Priority/Traffic class (8 bits)*	*Flow label (20 bits)*	
Payload length (16 bits)		*Header SICN (8 bits)*	*Hope limit (8 bits)*
Source address		*(128 bits)*	
Geographic Destination address		*(128 bits)*	
Metadata Addressing fields		*(128 bits)*	
Geographical addresses		*(128 bits)*	
Publisher ID addresses		*(128 bits)*	
Semantic addresses		*(128 bits)*	
...			

Fig. 4. SICN Header Format

Note that the number of Semantic addresses will be variable. The total number of addresses (Geographical, Publisher ID and Semantic) should be no more than 12. If you need to set the number of addresses to less than 12, it should complete the address set with zeroes, which indicates the end of the address list.

Routing schema. Conventional IP networks routing schemes works on network layer and do not take into consideration any aspect related to the content of the routed data. To reach a data destination, packets are labeled with Geographical Address (IPv6) that it is easy to reach by the help of routing tables and where these labels are learned dynamically by routing protocols or predefined statically by network administrators.

Literature proposed routing schemes [7] are based on Publisher/Subscriber scheme, which uses filters to match rendezvous points between Subscriber Interests, and Publisher advertisement compared to the conventional IP. These routing schemes work well with data type C and D where many Subscribers are interested in the published data. However, there is a shortage in dealing with the case of one Subscriber interested in the data from a specific Publisher (type A), i.e. when the Subscriber needs the data from a specified Publisher, whatever the data content. The proposed schemes in the literature would cost many network resources compared to the conventional IP network. Even in case of type B data if the Subscriber has a low reuse factor (frequency of usage), the conventional IP network will solve the problem in a better way. This fact is due to large amounts of filters that will be registered at the routers using the Publisher/Subscriber technique data. In addition, the latency time for the path, that is due to the path initiation required to match between the information and the Subscriber is inconvenient to provide quality of service (QoS) with type A data.

Routing scheme, should deliver data to Subscribers with an effective path cost (served from the nearest node that caches the data).

Fast routing is a necessity, which insists on designing a simple scheme that do not exhaust the router with much process complexity (e.g. solving first order logic filters or searching huge list of flat addresses).

Our proposed SICN takes into consideration the types of data usage. In addition, it can match between the Publisher knowledge and the Subscriber interested knowledge. Currently, the search engine holds this role. In other words, the whole network will work as a big routing search engine that matches Subscriber interest to data, and the Subscriber's interest to Publishers. This is done with the help of a three-dimension-naming scheme (Geographical address, User (Publisher/Subscriber) unique address, Semantic address) that is done in the extremes not in the core.

Routing tables. Routers will hold three tables with three address dimensions combined in them. The first one is the *Semantic-ID* that connects the Semantic address to the Publisher ID address. The second one is the *Geo-ID* that connects the Publisher ID address and the Geographical address and the third table is the *Geo-Semantic* matches the Semantic address to the Geographical address.

These three address dimensions will allow the matching between the Publisher and Subscriber based on naming scheme that includes any Publisher ID, Semantic or Geographical address in

the network and will be designed to include the four types of data and the four types of Subscriber's requests. A Subscriber interested in one of the three address dimensions can find a match to the other two address dimensions using the proposed routing tables. For example, an interest message containing only a Semantic address can easily be matched to Publisher IDs and their Geographical location using these tables. Considering another example where a subscriber having a phone call with a specific Publisher ID can follow the Geographical location of the Publisher using the second table.

Each table includes two parts. The first part, which is the address part (Publisher ID, Geographical and Semantic addresses) that names the data and are learnt or defined from the Publisher's advertisement. The second part of each table, which is the orientation part (cache (TTL) and Interface) that directs the data toward the Subscriber and are learnt from the Subscriber's interest message. The interface is an input-output port, which connects network nodes.

Conclusion. This article presents a new scheme in ICN. Through this project, we addressed the problem of Naming and Routing in the field of Information-Centric Networking where a new semantic-based scheme is proposed to solve the obstacles facing IP networks. We presented a new architecture scheme SICN and detailed its naming and a part of routing designs. An important contribution is classifying data into four types and classifying the Subscriber's request into four classes where the new system can cope with these different types and classes. In addition, three naming schemes were detailed. Furthermore, we designed the SICN Header format.

References

1. Jaber G., Patsei N. V. Information Centric Networking for web-based content distribution and manipulation. *Trudy BGTU* [Proceedings of BSTU], series 3, Physics and Mathematics. Informatics, 2017, no. 2, pp. 88–91.

2. Alzahrani B. A., Vassilakis V. G., Reed M. J. Key management in information centric networking. *Int. J. Comput. Networks Commun.*, 2013, vol. 5, no. 6, pp. 153–156.

3. Pepper R. Cisco Visual Networking Index (VNI) Global Mobile Data Traffic Forecast Update. *Tech. Rep*. Berlin, 2013, p. 245.

4. Olsen L. J. Services for substance abuse-affected families: The Project Connect experience. *Child Adolesc. Soc. Work J*., 1995, vol. 12, no. 3, pp. 183–196.

5. De Brito M. A. G., Galotto L., Sampaio L. P. Evaluation of the main MPPT techniques for photovoltaic applications. *IEEE Trans. Ind. Electron.*, 2013, vol. 60, no. 3, pp. 1156–1167.

6. Lee J.-C., Lim W.-S., Jung H.-Y. Scalable domain-based routing scheme for ICN. *Information and Communication Technology Convergence (ICTC): International Conference.* Jeju, 2014, pp. 770–774.

7. Navrotsky Y., Patsei N. Cashing Control and Optimization in Information-Content Networks. *Open Conference of Electrical, Electronic and Information Sciences (eStream): Proceedings of the Conference*. Vilnius, 2019, pp. 70–74.

8. Jaber G., Patsei N., Rahal F. Different Naming in Information-Centric Networks (ICN). *Scholars Journal of Engineering and Technology*, 2019, no. 7 (8), pp. 235–237.

9. UNL web community portal. Available at: http://www.unlweb.net/unlweb/ (accessed 18.11.2019).

10. UNL portal. Available at: http://www.undl.org/http://www.unlweb.net/unlweb/ (accessed 18.11.2019).

11. Uchida H., Zhu M. The universal networking language beyond machine translation. *International Symposium on Language in Cyberspace*. Seoul, 2001, pp. 26–27.

12. Alansary S., Nagi M., Adly N. The universal networking language in action in English-Arabic machine translation. *Proceedings of 9th Egyptian Society of Language Engineering Conference on Language Engineering (ESOLEC 2009)*. Cairo, 2009, pp. 23–24.

13. Hinden R., Deering S. IP Version 6 Addressing Architecture. Network Working Group. DOI: 10.17487/RFC4291. Available at: https://tools.ietf.org/html/rfc4291 (accessed 18.11.2019).

Information about the authors

Jaber Ghassan – PhD student. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: ghassanjaber@hotmail.com

Patsei Nataliya Vladimirovna − PhD (Engineering), Associate Professor, Head of the Department of Software Engineering. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: n.patsei@belstu.by

Rahal Fatima – PhD student. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: Fatimarahaljaber@gmail.com

Received after revision 19.11.2019

УДК 681.3.06

В. О. Берников
Белорусский государственный технологический университет

СРАВНИТЕЛЬНЫЙ АНАЛИЗ КРИПТОСТОЙКОСТИ СИММЕТРИЧНЫХ АЛГОРИТМОВ ШИФРОВАНИЯ

Проанализированы наиболее распространенные криптоаналитические атаки на блочные симметричные алгоритмы шифрования. Описаны основные методы, которые используются в атаках, выявлены преимущества и недостатки каждого метода. Рассмотрены и проанализированы наиболее известные в настоящее время симметричные криптосистемы на основе сравнения следующих характеристик: длина ключа должна составлять не менее 128 бит с возможностью быстрого расширения до 256 бит, длина обрабатываемого блока должна быть не менее 128 бит, структура каждого раунда алгоритма не должна иметь сложную математическую модель по причине продуктивности анализа в целом, а также данные алгоритмы шифрования должны быть устойчивыми к современным криптоаналитическим атакам. Приведена количественная оценка стойкости алгоритмов по следующим критериям: криптостойкость, запас криптостойкости, скорость расширения ключа, защита от атак по времени выполнения, реализация лавинного эффекта, возможность быстрого расширения ключа и возможность параллельных вычислений. Выявлены преимущества и недостатки каждого алгоритма шифрования. Выбран наиболее стойкий к взлому симметричный блочный алгоритм шифрования из рассмотренных.

Ключевые слова: симметричная криптосистема, криптостойкость, криптоаналитическая атака.

V. O. Bernikov
Belarusian State Technological University

COMPARATIVE ANALYSIS OF THE CRYPTOGRAPHIC RESISTANCE OF SYMMETRIC ALGORITHMS ENCRYPTION

The most common cryptanalytic attacks on block symmetric encryption algorithms are analyzed. The main methods that are used in attacks are described, the advantages and disadvantages of each method are identified. The most well-known symmetric cryptosystems are currently reviewed and analyzed based on a comparison of the following characteristics: the key length should be at least 128 bits with the ability to quickly expand to 256 bits, the length of the processed block should be at least 128 bits, the structure of each round of the algorithm should not have a complex mathematical model due to the productivity of the analysis as a whole, as well as these encryption algorithms must be resistant to modern cryptanalytic attacks. A quantitative assessment of the resistance of the algorithms is given according to the following criteria: cryptographic resistance, cryptographic strength margin, key expansion speed, protection against attacks at runtime, realization of an avalanche effect, the ability to quickly expand the key, and the possibility of parallel computing. The advantages and disadvantages of each encryption algorithm are revealed. The most resistant to cracking symmetric block encryption algorithm is selected from those considered.

Key words: symmetric cryptosystem, cryptographic resistance, cryptanalytic attack.

Введение. Задача обеспечения требуемой криптографической стойкости алгоритмов шифрования приобретает все большую актуальность в связи с развитием информационных технологий. Как известно, при помощи шифрования должны обеспечиваться следующие состояния безопасности: конфиденциальность, целостность и идентифицируемость передаваемой информации. Одним из наиболее продуктивных средств решения поставленной задачи является применение эффективных методов шифрования.

Для выбора соответствующего криптоалгоритма необходимо владеть математическим аппаратом, положенным в основу алгоритма, а также проанализировать возможность того или иного метода шифрования противостоять современным криптоаналитическим атакам. Далее важно выбрать критерии для оценки и анализа криптостойкости алгоритмов шифрования. Например, запас криптостойкости, скорость расширения ключа, защита от атак по времени выполнения, возможность быстрого расширения ключа и др.

В статье дана количественная оценка стойкости симметричных алгоритмов шифрования, а также выявлены их преимущества и недостатки. Данные криптосистемы были выбраны по причине того, что в настоящее время их ма-

тематический аппарат анализа и оценки стойкости лучше исследован по сравнению с асимметричными алгоритмами шифрования.

Основная часть. Как известно, симметричное шифрование – способ преобразования, в котором для зашифрования и расшифрования секретной информации используется один и тот же ключ [1]. Алгоритм шифрования выбирается сторонами до начала обмена сообщениями.

Криптостойкость – важнейшая характеристика для алгоритмов шифрования, которая обычно измеряется временем, необходимым для вскрытия того или иного метода шифрования при неких фиксированных ресурсах, имеющихся у злоумышленника. Нарушитель ставит перед собой следующие цели: нахождение открытого текста (при этом у него имеется криптограмма, но нет секретного ключа) или самого тайного ключа.

Симметричные криптосистемы обладают следующими достоинствами по сравнению с асимметричными криптосистемами:

1) скорость;

2) простота реализации (благодаря более простым операциям);

3) меньшая требуемая длина ключа для сопоставимой стойкости;

4) изученность (за счет большего возраста).

Необходимо также отметить и недостатки данных криптосистем:

– сложность управления ключами в большой сети. Означает квадратичное возрастание числа пар ключей, которые надо генерировать, передавать, хранить и уничтожать в сети;

– сложность обмена ключами. Для применения следует решить проблему надежной передачи ключей каждому абоненту, так как нужен секретный канал для передачи каждого ключа обеим сторонам.

Существует следующая классификация атак на симметричные алгоритмы шифрования [2]:

1) атака с известным открытым текстом. Предполагает у криптоаналитика некоторого количества пар текстов, каждая из которых представляет собой открытый текст и соответствующий ему шифртекст;

2) атака с выбранным открытым текстом. У криптоаналитика есть возможность выбора открытых текстов для получения соответствующих им шифртекстов;

3) адаптивная атака с выбором открытого текста. Криптоаналитик может не просто выбирать открытые тексты для зашифрования, но и делать это многократно с учетом результатов анализа ранее полученных данных;

4) атака с выбором шифртекста. Криптоаналитик может выбирать шифртексты и получать соответствующие им открытые тексты;

5) адаптивная атака с выбором шифртекста. Криптоаналитик может многократно выбирать шифртексты для их расшифрования с учетом предыдущих результатов.

Далее рассмотрим криптоаналитические методы, которые используются в атаках [3].

Метод «грубой силы» предполагает перебор всех возможных комбинаций ключа шифрования для нахождения искомого ключа. Защита от атак этого метода проста: увеличение размера ключа на 1 бит увеличит возможное количество ключей. Современная техника не позволяет в «лоб» атаковать 128-битный ключ полным перебором, однако данный метод можно использовать в контексте других методов криптоанализа.

Метод встречи посередине. Любые методы, способные вскрыть алгоритм шифрования быстрее, чем полный перебор всех возможных вариантов ключа шифрования, как правило, оперируют недостатками реализации того или иного метода шифрования. Примером данной атаки является вскрытие любого алгоритма шифрования, представляющего собой двойное шифрование данных с помощью какого-либо одного алгоритма. Во многих литературных источниках данный метод показан для взлома алгоритма Double DES, который в настоящее время не используется.

Дифференциальный криптоанализ. Данный метод основывается на анализе пар открытых текстов, между которыми существует определенная разность. При помощи этого метода вскрывается одноraундовый DES, а также сокращается количество комбинаций подбора секретного ключа для трехраундового DES с определенной вероятностью.

Линейный криптоанализ. Суть данного метода состоит в нахождении соотношений между открытым текстом, шифртекстом и ключом соответственно. Как и в дифференциальном криптоанализе, криптоаналитик находит некое одноraундовое соотношение и пытается распространить его на большее количество раундов. Во многих литературных источниках данный метод показан также на примере выявления закономерностей в работе алгоритма DES для поиска определенного количества начальных бит ключа, что по итогу заметно сокращает перебор оставшихся возможных комбинаций секретного ключа. Стоит отметить, что данный метод криптоанализа также продуктивен против алгоритмов RC5, NUSH и Noekeon.

Метод бумеранга является усилением дифференциального криптоанализа и состоит в использовании квартета (четырех вместо двух открытых текстов). Он представляет собой атаку с адаптивным выбором открытых текстов и

шифртекстов, которая на практике сложно применима. Данный метод был использован против CAST-256, MARS и SERPENT. Последние два алгоритма вскрываются только в вариантах с уменьшенным количеством раундов.

Сдвиговая атака. Уникальность атаки состоит в том, что ее успешность не зависит от количества раундов атакуемого алгоритма. Однако с помощью данной атаки можно вскрыть только те алгоритмы, раунды которых являются идентичными. Благодаря данной сдвиговой атаке был полностью раскрыт алгоритм шифрования TREYFER. Эта атака также применима к модифицированным симметричным алгоритмам DES и Blowfish, но не распространима на их полные версии. Стоит отметить, что несколько позже сдвиговая атака была усилена и применена на алгоритмы, в которых функции раундов не совсем идентичны, но имеют существенные сходства. Усиленная атака была использована для взлома нескольких вариантов алгоритма DES, а также 20-раундового стандарта шифрования ГОСТ.

Метод интерполяции применим к таким алгоритмам шифрования, которые используют достаточно простые алгебраические операции, в результате чего криптоаналитик может построить некий полином, который определяет взаимосвязь между шифртекстом и открытым текстом.

Невозможные дифференциалы. Основное отличие данного метода по сравнению с классическим дифференциальным анализом заключается в том, что тут используются дифференциалы с нулевой или минимальной вероятностью для того, чтобы сократить подмножество возможных ключей для выполнения дальнейшего перебора и нахождения секретного ключа. Этот метод нашел свое применение для вскрытия усеченных версий симметричных блочных алгоритмов шифрования, таких как IDEA.

Отметим, что это самые известные криптоаналитические атаки, которые предназначены для вскрытия или частичного взлома большинства симметричных криптосистем.

В данной статье рассмотрим следующие симметричные блочные алгоритмы шифрования, которые были участниками конкурса AES для выбора стандарта шифрования в США. Определим следующие критерии для отбора:

– 128-битный размер блока шифруемых данных;

– не менее трех поддерживаемых алгоритмом размеров ключей шифрования: 128, 192 и 256 бит;

– алгоритм должен быть стойким против современных криптоаналитических атак;

– структура и математическая модель алгоритма должны быть ясными и простыми, что облегчало изучение криптографической системы;

– должны отсутствовать слабые и эквивалентные ключи;

– скорость шифрования должна быть высокой;

– алгоритм должен предъявлять минимальные требования к оперативной памяти.

Многие симметричные системы шифрования были исключены из анализа по причине сложной математической модели и операций, которые были положены в основу структуры алгоритма, невозможности противостояния криптоаналитическим атакам, медленной скорости шифрования данных, а также невозможности реализации на разных платформах.

Для анализа криптостойкости были выбраны следующие симметричные блочные криптосистемы: AES, RC6, SERPENT и Twofish. Было обращено особое внимание на следующие компоненты каждого из алгоритмов: сложность структуры метода шифрования, способность быстрого расширения секретного ключа до 128, 192 и 256 бит.

В табл. 1 приведен результат сравнительного анализа эффективности данных криптосистем по выбранным критериям.

Таблица 1

Сравнительный анализ эффективности симметричных блочных криптосистем

Критерий	AES	RC6	SERPENT	Twofish
Криптостойкость	1	1	1	1
Запас криптостойкости	1	1	1	1
Скорость расширения ключа	1	0,5	0,5	0
Защита от атак по времени выполнения	1	0	1	0,5
Реализация лавинного эффекта	1	1	1	1
Возможность быстрого расширения ключа	0,5	0,5	1	1
Возможность параллельных вычислений	1	0,5	0,5	0,5
Результат	6,5	4,5	6	5

Анализ данных криптосистем производился по следующему принципу: если критерий не реализуем в определенном алгоритме, то выдвигается количественная единица для его оценки, равная 0; если критерий частично реализован в криптосистеме, то количественная единица принимает значение 0,5; следовательно, если критерий реализуем в методе шифрования без каких-либо ограничений, то коэффициент принимает значение 1.

Криптостойкость рассматриваемых симметричных шифров является достаточной. На ос-

новании многих литературных источников было выявлено, что для данных методов шифрования сложно реализуемы криптоаналитические атаки на полноценные или усеченные версии алгоритмов.

Под запасом стойкости понимается соотношение полного количества раундов и максимального из тех вариантов, против которого действуют какие-либо криптоаналитические атаки. Например, при помощи дифференциально-линейного криптоанализа вскрывается 11-раундовый SERPENT, тогда как в оригинальном алгоритме используется 32 раунда.

Защита от атак по времени выполнения заключалась в том, что скорость шифрования или расширения ключа не должна выходить за определенные установленные границы времени, в противном случае данный алгоритм будет подвержен данным атакам.

Лавинный эффект в указанных симметричных криптосистемах реализован в полной мере, поэтому все алгоритмы шифрования получили 1 количественную единицу за этот критерий.

На основании различных источников было установлено, что все данные симметричные криптосистемы поддерживают возможность быстрого расширения ключа, однако только SERPENT и Twofish реализуют такую возможность без каких-либо ограничений [4, 5]. Выявлено, что только алгоритм AES позволяет производить параллельные вычисления без ограничений, под которыми понимается одновременное выполнение операций внутри раунда и расширение ключа.

В ходе анализа было установлено, что алгоритм AES является наиболее стойким к взлому из рассмотренных с результатом 6,5 количественной единицы.

Проанализируем достоинства и недостатки данных алгоритмов шифрования, которые в итоге уступили алгоритму AES.

Алгоритм Twofish. Из преимуществ алгоритма можно выделить следующие:

1) процессы шифрования и дешифрования в алгоритме практически идентичны;

2) лучший из алгоритмов с точки зрения быстрого расширения ключа.

Отметим недостатки рассматриваемой криптосистемы:

– сложность структуры алгоритма затрудняет его анализ;

– сложная процедура расширения ключа;

– распараллеливание вычислений реализуемо с ограничениями.

Алгоритм SERPENT. Из достоинств алгоритма шифрования можно выделить следующие:

1) простая структура алгоритма, что значительно облегчает его анализ с целью нахождения возможных уязвимостей;

2) легко модифицируется для защиты от атак по времени выполнения, однако при этом снижается скорость.

Среди недостатков отметим следующие:

– самый медленный алгоритм в программных реализациях;

– процедуры шифрования и дешифрования различны, поэтому требуют различной реализации;

– распараллеливание вычислений реализуемо с ограничениями.

Алгоритм RC6. Из преимуществ данной симметричной криптографической системы можно выделить следующие:

1) простая структура алгоритма, что в итоге облегчает его анализ;

2) как и в алгоритме Twofish, процессы шифрования и дешифрования в алгоритме практически идентичны.

Отметим также недостатки рассматриваемого алгоритма:

– скорость шифрования зависит от того, поддерживает ли платформа 32-битное умножение и вращение на переменное число бит;

– достаточно сложно защищается от атак по времени выполнения;

– частично поддерживается быстрое расширение ключа;

– распараллеливание вычислений реализуемо с ограничениями.

Далее проанализируем симметричные криптосистемы AES и DES. Алгоритмы Twofish, SERPENT и RC6 не участвовали в данном сравнении по причине того, что эти криптосистемы имеют длину обрабатываемого блока и длину ключа такие же, как и у алгоритма AES. Стоит отметить, что рассмотренные в статье криптоалгоритмы отличаются между собой только типом архитектуры, количеством раундов и схемой генерации ключа.

В табл. 2 приведен сравнительный анализ алгоритма предшественника DES и нынешнего стандарта шифрования в США – AES.

Таблица 2

Сравнительный анализ алгоритмов AES и DES

Критерий	AES	DES
Длина блока, бит	128/192/256 в зависимости от длины ключа	64
Длина ключа, бит	128/192/256	56
Архитектура	SP – сеть «Квадрат»	Сеть Фейстеля
Число раундов	10/12/14 в зависимости от длины ключа	16
Схема генерации ключа	Умеренно сложная	Сложная

Анализ данных блочных симметричных алгоритмов шифрования показал, что нынешний

стандарт шифрования в США на основе алгоритма AES значительно превосходит своего предшественника и по длине обрабатываемого блока, и по длине ключа, что значительно повышает его криптостойкость. Отметим, что алгоритм DES не устойчив к различным криптоаналитическим атакам. Алгоритм Double DES взламывается за счет метода встречи посередине. Определенное количество раундов DES вскрывается на основе сдвиговой атаки, а также при помощи линейного и дифференциального криптоанализа [6].

В свою очередь, алгоритм AES устойчив к известным алгоритмам криптоанализа. Подчеркнем, что структура генерации ключа стала умеренно сложной по сравнению с алгоритмом DES.

Заключение. После проведения сравнительного анализа симметричных блочных алгоритмов шифрования по выбранным критериям было установлено, что AES является наиболее криптостойким алгоритмом с результатом в 6,5 количественной единицы. Единственный недостаток данного криптоалгоритма заключается в возможности расширения ключа лишь с некоторыми ограничениями.

Алгоритм Twofish был оценен в 5 количественных единиц из-за медленной скорости расширения ключа, а также частичной возможности быстрого расширения ключа и возможности параллельных вычислений.

Алгоритм RC6 получил 4,5 количественной единицы по причине невозможности защиты от атак по времени выполнения. Возможность быстрого расширения ключа и возможность параллельных вычислений реализуемы с ограничениями.

В свою очередь, алгоритм SERPENT был наиболее близок к показателям AES с результатом в 6 количественных единиц. Скорость расширения ключа относительно медленная, а возможность быстрого расширения ключа и возможность параллельных вычислений реализуемы в данной криптосистеме с ограничениями.

Литература

1. Урбанович П. П. Защита информации методами криптографии, стеганографии и обфускации. Минск: БГТУ, 2016. 220 с.

2. Брассар Ж. Современная криптология. М.: Полимед, 1999. 176 с.

3. Панасенко С. П. Алгоритмы шифрования. Специальный справочник. СПб.: БХВ-Петербург, 2009. 576 с.

4. Аграновский А. В., Хади Р. А. Практическая криптография: Алгоритмы и их программирование. М.: СОЛОН-Р, 2002. 257 с.

5. Грушо А. А., Тимошина Е. Е., Применко Э. А. Анализ и синтез криптоалгоритмов. Курс лекций. Йошкар-Ола: Изд-во МФ МОУ, 2000. 110 с.

6. Бабенко Л. К., Ищукова Е. А. Современные алгоритмы блочного шифрования и методы их анализа. М.: Гелиос АРВ, 2006. 258 с.

References

1. Urbanovich P. P. *Zashchita informatsii metodami kriptografii, steganografii i obfuskatsii* [The protection of information based on the methods by cryptography, steganography and obfuscation]. Minsk, BGTU Publ., 2016. 220 p.

2. Brassar Zh. *Sovremennaya kriptologiya* [Modern cryptology]. Moscow, Polimed Publ., 1999. 176 p.

3. Panasenko S. P. *Algoritmy shifrovaniya. Spetsial'nyy spravochnik* [Encryption Algorithms. Special reference]. St. Petersburg, BKhV-Peterburg Publ., 2009. 576 p.

4. Agranovsky A. V., Hadi R. A. *Prakticheskaya kriptografiya: Algoritmy i ikh programmirovaniye* [Practical cryptography: Algorithms and their programming]. Moscow, SOLON-R Publ., 2002. 257 p.

5. Grusho A. A., Timoshina E. E., Primenko E. A. *Analiz i sintez kriptoalgoritmov. Kurs lektsiy* [Analysis and synthesis of cryptographic algorithms. Lecture course]. Yoshkar-Ola, Izdatel'stvo MF MOU Publ., 2000. 110 p.

6. Babenko L. K., Ishchukova E. A. *Sovremennyye algoritmy blochnogo shifrovaniya i metody ikh analiza* [Modern block cipher algorithms and methods for their analysis]. Moscow, Gelios ARV Publ., 2006. 258 p.

Информация об авторе

Берников Владислав Олегович – аспирант, ассистент кафедры информационных систем и технологий. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: vladbernikovronaldo@gmail.com

Information about the author

Bernikov Vladislav Olegovich – PhD student, assistant lecturer, the Department of Information Systems and Technology. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: vladbernikovronaldo@gmail.com

Поступила после доработки 14.11.2019

АЛГОРИТМИЗАЦИЯ И ПРОГРАММИРОВАНИЕ

УДК 316.776

Я. Ю. Навроцкий, Н. В. Пацей
Белорусский государственный технологический университет

АЛГОРИТМЫ МАРШРУТИЗАЦИИ ИМЕНОВАННЫХ ОБЪЕКТОВ В ИНФОРМАЦИОННО-ОРИЕНТИРОВАННЫХ СЕТЯХ

Работа посвящена проблеме проектирования информационно-ориентированных сетей (ICN – Information-Centric Networks), в частности одному из самых важных вопросов – публикации, поиска и доставки именованных объектов сети. Пошагово описаны алгоритмы маршрутизации, а также схемы, используемые при именовании и разрешении имен для четырех оставшихся на сегодняшний день актуальных проектов архитектур информационно-ориентированных сетей: Data-Oriented Network Architecture (DONA), Content-Centric Networking/Named Data Networking (CCN/NDN), Scalable and Adaptive Internet Solutions (SAIL) и MobilityFirst (MF). Пошагово и единообразно описан процесс последовательного обмена сообщениями между «издателем» и «подписчиком», а также работа блоков разрешения имен для выбранных сетей. Выполнен анализ производительности рассматриваемых алгоритмов на основе сценария получения объекта, рассчитана асимптотическая сложность. Указаны недостатки каждой стратегии при практическом использовании. Установлено, что более производительными являются архитектуры DONA и CCN/NDN. Наименьшая производительность при поиске и доставке локатора будет в MF-сети.

Ключевые слова: маршрутизация, информационно-ориентированные сети, сообщение, сервер, объект, разрешение имен, хэш.

Ya. Yu. Navrotsky, N. V. Patsei
Belarusian State Technological University

ALGORITHMS OF NAMEED OBJECT ROUTING IN INFORMATION-CENTRIC NETWORKS

The article is devoted to the problem of designing Information-Centric Networks (ICN), in particular, one of the most important issues – the publication, search and delivery of named objects. The step-by-step routing algorithms are described, as well as the schemes used in naming and resolving names for the four remaining current projects of Information-Centric Networks architecture: Data-Oriented Network Architecture (DONA), Content-Centric Networking/Named Data Networking (CCN/NDN), Scalable and Adaptive Internet Solutions (SAIL) and MobilityFirst (MF). The process of sequential messaging between “publisher” and “subscriber”, as well as the operation of name resolution nodes for the selected networks, is described in detail and in a uniform manner. The performance analysis of the described algorithms based on the scenario of obtaining the object is performed, the asymptotic complexity is calculated. The shortcomings of each strategy in practical use are indicated. It is established that the DONA and CCN/NDN architectures are more productive. The smallest performance in locating and delivering a locator will be in the MF network.

Key words: routing, information-centric networks, message, server, object, name resolution, hash.

Введение. Доставка и управление именованной информацией является главным применением сети интернет. В традиционной хост-ориентированной архитектуре появляются новые технологии доставки контента, в основе которых лежит принцип получения данных по имени без учета расположения сервера. Возрастающий интернет-трафик вынуждает использовать набор различных технологий для обеспечения кэширования, репликации и доставки контента. Невозможно однозначно и безопасно идентифицировать информацию, не зная канала ее получения, это вынуждает одновременно использовать несколько подходов доставки данных, что приводит к потере эффективности. Информационно-ориентированные сети (ICN – Information-Centric Networks) как следующее поколение сетей должны решить все эти проблемы. За последние десятилетия было предложено множество различных вариантов архитектур информационных сетей. Ответа на вопрос, какая из архитектур будет использоваться в

качестве основы, пока нет. Рассмотрим подробнее алгоритмы публикации/подписки данных в четырех проектах ICN: Data-Oriented Network Architecture, Content-Centric Networking/Named Data Networking, Scalable and Adaptive Internet Solutions, MobilityFirst. Проанализируем производительность алгоритмов маршрутизации на примере сценария извлечения данных.

Основная часть. ***Data-Oriented Network Architecture (DONA).*** В DONA каждый информационный объект или сервис связан с владельцем [1]. Имя объекта формируется из криптографического хэша публичного ключа владельца P и его уникального названия L. Имена при этом являются неструктурированными, глобально уникальными, независимыми от уровня приложения и местоположения. Для неизменяемых данных название может быть криптографическим хэшем самого информационного объекта. Предполагается, что пользователи, заинтересованные в информационном объекте, получат его имя через определенные механизмы, например поисковую систему. В отличие от структурированных DNS-имен, flat-имена в DONA не содержат административной информации, что позволяет сопоставлять имена DONA с пользовательскими именами [1].

Разрешение имен в DONA обеспечивается специализированными серверами, называемыми обработчиками разрешений (Resolution Handlers – RH), при этом автономная система в DONA состоит как минимум из одного логического сервера RH. Под автономной системой (Autonomous System – AS) понимают множество IP-сетей и маршрутизаторов, управляемых одним или несколькими операторами, имеющими единую политику маршрутизации. Все серверы RH связаны между собой, образуя иерархический сервис разрешения имен, который находится вверху иерархии связей междоменной маршрутизации, что позволяет сервисам разрешения имен и маршрутизации данных соблюдать установленные правила маршрутизации между автономными системами.

Публикация объекта в сети осуществляется с помощью сообщения *REGISTER* (рис. 1).

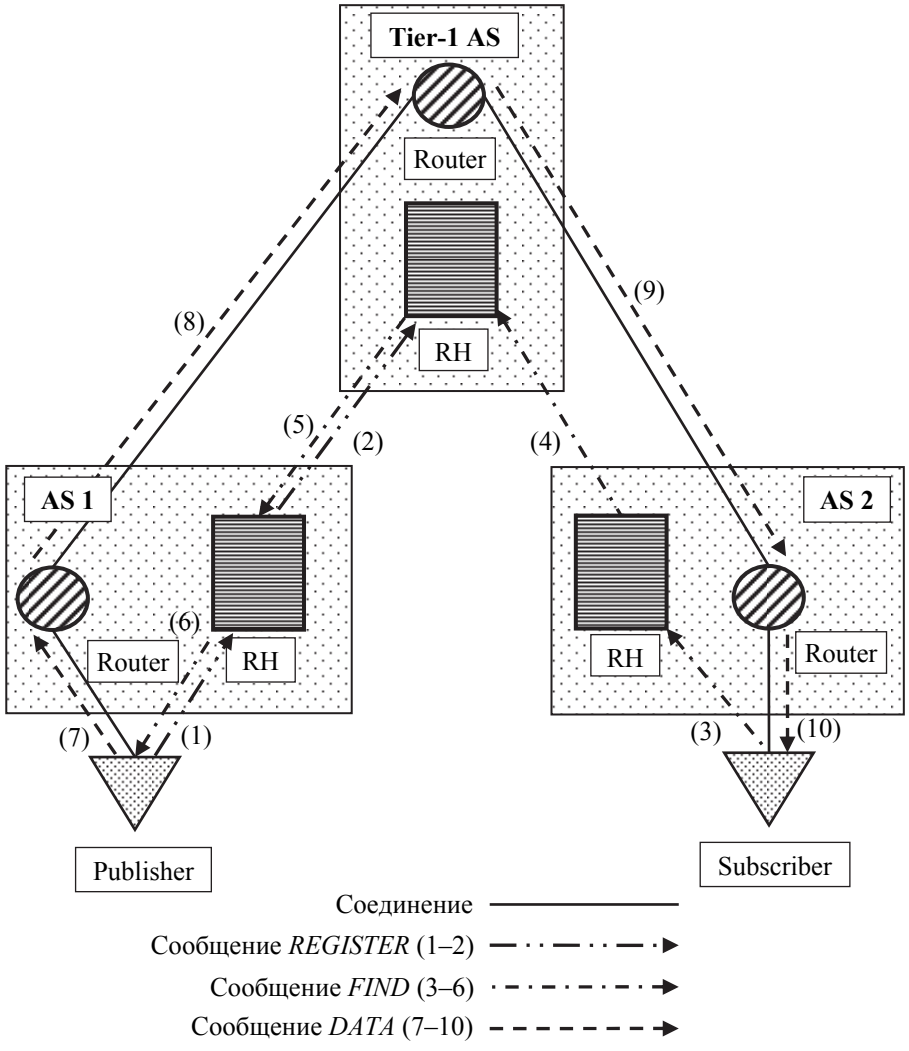


Рис. 1. Схема маршрутизации объектов для DONA

Издатель отправляет сообщение *REGISTER* с именем объекта на локальный сервер RH. Затем сервер RH распространяет это сообщение на серверы RH в своих родительских и пиринговых доменах, следуя установленным правилам маршрутизации (стрелки 2–3 на рис. 1). Каждый сервер RH, получивший такое сообщение, сохраняет у себя сопоставление между именем объекта и адресом сервера RH, который переадресовал ему сообщение *REGISTER*. Конечной точкой сообщения *REGISTER* является RH провайдера верхнего уровня (tier-1). Все провайдеры верхнего уровня равны и связаны между собой, что позволяет серверам RH в каждой из них получать информацию о регистрации объектов во всей сети. Издатели также могут формировать сообщения-шаблоны *REGISTER*, чтобы указать иерархии RH на то, что они могут предоставить все возможные данные в соответствии с определенным шаблоном, например, предоставить все версии определенного документа.

Для получения данных клиент/подписчик отправляет сообщение *FIND* на локальный сервер RH, который передает это сообщение вверх по сети, до тех пор, пока не будут найдены соответствующие записи о регистрации объекта (стрелки 4–5 на рис. 1). После этого запрос маршрутизируется согласно указателям, созданным при регистрации (стрелки 6–7 на рис. 1). Поскольку провайдеры tier-1 знают обо всех объектах в сети, этот процесс гарантированно завершится успешно в случае, если запрошенное имя существует. Подписчики также могут отправлять сообщения-шаблоны *FIND* для запроса данных с определенным названием.

В DONA маршрутизация данных может быть разделена или связана с сервисом разрешения имен. В варианте с разделением, если сообщение *FIND* достигает издателя, данные могут быть напрямую отправлены подписчику с использованием обычной IP-маршрутизации и передачи. Фактическая передача данных будет осуществляться в соответствии с установленными правилами маршрутизации сетевого трафика между автономными системами издателя и подписчика. В другом случае в сообщение *FIND* записываются адреса серверов RS, которые оно проходит с учетом последовательности систем AS. Когда запрос достигает издателя, данные о пройденном пути используются в обратном порядке, для доставки данных подписчику (стрелки 8–11 на рис. 1). В таком варианте устраняется необходимость в глобальных IP-адресах и больших таблицах маршрутизации типа Border Gateway Protocol (BGP), поскольку все данные о пути имеют локальное значение. Такой вариант передачи обеспечивает симметричную маршрутизацию между запросами и ответами.

DONA также поддерживает каналы многоадресной передачи данных. Реализуется это с помощью кэширования сообщения *FIND* в серверах RH. Это позволяет клиенту подписаться на обновления определенной информации на время срока действия сообщения *FIND*. В случае если сервер RH получает еще одно сообщение *FIND* с запросом на ту же информацию, то записи об ожидании данных объединяются в единую запись с несколькими указателями пути для ответа, тем самым создавая дерево многоадресной доставки (multicast distribution tree). Такая передача данных может работать только при совместной работе систем маршрутизации и именования, так как требуется, чтобы данные следовали по обратному пути системы AS, выбранному сообщениями *FIND*.

Content-Centric Networking/Named Data Networking (CCN/NDN). Имена в NDN являются иерархическими и похожи на URL. Например, имя может выглядеть как /wiki.org/icn/ccn.html. Однако имена NDN не являются URL-адресами. Каждая часть имени может быть чем угодно – это понятное пользователю название или же значение хэша. В NDN поиск информационного объекта осуществляется путем сопоставления введенного запроса с названием объекта, где запрос – префикс названия объекта. Например, запрос на /wiki.org/icn/ccn.html может вернуть объект /wiki.org/icn/ccn.html/_v1/_s1. Это означает, что получен первый сегмент первой версии запрошенных данных. После получения такого объекта пользователь может запросить следующий сегмент данных, явно указав это в имени. Способ сегментации информационных объектов и правило сопоставления префиксов будет известно уровню приложения пользователя. Правило сопоставления префиксов позволяет уровню приложения исследовать доступность данных по сети, а также пользователю запрашивать данные, которые еще не были опубликованы. В этом случае издатель указывает, что может отвечать на запросы с определенным префиксом. После создания объектов они будут возвращены клиентам с полными именами. Данный подход может быть использован в случае, когда информационные объекты генерируются динамически, поэтому их полные имена не могут быть известны заранее, например голосовые конференции [2].

В NDN для получения данных клиент отправляет сообщение *INTEREST*, в ответ на которое получает сообщение *DATA* (рис. 2). Оба типа сообщений содержат имя запрошенного/переданного информационного объекта. Все сообщения пересылаются последовательно при помощи маршрутизаторов контента (Content Router – CR), при этом каждый маршрутизатор CR содержит три структуры данных: таблицу

маршрутизации информации (или информационную базу пересылки – Forwarding Information Base – FIB), таблицу ожидания интересов (Pending Interest Table – PIT) и хранилище контента (Content Store – CS). FIB хранит сопоставление имени информационного объекта и адреса, по которому можно его запросить, маршрутизатор контента использует FIB для пересылки сообщений *INTEREST*. Таблица PIT хранит информацию о том, кто какие данные ожидает. При поступлении запроса на маршрутизатор в PIT создается запись с адресом отправителя сообщения *INTEREST* и ожидаемого для него сообщения *DATA*. Хранилище контента является локальным кэшем для информационных объектов.

При поступлении сообщения *INTEREST* маршрутизатор сначала проверяет хранилище контента на наличие запрошенного объекта. Если объект найден, он отправляется клиенту в сообщении *DATA*, а сообщение *INTEREST* удаляется. В случае если объекта в хранилище нет, маршрутизатор выполняет сопоставление префиксов имен в своей базе FIB для определения маршрута пересылки сообщения *INTEREST*. Если запись найдена в базе FIB, маршрутизатор создает запись в таблице PIT и передает сообщение *INTEREST* на маршрутизатор CR, который был указан базой FIB. Если запись в PIT по запрошенному объекту уже существует, маршрутизатор добавляет к ней адрес маршрутизатора, с которого поступил запрос, а сообщение *INTEREST* удаляется (стрелки 1–3 на рис. 2).

Это позволяет эффективно формировать дерево многоадресной передачи (a multicast tree) информационных объектов.

Информационный объект возвращается подписчику в сообщении *DATA* в последовательном режиме, основываясь на состоянии таблицы PIT. Когда маршрутизатор CR получает сообщение *DATA*, он сначала кэширует информационный объект в хранилище CS, а затем на основании таблицы PIT передает сообщение вниз по сети. Сообщение *DATA* дублируется на каждого клиента в записи PIT, что обеспечивает многоадресную доставку. После отправки сообщения *DATA* маршрутизатор удаляет запись из таблицы PIT (стрелки 4–6 на рис. 2). В случае если при получении сообщения *DATA* в таблице PIT нет соответствующих записей, маршрутизатор удаляет полученное сообщение.

В NDN разрешение имен и маршрутизация данных связаны. Маршрутизация сообщений *DATA* осуществляется согласно данным таблицы PIT. Заполнение таблицы FIB в NDN может выполняться по протоколам распределенной маршрутизации, таким как OSPF [3], в которых маршрутизаторы CR используют префиксы имен, а не диапазоны IP-адресов. Маршрутизаторы CR могут иметь множество адресов для одного префикса в записях FIB. В этом случае стратегический уровень может выбрать отправку сообщения *INTEREST* либо сразу на все адреса, либо только на интерфейс, который демонстрировал лучшую производительность.

Scalable and Adaptive Internet Solutions (SAIL). В архитектуре SAIL имена информационных объектов одновременно являются неструктурированными и иерархическими. SAIL определяет схему URI-имен как *ni://A/L*, в которой имена состоят из авторитетной части *A*, связанной с владельцем информации, и локальной части *L*. Каждая часть может быть хэшем, что позволит проводить самосертификацию (проверка подлинности полученной информации), или любым другим видом строки, что позволит использовать их как обычные URL. Для поиска нужного объекта в SAIL служат неструктурированные имена. В то же время для маршрутизации используются иерархические имена (как в CCN/NDN) [4, 5].

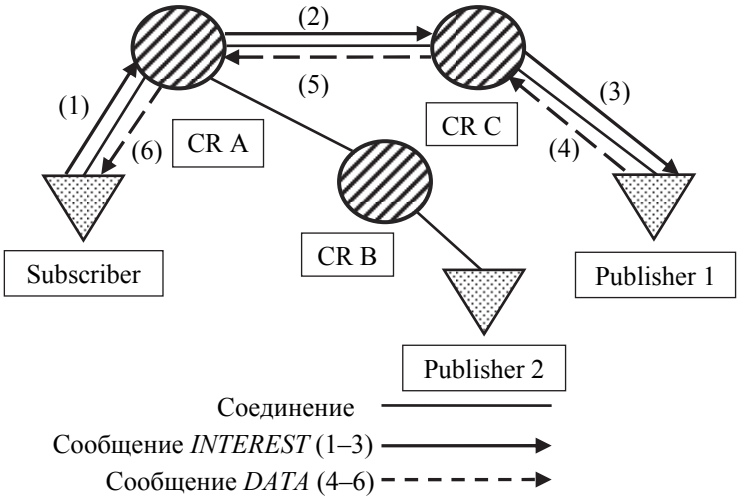


Рис. 2. Схема маршрутизации объектов для NDN

Разрешение имен и маршрутизация данных могут быть связаны, разделены или работать в гибридном варианте. В случае раздельной реализации система разрешения имен (Name Resolution System – NRS) преобразует имя объекта в локатор, который в дальнейшем необходим для получения объекта. Локатор содержит IP-адрес именного объекта. Система NRS может быть разновидностью хэш-таблицы DHT, многоуровневой DHT или иерархической SkipNet. В случае реализации многоуровневой DHT каждая область сети поддерживает свою собственную локальную систему NRS для обработки разрешения *L*-части, в то время как глобальная система NRS обрабатывает разрешение *A*-части (рис. 3). Издатель публикует информационный объект, отправляя сообщение *PUBLISH* с локатором в локальную систему NRS (стрелка 1 на рис. 3), которая сохраняет сопоставление *L*-части с локатором. Затем локальная система NRS объединяет все *L*-части, входящие в эту область (*A*), и передает их в фильтр Блума, результат добавляется в сообщение *PUBLISH* и отправляется в глобальную систему NRS (стрелка 2 на рис. 3).

Глобальная система NRS сохраняет сопоставление между областью (*A*) с учетом фильтра Блума и локальной системой NRS. Для запроса данных подписчик отправляет сообщение *GET* в свою локальную систему NRS, которая обращается к глобальной системе NRS (стрелки 3–4 на рис. 3) для запроса локатора объекта (стрелки 4–5 на рис. 3). Затем подписчик отправляет сообщение *GET* издателю, используя возвращенный локатор (стрелки 7–9 на рис. 3), и издатель отвечает информационным объектом в сообщении *DATA* (стрелки 10–12 на рис. 3) [6, 7].

В совместной реализации протокол маршрутизации используется для публикации имен объектов и заполнения таблиц маршрутизации CR. Подписчик отправляет сообщение *GET* на свой локальный маршрутизатор CR, который передает его последовательно, по направлению к издателю (стрелки a–c на рис. 3). Когда информационный объект найден, он возвращается с помощью сообщения *DATA* по обратному пути, пройденному сообщением *GET* (стрелки d–f на рис. 3).

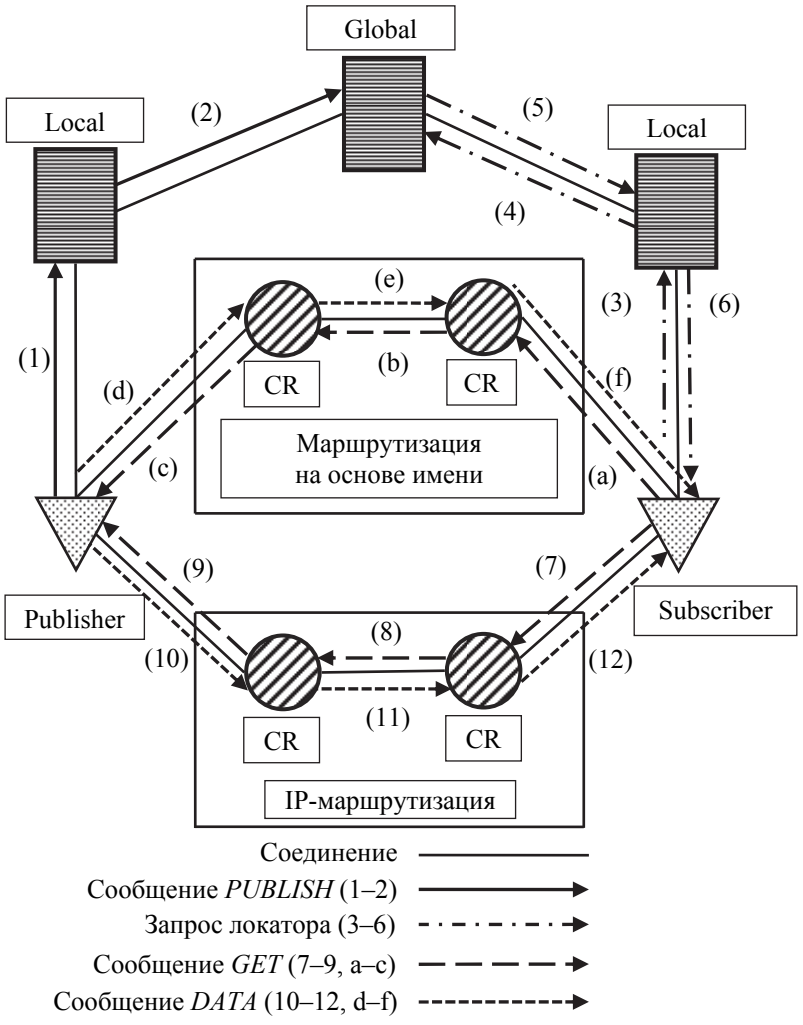


Рис. 3. Схема маршрутизации объектов для SAIL

В отличие от NDN, где информация, хранимая в маршрутизаторе CR, используется для составления обратного пути, в SAIL сообщения *GET* содержат информацию о пройденном пути, которая затем служит для доставки данных к подписчику.

В гибридном режиме работы система NRS возвращает подсказки маршрутизации, которые являются частью локатора и могут использоваться для передачи сообщения *GET* по одному из направлений. Таким образом, отправка сообщения *GET* может начинаться с получения подсказок маршрутизации от системы NRS для доставки сообщения *GET* как можно ближе к источнику, а затем, используя информацию маршрутизации на основе имени, хранящуюся в маршрутизаторе CR, достигать своего места назначения. Также возможен вариант, когда отправка сообщения *GET* может начинаться с информации о маршрутизации на основе имени, хранящейся в маршрутизаторе CR, и прибегать к системе NRS для дальнейших подсказок маршрутизации.

MobilityFirst (MF). В архитектуре MobilityFirst каждому сетевому объекту назначается глобальный уникальный идентификатор (GUID) через глобальную службу имен, которая переводит понятные человеку имена в GUID. Каждое устройство в MF должно получить GUID для себя, своих информационных объектов и своих сервисов. Если объект доступен из нескольких мест в сети, то все его копии будут иметь одинаковый идентификатор GUID. Поскольку все сущности именованы, поддерживается как доставка данных на основе имени, так и хост-ориентированная доставка [8]. Любое взаимодействие начинается с преобразования GUID в сетевой адрес. Преобразование выполняется за один или более шагов, и отвечает за это глобальная служба разрешения имен (Global Name Resolution Service – GNRS) (рис. 4).

Для публикации данных издатель запрашивает у службы именования идентификатор GUID, а затем регистрирует его и свой сетевой адрес в службе GNRS (стрелка 1 на рис. 4). Для получения данных клиент отправляет сообщение *GET* на локальный маршрутизатор CR. Сообщение *GET* содержит GUID-идентификатор запрашиваемого объекта, а так же GUID-идентификатор клиента, на который будет отправлен ответ (стрелка 2 на рис. 4). CR выполняет маршрутизацию только на основе фактических сетевых адресов, например IP-адресов, следовательно, он запрашивает службу GNRS для сопоставления GUID-идентификатора запрошенного объекта с его сетевым адресом (стрелка 3 на рис. 4).

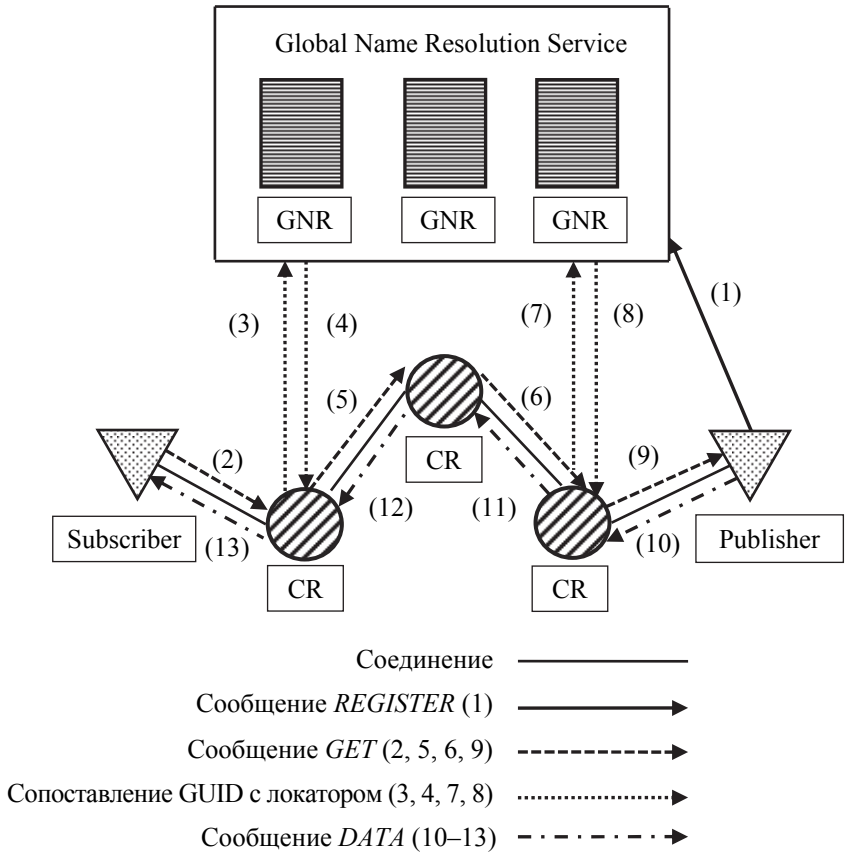


Рис. 4. Схема маршрутизации объектов для MF

Служба GNRS присылает на такой запрос (стрелка 4 на рис. 4) набор сетевых адресов. В зависимости от запроса может прислать целый маршрут, часть целого маршрута и/или промежуточные сетевые адреса. CR выбирает один из сетевых адресов, добавляет его в сообщение *GET*, которое затем пересылает, используя обычные таблицы маршрутизации (стрелки 5–6 и 9 на рис. 4). Сообщение *GET* включает в себя как идентификатор GUID запрашиваемого объекта, так и его сетевой адрес, это позволяет любому маршрутизатору CR обратиться к службе GNRS для получения обновленного списка сетевых адресов для запрошенного объекта (стрелки 7–8 на рис. 4), если, например, сообщение *GET* не может быть доставлено издателю. По такой же схеме издатель отправляет сообщение ответа подписчику (стрелки 10–13 на рис. 4) [9, 10].

Фактическая маршрутизация выполняется на основе сетевых адресов, а служба GNRS используется только для сопоставления идентификаторов GUID с сетевыми адресами. Для статических сервисов или данных MobilityFirst может преобразовывать каждый идентификатор GUID в сетевой адрес только один раз, как в DNS, и работать только на основе сетевых адресов, игнорируя идентификатор GUID. Для более динамичных сервисов идентификатор GUID может преобразовываться несколько раз; первый маршрутизатор (опционально и другие) отправляет запрос в GNRS на получение сетевых адресов, привязанных к данному GUID, и затем принимает решение о переадресации на основе ответа от службы GNRS. Таким образом, переадресация может быть быстрой, когда служба GNRS не принимается во внимание, или медленной, когда маршрутизаторы обращаются к службе GNRS для получения обновленного списка сетевых адресов. Позднее связывание особенно полезно при большой мобильности клиента или издателя. Следует отметить, что каждое сообщение доставляется отдельно, т. е. сообщение *GET* и информационный объект, отправляемый в ответ на него, индивидуально маршрутизируются на основе их GUID-назначения, поэтому разрешение имен и маршрутизация данных в MobilityFirst разделены.

Анализ производительности. Для сравнения алгоритмов маршрутизации рассмотрим сценарий получения данных каждой сети. Сценарии передачи сообщения исключают идеальную ситуацию для каждой из сетей. Например, в случае с CCN/NDN опустим момент извлечения данных из кэша при получении сообщения *INTEREST*. Тогда сценарий принимает следующий вид: сообщение *INTEREST* проходит полный путь подписчик – издатель – подписчик. При таких условиях ключевую роль будут играть время доставки сообщения между маршрутизаторами, количество пройденных узлов и количество выполняемых операций в узлах сети. В сетях DONA, MF и SAIL данные о пройденном маршруте сохраняются в сообщении *FIND*, а в CCN/NDN данные хранятся в таблице PIT. В первых трех случаях способ хранения данных можно рассматривать с позиции очереди или стека, в CCN/NDN – хэш-таблицы. Для упрощения анализа приведем все к хэш-таблице. Время доставки сообщения между узлами и количество маршрутизаторов в сети примем за постоянную величину, одинаковую для всех сетей.

Особенности иерархической структуры сети DONA позволяют сразу провести маршрутизацию сообщения от подписчика к издателю. При восходящем пути на каждом узле будет выполнена операция записи в сообщении *FIND* информации о пройденном маршруте. При нисходящем пути запись будет удалена. Асимптотическую сложность удаления и вставки обозначим *Od* и *Oc* соответственно, время запроса локатора – *tl*, количество узлов – *n*. Тогда сложность получения данных будет: $Og = tl + 2n(Od + Oc)$.

Извлечение данных в сети CCN/NDN потребует обращения к таблицам FIB и PIT. Так как используется одинаковая структура для хранения информации, разница будет в том, где она хранится, то сложность извлечения остается прежней. При этом в обоих случаях величина $tl = 0$.

В MF используется GNRS для получения локатора при каждой отправке сообщения, поэтому *tl* будет отлично от нуля и пренебречь им нельзя. Похожий алгоритм используется в сети SAIL, где за получение локатора отвечает NRS. Однако в случае с SAIL локатор будет запрошен только один раз, а в случае с MF локатор запрашивается каждый раз на пути от подписчика к издателю и обратно. В итоге сложность алгоритма извлечения данных в MF будет: $Og = 2tl + 2n(Od + Oc)$.

Для анализа будем использовать модель экспериментальной сети Abilene [11]. Сеть имеет 22 источника данных, 172 маршрутизатора, 139 edge-узлов и 139 клиентов. Если пренебречь такими характеристиками, как потеря пакетов, задержка и пропускная способность, и принять за единицу производительность сети DONA в сценарии извлечения данных, то можно рассчитать относительную производительность (Λ) для остальных сетей, которая представлена в таблице ниже.

Относительная производительность сетей

DONA	CCN/NDN	SAIL	MF
1,00	1,05	0,94	0,89

С учетом введенных обобщений и ограничений получаем практически одинаковую производительность для CCN/NDN и DONA, наибольшую из всех рассмотренных сетей. Менее производительной будет сеть SAIL, на последнем месте – MF.

Заключение. Как видно из описания схем маршрутизации, в ICN информация передается более эффективным образом и расположена ближе к пользователям, чем в TCP/IP. В статье представлен обзор схем именования и механизмов маршрутизации четырех проектов ICN. При маршрутизации используются два разных подхода: маршрутизация на основе имен и на основе разрешения имен.

Схема, основанная на разрешении имен, значительно снижает накладные расходы при публикации контента, особенно для больших размеров сети, и скорость публикации контента также высока. Однако она требует дополнительного запроса в процессе поиска контента. Основной же проблемой flat-именования является масштабируемость, поэтому для большого количества сетевых объектов она не подходит. Очевидно, при построении сценариев маршрутизации требуется компромиссное решение.

При построении схем маршрутизации предъявляют ряд других требований: масштабируемость (архитектуры ICN должны быть способны обслуживать огромное количество объектов); эффективный поиск ближайшей копии для сокращения междоменного трафика; гарантия доставки контента с минимальными задержками (высокая скорость поступления запросов вызовет переполнение таблиц, а запросы подписчиков приведут к увеличению скорости ретрансляции и краху всей сети); безопасность и фильтрация (злоумышленники могут создавать искусственные запросы для заполнения таблиц на маршрутизаторах); единая точка отказа (присуще всем рассмотренным архитектурам). Ни одна из представленных здесь архитектур явно не указала режим удаления контента или обновление метаданных.

Литература

1. A data-oriented (and beyond) network architecture / T. Koponen [et al.] // ACM SIGCOMM. 2007. P. 181–192.

2. VoCCN: Voice over content-centric networks / V. Jacobson [et al.] // ACM ReArch Workshop. 2009. P. 20–65.

3. Networking named content / V. Jacobson [et al.] // ACM CoNEXT. 2009. P. 1–12.

4. The Network of Information: Architecture and applications [Electronic resource] // SAIL Project. URL: https://sail-project.eu/deliverables (date of access: 18.09.2019).

5. Final NetInf Architecture [Electronic resource] // SAIL Project. URL: https://sail-project.eu/deliverables (date of access: 18.09.2019).

6. MDHT: a hierarchical name resolution service for information-centric networks / M. D'Ambrosio [et al.] // ACM Workshop on Information-Centric Networking (ICN). 2011. P. 584–587.

7. Dannewitz C., D'Ambrosio M., Vercellone V. Hierarchical DHTbased name resolution for information-centric networks // Computer Communications. 2012. vol. 36, no. 7. P. 736–749.

8. Baid A., Vu T., Raychaudhuri D. Comparing alternative approaches for networking of named objects in the future Internet // IEEE Workshop on Emerging Design Choices in Name-Oriented Networking (NOMEN). 2012. P. 89–93.

9. Map: A shared hosting scheme for dynamic identifier to locator mappings in the global Internet / T. Vu [et al.] // IEEE International Conference on Distributed Computing Systems (ICDCS). 2012. P. 1–92.

10. Jaber G., Patsei N., Rahal F. A Survey: Routing Schemes in Information-Centric Networks (ICN) // Scholars Journal of Engineering and Technology. 2019. No. 7 (8). P. 229–234.

11. Internet2: [site]. URL: https://www.internet2.edu/products-services/advanced-networking/ (date of access: 18.09.2019).

References

1. Koponen T., Chawla M., Chun B., Ermolinskiy A., Kim K. H., Shenker S., Stoica I. A data-oriented (and beyond) network architecture. *ACM SIGCOMM*, 2007, pp. 181–192.

2. Jacobson V., Smetters D. K., Briggs N. H., Plass M. F., Stewart P., Thornton J. D., Braynard R. L. VoCCN: Voice over content-centric networks. *ACM ReArch Workshop*, 2009, pp. 20–65.

3. Jacobson V., Smetters D. K., Thornton J. D., Plass M. F., Briggs N. H., Braynard R. L. Networking named content. *ACM CoNEXT*, 2009, pp. 1–12.

4. The Network of Information: Architecture and applications. *SAIL Project*. Available at: https://sail-project.eu/deliverables (accessed 18.09.2019).

5. Final NetInf Architecture. *SAIL Project*. Available at: https://sail-project.eu/deliverables (accessed 18.09.2019).

6. D'Ambrosio M., Dannewitz C., Karl H., Vercellone V. MDHT: a hierarchical name resolution service for information-centric networks. *ACM Workshop on Information-Centric Networking (ICN)*, 2011, pp. 584–587.

7. Dannewitz C., D'Ambrosio M., Vercellone V. Hierarchical DHTbased name resolution for information-centric networks. *Computer Communications*, 2012, vol. 36, no. 7, pp. 736–749.

8. Baid A., Vu T., Raychaudhuri D. Comparing alternative approaches for networking of named objects in the future Internet. *IEEE Workshop on Emerging Design Choices in Name-Oriented Networking (NOMEN)*, 2012, pp. 89–93.

9. Vu T., Baid A., Zhang Y., Nguyen T., Fukuyama J., Martin R., Raychaudhuri D. Map: A shared hosting scheme for dynamic identifier to locator mappings in the global Internet. *IEEE International Conference on Distributed Computing Systems (ICDCS)*, 2012, pp. 1–92.

10. Jaber G., Patsei N., Rahal F. A Survey: Routing Schemes in Information-Centric Networks (ICN). *Scholars Journal of Engineering and Technology*, 2019, no. 7 (8), pp. 229–234.

11. Internet2. Available at: https://www.internet2.edu/products-services/advanced-networking/ (accessed 18.09.2019).

Информация об авторах

Навроцкий Ярослав Юрьевич – аспирант. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: yaroslav.navrotskiy.yn@mail.ru

Пацей Наталья Владимировна – кандидат технических наук, доцент, заведующая кафедрой программной инженерии. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: n.patsei@belstu.by

Information about the authors

Navrotsky Yaroslav Yur'yevich − PhD student. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: yaroslav.navrotskiy.yn@mail.ru

Patsei Nataliya Vladimirovna – PhD (Engineering), Associate Professor, Head of the Department of Software Engineering. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: n.patsei@belstu.by

Поступила после доработки 29.10.2019

УДК 519.86

Н. Н. Буснюк
Белорусский государственный технологический университет

ИССЛЕДОВАНИЕ ВЗАИМОЗАВИСИМОСТИ СТОИМОСТИ И ДЛИТЕЛЬНОСТИ ПРОЕКТА В СЕТЕВЫХ ЗАДАЧАХ

В зависимости от количества и качества имеющихся у организации трудовых ресурсов и сложности проекта все задачи сетевого планирования можно разбить на виды, к каждому из которых затем разрабатывать специальные эффективные методы решения.

Проект, которому соответствует сетевой граф, характеризуется двумя качественными показателями (критериями) – длительностью выполнения проекта и стоимостью проекта. Первый показатель равен длине критического пути (если веса дуг представляют продолжительности работ), второй характеризуется суммой весов всех дуг графа (если длительность выполнения работы прямо пропорциональна затратам на ее выполнение). Для случая, когда количество работников совпадает с количеством работ, второй показатель будет оптимален (минимален) при расстановке работников в соответствии с решением задачи о назначениях – нахождением совершенного паросочетания минимального веса в полном двудольном графе. Такое оптимальное решение находится точно за полиномиальное время. Возникает вопрос, насколько таким способом найденное решение близко к решению задачи по первому критерию.

В статье доказана теорема о том, что расстановка рабочих на работы в соответствии с оптимальным (минимальным) решением задачи о назначениях дает сколь угодно плохое решение задачи сетевого планирования, а также приведены примеры сетей для некоторых частных случаев дискретной задачи сетевого планирования.

Ключевые слова: сетевой график, критический путь, задача сетевого планирования, задача о назначениях, оптимизация, длительность выполнения проекта, стоимость проекта.

N. N. Busnyuk
Belarusian State Technological University

RESEARCH OF INTERDEPENDENCE OF COST AND PROJECT DURATION IN NETWORK TASKS

Depending on the quantity and quality of the organization's labor resources and the complexity of the project, all network planning tasks can be divided into types, and then special effective methods of solution can be developed for each of them.

The project, which corresponds to the network graph, is characterized by two qualitative indicators (criteria) – the duration of the project and the cost of the project. The first indicator is equal to the length of the critical path (if the weight of the arc represents the duration of the work), the second is characterized by the sum of the weights of all the arcs of the graph (if the work duration is directly proportional to the cost of its implementation). For the case when the number of employees coincides with the number of jobs, the second indicator will be optimal (minimum) when arranging the employees in accordance with the solution of the assignment problem – finding the perfect matching of minimum weight in a full bipartite graph. Such an optimal solution can be found exactly in polynomial time. The question arises of how close the solution found to the solution of the problem by the first criterion.

The article proved the theorem that the placement of workers in accordance with the optimal (minimum) solution to the assignment problem gives an arbitrarily poor solution to the network planning problem, as well as examples of networks for some special cases of the discrete network planning problem.

Key words: network chart, critical path, network planning task, assignment task, optimization, project duration, project cost.

Введение. В зависимости от количества и качества имеющихся у организации трудовых ресурсов и сложности проекта все задачи сетевого планирования (ЗСП) можно разбить на виды, к каждому из которых затем разрабатывать специальные эффективные методы решения. В [1] сетевые дискретные задачи были разделены на 4 группы:

1А. $m \geq n$, a_{ij} – константы по всем i,
2А. $m < n$, a_{ij} – константы по всем i,
1Б. $m \geq n$, a_{ij} – переменные по i,
2Б. $m < n$, a_{ij} – переменные по i,

где n – количество работ; m – количество работников; a_{ij} – элемент матрицы A длительностей выполнения работы j работником i.

Если нет разницы, кого из работников назначать на любую из работ (т. е. их качественные показатели одинаковы), и работников в достаточном количестве при любой структуре сетевого графа, то такая постановка задачи равносильна случаю 1А. Данная задача сетевого планирования решается однократным применением любого известного метода для решения *задачи сетевого планирования* (нахождение критического пути в сетевом графе) [2].

Если качественные показатели работников разные и работников достаточно для того, чтобы любая работа могла начаться сразу после завершения всех предшествующих ей работ (т. е. нет простоев работ), то критический путь в сети будет разным в зависимости от способа расстановки (назначений) работников на работы. Такая ситуация равносильна случаю 1Б.

Проект, которому соответствует сетевой граф, характеризуется двумя качественными показателями (критериями) – длительностью выполнения проекта и стоимостью проекта. Первый показатель равен длине критического пути (если веса дуг представляют продолжительности работ), второй характеризуется суммой весов всех дуг графа (если длительность выполнения работы прямо пропорциональна затратам на ее выполнение). Для случая $m = n$ второй показатель будет оптимален (минимален) при расстановке работников в соответствии с решением *задачи о назначениях* – нахождением совершенного паросочетания минимального веса в полном двудольном графе. Такое оптимальное решение находится точно за полиномиальное время. Возникает вопрос, насколько таким способом найденное решение близко к решению задачи по первому критерию.

Целью данной статьи является оценка длины критического пути при расстановке работников в соответствии с решением задачи о назначениях для произвольного сетевого графа.

Основная часть. Легко привести примеры сетей, когда расстановка работников в соответствии с минимальным решением задачи о назначениях приводит к критическому пути наименьшей длины. Первый пример – сеть содержит лишь один путь, т. е. в каждое промежуточное событие входит одна работа и из него выходит также одна работа (рис. 1).

Рис. 1. Сеть состоит из одного пути

Второй пример – сеть, в которой все дуги параллельны, т. е. сеть содержит n путей (рис. 2).

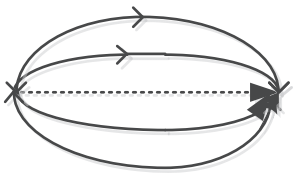


Рис. 2. Сеть содержит n путей

Теорема. При расстановке рабочих на работы в соответствии с минимальным решением задачи о назначениях в общем случае можно получить сколь угодно длинный критический путь.

Доказательство. Для простоты изложения допустим, что n – четное число, т. е. $n = 2r$. В качестве сети G возьмем граф, в котором r параллельных дуг соединяют источник со стоком, и еще есть путь длины r дуг (рис. 3).

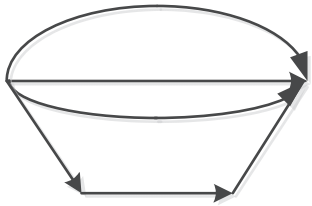


Рис. 3. Пример сети для $r = 3$

Занумеруем дуги сети следующим образом. Путям длины 1 присвоим номера от 1 до r, дугам $r + 1$ пути – номера от $r + 1$ до $2r$.

В качестве матрицы A возьмем матрицу следующего вида:

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ & & \dots & \\ 1 & 1 & \dots & 1 \\ \alpha & \alpha & \dots & \alpha \\ & & \dots & \\ \alpha & \alpha & \dots & \alpha \end{bmatrix}.$$

Все элементы первых r строк равны 1, элементы оставшихся r строк равны α, где α – сколь угодно большое натуральное число. Тогда любое паросочетание, задаваемое такой матрицей, будет минимальным веса $\alpha(r + 1)$. Если на первые r работ назначить работников с номерами от 1 до r, на остальные – оставшихся работников, то в сети получится критический путь длины αr. Если же назначить работников с номерами от 1 до r на работы с номерами от $r + 1$ до $2r$, то получим в сети пути длины r и α. Таким образом, при той же стоимости проекта время его выполнения сократилось в r (либо α) раз. Поскольку r и α – любые целые положительные числа, то теорема доказана.

(По определению сетевой граф не содержит параллельных дуг. Мы здесь рассматриваем

более общий случай. От сети с параллельными дугами можно перейти к обычному сетевому графу, разбив каждую из параллельных дуг на две, с узлом посередине (т. е. применить операцию гомеоморфизма). В нашем примере можно таким образом добавить $2r$ узлов, длина всех путей возрастет вдвое, а величина оценки сохранится (уменьшение в r либо α раз).)

Если же брать в расчет меньшее количество работников (это всегда возможно при повторных назначениях), то полученный в рассмотренном примере худший по длине критический путь может быть недостижим.

Например, если назначать лишь первых двух работников на все работы, то худший случай достижим при как можно большем простое работ на $r + 1$ пути. Если два работника выполнят вначале работы, соответствующие первым r дугам, то весь проект будет выполнен за $1{,}5r$. Если же сразу один работник возьмется за работы на $r + 1$ пути, а второй – за остальные работы, то весь проект будет выполнен за время r. Данный пример иллюстрирует, что, во-первых, увеличение количества работников в общем случае не приводит к получению лучшего решения, и, во-вторых, что уменьшение количества задействованных работников, влекущее возникновение простоев некоторых работ, тем не менее, может привести к оптимальному решению.

Таким образом, показали, что при переменных a_{ij} различным назначениям работников будут соответствовать различные решения (критические пути).

Если решать задачу методом итераций, улучшая начальный план, то не очевидно, какой начальный план (назначения работников) будет наиболее эффективным в смысле минимизации количества итераций. Скорость получения оптимума зависит от структуры сети. Если проверять каждое назначение, то в случае $m = n$ затратим $O(n!*L)$ операций, где L – сложность решения ЗСП.

Перейдем к рассмотрению случаев 2А и 2Б.

В [1] показано, что простоев работ для случая 2А можно избежать, если $m \geq g$, где g – максимально возможное число одновременно выполняемых работ в сети G. При таком варианте имеем классическую ЗСП.

Можно привести пример, когда для одних и тех же сети G и величины $m < g$ в случае 2А простой работы неизбежен, а в случае 2Б простоя можно избежать (рис. 4). При таких постановках задач в случае 2Б решением задачи будет длина критического пути графа G, а в случае 2А в сеть G нужно добавлять новые дуги, соответствующие времени простоя, и повторно находить критические пути.

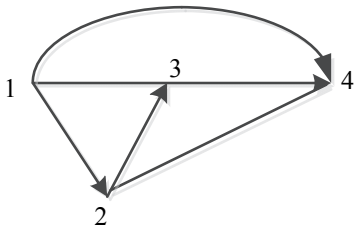


Рис. 4. Сеть из шести работ и четырех событий

Пусть работы занумерованы следующим образом (таблица).

Дуги и соответствующие им номера работ

Номер работы	1	2	3	4	5	6
Дуга графа	14	12	13	23	24	34

И пусть матрица A в случае 2А имеет вид

$$\begin{bmatrix} 3 & 1 & 2 & 3 & 4 & 2 \\ 3 & 1 & 2 & 3 & 4 & 2 \\ 3 & 1 & 2 & 3 & 4 & 2 \end{bmatrix}.$$

При такой матрице назначений (длительностей работ) работы 12 и 13 выполнятся раньше работы 14, если все три работы начать одновременно (в момент времени $t = 0$). В момент времени $t = 1$ одна из работ 23 или 24 будет простаивать минимум одну единицу времени, пока закончится работа 13. Если работник с этой работы перейдет на работу 23 (24), начнется простой работы 34. Если же работник после завершения работы 13 приступит к работе 34, одна из работ 23 или 24 будет уже простаивать минимум две единицы времени. Номера работников не играют роли, кто из них на какую работу назначен.

В случае же 2Б на этом же сетевом графе, когда производительности работников разные, простоев можно избежать.

Пусть матрица A в случае 2Б имеет вид

$$\begin{bmatrix} 3 & 1 & 2 & 3 & 4 & 2 \\ 1 & 2 & 2 & 3 & 2 & 2 \\ 2 & 3 & 2 & 1 & 3 & 4 \end{bmatrix}.$$

Здесь уже играет роль, какого работника на какую работу назначать. В момент времени $t = 0$ все три работника приступят к выполнению трех работ. Если работник i назначен на работу jk, будем применять обозначение $i \in jk$.

Пусть $2 \in 14$, $1 \in 12$, тогда $3 \in 13$. В момент $t = 1$ работники 1 и 2 закончат текущие работы и могут приступить к работам 23 и 24. А работу 34 может выполнить работник 3 сразу же после завершения работы 13.

Приведенный пример показывает, что в более сложной задаче 2Б (по количеству перебора всех вариантов назначений работников) крити-

ческий путь находится за одну итерацию (проход графа) [3]. В случае 2А критический путь нужно искать каждый раз для модифицированной сети.

Заключение. Таким образом, оптимизация стоимости проекта за счет сокращения времени выполнения работ и увеличения числа работников, выполняющих эти работы, может создать в сети критический путь, удлиняющий время выполнения всего проекта (по сравнению с достижимым минимальным временем) в α раз, где α – количество работ, либо длительность работы.

Поэтому представляется интересным создание методов поиска локально-оптимальных решений одновременно по обоим критериям: стоимости проекта и длительности проекта. Эти методы должны учитывать такие практические ситуации, когда в некоторые моменты времени простаивают работы (т. е. имеем параллельные работы); простаивают работники (количество параллельных работ уменьшилось); один и тот же работник может назначаться последовательно на несколько работ проекта [1, 4, 5].

Литература

1. Буснюк Н. Н. Разновидности задачи сетевого планирования, некоторые методы их решения и алгоритмические оценки // Труды БГТУ. Сер. 3, Физ.-мат. науки и информатика. 2019. № 2. С. 101–104.

2. Плескунов М. А. Задачи сетевого планирования. Екатеринбург: Урал. ун-т, 2014. 92 с.

3. Буснюк Н. Н., Черняк А. А. Математическое моделирование. Минск: Беларусь, 2014. 216 с.

4. Буснюк Н. Н., Новиков В. А. Метод оптимального решения задачи о назначениях в сетевом планировании // Труды БГТУ. 2016. № 6: Физ.-мат. науки и информатика. С. 170–172.

5. Буснюк Н. Н., Новиков В. А. Метод решения задачи сетевого планирования при ограниченных трудовых ресурсах // Труды БГТУ. Сер. 3, Физ.-мат. науки и информатика. 2017. № 2. С. 126–128.

References

1. Busnyuk N. N. Varieties of the network planning problem, some methods of their solution and algorithmic estimates. *Trudy BGTU* [Proceedings of BSTU], series 3, Physics and Mathematics. Informatics, 2019, no. 2, pp. 101−104 (In Russian).

2. Pleskunov M. A. *Zadachi setevogo planirovaniya* [Network planning problems]. Ekaterinburg, Ural'skiy universitet Publ., 2014. 92 p.

3. Busnyuk N. N., Chernyak A. A. *Matematicheskoye modelirovaniye* [Mathematical modeling]. Minsk, Belarus' Publ., 2014. 216 p.

4. Busnyuk N. N., Novikov V. A. Optimal solution method of assignment problem in network planning. *Trudy BGTU* [Proceedings of BSTU], 2016, no. 6: Physics and Mathematics. Informatics, pp. 170−172 (In Russian).

5. Busnyuk N. N., Novikov V. A. Solution method of network planning task with restricted labor resources. *Trudy BGTU* [Proceedings of BSTU], series 3, Physics and Mathematics. Informatics, 2017, no. 2, pp. 126−128 (In Russian).

Информация об авторе

Буснюк Николай Николаевич − кандидат физико-математических наук, доцент, доцент кафедры информационных систем и технологий. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: busnnn@belstu.by

Information about the author

Busnyuk Nikolay Nikolaevich − PhD (Physics and Mathematics), Associate Professor, Assistant Professor, the Department of Information Systems and Technology. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: busnnn@belstu.by

Поступила после доработки 19.11.2019

УДК 004.021

И. А. Литвинович, А. С. Наркевич
Белорусский государственный технологический университет

РАЗРАБОТКА И ОПТИМИЗАЦИЯ АЛГОРИТМОВ ПОИСКА ПРОФИЛЕЙ ПОЛЬЗОВАТЕЛЕЙ СОЦИАЛЬНОЙ СЕТИ ПО ФОТОГРАФИИ

В статье рассмотрены алгоритмы определения принадлежности профиля пользователя социальной сети к одному из лиц, найденному на фотографиях данного профиля. Этот процесс называется построением пользовательского профиля и направлен на нахождение однозначного соответствия лица профилю. Также в статье приведен поэтапный обзор разработанного алгоритма построения пользовательского профиля. Данный алгоритм основывается на использовании сверточной нейронной сети FaceNet для обработки фотографий и нахождения лиц, а также алгоритма кластеризации найденных лиц. Узким местом разработанного алгоритма построения профилей является кластеризация. Для получения максимальной производительности и точности были исследованы несколько популярных алгоритмов кластеризации. Сделан обзор самых популярных алгоритмов, произведены замеры производительности и надежности каждого из них. В качестве наиболее оптимального выбран алгоритм DBSCAN. Область применения алгоритма построения пользовательского профиля, описанного в статье, достаточно обширна, однако основной целью является подготовка набора пользовательских данных для последующего поиска пользовательских профилей по фотографии в определенной социальной сети. Разработанный алгоритм был успешно применен и хорошо себя зарекомендовал.

Ключевые слова: распознавание лиц, кластеризация, алгоритм, социальная сеть, характеристический вектор.

I. A. Litvinovich, A. S. Narkevich
Belarusian State Technological University

DEVELOPMENT AND OPTIMIZATION OF ALGORITHMS FOR SEARCHING PROFILES OF SOCIAL NETWORK USERS BY PHOTO

This article discusses algorithms for determining whether a profile of a social network user belongs to one of the individuals found in the photos of this profile. This process is called building a user profile and is aimed at finding a one-to-one correspondence of a person to a profile. This article also provides a step-by-step overview of the developed algorithm for constructing a user profile. This algorithm is based on the use of the FaceNet convolutional neural network for processing photos and finding faces, as well as the algorithm for clustering found faces. The bottleneck of the developed algorithm for constructing profiles is clustering. To obtain maximum performance and accuracy, several popular clustering algorithms were investigated. A review of the most popular algorithms is made, the performance and reliability of each of them are measured. The DBSCAN algorithm was chosen as the most optimal. The scope of the algorithm for constructing a user profile described in this article is quite extensive, however, the main goal is to prepare a set of user data for the subsequent search for user profiles from a photograph in a certain social network. The developed algorithm was successfully applied and proved to be good in the next stage of my work.

Key words: face recognition, clustering, algorithm, social network, characteristic vector.

Введение. Основная цель работы – это точное определение принадлежности профиля одному из лиц, найденных на фотографиях. Профиль в социальной сети – это личная страница зарегистрированного пользователя с указанием личной информации о нем, включая фото профиля (или аватар − пользовательское фото), сведения о друзьях, статусах, группах, сообществах, записи на стене, фотографии и пр. Основная сложность определения однозначного соответствия лица профилю заключается в том, что пользователи социальных сетей часто выкладывают групповые фотографии, а также фотографии знаменитостей, друзей, кумиров, а также фотографии, на которых людей нет. Для достижения цели важно решить две задачи:

1) выбор критериев, по которым необходимо отбирать фотографии для обработки;

2) выбор максимально эффективного алгоритма кластеризации лиц, найденных на фотографиях.

Кластеризация – это процесс объединения объектов с одинаковыми характеристиками в одну группу, а с различными характеристиками в другие группы. После отбора необходимых фотографий, проведения обработки данных фотографий и кластеризации результатов следует вычислить центроид лидирующего кластера,

который определяется на основании векторов, входящих в него. Центроид – точка, которая является центром кластера. В итоге объекты в одном кластере имеют схожие характеристики, что означает, что отдельный человек попадает в свой кластер.

Основная часть. В общем случае алгоритм построения пользовательского профиля содержит три этапа, каждый из которых имеет свои входные и выходные данные.

На первом этапе выполняется скачивание фотографий из пользовательского аккаунта. Входом для данного этапа является список всех фотографий пользователя, включая всю метаинформацию, такую как разрешение фотографий, количество лайков, количество комментариев, является ли фотография аватаркой.

На втором этапе происходит нахождение всех лиц на уже выбранных на предыдущем этапе фотографиях. Для распознавания лиц используется нейронная сеть FaceNet. FaceNet – это технология Google, опубликованная в 2015 г., разработанная Ф. Шрофом, Д. Калиниченко и Д. Филибином и описанная в [1]. Для распознавания предназначена обученная глубокая сверточная нейронная сеть, которая возвращает 128-размерный вектор признаков, отлично классифицирующийся. Евклидово расстояние в 128-мерном пространстве используется в качестве критерия для измерения похожести лиц.

На третьем этапе производится кластеризация всех найденных на предыдущем этапе векторов для получения некоторого количества кластеров, из которых по определенным признакам выбирается лидирующий кластер и вычисляется его центроид, представляющий собой вектор, который характеризует все лица из заданного кластера. Лидирующий кластер – это тот кластер, лица на котором принадлежат владельцу профиля. Лидирующий кластер определяется на основании размера кластера, т. е. лидирующим выбирается кластер с наибольшим количеством входящих в него векторов. После получения единого вектора (центроида) он сохраняется в базе данных с привязкой к текущему пользовательскому аккаунту. Этот вектор может быть использован как векторное представление лица владельца аккаунта.

Кластеризация применяется к математическим (или в данном случае векторным) представлениям всех лиц. Различия в этих векторах определяются на основании расстояния между точками, которым соответствуют векторы. Расстояние можно рассчитать, взяв евклидово расстояние между двумя векторами. Евклидово расстояние между двумя точками, которым соответствуют радиус-векторы $p = \{p_1, p_2, ..., p_{128}\}$ и $q = \{q_1, q_2, ..., q_{128}\}$, имеющие 128 компонент, вычисляется следующим способом:

$$d(p,q) = \sqrt{\sum_{i=1}^{128} (q_i - p_i)^2}.$$

Для кластеризации существует несколько подходов с различной производительностью. Цель этого исследования – оценка различных алгоритмов кластеризации, описанных ниже.

Пороговая кластеризация – это подход к кластеризации, предложенный в [1]. Добавление нового вектора в кластер оценивается на основании уже кластеризованных векторов с учетом расстояния между двумя точками, которые ему соответствуют. Если расстояние между новым лицом и его ближайшим соседом в наборе, лицо которого уже кластеризовано, меньше порогового значения, заданного пользователем, то лицо добавляется в существующий кластер. Если для данного лица все расстояния ниже порога, т. е. совпадений нет, то необходимо создать новый кластер. Пороговое значение играет важную роль в этом подходе. Выбор низкого порогового значения приводит к множеству ложных негативных срабатываний: пара лиц, которые имеют расстояние выше порога, но получены от разных людей. Поэтому во время эксперимента крайне важно осторожно указывать этот параметр, чтобы получить желаемый результат.

В методе кластеризации Mean Shift, описанном Команичиу и Меером [2], каждый характеристический вектор представлен в евклидовом пространстве. Основное распределение оценивается с помощью подхода, называемого оценкой плотности ядра. Это работает путем размещения ядра в каждой точке набора данных и перемещения каждой точки в направлении ее изменения. На примере кандидата x_i правило обновления для итерации t выглядит следующим образом:

$$x_i^{t+1} = x_i^t + m(x_i^t),$$

где $m(x_i)$ – средний вектор сдвига, который вычисляется для каждого лица и указывает на область максимального увеличения плотности точек:

$$m(x_i) = \frac{\sum_{x_j \in N(x_i)} K(x_j - x_i) x_i}{\sum_{x_j \in N(x_i)} K(x_j - x_i)},$$

здесь $N(x_i)$ – окрестность выборок на заданном расстоянии вокруг x_i; функция $K(x_j - x_i)$ определяет вес каждого элемента.

Преимущество этого подхода состоит в том, что он является непараметрическим алгоритмом, так как он не делает предположений о данных.

Например, (в отличие от *k*-средних) количество кластеров (или прототипов) не указывается.

Пространственная кластеризация данных с шумом на основе плотности Density-Based Spatial Clustering of Application with Noise (DBSCAN) – это алгоритм кластеризации данных, предложенный М. Эстером, Х.-П. Кригелем, Ю. Сандером и С. Сюй в 1996 г. Это также непараметрический подход. Не нужно заранее указывать точное количество кластеров. Вместо этого, учитывая набор точек данных (или вложений), DBSCAN группирует точки, которые лежат близко друг к другу на основе евклидова расстояния. Алгоритм DBSCAN требует двух параметров: минимального расстояния между двумя точками, которые могут быть сгруппированы вместе, и минимального расстояния для формирования плотной области. В случае кластеризации лиц минимальное количество точек для формирования плотной области должно быть равно 1: если это число равно 1, то лицо без близких соседей может образовывать кластер с одним элементом. Расстояние между двумя точками, которые можно сгруппировать, может варьировать, и наилучшее значение должно быть получено во время экспериментов.

Для сравнения подходов кластеризации был выбран набор данных под названием Labeled Faces in the Wild (LFW), собранный исследователями из Университета Массачусетса. Этот набор данных позволяет тестировать методы в маркированной базе данных. Маркировка означает, что личность человека на изображении известна. LFW содержит 13 233 изображения 5749 человек, где 1680 человек имеют два или более изображений, остальные имеют не более одного.

Для сравнения производительности различных алгоритмов кластеризации могут быть использованы различные оценки. Одной из таких мер качества кластеров является попарная *F*-мера, которая использовалась в работе.

Введем следующие определения. Рассмотрим два набора меток: L и C. Набор $L = \{l_1, l_2, ..., l_n\}$ содержит фактические метки для каждого лица, используемого в кластеризации. Набор $C = \{c_1, c_2, ..., c_n\}$ является выходом алгоритма кластеризации для каждого лица.

Количество истинных позитивных срабатываний (*TP*) или чувствительности состоит из пар лиц (i, j), которые правильно сгруппированы в один кластер. *TP* определяется как:

$$TP = |(i, j)|, \text{ где } c_i = c_j \text{ и } l_i = l_j.$$

Количество ложных срабатываний (*FP*) состоит из пар лиц (i, j), которые неправильно сгруппированы в одном кластере. Количество ложных срабатываний вычисляется как:

$$FP = |(i, j)|, \text{ где } c_i = c_j \text{ и } l_i \neq l_j.$$

Количество позитивных срабатываний при ожидаемом негативном (*TN*) состоит из пар, которые сгруппированы в кластер, хотя должны находиться в разных кластерах. Количество истинных негативных срабатываний рассчитывается как:

$$TN = |(i, j)|, \text{ где } c_i \neq c_j \text{ и } l_i \neq l_j.$$

Количество ложных негативных срабатываний (*FN*) состоит из пар граней (i, j), которые неправильно сгруппированы в разные кластеры. Количество ложных негативных срабатываний находится как:

$$FN = |(i, j)|, \text{ где } c_i \neq c_j \text{ и } l_i = l_j.$$

Попарная точность (*P*) определяется как отношение пар, которые правильно сгруппированы в один и тот же кластер (*TP*), ко всем парам, которые фактически были сгруппированы в один и тот же кластер с помощью алгоритма кластеризации (*TP* + *FP*). Поэтому попарная точность вычисляется как:

$$P = \frac{TP}{TP + FP}.$$

Полнота (*R*) – это доля пар, которые правильно сгруппированы в один кластер (*TP*) по всем парам одного кластера (*TP* + *FN*). Полнота рассчитывается как:

$$R = \frac{TP}{TP + FN}.$$

F-мера (*F*-measure) – характеристика, которая позволяет дать оценку одновременно по точности и полноте:

$$F = 2\frac{P \cdot R}{P + R}.$$

F-мера эффективно определяет окончательную кластеризацию, выполняемую алгоритмом кластеризации. Если алгоритм кластеризации создает единый кластер для каждого отдельного лица, точность высокая, но полнота крайне низкая. В этом случае *F*-мера задает низкую оценку производительности. Если алгоритм кластеризации создает один кластер, содержащий все лица, полнота высокая, но точность низкая. *F*-мера также указывает на плохую производительность в этом случае.

Одна из целей кластеризации – кластеризация с осторожностью, т. е. желательно кластеризовать лица только тогда, когда есть уверенность, что изображения содержат одного и того же человека. Нежелательный эффект *F*-меры заключается в том, что он может увеличиваться при росте количества ложных срабатываний.

Так как целесообразно уменьшить количество ложных срабатываний, поэтому нужно установить лимит ложных срабатываний; он не может быть больше чем 1% от количества изображений в наборе данных.

F-мера получается путем оценки истинных или ложных положительных результатов и истинных или ложных отрицательных значений.

Каждый алгоритм кластеризации имеет определенный параметр, который может варьировать, что приводит к различным результатам кластеризации. Для получения значений в таблице использовались следующие параметры:

– пороговая кластеризация: порог – 0,49 ед.;

– среднее смещение: пропускная способность – 0,38 ед.;

– DBSCAN: расстояние – 0,40 ед.

Результаты работы алгоритмов

Метод	*F*, ед.	Количество, шт.	*FP*, ед.	*P*, ед.	*R*, ед.
Разбиение вручную	1,0	4935	0	1,0	1,0
Mean Shift	0,12	4701	10	0,95	0,06
DBSCAN	0,14	4850	7	0,99	0,08
Пороговая кластеризация	0,13	4840	11	0,98	0,08

Заключение. В наборе данных Labeled Faces in the Wild алгоритм кластеризации DBSCAN показал сопоставимую производительность с алгоритмом Mean Shift. Для большого набора данных, использованного в эксперименте, производительность аналогична Mean Shift, но кластеризация DBSCAN, исходя из результатов, приведенных в таблице, лучше с точки зрения точности, полноты и небольшого количества ложных срабатываний. Также DBSCAN продемонстрировал лучшую *F*-меру (0,14 ед.), чем пороговая кластеризация (0,13 ед.). Учитывая это, делается вывод, что DBSCAN является алгоритмом кластеризации, который хорошо работает при кластеризации лиц и, следовательно, является приоритетным при выборе алгоритма кластеризации для построения пользовательского профиля.

Литература

1. Schroff F., Kalenichenko D., Philbin J. FaceNet: A Unified Embedding for Face Recognition and Clustering // CoRR abs/1503.03832 (2015). arXiv: 1503.03832. URL: http://arxiv.org/abs/1503.03832 (date of access: 02.10.2019).

2. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments / Gary B. Huang [et al.]. Amherst: University of Massachusetts, 2007. P. 1–10.

References

1. Schroff F., Kalenichenko D., Philbin J. FaceNet: A Unified Embedding for Face Recognition and Clustering. *CoRR abs/1503.03832 (2015). arXiv: 1503.03832*. Available at: http://arxiv.org/abs/1503.03832 (accessed 02.10.2019).

2. Gary B. Huang, Ramesh M., Berg T., Learned-Miller E. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. Amherst, University of Massachusetts, 2007. P. 1–10.

Информация об авторах

Литвинович Игорь Алексеевич – магистрант. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: ihar.litvinovich@ gmail.com

Наркевич Аделина Сергеевна − старший преподаватель кафедры программной инженерии. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: nas@belstu.by

Information about the authors

Litvinovich Ihar Alekseevich – Master's degree student. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: ihar.litvinovich@gmail.com

Narkevich Adelina Sergeevna − Senior Lecturer, the Department of Software Engineering. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: nas@belstu.by

Поступила после доработки 13.11.2019

ИНФОРМАЦИЯ ОТ РЕДАКЦИОННОЙ КОЛЛЕГИИ СЕРИИ 3 НАУЧНОГО ЖУРНАЛА «ТРУДЫ БГТУ»

В статье А. М. Тимофеева «Скорость передачи информации однофотонного канала связи с приемным модулем на основе счетчика фотонов с мертвым временем продлевающегося типа», опубликованной в 2019 г. в журнале № 2 на с. 79–86, при создании ее оригинала-макета произошел сдвиг горизонтальной шкалы на рис. 1, 2 (см. на с. 82–83) в сторону меньших значений средней скорости счета, что привело к несоответствию текста и рисунков. Ниже приведены исправленные рисунки, представленные автором статьи.

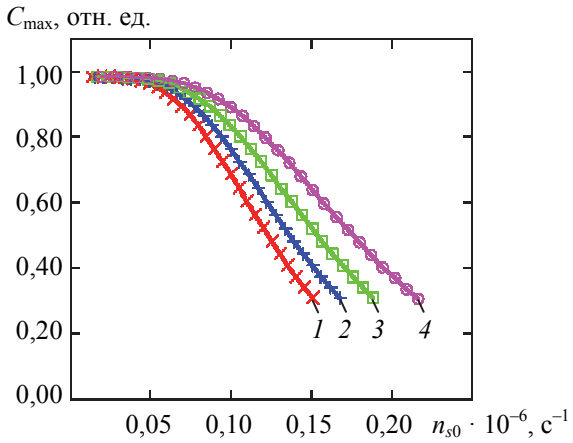


Рис. 1. Зависимость пропускной способности канала связи от средней скорости счета сигнальных импульсов n_{s0}: $N_1 = 1$, $N_2 = 7$, $n_t = 10^3$ с$^{-1}$, $\tau_b = 100$ мкс, средняя длительность мертвого времени: *1* – $\tau_d = 0$; *2* – $\tau_d = 5$ мкс; *3* – $\tau_d = 10$ мкс; *4* – $\tau_d = 15$ мкс

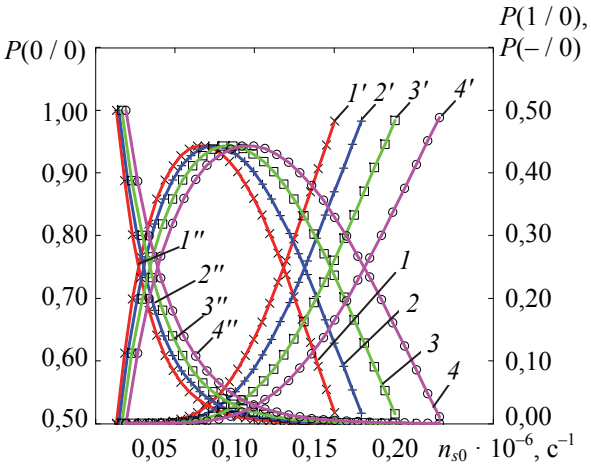


Рис. 2. Зависимости переходных вероятностей $P(0 / 0)$ (кривые *1–4*), $P(1 / 0)$ (кривые *1′–4′*) и $P(– / 0)$ (кривые *1″–4″*) от средней скорости счета сигнальных импульсов n_{s0}: $N_1 = 1$, $N_2 = 7$, $n_t = 10^3$ с$^{-1}$, $\tau_b = 100$ мкс, средняя длительность мертвого времени: *1* – $\tau_d = 0$; *2* – $\tau_d = 5$ мкс; *3* – $\tau_d = 10$ мкс; *4* – $\tau_d = 15$ мкс

Редакционная коллегия серии 3 журнала «Труды БГТУ» приносит свои извинения автору статьи А. М. Тимофееву за допущенную техническую ошибку.

СОДЕРЖАНИЕ

CONTENTS

Редактор *Е. С. Ватеичкина*
Компьютерная верстка *О. А. Солодкевич*
Корректор *Е. С. Ватеичкина*

Подписано в печать 13.03.2020. Формат 60×84 $^{1}/_{8}$.
Бумага офсетная. Гарнитура Таймс. Печать ризографическая.
Усл. печ. л. 11,5. Уч.-изд. л. 12,5.
Тираж 100 экз. Заказ .

Издатель и полиграфическое исполнение:
УО «Белорусский государственный технологический университет».
Свидетельство о государственной регистрации издателя,
изготовителя, распространителя печатных изданий
№ 1/227 от 20.03.2014.
Ул. Свердлова, 13а, 220006, г. Минск.