

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Том 21
2015
№ 8

ТЕОРЕТИЧЕСКИЙ И ПРИКЛАДНОЙ НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

Издается с ноября 1995 г.

УЧРЕДИТЕЛЬ
Издательство "Новые технологии"

СОДЕРЖАНИЕ

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ

Мочалов И. А., Хрисат М. С., Шихаб Еддин М. Я. Нечеткие уравнения в частных производных в задачах управления 563

ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ И СЕТИ

Аряшев С. И., Бобков С. Г., Зубковский П. С., Ивасюк Е. В. Реализация блока компенсированного суммирования для повышения точности аппаратных вычислений. 570

Шыхалиев Р. Г. Повышение эффективности мониторинга компьютерных сетей на основе оптимизации системы поллинга 576

ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ И ИЗОБРАЖЕНИЙ

Кухаренко Б. Г., Солнцева-Чалей М. О. Применение моделей непрерывного профиля для анализа результатов кластеризации многомерных траекторий . 585

Вольнская А. В., Сапожников И. В. Моделирование линейного мультипараметрического модема для каналов передачи информации с низким отношением сигнал/помеха 590

СИСТЕМЫ СВЯЗИ И ТЕЛЕКОММУНИКАЦИИ

Мильковский А. Г., Данилюк А. Ю., Карутин С. Н., Глухов П. Б. Основы обеспечения целостности системы ГЛОНАСС. 595

Перепелкин Д. А. Концептуальный подход динамического формирования трафика программно-конфигурируемых телекоммуникационных сетей с балансировкой нагрузки 602

БЕЗОПАСНОСТЬ ИНФОРМАЦИИ

Маркин Д. О., Комашинский В. В., Баранов И. Ю. Модель управления профилем защиты мобильного устройства при доступе к услугам с разным уровнем конфиденциальности. 611

Капулин Д. В., Дрозд О. В. Устройство аппаратного шифрования производственных данных. 618

ГЕОИНФОРМАЦИОННЫЕ СИСТЕМЫ

Бешенцев А. Н. Геоинформационная система долгосрочного мониторинга природопользования 625

Журнал в журнале НЕЙРОСЕТЕВЫЕ ТЕХНОЛОГИИ

Осипов В. Ю. Ассоциативно-пространственная адресация к памяти рекуррентных нейронных сетей 631

Лёзина И. В. Сравнительный анализ погрешности прогноза биржевых индексов методами экспоненциального сглаживания и нейросетевого моделирования 637

Главный редактор:
СТЕМПКОВСКИЙ А. Л.,
акад. РАН, д. т. н., проф.

Зам. главного редактора:
ИВАННИКОВ А. Д., д. т. н., проф.
ФИЛИМОНОВ Н. Б., д. т. н., с.н.с.

Редакционный совет:
БЫЧКОВ И. В., акад. РАН, д. т. н.
ЖУРАВЛЕВ Ю. И.,
акад. РАН, д. ф.-м. н., проф.
КУЛЕШОВ А. П.,
акад. РАН, д. т. н., проф.
ПОПКОВ Ю. С.,
чл.-корр. РАН, д. т. н., проф.
РУСАКОВ С. Г.,
чл.-корр. РАН, д. т. н., проф.
РЯБОВ Г. Г.,
чл.-корр. РАН, д. т. н., проф.
СОЙФЕР В. А.,
чл.-корр. РАН, д. т. н., проф.
СОКОЛОВ И. А., акад.
РАН, д. т. н., проф.
СУЕТИН Н. В., д. ф.-м. н., проф.
ЧАПЛЫГИН Ю. А.,
чл.-корр. РАН, д. т. н., проф.
ШАХНОВ В. А.,
чл.-корр. РАН, д. т. н., проф.
ШОКИН Ю. И.,
акад. РАН, д. т. н., проф.
ЮСУПОВ Р. М.,
чл.-корр. РАН, д. т. н., проф.

Редакционная коллегия:
АВДОШИН С. М., к. т. н., доц.
АНТОНОВ Б. И.
БАРСКИЙ А. Б., д. т. н., проф.
ВАСЕНИН В.А., д. ф.-м. н., проф.
ВИШНЕКОВ А. В., д. т. н., проф.
ГАЛУШКИН А.И., д. т. н., проф.
ДИМИТРИЕНКО Ю. И., д. ф.-м. н., проф.
ДОМРАЧЕВ В.Г., д. т. н., проф.
ЗАБОРОВСКИЙ В. С., д. т. н., проф.
ЗАГИДУЛЛИН Р. Ш., к. т. н., доц.
ЗАРУБИН В. С., д. т. н., проф.
КАРПЕНКО А. П., д. ф.-м. н., проф.
КОЛИН К. К., д. т. н., проф.
КУЛАГИН В. П., д. т. н., проф.
КУРЕЙЧИК В. М., д. т. н., проф.
КУХАРЕНКО Б. Г., к. ф.-м. н., доц.
ЛЬВОВИЧ Я. Е., д. т. н., проф.
МИХАЙЛОВ Б. М., д. т. н., проф.
НЕЧАЕВ В. В., к. т. н., проф.
ПОЛЕЩУК О. М., д. т. н., проф.
СОКОЛОВ Б. В., д. т. н., проф.
ТИМОНИНА Е. Е., д. т. н., проф.
УСКОВ В. Л., к. т. н. (США)
ФОМИЧЕВ В. А., д. т. н., проф.
ЧЕРМОШЕНЦЕВ С. Ф., д. т. н., проф.
ШИЛОВ В. В., к. т. н., доц.

Редакция:
БЕЗМЕНОВА М. Ю.
ГРИГОРИН-РЯБОВА Е. В.
ЛЫСЕНКО А. В.
ЧУГУНОВА А. В.

Информация о журнале доступна по сети Internet по адресу <http://novtex.ru/IT>.

Журнал включен в систему Российского индекса научного цитирования.

Журнал входит в Перечень научных журналов, в которых по рекомендации ВАК РФ должны быть опубликованы научные результаты диссертаций на соискание ученой степени доктора и кандидата наук.

INFORMATION TECHNOLOGIES

INFORMACIONNYYE TEHNOLOGII

Vol. 21
2015
No. 8

THEORETICAL AND APPLIED SCIENTIFIC AND TECHNICAL JOURNAL

Published since November 1995

ISSN 1684-6400

CONTENTS

INTELLIGENT SYSTEMS AND TECHNOLOGIES

- Mochalov I. A., Khrisat M. S., Shihab Eddin M. Y.** Fuzzy Partial Differential Equation in the Task of Control 563

COMPUTING SYSTEMS AND NETWORKS

- Aryashev S. I., Bobkov S. G., Zubkovskiy P. S., Ivasyuk E. V.** Development of Compensated Addition Hardware Module to Improve Calculation Accuracy 570
- Shikhaliyev R. H.** Improving the Monitoring Efficiency of Computer Networks Based on the Polling System Optimization 576

DIGITAL PROCESSING OF SIGNALS AND IMAGES

- Kukhareenko B. G., Solntseva-Chaley M. O.** Applying Continuous Profile Models for Analysis of Multidimensional Trajectory Clustering Results 585
- Volynskaya A. V., Sapozhnikov I. V.** Modelling of the Linear Multiparametrical Model for Channels of Information Transfer with the Low Relation a Signal/Noise . 590

CONNECTIONS AND TELECOMMUNICATIONS SYSTEMS

- Milkovskiy A. G., Danilyuk A. Yu., Karutin S. N., Glukhov P. B.** Ensuring GLONASS Integrity 595
- Perepelkin D. A.** Conceptual Approach of Dynamic Traffic Formation of Software-Defined Telecommunication Networks with Load Balancing 602

CRYPTOSAFETY INFORMATION

- Markin D. O., Komashinskij V. V., Baranov I. Yu.** Mobile Device Security Profile Management Model Using Access to Services with Different Privacy Level 611
- Kapulin D. V., Drozd O. V.** Hardware-Based Encryption Production Data Device . . 618

GEOINFORMATION SYSTEMS

- Beshentsev A. N.** The Geoinformation System of Long-Term Monitoring of Nature Use 625

Journal-in-journal

NEUROTECHNOLOGIES

- Osipov V. Yu.** Associative and Spatial Addressing to Memory of Recurrent Neural Networks 631
- Lyožina I. V.** Comparative Analysis of Exchange Index Forecast Error by Using Methods of Exponential Smoothing and Neural Network Modeling 637

Editor-in-Chief:

Stempkovsky A. L., Member of RAS,
Dr. Sci. (Tech.), Prof.

Deputy Editor-in-Chief:

Ivannikov A. D., Dr. Sci. (Tech.), Prof.
Filimonov N. B., Dr. Sci. (Tech.), Prof.

Chairman:

Bychkov I. V., Member of RAS,
Dr. Sci. (Tech.), Prof.
Zhuravljov Yu. I., Member of RAS,
Dr. Sci. (Phys.-Math.), Prof.
Kuleshov A. P., Member of RAS,
Dr. Sci. (Tech.), Prof.
Popkov Yu. S., Corresp. Member of RAS,
Dr. Sci. (Tech.), Prof.
Rusakov S. G., Corresp. Member of RAS,
Dr. Sci. (Tech.), Prof.
Ryabov G. G., Corresp. Member of RAS,
Dr. Sci. (Tech.), Prof.
Soifer V. A., Corresp. Member of RAS,
Dr. Sci. (Tech.), Prof.
Sokolov I. A., Member of RAS,
Dr. Sci. (Phys.-Math.), Prof.
Suetin N. V.,
Dr. Sci. (Phys.-Math.), Prof.
Chaplygin Yu. A., Corresp. Member of RAS,
Dr. Sci. (Tech.), Prof.
Shakhnov V. A., Corresp. Member of RAS,
Dr. Sci. (Tech.), Prof.
Shokin Yu. I., Member of RAS,
Dr. Sci. (Tech.), Prof.
Yusupov R. M., Corresp. Member of RAS,
Dr. Sci. (Tech.), Prof.

Editorial Board Members:

Avdoshin S. M., Cand. Sci. (Tech.), Ass. Prof.
Antonov B. I.
Barsky A. B., Dr. Sci. (Tech.), Prof.
Vasenin V. A., Dr. Sci. (Phys.-Math.), Prof.
Vishnekov A. V., Dr. Sci. (Tech.), Prof.
Galushkin A. I., Dr. Sci. (Tech.), Prof.
Dimitrienko Yu. I., Dr. Sci. (Phys.-Math.), Prof.
Domrachev V. G., Dr. Sci. (Tech.), Prof.
Zaborovsky V. S., Dr. Sci. (Tech.), Prof.
Zagidullin R. Sh., Cand. Sci. (Tech.), Ass. Prof.
Zarubin V. S., Dr. Sci. (Tech.), Prof.
Karpenko A. P., Dr. Sci. (Phys.-Math.), Prof.
Kolin K. K., Dr. Sci. (Tech.)
Kulagin V. P., Dr. Sci. (Tech.), Prof.
Kureichik V. M., Dr. Sci. (Tech.), Prof.
Kukhareenko B. G., Cand. Sci. (Phys.-Math.)
Ljvovich Ya. E., Dr. Sci. (Tech.), Prof.
Mikhailov B. M., Dr. Sci. (Tech.), Prof.
Nechaev V. V., Cand. Sci. (Tech.), Ass. Prof.
Poleschuk O. M., Dr. Sci. (Tech.), Prof.
Sokolov B. V., Dr. Sci. (Tech.)
Timonina E. E., Dr. Sci. (Tech.), Prof.
Uskov V. L. (USA), Dr. Sci. (Tech.)
Fomichev V. A., Dr. Sci. (Tech.), Prof.
Chermoshentsev S. F., Dr. Sci. (Tech.), Prof.
Shilov V. V., Cand. Sci. (Tech.), Ass. Prof.

Editors:

Bezmenova M. Yu.
Grigorin-Ryabova E. V.
Lysenko A. V.
Chugunova A. V.

Complete Internet version of the journal at site: <http://novtex.ru/IT>.

According to the decision of the Higher Certifying Commission of the Ministry of Education of Russian Federation, the journal is inscribed in "The List of the Leading Scientific Journals and Editions wherein Main Scientific Results of Theses for Doctor's or Candidate's Degrees Should Be Published"

УДК 517.97

И. А. Мочалов, д-р техн. наук, проф. МГТУ им. Н. Э. Баумана,
М. С. Хрисат, аспирант, e-mail: mohd.khirisat@fet.edu.yo, **М. Я. Шихаб Еддин**, аспирант,
Российский университет дружбы народов

Нечеткие уравнения в частных производных в задачах управления

Рассматриваются нечеткие уравнения в частных производных первого и второго порядков и алгоритм получения для них соответствующих нечетких решений S (Seikkala) и BF (Buckley-Feuring) типов. Синтезированы нечеткие оптимальные регуляторы S и BF типов.

Ключевые слова: нечеткие уравнения в частных производных, S (Seikkala) решение, BF (Buckley-Feuring) решение

Введение

В инженерно-исследовательских работах часто возникает задача решения уравнений в частных производных первого и второго порядков. Применительно к теории управления уравнения первого порядка появляются при решении задач оптимального управления и аналитического конструирования оптимальных регуляторов с использованием метода динамического программирования (метод Беллмана) [1, 2], а также при математическом конструировании (синтезе) оптимальных по критерию обобщенной работы линейных регуляторов, предложенному А. А. Красовским [3]. Уравнения второго порядка различных типов появляются, как правило, при решении разнообразных задач математической физики [4]. Среди множества этих задач выделим так называемые элементарные уравнения второго порядка [5], решения которых не выражаются в виде рядов различных типов: Фурье, Бесселя, Лежандра и т. д. К другим типам уравнений отнесем те из них, решения которых представляются в форме различных функциональных рядов [6]. Такое разделение уравнений обусловлено нерешенностью в настоящее время представления рядов в нечеткой форме. Например, неясно, каким образом определяется функция принадлежности символа " ∞ " при исследовании расходимости рядов.

Четкие уравнения имеют различные коэффициенты, определяемые экспериментальным путем, и краевые условия (начальные и граничные условия), которые обычно измеряются датчиками. Очевидно, что экспериментальные данные и измерительная информация характеризуются различного рода возмущениями, которые вносят неопределенность в математические модели процессов, что может при-

вести к значительным ошибкам в принятии решений по этим моделям.

В настоящее время теория нечетких множеств находит широкое применение в различных областях науки и техники [7–9], что часто приводит к значительному экономическому эффекту. В связи с этим одним из путей учета неопределенностей в уравнениях в частных производных является представление экспериментальной и измерительной информации в нечетких терминах. Такая трактовка возмущений приводит к нечетким уравнениям в частных производных, решению которых посвящена настоящая статья.

Базовые определения и обозначения

К ним относятся понятия: функция принадлежности; нечеткое треугольное число; нечеткая функция; нечеткие производные; нечеткая начальная задача.

Функция принадлежности. В теории нечетких множеств одним из базовых понятий является принадлежность элемента x некоторому множеству $X(x \in X \subset R_1)$. В зависимости от типа множества (четкое или нечеткое) обозначение $x \in X$ формализуется с помощью:

- характеристической функции $r^*(x)$, $x \in X \subset R_1$, $r \in \{0; 1\}$ для четкого элемента x ;
 - функции принадлежности $r(x)$, $x = x_H \in X$, $r \in [0; 1]$ для нечеткого элемента x_H .
- Они определяются следующим образом:

$$r^*(x) = \begin{cases} 1, & x \in X; \\ 0, & x \notin X; \end{cases}$$

$$r(x) = \begin{cases} r(x) \in [0; 1]; \\ \bar{r}(x) \in [0; 1]; \end{cases}$$

где $r(x)$ — многозначная функция с $\underline{r}(x)$, $\bar{r}(x)$ — "левой" и "правой" однозначными функциями (ветвями) соответственно относительно $r(x = x_H) = 1$. В уровневой форме с применением обратного отображения $r^{-1}(x)$ используется представление

$$x_H = r^{-1}(x) = x(r) = (\underline{x}(r), \bar{x}(r)/r \in [0; 1]).$$

В символической форме имеем:

$$x \in X \subset R_1 \begin{cases} \nearrow r^*(x) = \{0; 1\} \text{ при } x \text{ — четкий элемент } X; \\ \searrow r(x) = [0; 1] \text{ при } x \text{ — нечеткий элемент } X. \end{cases}$$

Совокупность нечетких элементов $\{x_H\}$ задают нечеткое множество X_H . Для обозначения приняты представления

$$x_H \Leftrightarrow r(x), r \in [0; 1] \Leftrightarrow$$

$$\Leftrightarrow (r(x), \bar{r}(x), r \in [0; 1]) \Leftrightarrow (\underline{x}(r), \bar{x}(r)/r \in [0; 1]),$$

где $\underline{x}(\cdot)$, $\bar{x}(\cdot)$ — обратные отображения соответственно для $\underline{r}(x)$, $\bar{r}(x)$.

Нечеткое треугольное число. Оно определяется посредством трех чисел $a_1 < a_2 < a_3$, $a_i \in R_1$, $i = 1, 2, 3$. График $r(x)$ для x_H в координатной плоскости (x, r) имеет форму треугольника с основанием (support) $\text{supp } x_H = [a_1, a_3]$, а высота его исходит из точки с координатами $(x = a_2, r = 0)$. Полагают что если $a_1 \geq 0$, то $x_H \geq 0$; если $a_3 \leq 0$, то $x_H \leq 0$. Арифметические операции $+$, $-$, \times , $:$ (будем обозначать $*$) для $x_{1H} = (a_{11}|a_{12}|a_{13})$ и $x_{2H} = (a_{21}|a_{22}|a_{23})$ задаются в виде

$$x_H = x_{1H} * x_{2H} = (a_{11} * a_{21}|a_{12} * a_{22}|a_{13} * a_{23}).$$

Нечеткая функция (отображение). Это отображение нечеткой области (множество) X в нечеткую область (множество) значений Y с функцией принадлежности $r_y(x)$, т. е. $X \in x \xrightarrow{r_y(x)} y \in Y$, где в общем случае x, y — векторы.

Нечеткие производные. Пусть для простоты имеем нечеткую функцию одного переменного $y_H(x) \equiv y(x, r)$, $x, y \in R_1$, $r \in [0, 1]$, т. е. $y_H(x) = (\underline{y}(x, r), \bar{y}(x, r)/r \in [0, 1])$, которая при любом $x \in I \subset R_1$ определяет нечеткое треугольное число.

Согласно общему подходу при определении производной от некоторой функции в заданном пространстве необходимо в нем задать операции "—" (вычитание или существование обратного элемента), "×" (умножение) на константу и "lim" (пределный переход относительно заданной метрики). Применение этого общего подхода к различным метрическим пространствам приводит к разнообразным нечетким производным [5]. В частности, в теории нечетких дифференциальных уравнений обычно рассматриваются следующие типы нечетких производных:

(i) производная Готшела—Воксмана (Goestshel—Voxman) — $\dot{y}_H^{GV}(x)$;

- (ii) производная Сейккалы (Seikkala) — $\dot{y}_H^S(x)$;
- (iii) производная Дубоиса—Праде (Dubois—Prade) — $\dot{y}_H^{DP}(x)$;
- (iv) производная Пури—Ралеску (Puri—Ralescu) — $\dot{y}_H^{PR}(x)$;
- (v) производная Кэндела—Фридмана—Минга (Kandel—Friedman—Ming) — $\dot{y}_H^{KFM}(x)$.

В работе [5] доказывается, что все перечисленные производные равны между собой.

Нечеткая начальная задача. Она рассматривается для производных $\dot{y}_H^{PR}(x)$, $\dot{y}_H^S(x)$, $\dot{y}_H^{KFM}(x)$, а для производных $\dot{y}_H^{GV}(x)$, $\dot{y}_H^{DP}(x)$, как правило, не рассматривается. Это обусловлено тем, что в последних двух случаях возможна ситуация, когда эти производные для какого-то $x = x^*$ не выражаются нечеткими треугольными числами, т. е. один из углов их функций принадлежности относительно основания больше 90° , и тогда эти производные не существуют. В трех предыдущих случаях эти производные всегда существуют, так как в случае, отмеченном ранее, используются нечеткие "слабые" числа [10].

Пусть имеем четкую начальную задачу

$$\dot{y}_x = f(x, y, k), y(x = 0) = c, \quad (1)$$

где $k = (k_1, \dots, k_n)$ — вектор констант; $x \in I \subset R_1$. Полагается, что $f(\cdot)$ удовлетворяет условиям существования и единственности решения четкой задачи (1), которую обозначим как $y = g(x, k, c)$, $c = \text{const}$: $c \in R_1$.

Константы $k_i, c, i = \overline{1, n}$, всегда являются неточными заданными (неопределенными) и моделируются нечеткими треугольными числами. Тогда проводим замену $k_i \rightarrow k_{iH}$, $c \rightarrow c_H$, где индекс "H" — символ нечеткости. В результате получим нечеткую начальную задачу:

$$\dot{y}_H(x) = f(x, y_H, k_H), y_H(x = 0) = c_H, \quad (2)$$

где $\dot{y}_H(x)$ — некоторое определение производной для нечеткой функции $y_H(x)$. Здесь необходимо получить решение (2) $y_H = g(x, k_H, c_H)$, которое для любого x является нечетким числом.

Типы решений

В соответствии с [5] имеют место два типа решений для нечеткой начальной задачи (2):

BFS (Buckley—Feuring solution) — $y_H^{\text{BFS}}(x)$;

SS (Seikkala solution) — $y_H^{\text{SS}}(x)$.

Между ними имеется следующая взаимосвязь:

$$\left. \begin{matrix} \exists \text{BFS} \Rightarrow \exists \text{SS} \\ \exists \text{SS} \not\Rightarrow \exists \text{BFS} \end{matrix} \right\} \Rightarrow \exists \text{BFS} \not\Leftrightarrow \exists \text{SS},$$

поэтому первоначально ищется BFS и, если оно не существует, то ищется SS. Достаточным условием существованием BFS является цепочка соотношений:

$$\dot{f}_y > 0, \dot{g}_c > 0; (\dot{f}_{k_i})(\dot{g}_{k_i}) > 0, i = \overline{1, n}, \quad (3)$$

которые формализуют условия одновременного возрастания (убывания) $g(\cdot), f(\cdot)$ относительно параметров k_i, c нечеткой начальной задачи (2).

Алгоритм решения нечеткого дифференциального уравнения

Первоначально решается четкое дифференциальное уравнение (1) и находится $g(x, k, c)$. Для $g(\cdot), f(\cdot)$ проверяются условия (3). Если они выполняются, то после фаззификации путем замены четких параметров (1) на нечеткие формулируется задача (2) и для нее находится BFS. Если хотя бы одно из условий (3) не выполняется, то BFS не существует и ищется SS, которое существует всегда, если существует решение (1).

Нечеткие уравнения в частных производных

Для уравнений первого порядка алгоритм решения, изложенный в [10], основан на решении четких уравнений [6], т. е. исходное четкое уравнение в частных производных модифицируется в четкие обыкновенные дифференциальные уравнения.

Уравнения первого порядка. Пусть имеем

$$\begin{aligned} \varphi(D_x)U(x) = B(x, k) &\Leftrightarrow \sum_{i=1}^n a_i(x)U_{x_i}(x) = B(x, k) \Leftrightarrow \\ &\Leftrightarrow a_1(x)\frac{\partial U}{\partial x_1} + \dots + a_n(x)\frac{\partial U}{\partial x_n} = B(x, k), \end{aligned} \quad (4)$$

где $\varphi(\cdot) = \sum_i a_i(\cdot)D_{x_i}$ — дифференциальный оператор полиномиального типа; $D_{x_i} = \frac{\partial}{\partial x_i}$; $k = (k_1, \dots, k_n)$ — вектор четких констант; $x = (x_1, \dots, x_n)$; U — некоторая функция.

Решаем уравнение (4) методом Лагранжа [11]. Для этого составляется вспомогательная система четких обыкновенных дифференциальных уравнений

$$\sum_{i=1}^n \bar{a}_i^{-1}(\cdot)dx_i = b^{-1}(x)dU,$$

из которой находятся n независимых интегралов

$$\Psi_i(x, k_1, \dots, k_n) = c_i, i = \overline{1, n},$$

где c_i — четкие константы интегрирования. Общее решение (4) в неявном виде равно:

$$F(\Psi_1, \dots, \Psi_n, k, c),$$

где $F(\cdot)$ — некоторая дифференцируемая функция. В частности, если U входит только в один из первых интегралов, то общее решение может быть записано в виде

$$\Psi_n(x_1, \dots, x_n, U) = f(\Psi_1, \dots, \Psi_{n-1}, k, c),$$

где f — произвольная дифференцируемая функция. Разрешив последнее уравнение относительно U , получим общее решение исходного уравнения в явном виде $U(x, k, c)$, где c находится из заданных условий,

$$U(x_1, \dots, x_n, k, c).$$

Условие (3) определяют существование BFS или SS. BFS находится из обыкновенных дифференциальных уравнений, а SS — из соответствующей системы дифференциальных уравнений с фаззификацией решений.

Уравнения второго порядка элементарного типа.

Для простоты пусть имеем вектор переменных $x = (x_1, x_2)$ и

(i) уравнение $\varphi(D_{x_1}, D_{x_2})U(x_1, x_2) = F(x_1, x_2, k)$, $k = (k_1, \dots, k_n)$ — вектор констант $k_i \in I_i \subset R_1, i = \overline{1, n}$;

(ii) граничные условия могут быть в различных формах:

- $U(x_1 = 0, x_2) = c_1$;
- $U(x_1, x_2 = 0) = c_2$;
- $U(x_1 = M, x_2) = c_3$;
- $U(x_1 = 0, x_2) = g(x_1, c_4)$;
- $U(x_1, x_2 = 0) = f_1(x_1, c_5)$;
- ..., $U_{x_1}(x_1, x_2 = 0) = f_2(x_1, c_6)$;
- $U_{x_2}(x_1 = 0, x_2) = g_2(x_2, c_7, c_8), \dots$.

Тогда при наличии (i), (ii) имеем четкое решение

$$U(x_1, x_2) = G(x_1, x_2, k, c),$$

где k, c — четкие векторы констант. Так же, как и в уравнениях первого порядка, проверяются условия (3) и далее для соответствующего нечеткого уравнения находятся BFS или SS.

Примеры

1. **Уравнение первого порядка.** Исследовать тип нечеткого решения для уравнения

$$\frac{\partial U}{\partial x_1} + \frac{\partial U}{\partial x_2} = k_1 x_1 + k_2$$

при условиях $x_2 = \lambda_1 x_1$; $U = \lambda_2 x_1^2$, где k_i, λ_i — const, $i = 1, 2$.

В канонической форме имеем уравнение

$$a_1 \frac{\partial U}{\partial x_1} + a_2 \frac{\partial U}{\partial x_2} = b, a_1 = a_2 = 1; b = k_1 x_1 + k_2,$$

поэтому в соответствии с методом Лагранжа для исходного линейного уравнения имеем вспомогательную систему

$$\left. \frac{dx_1}{a_1} \right|_{a_1=1} = \left. \frac{dx_2}{a_2} \right|_{a_2=1} = \left. \frac{dU}{b} \right|_{b=k_1x_1+k_2} \Leftrightarrow \\ \Leftrightarrow \frac{dx_1}{1} = \frac{dx_2}{1} = \frac{dU}{k_1x_1+k_2}.$$

Из первых двух уравнений системы имеем первый интеграл

$$\frac{dx_1}{1} = \frac{dx_2}{1} \Rightarrow x_1 = x_2 + c_1,$$

c_1 — константа интегрирования.

Из первого и последнего уравнений системы имеем второй интеграл

$$\frac{dx_1}{1} = \frac{dU}{k_1x_1+k_2} \Rightarrow (k_1x_1+k_2)dx_1 = dU \Rightarrow$$

$$\Rightarrow 0,5k_1x^2 + k_2x_1 = U + c_2 \Rightarrow 0,5k_1x_1^2 + k_2x_1 - c_2,$$

где c_2 — константа интегрирования. Общее решение равно

$$\begin{cases} \Psi_1(x_1, x_2) = c_1 \\ \Psi_2(x_1, x_2, U) = c_2 \end{cases} \Rightarrow \begin{cases} x_1 - x_2 = c_1; \\ 0,5k_1x^2 + k_2x_1 - U = c_2 \end{cases}$$

или в явной форме

$$U = 0,5k_1x^2 + k_2x_1 - c_2 \Big|_{x_1=x_2+c_1} = \\ = 0,5k_1(x_2 + c_1)^2 + k_2(x_2 + c_1) - c_2.$$

Находим c_1, c_2 из заданных условий, которые в параметрической форме имеют вид

$$x_1 = t, x_2 = \lambda_1 t, U = \lambda_2 t^2.$$

Подставляем их в первый и второй интегралы

$$x_1 - x_2 \Big|_{\substack{x_1=t \\ x_2=\lambda_1 t}} = c_1 \Rightarrow t - \lambda_1 t = c_1 \Rightarrow t = c_1(1 - \lambda_1)^{-1};$$

$$0,5k_1(x_2 + c_1)^2 + k_2(x_2 + c_1) - U \Big|_{\substack{x_1=t \\ x_2=\lambda_1 t \\ U=\lambda_2 t^2}} = c_2 \Rightarrow \\ \Rightarrow 0,5k_1(\lambda_1 t + c_1)^2 + k_2(\lambda_1 t + c_1) - \lambda_2 t^2 = c_2.$$

Исключаем t из последних двух соотношений, тогда в неявной форме получим:

$$F(x_1, x_2, U, k_1, k_2, \lambda_1, \lambda_2) = 0 \Leftrightarrow \\ \Leftrightarrow 0,5k_1[\lambda_1(1 - \lambda_1)^{-1}(x_1 - x_2) + c_1]^2 + \\ + k_2x_1 + U]^2 + k_2[\lambda_1(1 - \lambda_1)^{-1}(x_1 - x_2) + \\ + (0,5k_1x_1^2 + k_2x_1 + U)] - \lambda_2(1 - \lambda_1)^{-2}(x_1 - x_2)^2 - \\ - 0,5k_1x_1^2 - k_2x_1 + U = 0, \lambda_1 \neq 1, \lambda_2 \neq 1.$$

Для дальнейшего анализа упростим полученное выражение, положив $k_1 = 0, \lambda_1, \lambda_2 = 0,5$, тогда получим:

$$U(x_1, x_2) = 0,5(1 - k_2)^{-1}(x_1 - x_2)^2 - \\ - k_2(1 - k_2)^{-1}(x_1 - x_2) + k_2x_1, k_2 \neq 1. \quad (5)$$

Применительно к нечетким уравнениям в частных производных условие (3) существования BFS имеет вид:

$$\dot{F}_{k_i} \dot{G}_{k_i} > 0.$$

Расчеты дают:

- $\Psi(Dx_i)U(x_1, x_2) = F(x_1, x_2, k_2) = k_2 > 0 \Rightarrow \\ \Rightarrow \dot{F}_{k_2} = 1 > 0;$
- $\dot{G}_{k_2} = \frac{\partial}{\partial k_2} [a(1 - k_2)^{-1} + bk_2(1 - k_2)^{-1} + ck_2] = \\ = a(1 - k_2)^{-2} - b(1 - k_2)^{-2} + c, \\ a = 0,5(x_1 - x_2)^2; b = (x_1 - x_2); c = x_1, \text{ поэтому,} \\ \text{если } b = x_1 - x_2 > 0 \Leftrightarrow x_1 > x_2; \\ c|_{c=x_1} \Leftrightarrow x_1 > 0, \text{ то } \dot{G}_{k_2} > 0.$

Таким образом, $\dot{F}_{k_i} \dot{G}_{k_i} > 0$, значит, BFS существует, в противном случае SS существует. Находим BFS. Для этого в исходном уравнении проводится фаззификация путем замены $k_2 \rightarrow k_{2H}, U \rightarrow U_H$, где полагается, что $k_{2H} = (\underline{k}_2(r), \bar{k}_2(r) / r \in [0; 1])$. BFS следует из формулы (5):

$$U_H^{\text{BFS}}(x_1, x_2) = \\ = (\min_{k_{2H}} U(x_1, x_2, k_{2H}), \max_{k_{2H}} U(x_1, x_2, k_{2H})), \min_{k_{2H}} U(\cdot) = \\ = 0,5(1 - \bar{k}_2(r))^{-1}(x_1 - x_2)^2 - \\ - \bar{k}_2(r)(1 - \bar{k}_2(r))^{-1}(x_1 - x_2) + \underline{k}_2(r)x_1; \\ \max_{k_{2H}} U(\cdot) = 0,5(1 - \underline{k}_2(r))^{-1}(x_1 - x_2)^2 - \\ - \underline{k}_2(r)(1 - \underline{k}_2(r))^{-1}(x_1 - x_2) + \bar{k}_2(r)x_1;$$

SS следует из уравнений:

$$\begin{cases} \frac{\partial U}{\partial x_1} + \frac{\partial U}{\partial x_2} = \underline{k}_2(r); \\ x_1 = \underline{\lambda}_1(r)x_2; U = \underline{\lambda}_2(r)x_1^2; \\ \frac{\partial \bar{U}}{\partial x_1} + \frac{\partial \bar{U}}{\partial x_2} = \bar{k}_2(r); \\ x_1 = \bar{\lambda}_1(r)x_2; U = \bar{\lambda}_2(r)x_1^2, \end{cases}$$

откуда оно при $\underline{\lambda}_i(\cdot) = \bar{\lambda}_i(\cdot) = 0,5, i = 1, 2$, получается из (5) путем замены k_2 соответственно на $\underline{k}_2(r), \bar{k}_2(r)$:

$$U(x_1, x_2, k_2) = (\underline{U}(x_1, x_2, \underline{k}_2(r), \bar{k}_2(r)), \\ \bar{U}(x_1, x_2, \underline{k}_2(r), \bar{k}_2(r))/r \in [0; 1]).$$

2. **Элементарное уравнение второго порядка** [10]. Исследовать тип нечеткого решения для уравнения

$$\begin{cases} \frac{\partial^2 U}{\partial x_1 \partial x_2} = k_1, k_1 \in j_1 = [0; M_3], M_3 > 0; \\ I_1 = [0; M_1], M_1 > 0; I_2 = [0; j_1], (x, y) \in I_1 \times I_2; \\ U(x_1 = 0, x_2) = c_1 \sin x_2, c_1 \in L_1 = [0; M_4], M_4 > 0; \\ U(x_1, x_2 = 1) = c_2 x_1^2, c_2 \in L_2 = [0; M_5], M_5 > 0. \end{cases}$$

Находим четкое решение:

$$\begin{aligned} \frac{\partial^2 U}{\partial x_1 \partial x_2} = k_1 &\Rightarrow \frac{\partial U}{\partial x_1} = k_1 x_2 + a \Rightarrow \frac{\partial U}{\partial x_1} \Big|_{x_2=1} = c_2 x_1^2 \Rightarrow \\ &\Rightarrow k_1 \cdot 1 + a = c_2 x_1^2 \Rightarrow a = c_2 x_1^2 - k_1 \Rightarrow \\ &\Rightarrow \frac{\partial U}{\partial x_1} = k_1 x_2 + \underbrace{c_2 x_1^2 - k_1}_a \Rightarrow \frac{\partial U}{\partial x_1} = k_1(x_2 - 1) + c_2 x_1^2 \Rightarrow \\ &\Rightarrow U(x_1, x_2) = k_1 x_1(x_2 - 1) + 3^{-1} c_2 x_1^3 + b \Rightarrow \\ &\Rightarrow U|_{x_1=0} = b \Rightarrow b = c_1 \sin x_2 \Rightarrow \\ &\Rightarrow U(x_1, x_2) = k_1 x_1(x_2 - 1) + 3^{-1} c_2 x_1^3 + c_1 \sin x_2. \end{aligned}$$

После фаззификации ($k_1 \rightarrow k_{1H}; c_1 \rightarrow c_{1H}; c_2 \rightarrow c_{2H}; U(\cdot) \rightarrow U_H(\cdot)$) получим нечеткое решение

$$U_H(x_1, x_2) = k_{1H} x_1(x_2 - 1) + 3^{-1} c_{2H} x_1^3 + c_{1H} \sin x_2, \quad (6)$$

где k_{1H}, c_{1H}, c_{2H} — нечеткие положительные треугольные константы:

$$k_{1H} = (\underline{k}_1(r), \bar{k}_1(r)/r \in [0; 1]); c_{1H} = (\underline{c}_1(r), \bar{c}_1(r)/r \in [0; 1]); c_{2H} = (\underline{c}_2(r), \bar{c}_2(r)/r \in [0; 1]).$$

Исследуем типы решений. Пусть $0 \leq x_2 < 1$, тогда

$$\frac{\partial U}{\partial k_1} = 1 \cdot x_1(x_2 - 1) < 0, \text{ но } \frac{\partial F}{\partial k_1} = 1 > 0,$$

поэтому (6) не является BFS, но является SS. Если теперь

$$1 < x_2 \leq \pi, \text{ тогда } \frac{\partial U}{\partial k_1} = 1 \cdot x_1(x_2 - 1) > 0, \frac{\partial F}{\partial k_1} > 0,$$

поэтому (6) является BFS.

При $0 \leq x_2 < 1$ SS находится из уравнений:

$$\begin{cases} \frac{\partial^2 U}{\partial x_1 \partial x_2} = \underline{k}_1(r); & \frac{\partial^2 U}{\partial x_1 \partial x_2} = \bar{k}_1(r); \\ U(0, x_2) = \underline{c}_1(r) \sin x_2; & U(0, x_2) = \bar{c}_1(r) \sin x_2; \\ U(x_1, x_2 = 1) = \underline{c}_2(r) x_1^2; & U(x_1, x_2 = 1) = \bar{c}_2(r) x_1^2. \end{cases}$$

В результате из (6) получим SS:

$$U_H(x_1, x_2) = (\underline{U}(x_1, x_2, k_{1H}, c_{1H}, c_{2H}), \bar{U}(x_1, x_2, k_{1H}, c_{1H}, c_{2H})),$$

где $k_{1H}, c_{iH}, i = 1, 2$, — нечеткие константы, которые были определены ранее.

Синтез нечеткого оптимального регулятора

При решении проблемы оптимального управления применительно к задаче синтеза нечетких оптимальных линейных регуляторов появляются нечеткие нелинейные или линейные уравнения в частных производных первого порядка. Как известно [12], оптимальный регулятор в задаче

$$\dot{x} = Ax + By, x(0) = x_0,$$

$$I = 0,5 \int_{t_0}^{t_1} [x^T Sx + y^T Qy] dt + 0,5 x^T(t_1) \Lambda x(t_1) \rightarrow \min$$

определяется из уравнения Беллмана

$$\max_y \{ \dot{u}_t + (u_x^T [Ax + By] - 0,5 [x^T Sx + y^T Qy]) \} = 0;$$

$$u(t_1, x) = 0,5 x^T \Lambda x,$$

откуда

$$y(t, x) = \arg \max_y (u_x^T By - 0,5 y^T Qy),$$

где A, B — матрицы с $\dim A = (n \times n), \dim B = (n \times q)$; S, Λ — неотрицательно определенные симметричные матрицы с $\dim S = \dim \Lambda = (n \times n)$; Q — положительно определенная симметричная матрица с $\dim Q = (q \times q)$.

Это приводит к следующей структуре оптимального управления:

$$y_{\text{опт}} = Q^{-1} B^T \dot{\varphi}_x,$$

где $\varphi(t, x) = 0,5 x^T Kx, K$ — неизвестная симметричная матрица с $\dim K = (n \times n)$, которая находится из матричного дифференциального уравнения Риккати

$$\dot{K}_t = -A^T K - KA - KBQ^{-1} B^T K, K(t = t_1) = -\Lambda. \quad (7)$$

Нечеткая задача синтеза оптимального регулятора появляется, когда A, B, S, Q, Λ являются нечеткими параметрами, возникающими при неопре-

деленности в их задании и приводящие к нечеткому дифференциальному уравнению в частных производных. Определим тип нечеткого оптимального регулятора (типа S или BF) в простейшем случае.

Пусть имеем нечеткую модель электропривода

$$T_H \dot{x} = -x + y, x(t=0) = x_0$$

и функционал $I = 0,5 \int_{t_0=0}^{t_1=1} y^2(t)dt + 0,5x^2(t_1 = 1)$.

Здесь T_H — нечеткая электромеханическая постоянная времени с функцией принадлежности треугольной формы. Необходимо найти оптимальное управление $y_{\text{опт}}$, обеспечивающее $\min_y I$.

Из нечеткого уравнения Риккати, которое появляется в результате решения нечеткого нелинейного дифференциального уравнения в частных производных при

$$A = -T_H^{-1}; B = T_H^{-1}; Q = 1; S = 0; \Lambda = \lambda_H,$$

имеем из формулы (7)

$$\dot{K} = f(t, k, T_H, \lambda_H) = 2T_H^{-1}K - K^2, K(t_1 = 1) = -\lambda_H. (8)$$

Определим тип решения полученной нечеткой начальной задачи. Для этого находим решение соответствующей четкой начальной задачи

$$\dot{K} = 2T^{-1}K - K^2, K(1) = -\lambda \Rightarrow$$

$$\Rightarrow -\int (K^2 - 2T^{-1}K)^{-1} dt \Rightarrow K = 2T^{-1}(1 - ce^{2T^{-1}t})^{-1},$$

где c — константа интегрирования, определяемая из начальных условий:

$$-\lambda = 2T^{-1}(1 - ce^{2T^{-1}})^{-1} \Rightarrow$$

$$\Rightarrow c = e^{2T^{-1}} + 2T^{-1}\lambda^{-1}e^{2T^{-1}}.$$

В результате получим:

$$K = g(t, T, \lambda) = 2T^{-1} \left[1 - \frac{(e^{2T^{-1}} + 2T^{-1}\lambda^{-1}e^{2T^{-1}})}{c(\lambda)} e^{2T^{-1}t} \right]^{-1}.$$

Для исследования типов решений (8) необходимо определение знаков следующих функций [11]: $\dot{f}_k(t, K, T, \lambda)$; $\dot{g}_\lambda(t, T, \lambda)$. Вычисления дают:

$$\dot{f}_k = \frac{\partial}{\partial K}(2T^{-1}K - K^2) = \begin{cases} T^{-1} - K > 0, K(t) < T^{-1}; \\ T^{-1} - K \leq 0, K(t) \geq T^{-1}; \end{cases}$$

$$\dot{g}_\lambda = ab > 0,$$

где $a = -2T^{-1}(1 - c(\lambda)e^{2T^{-1}t})^{-2}$, $b = -2T^{-1}e^{-2T^{-1}}\lambda^{-2}$.

В результате имеем следующие варианты существования нечетких решений (8):

- вариант 1, когда $K(t) \geq T^{-1}$, тогда $\dot{f}_k > 0$, $\dot{g}_\lambda > 0$ и существует BF решение (8);
- вариант 2, когда $K(t) < T^{-1}$, тогда $\dot{f}_k \leq 0$, $\dot{g}_\lambda > 0$ и не существует BF решение (8), но существует его SS.

Пусть $K(t) < T^{-1}$, тогда имеем:

$$K_H^{\text{BFS}}(t) = (\min_r K(t, r), \max_r K(t, r) | r \in [0; 1]).$$

Нечеткое BF оптимальное управление будет равно:

$$u_H^{\text{BFS}}(t, x) = Q^{-1} \cdot B^T \cdot K_H^{\text{BFS}}(t)x \Big|_{B=T^{-1}}^{Q=1} = (\min_r u(t, r), \max_r u(t, r) | r \in [0; 1]).$$

При $K(t) \geq T^{-1}$ имеем:

$$\dot{K}_H^S = 2T_H^{-1}K_H - K_H^2 \Leftrightarrow K_H(1) = \lambda_H$$

$$\Leftrightarrow \begin{cases} \left(\begin{smallmatrix} \dot{K} \\ \dot{\bar{K}} \end{smallmatrix} \right) = \begin{pmatrix} 2T^{-1} & 0 \\ 0 & 2\bar{T}^{-1} \end{pmatrix} \begin{pmatrix} K \\ \bar{K} \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} K^2 \\ \bar{K}^2 \end{pmatrix}; \\ K(t=1) = \underline{\lambda}, \bar{K}(t=1) = \bar{\lambda}, \end{cases}$$

где $T_H = (T(r)|T|\bar{T}(r))$; $\lambda_H = (\underline{\lambda}(r)|\lambda|\bar{\lambda}(r))$, $r \in [0; 1]$ — нечеткие треугольные числа. Из решения системы получим \underline{K} , \bar{K} , поэтому $K_H^{\text{SS}}(t)$ будет равно

$$K_H^{\text{SS}}(t) = (\underline{K}(t, r), \bar{K}(t, r) | r \in [0; 1])$$

и нечеткое S оптимальное управление будет равно

$$u_H^{\text{SS}}(t, x) = Q^{-1} \cdot B^T \cdot K_H^{\text{SS}}(t)x \Big|_{B=T^{-1}}^{Q=1} = (\underline{u}(t, r), \bar{u}(t, r) | r \in [0; 1]).$$

Заключение

Нечеткие уравнения второго порядка эллиптического, гиперболического, параболического типов решаются численными методами [13, 14]. В этом случае удастся избежать представления нечетких решений в виде различных типов рядов, что в численной форме эквивалентно представлению их конечными нечеткими суммами.

Даются базовые определения теории нечетких множеств: функция принадлежности; нечеткое треугольное число; нечеткое отображение; нечеткие производные; нечеткая начальная задача.

Рассматриваются нечеткие уравнения в частных производных первого и второго порядков. В первом случае используется метод Лагранжа, который, как известно, эквивалентен решению нечетких обыкновенных уравнений, что приводит к получению BFS (Buckley-Feuring solution) и SS(Seikkala solution) для уравнений в частных производных первого по-

рядка. Для второго случая рассматриваются элементарные уравнения, для которых получают BFS и SS.

Приводятся простейшие примеры получения BFS и SS для нечетких уравнений в частных производных первого и второго порядков.

Рассмотрена задача синтеза нечеткого оптимального регулятора в виде BF и S типов.

Список литературы

1. **Ванько В. И., Ермошина О. В., Кувыркин Г. Н.** Вариационное исчисление и оптимальное управление. М.: Изд-во МГТУ им. Н. Э. Баумана, 1999.
2. **Афанасьев В. Н., Колмановский В. Б., Носов В. Р.** Математическая теория конструирования систем управления. М.: Высшая школа, 2003.
3. **Справочник** по теории автоматического управления / Под ред. А. А. Красовского. М.: Наука, 1981.
4. **Мартинсон Л. К., Малов Ю. И.** Дифференциальные уравнения математической физики. М.: Изд-во МГТУ им. Н. Э. Баумана, 2006.
5. **Buckley J. J., Feuring T.** Fuzzy differential equations // Fuzzy sets and systems. 2000. N 110. P. 43—54.

6. **Шарма Дж. Н., Сингх К.** Уравнения в частных производных для инженеров. М.: Техносфера, 2002.

7. **Асмолова Ю. Е., Мочалов И. А.** Элементы нечеткого вариационного исчисления // Вестник Российского университета дружбы народов. 2010. № 4. P. 37—43.

8. **Деменков Н. П., Мочалов И. А.** Нечеткая интерполяция // Электронное научно-техническое издание "Наука и образование". 2012. № 2.

9. **Деменков Н. П., Мочалов И. А.** Нечеткие сплайны // Вестник МГТУ им. Н. Э. Баумана. Сер. Приборостроение. 2012. № 2 (87). P. 48—58.

10. **Buckley J. J., Feuring T.** Introduction to fuzzy partial differential equations // Fuzzy sets and systems. 1999. N 105. P. 241—248.

11. **Филиппов А. Ф.** Сборник задач по дифференциальным уравнениям. М.: Наука, 1973.

12. **Пантелеев А. С., Бортакровский А. С.** Теория управления в примерах и задачах. М.: Высшая школа, 2003.

13. **Kermani M. A., Saburi F.** Numerical method for fuzzy partial differential equations // Applied mathematical sciences. 2007. Vol. 1, N 27. P. 1299—1309.

14. **Allahviranloo T., Kermani M. A.** Numerical methods for fuzzy linear partial differential equations under new definition for derivative // Iranian journal of fuzzy systems, 2010. V. 7, N 3. P. 33—50.

I. A. Mochalov, Professor Bauman Moscow State Technical University,

M. S. Khrisat, Graduate Student, e-mail: mohd.khrisat@fet.edu.yo,

M. Ya. Shihab Eddin, Graduate Student, Russian Peoples' Friendship University, Moscow

Fussy Partial Differential Equation in the Task of Control

Provides an overview on the use of partial differential equations in optimal control problems, it is noted that they are used in the methods of Bellman and Krasovsky. Second-order equations are presented in the form of two classes: Elementary and equations whose solutions are represented as different functional series. These equations have different coefficients, which are represented as fuzzy variables. In this connection, is formed and solved one-point boundary value problem on the synthesis of fuzzy optimal control method Bellman.

Next article provides basic definitions and notation of the theory of fuzzy sets, which are used for solving the fuzzy-point problem. These include the notions of membership functions; types of fuzzy numbers; fuzzy function; types of fuzzy derivatives; types of fuzzy initial value problems and algorithms for their solution.

Solved fuzzy partial differential equations of the first order and second-order elementary, then we solve the problem of synthesis of fuzzy optimal linear regulator. In accordance with the types of solutions of the initial problem of fuzzy controllers are synthesized BF (Buckley-Feuring) and S (Seikkala) types.

It is noted that the fuzzy equation of elliptic, hyperbolic and parabolic types are solved by numerical methods. This avoids making the fuzzy representation in the form of various types of rows.

The article ends with conclusions. A reference contains 15 sources. 4 of them are foreign-source.

Keywords: fuzzy systems, fuzzy partial differential equations, fuzzy regulators BF (Buckley-Feuring), fuzzy regulator S (Seikkala)

Reference

1. **Van'ko V. I., Ermoshina O. V., Kuvyrkin G. N.** Variacionnoe ischislenie i optimal'noe upravlenie. M.: Izd-vo MGTU im. N. Je. Baumana, 1999.
2. **Afanas'ev V. N., Kolmanovskij V. B., Nosov V. R.** Matematicheskaja teorija konstruirovaniya sistem upravlenija. M.: Vysshaja shkola, 2003.
3. **Spravochnik** po teorii avtomaticheskogo upravlenija. Pod red. A. A. Krasovskogo. M.: Nauka, 1981.
4. **Martinson L. K., Malov Ju. I.** Differencial'nye uravnenija matematicheskoi fiziki. M.: Izd-vo MGTU im. N. Je. Baumana, 2006.
5. **Buckley J. J., Feuring T.** Fuzzy differential equations. Fuzzy sets and systems. 2000, no. 110, pp. 43—54.
6. **Sharma Dzh. N., Singh K.** Uravnenija v chastnyh proizvodnyh dlja inzhenerov. M.: Tehnosfera, 2002.
7. **Asmolova Ju. E., Mochalov I. A.** Jelementy nechetkogo variacionnogo ischislenija. Vestnik rossijskogo universiteta družby narodov. 2010, no. 4, pp. 37—43.

8. **Demencov N. P., Mochalov I. A.** Nечetkaja interpoljacija. Jelektronnoe nauchno-tehnicheskoe izdanie "Nauka i obrazovanie". Fevral' 2012, no. 2.

9. **Demencov N. P., Mochalov I. A.** Nечetkie splajny. Vestnik MGTU im. N. Je. Baumana, Priborostroenie. 2012, no. 2 (87), pp. 48—58.

10. **Buckley J. J., Feuring T.** Introduction to fuzzy partial differential equations. Fuzzy sets and systems. 1999, no. 105, pp. 241—248.

11. **Filippov A. F.** Sbornik zadach po differencial'nym uravnenijam. M.: Nauka, 1973.

12. **Panteleev A. S., Bortakovskij A. S.** Teorija upravlenija v primerah i zadachah. M.: Vysshaja shkola. 2003.

13. **Kermani M. A., Saburi F.** Numerical method for fuzzy partial differential equations. Applied mathematical sciences. 2007, vol. 11, no. 27, pp. 1299—1309.

14. **Allahviranloo T., Kermani M. A.** Numerical methods for fuzzy linear partial differential equations under new definition for derivative. Iranian journal of fuzzysystems. 2010, vol. 7, no. 3, pp. 33—50.

УДК 004.315

С. И. Аряшев, канд. техн. наук, зав. отд., С. Г. Бобков, д-р техн. наук, зам. дир-ра,
П. С. Зубковский, зав. сектором, Е. В. Ивасюк, науч. сотр.,
Научно-исследовательский институт системных исследований РАН, e-mail: zubkovsky@cs.niisi.ras.ru

Реализация блока компенсированного суммирования для повышения точности аппаратных вычислений

Приведено обоснование необходимости разработки арифметических блоков, реализующих операции компенсированного сложения и умножения для ускорения трансцендентных функций в инженерных расчетах газовой и термодинамики, а также описана предложенная авторами модернизация блока вещественного суммирования для выполнения макроразрядных операций компенсированного сложения в библиотеке CRLIBM.

Ключевые слова: компенсированное сложение, трансцендентные функции, аппаратное ускорение

Введение

При выполнении сложных инженерных расчетов, таких как, например, моделирование химической и газодинамической фаз процессов горения и детонации, требования к арифметической производительности микропроцессора не ограничиваются только основными арифметическими операциями — сложением и умножением. Эффективная аппаратная поддержка векторных, матричных, комплексных и трансцендентных операций над вещественными числами одинарной и двойной точности оказывает большое влияние на производительность микропроцессора.

Результаты профилирования [1] программного кода одного из вариантов реализации задачи горения были получены для выделения функций, вычисление которых с применением стандартного блока вещественной арифметики занимает наибольшее время. От реализации аппаратной поддержки этих функций можно получить наибольший выигрыш в производительности всей задачи в целом. Помимо матрично-векторных операций, в первых строчках результатов профилирования оказались функции вычисления экспоненты и логарифма для вещественных чисел двойной точности.

Полная аппаратная поддержка трансцендентных функций в микропроцессоре возможна в том случае, если допускается большая погрешность вычислений и формат вещественной арифметики одинарной точности. Если требуется получение результата в формате *double* с погрешностью не более $0.5ulpr$, то целесообразно прибегнуть к программно-аппаратному подходу, так как затраты для полностью

аппаратной реализации оказываются чрезвычайно велики.

Несмотря на то что стандарт IEEE754 ничего не декларирует об элементарных (трансцендентных) функциях, существуют программные библиотеки, обеспечивающие вычисление таких функций с ошибкой не более $0.5ulpr$ мантиссы формата *double*. Это гарантирует стабильность результатов программного обеспечения, позволяя переносить выполняемый код с платформы на платформу. Следовательно, в качестве альтернативы традиционному аппаратному ускорению вычисления элементарных функций существует способ обеспечения аппаратного ускорения части алгоритмических "примитивов", задействованных в программной библиотеке. Речь идет о компенсированных функциях сложения и умножения, которые являются основным строительным элементом библиотек двойной/тройной внутренней точности (*double-double*, *triple-double*).

Например, библиотека CRLIBM [2] реализует вычисление элементарных функций с корректным округлением. Библиотека написана на языке C, поддерживает точность промежуточных вычислений на уровне *double-double/triple-double*, а также обеспечивает поддержку аппаратных операций *FMA (Fused Multiply-Add)* [3]. Необходимость оперирования с переменными, представленными двумя/тремя числами типа *double*, делает необходимым использование компенсированных операций умножения и сложения. Простейшими являются операции сложения и умножения (*Add12*, *Mul12*, *Add12Cond*, *Mul12Cond*), получающие на вход по два числа формата *double*, результатом операций является также пара чисел, сумма которых дает

точный результат. Более сложные компенсированные операции, складывающие и умножающие числа в формате *double-double/triple-double*, основаны на операциях *Add12*, *Mul12*, *Add12Cond*, *Mul12Cond*, следовательно, последние являются основным строительным элементом библиотеки, от быстродействия которых зависит скорость вычисления.

Разработка аппаратных блоков вещественной арифметики, реализующих "нижний" уровень компенсированных операций, позволит ускорить выполнение функций библиотеки компенсированных вычислений путем переноса вычислений со стандартного блока *FPU* на специализированный вычислитель. Ниже приведено описание предложенной авторами модернизации блока вещественного сумматора, позволяющей реализовать компенсированное сложение *Add12*, *Add12Cond* посредством одной операции.

Вычисление ошибки округления

Так как алгоритм вещественного сложения/вычитания включает этап округления, результат операции содержит ошибку, которая равна разности округленного числа и точного. Согласно стандарту IEEE754 значение ошибки не превышает $0,5 \text{ ulp}$ мантииссы в режиме округления *RNE* (к ближайшему/четному) и 1 ulp в остальных режимах (к плюс/минус бесконечности, к нулю). Известно, что в режиме округления *RNE* ошибка представима вещественным числом в формате, в котором проводится основная операция [4].

Данная ошибка может быть определена как программным путем, так и полностью аппаратным. Программный подход применим в случае, если ничего неизвестно про внутреннее устройство модуля сложения/вычитания. Существуют два компенсированных алгоритма суммирования пары вещественных чисел. Первый алгоритм связан со сравнением чисел и требует сортировки [5]:

$$\begin{aligned} \text{function}[x, y] &= \text{FastTwoSum}(a, b), \\ x &= fl(a + b), y = err \end{aligned} \quad (1)$$

$$\begin{cases} |a| > |b|: x = fl(a + b), y = fl(fl(a - x) + b) \\ |a| < |b|: x = fl(a + b), y = fl(fl(b - x) + a). \end{cases}$$

Операций в цепочке всего три, однако требуется сравнение, которое может привести к значительной потере производительности вследствие ошибок предсказания условных переходов.

Второй алгоритм [5] не требует сравнения исходных операндов, однако включает большее число операций (5 операций в цепочке, 6 всего):

$$\begin{aligned} \text{function}[x, y] &= \text{TwoSum}\{a, b\}, \\ x &= fl\{a + b\}, y = err \\ z &= fl(x - a), \\ y &= fl(fl(a - fl(x - z)) + fl(b - z)). \end{aligned} \quad (2)$$

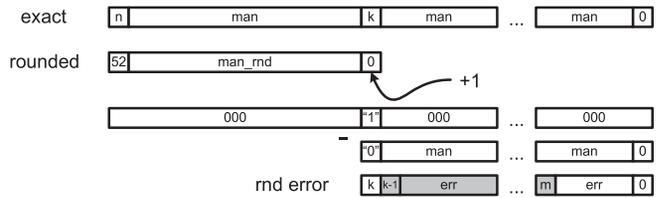


Рис. 1. Иллюстрация процесса округления и вычисления ошибки округления

Таким образом, обеспечив микропроцессор аппаратными блоками, вычисляющими погрешность округления простых операций типа сложения и умножения, можно получить существенный выигрыш по быстродействию на задачах, содержащих компенсированные операции.

На практике чаще применяется второй алгоритм, потому что он не содержит в своем составе операций сравнения, которые могут быть преобразованы в команды условного перехода, несмотря на то что его задержка составляет пять вещественных операций против трех у первого алгоритма.

Этап округления результата вещественной операции сложения обеспечивает выделение необходимой разрядности точной мантииссы результата, ее инкрементирование в зависимости от действующего режима округления (*rounding mode*) и от присутствия значащих битов в отбрасываемой части (*sticky bit*). На рис. 1 проиллюстрирован процесс округления точного результата и последующего вычисления ошибки округления. Часть $[n:k]$ точной мантииссы результата определяется действующим рабочим режимом точности: $[52:0]$ для *double*, $[23:0]$ для *single*. Для формирования ошибки округления отбрасываемая часть мантииссы должна быть корректно сформирована, особенно в части $[k - 1, m]$, разрядность которой равна разрядности основной части мантииссы.

Модернизация аппаратного модуля сумматора

Быстродействующие алгоритмы сложения/вычитания в настоящее время строятся по принципу разделения вычислительного процесса на две параллельные ветви, работающие при различных значениях разности экспонент операндов. Так, алгоритм, описанный в работе [5], делится на две части, называемые *n_path* и *far_path*.

Ниже приведены характеристики ветви *near_path*.

1. Разница экспонент равна одному из значений $\delta \in \{-1, 0, 1\}$ и вычисляется вычитанием двух младших битов экспонент исходных чисел. Сдвиг выравнивания определяется значением в одну позицию.

2. Происходит эффективное вычитание (SEFF), поэтому меньший операнд инвертируется. Используется представление *one's complement* отрицательного числа.

3. Разница мантиисс (после перестановки и предварительного сдвига) лежит в диапазоне $fsum \in (-2, 2)$

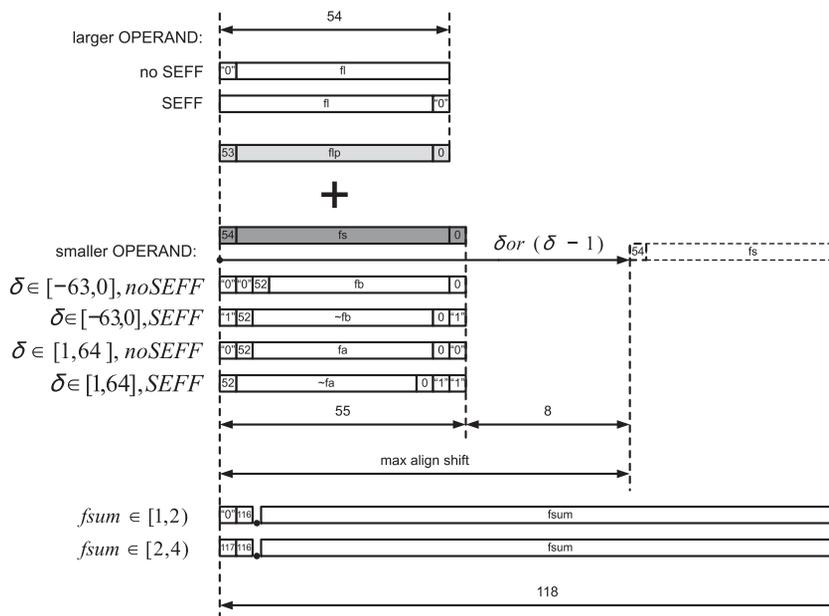


Рис. 2. Сложение выровненных мантисс в ветви *far_path*

и, следовательно, может быть точно представлена 52 битами правее десятичной точки. Поэтому нет необходимости проводить округление.

Характеристики ветви *far_path* следующие.

1. Происходит эффективное сложение.

2. Происходит эффективное вычитание с разностью мантисс $fsum \in [2, 4)$ после предварительного сдвига.

3. Абсолютное значение разности экспонент $|\delta| \geq 2$.

Методика разделения на параллельные ветви обеспечивает выигрыш по быстродействию, позволяя упростить или убрать некоторые стадии простейшего алгоритма в каждой из ветвей. Более того, полноценное округление выполняется в ветви *far_path*, так как происходит существенное увеличение внутренней разрядности мантиссы.

Процесс выравнивания мантисс ветви *far_path* проиллюстрирован на рис. 2. Схема вычисления сдвига выравнивания алгоритма имеет следующие особенности:

- расчет сдвига выравнивания выполняется в коде *one's complement*:

$$\delta = \begin{cases} \delta - 1, & \text{if } \delta > 0; \\ \delta, & \text{if } \delta \leq 0, \end{cases}$$

- сдвиг ограничен значением 64 (6 бит): 63 + предварительный сдвиг, при фактическом значении $\delta > 64$ сдвиг равен 64.

На рис. 3 показан процесс округления точной мантиссы ветви *far_path*.

Как видно из иллюстраций на рис. 2 и 3, формирование отбрасываемой части мантиссы осуществляется способом, который не подходит для вычисления ошибки округления ввиду ограничения максимального сдвига выравнивания меньшего слагаемого значением, равным 64. Отбрасываемая часть при таком подходе сигнализирует только о наличии потери точности при округлении. Это сделано для унификации поля выровненного значения меньшего операнда для сложения и вычисления признака потери точности (*sticky-bit*). Поэтому для вычисления ошибки округления необходимо корректное формирование отбрасываемой части

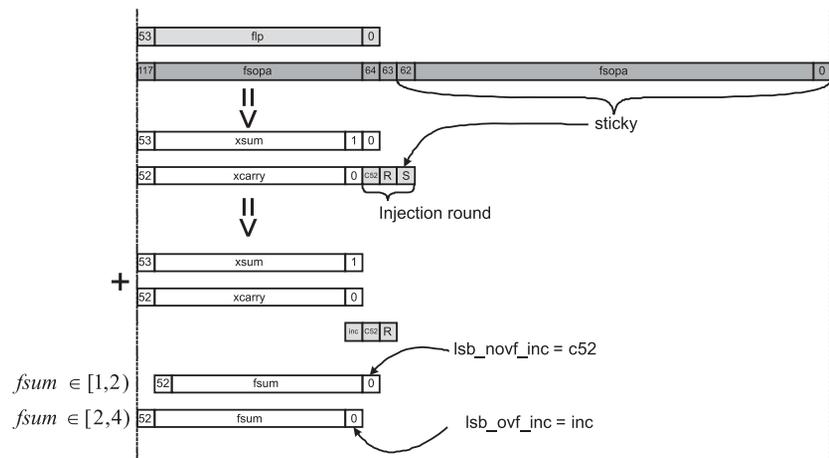


Рис. 3. Иллюстрация процесса округления в ветви *far_path*

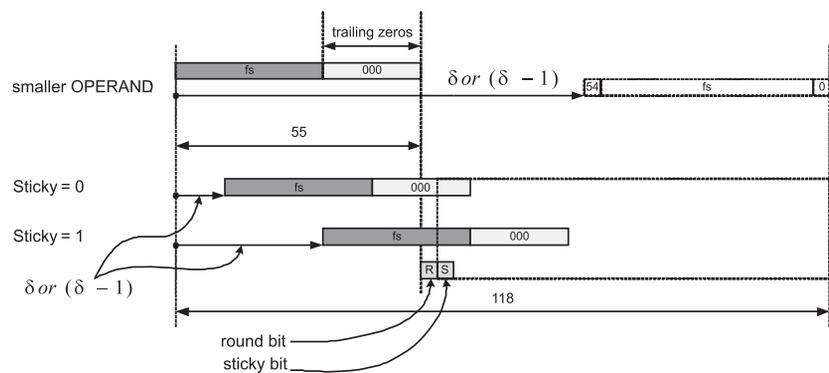


Рис. 4. Предсказание *sticky bit*

мантиссы, по крайней мере, следующих $[k - 1, m]$ бит (см. рис. 1). Для этого необходимо увеличить порог ограничения максимального сдвига выравнивания до 127 бит.

Чтобы не усложнять внутреннее устройство оригинального алгоритма и сохранить ширину промежуточного значения мантиссы в ветви *far_path* на уровне 118 бит, не теряя при этом возможности корректно вычислять *sticky-bit*, необходимо заменить его прямое вычисление (межбитовое *OR* $fsopa[62:0]$) на предсказание. Для этого необходимо сравнить значение сдвига со значением младших нулей мантиссы меньшего числа *is* (рис. 4):

$$Sticky = \begin{cases} 1, & fs_{tzer} < ashift, \\ 0, & fs_{tzer} \geq ashift. \end{cases}$$

В табл. 1 приведены выражения для предсказания *sticky-bit*. $Fa_tznum[5:0]$, $Fb_tznum[5:0]$ — значения младших нулей мантисс операндов *a* и *b*, $delta[11:0]$ — значение разности экспонент. *Sign_med* — знак разности экспонент для случая $|\delta| < 64$, **SEFF** — признак эффективного вычитания.

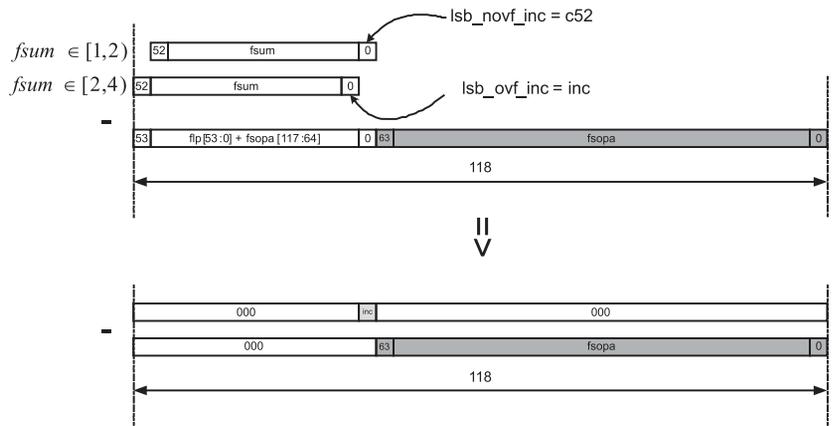


Рис. 5. Упрощение процесса вычитания при вычислении ошибки

Как было отмечено выше, ошибку округления можно вычислить как разность округленного числа и точного. Округление, в общем смысле, представляет собой инкрементирование определенной части мантиссы точного результата (определяемой разрядностью рабочего формата). В зависимости от битов, находящихся в этой части, инкрементирование может привести к распространению переноса до старших битов и привести к переполнению, а затем к корректировке экспоненты. Для упрощения аппаратного алгоритма вычисления ошибки целесообразно выполнять вычитание точного результата не из округленного числа, а вычитание отбрасываемой части мантиссы из константы округления — бита инкремента, что позволит не проводить вычитание старших полей мантисс, представляющих результат (рис. 5).

Значение выровненной мантиссы *fsopa* может быть отрицательным (*one's complement*), когда действует эффективное вычитание, поэтому действия, необходимые для вычисления ошибки округления, будут различны для разных сочетаний значений **SEFF** и **RND_INC**, табл. 2 и 3.

Модернизация упомянутого модуля будет заключаться в модификации одной из ветвей работы алгоритма — *far_path*. Список изменений следующий.

1. Увеличение порога ограничения максимального сдвига выравнивания до 127, для корректного формирования следующих 53 бит ошибки после *lsb* мантиссы основного результата.

2. Замена прямого вычисления *sticky-bit* (межбитовое *OR* $fsopa[62:0]$) на предсказание для повышения быстродействия: дополнение аппаратуры модуля счетчиками младших нулей операндов *a* и *b*, добавление компараторов определения потери точности.

3. Дополнение аппаратуры модуля инкрементом отбрасываемой части мантиссы для завершения двоичного дополнения до 2. Вычисления внутри оригинального алгоритма проводятся в форме *one's complement*.

Таблица 1

Предсказание *Sticky-bit*

<i>SIGN_MED</i>	<i>SEFF</i>	<i>Sticky-bit condition</i> (<i>Sticky-bit</i> = 1 if condition is negative)
0	0	$\{6'b0, fb_tznum[5:0]\} + \sim delta[11:0] + 12'b01$
0	1	$\{6'b0, fb_tznum[5:0]\} + \sim delta[11:0] + 12'b10$
1	0	$\{6'b0, fa_tznum[5:0]\} + delta[11:0] + 12'b10$
1	1	$\{6'b0, fa_tznum[5:0]\} + delta[11:0] + 12'b11$

Таблица 2

Выражения для вычисления ошибки округления с учетом знака

<i>RND_INC</i>	<i>SEFF</i>	<i>RND_ERR</i>
0	0	$-(fsopa[63:0]) = \sim fsopa[63:0] + 64'b1$
0	1	$-(fsopa[63:0]) + 64'b1 = fsopa[63:0]$
1	0	$\{1, 64'b0\} - \{1'b0, fsopa[63:0]\} =$ $= \{1, 64'b0\} + \{1'b1, \sim fsopa[63:0]\} + 65'b1$
1	1	$\{1, 64'b0\} - \{1'b0, (fsopa[63:0] + 64'b1)\} =$ $= \{1, 64'b0\} + \{1'b1, fsopa[63:0]\}$

Таблица 3

Выражения для вычисления модуля ошибки округления

<i>RND_INC</i>	<i>SEFF</i>	$ RND_ERR $
0	0	$fsopa[63:0]$
0	1	$fsopa[63:0] + 64'b1$
1	0	$\sim fsopa[63:0] + 64'b1$
1	1	$\sim fsopa[63:0]$

4. Дополнение аппаратуры модуля подблоком предсказания отбрасываемой инкрементированной части мантиссы и счетчиком старших нулей предсказанной суммы для ускорения процесса нормализации мантиссы ошибки.

5. Дополнение аппаратуры модуля подблоками основной и постнормализации значения ошибки.

Модернизированная структурная схема аппаратной части *far_path* алгоритма приведена на рис. 6. Жирными линиями отмечены модификации, отличающиеся от оригинального алгоритма. Предложенные модификации оригинального блока нацелены на достижение новой функциональности и сохранение быстродействия компенсированного сложения на уровне обычной операции.

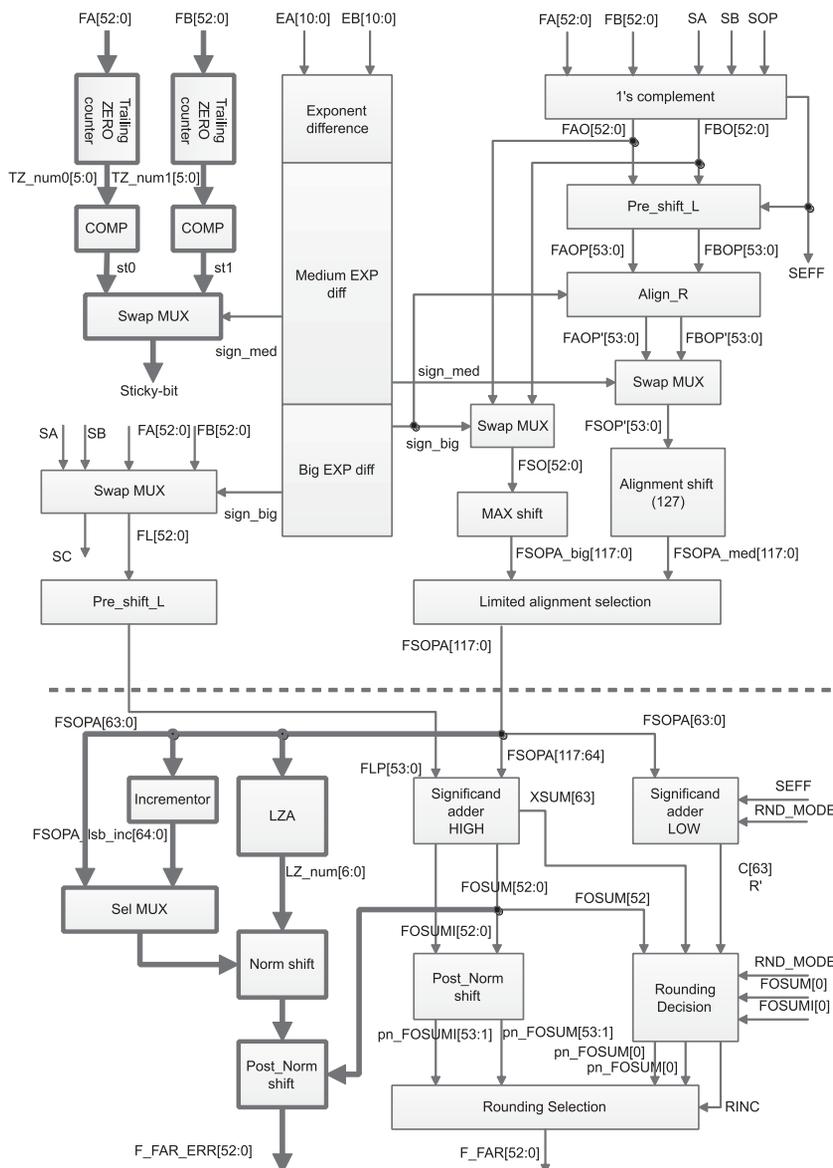


Рис. 6. Структурная схема модернизированной аппаратной *far_path* части алгоритма сложения вещественных чисел

Оценка влияния на быстродействие библиотеки

Оценить ускорение вычислений можно на примере вычисления функции библиотеки *exp*. Корректность вычисления экспоненты доказана авторами с помощью программ автоматической верификации. Стадия приведения аргумента осуществляется следующей последовательностью действий:

$$Mul12(\&s1, \&s2, msLog2DivLh, kd);$$

$$s3 = kd * msLog2Div2Lm;$$

$$s4 = s2 + s3;$$

$$s5 = x + s1;$$

$$Add12Cond = (rh, rm, s5, s4);$$

Mul12 — компенсированное умножение; *Add12Cond* — компенсированное сложение. Эти операции осуществляют приведение аргумента по формуле $r = x - k \cdot C$, где x — входной операнд; $C = \ln 2 / 4096$; $k = (x \cdot 4096 / \ln 2)$. В данном случае константа C представлена двумя числами в формате *double*: *msLog2Div2Lh* и *msLog2Div2Lm*.

Компенсированное сложение было описано выше и состоит из шести вещественных операций сложения. Компенсированное умножение является существенно более сложной операцией и содержит в своем составе 17 вещественных операций:

$$function[x, y] = TwoMul(a, b),$$

$$x = fl(a*b), y = err,$$

$$ap = fl(a*c), bp = fl(b*c),$$

$$a1 = fl(fl(a - ap) + ap),$$

$$b1 = fl(fl(b - bp) + bp),$$

$$a2 = fl(a - a1), b2 = fl(b - b1),$$

$$y = fl(fl(fl(fl(fl(a1*b1) - x) + fl(a1*b2)) + fl(a2*b1)) + fl(a2*b2)).$$

Замена этих двух макрофункций на модифицированные операции сложения и умножения, длительность которых равна обычным операциям, повысит производительность вычислений. Принимая во внимание, что макрокоманды компенсированного сложения и умножения заменяются компилятором на 6 и 17 операций типа сложения и умножения, можно сделать вывод, что приведенная часть кода будет преобразована в 26 операций. При реализации аппаратного компенсированного сложения данная часть будет занимать 21 операцию, при реализации еще и компенсированного умножения — 5 операций.

Выигрыш от аппаратной поддержки компенсированного сложения при вычислении экспоненты достигает 20 %, а реализация аппаратного компенсированного умножения увеличивает этот выигрыш до 5 раз. Таким образом, создание специализированного сопроцессора, базирующегося на блоках компенсированных арифметических операций и выполняющего набор основных трансцендентных функций для инженерных расчетов в сфере газовой и термодинамики, является перспективным направлением.

1. **Бобков С. Г., Аряшев С. И., Барских М. Е., Zubkovskiy П. С., Ивасюк Е. В.** Высокопроизводительные расширения архитектуры универсальных микропроцессоров для ускорения инженерных расчетов // Информационные технологии. 2014. № 6. С. 27–37.
2. **CRLIBM Documentation.** URL: <http://lipforge.ens-lyon.fr/www/crlibm/documentation.html>. (Дата обращения: 30.06.2014).
3. **Аряшев С. И., Zubkovskiy П. С., Ивасюк Е. В.** Реализация модуля операции "Умножение с накоплением" повышенной точности и быстродействия // 9-я Российская научно-техническая конференция "Электроника, микро- и наноэлектроника": сб. науч. трудов. — М.: МИФИ, 2007. С. 174–178.
4. **Dekker T. J.** A floating-point technique for extending the available precision // Numer. Math. 1971. N. 18. P. 224–242.
5. **Knuth D. E.** The Art of Computer Programming, Vol. 2, Seminumerical Algorithms. Reading, MA; USA: Addison-Wesley, Third edition, 1998.
6. **Seidel P. M., Even G.** Delay-optimized implementation of IEEE floating-point addition // Computers, IEEE Transactions. 2004. N. 53 (2). P. 97–113.

S. I. Aryashev, Head of Department, **S. G. Bobkov**, Deputy Director,
P. S. Zubkovskiy, Head of Sector, **E. V. Ivasyuk**, Researcher
 SRISA RAS, Moscow, e-mail: zubkovsky@cs.niisi.ras.ru

Development of Compensated Addition Hardware Module to Improve Calculation Accuracy

The article presents a development reasoning for arithmetic units that implement the compensated operations of addition and multiplication to accelerate elementary functions in scientific calculations of gas dynamics and thermodynamics.

Full hardware support for transcendental functions in the microprocessor is allowed in the case of a large calculation errors and the single-precision real arithmetic format. If double-precision result with an error of no more than 0,5 ulp is required, it is advisable to use a hardware-software approach, since the cost for a fully hardware implementation are extremely high.

Despite the fact that the standard IEEE754 did not declare anything about elementary (transcendental) functions, there are software libraries that provide the calculation of these functions with an error of not more than 0,5 ulp mantissa format double. This ensures the stability of the software results, enabling cross-platform compatibility. Therefore, there is a method for providing hardware accelerated some algorithmic "primitives", involved in a software library, as an alternative to conventional hardware acceleration calculating elementary functions. This is the compensated functions of addition and multiplication, which are the basic building block of libraries with double/triple internal precision (double-double, triple-double).

For example, the library CRLIBM implements the calculation of elementary functions with correct rounding. The library is written in C and it is support the double-double/triple-double accuracy of intermediate calculations, and also provides support for hardware operations FMA (Fused Multiply-Add).

Development of hardware floating-point units that implement the compensated lower-level operations will accelerate the implementation of compensated library functions allowing the use of specialized arithmetic unit instead of standard FPU.

Authors described the proposed improvement of the floating-point addition module to perform the addition of macro compensated CRLIBM.

The benefits of adding the compensated hardware-unit support in the exponent function calculation reached 20 percent, and implementation of compensated hardware multiplier provides the acceleration up to 5 times. Thus, the creation of a specialized coprocessor based on compensated arithmetic units and supported a set of basic transcendental functions operations for scientific calculations in the field of thermodynamics and gas is a promising direction.

Keywords: compensated addition, elementary functions, hardware acceleration

References

1. **Bobkov S. G., Aryashev S. I., Barskyh M. E., Zubkovskiy P. S., Ivasyuk E. V.** Informacionnye tehnologii, 2014, no. 6, pp. 27–37 (in Russian).
2. <http://lipforge.ens-lyon.fr/www/crlibm/documentation.html> (accessed 30.06.2014).
3. **Elektronika**, mikro- i nanojelektronika. 9-ja Rossijskaja nauchno-tehnicheskaja konferencija, 2007, pp. 174–178 (in Russian).

4. **Dekker T. J.** A floating-point technique for extending the available precision. Numer. Math., 1971, no. 18, pp. 224–242.
5. **Knuth D. E.** The Art of Computer Programming, Vol. 2, Seminumerical Algorithms. Addison-Wesley, Reading, MA, USA, 1998, 784 p.
6. **Seidel P. M., Even G.** Delay-optimized implementation of IEEE floating-point addition, Computers, IEEE Transactions on, 2004, no. 53 (2), pp. 97–113.

Р. Г. Шыхалиев, канд. техн. наук, вед. науч. сотр., ramiz@science.az
Институт Информационных Технологий НАНА, г. Баку

Повышение эффективности мониторинга компьютерных сетей на основе оптимизации системы поллинга

Предлагается методология повышения эффективности мониторинга компьютерных сетей (КС), которая позволяет минимизировать время, затрачиваемое на мониторинг. Основной целью при этом является оптимизация мониторинга сетевых узлов КС при заданных сетевых ресурсах. Для достижения этой цели предлагается использовать модель системы поллинга, оптимизируя которую можно достичь оптимизации мониторинга КС.

Ключевые слова: компьютерные сети, сетевой мониторинг, поллинг сетевых узлов, SNMP-поллинг, системы поллинга, дисциплина поллинга, среднее время ожидания

Введение

Непрерывно растущий масштаб и сложность современных компьютерных сетей (КС) увеличивают потребность в их автоматическом управлении. При этом время, затрачиваемое на принятие решений по управлению сетью, имеет существенное значение, особенно при нарушении безопасности сети и возникновении аварийных ситуаций в критических системах.

Ключевым компонентом автоматического управления КС является мониторинг в режиме реального времени, одной из задач которого является измерение характеристик сетевых узлов и каналов связи, позволяющих контролировать основные параметры сети. При этом для получения значений характеристик сетевых узлов и каналов связи может быть использован непосредственный поллинг узлов КС в режиме реального времени. Исходя из этого, мониторинг КС может быть смоделирован как система поллинга [1], что позволит в режиме реального времени непрерывно обновлять данные мониторинга и обеспечивать их актуальность.

Как правило, мониторинг КС осуществляется только в отношении основных (или критических) сетевых узлов [2], к которым относятся серверы, маршрутизаторы, коммутаторы, хабы и межсетевые экраны. Кроме основных сетевых узлов, при необходимости также может быть осуществлен мониторинг и других сетевых узлов, например рабочих станций, терминалов и т. д. Конечно, самой лучшей стратегией мониторинга КС был бы постоянный опрос всех ее узлов в реальном времени в целях получения всеобъемлющей информации о состоянии КС. Однако вследствие распространения задержки при поллинге монитором узлов сети и увеличения общего времени обработки всех запросов увеличивается время реагирования сетевой управляющей системы на возможные аварийные и аномальные ситуации, что особенно актуально в больших КС. В итоге снижается эффективность систем управле-

ния КС, т. е. эффективность систем управления сетями зависит не только от эффективности алгоритмов управления, но и от оптимальности мониторинга сети. Таким образом, оптимизация мониторинга КС, т. е. оптимизация поллинга узлов КС, является одной из главных задач по управлению КС. Следовательно, необходимо разработать новый подход к мониторингу КС для оптимизации его процесса. При этом оптимизация процесса мониторинга позволит уменьшить время реагирования сетевой управляющей системы на инциденты и аварийные ситуации. Исходя из этого, в данной работе для мониторинга КС предлагается использовать теорию систем поллинга.

В литературе имеются ряд работ, которые посвящены повышению эффективности мониторинга сетей путем оптимизации тех или иных аспектов поллинга. Например, в работе [3] авторы предлагают динамическую схему поллинга, которая позволяет оптимизировать мониторинг сети посредством оптимизации частоты опроса узлов сети. Предлагаемая схема поллинга позволяет определить частоту опроса для каждого узла сети, и для этого используются метод анализа иерархий и метод ветвей и границ. При этом основной целью является обеспечение эффективности управления сетью при достаточно высокой производительности сети.

В работе [4] для адаптивного мониторинга масштабируемых корпоративных систем авторы предлагают подход оптимизации плана поллинга, который позволяет изменять план мониторинга. Адаптивный мониторинг является перспективным методом автоматизации конфигурации сервера мониторинга корпоративных систем с динамической реконфигурацией. Однако генерация плана поллинга является NP-трудной задачей, и авторы предлагают приближенный алгоритм.

В работе [5] для повышения эффективности мониторинга и управления КС с помощью SNMP-поллинга авторы предлагают подход, основанный на эффективной передаче больших и динамически

изменяющихся MIB-баз между центром управления сетью и SNMP-агентами. Этот механизм, известный как GetModify, позволяет передавать данные, измененные только в интервалах между опросами SNMP-агентов монитором. Причем может быть вычислена скорость передачи данных, а также число пакетов и количество информации, переданной за единицу времени. В работе [6] авторами рассмотрена задача структурного синтеза систем мониторинга. Для этого предлагается система поддержки принятия решений по выбору структуры системы мониторинга и дисциплины поллинга узлов фрагмента КС.

Анализ работ, указанных выше, показал, что, несмотря на большой вклад в повышение эффективности мониторинга сетей, эти работы основываются на архитектурном решении задачи оптимизации тех или иных аспектов поллинга. Причем не рассматриваются характеристики самой системы поллинга, такие как интенсивность поступления заявок (данных мониторинга), среднее время ожидания заявок в очереди, время переключения сервера между очередями и т. д. Исходя из этого, в данной работе для повышения эффективности мониторинга КС предлагается оптимизировать саму систему поллинга, и поскольку одним из основных показателей оптимальности систем поллинга является минимальное среднее время ожидания заявок в каждой очереди, то для оптимизации поллинга узлов КС решено минимизировать среднее время ожидания заявок в очередях. На наш взгляд, это был бы единый подход для повышения эффективности мониторинга КС на основе системы поллинга.

1. Методы мониторинга компьютерных сетей

На сегодняшний день для сбора необходимой информации о состоянии КС в основном используется мониторинг. К задачам мониторинга относят: сбор значений характеристик каналов передачи и коммутирующего оборудования КС; выявление в КС аномальных ситуаций, а также узких мест; прогнозирование последствий изменений в топологии КС; мониторинг поведения пользователей КС и т. д. [7].

Существуют несколько способов мониторинга, которые позволяют минимизировать влияние процесса мониторинга на функционирование КС [8–10]. Один из способов основывается на ограничении трафика, генерируемого системой сетевого мониторинга. Метод в основном заключается в том, что сетевые администраторы в системе сетевого мониторинга могут устанавливать оптимальный предельный уровень для объема трафика, генерируемого ею. В результате трафик системы сетевого мониторинга будет минимизирован, и достоверность результатов повысится. Другой способ заключается в мониторинге не всех узлов КС, а только критических, таких как маршрутизаторы или серверы. Этот метод целесообразно применять в больших КС, где число

узлов измеряется тысячами и мониторинг всех узлов был бы бесполезным.

При мониторинге КС важен мониторинг не только сетевых устройств, но и приложений, выполняемых в них. Например, системному администратору необходимо быть уверенным, что операционные системы сетевых устройств (например, серверов, маршрутизаторов, коммутаторов и т. д.) сконфигурированы корректно и функционируют правильно. Кроме того, мониторинг приложений в основном используется для проверки текущего состояния сетевых узлов, например серверов. В зависимости от типа узла может быть осуществлен мониторинг специфических для данного узла характеристик. На основании этих характеристик можно определить состояние узла, а также факт превышения значения той или иной характеристики критического уровня. Вместе с тем при мониторинге КС должно быть принято решение о применении централизованного (т. е. мониторинг в одной точке) или распределенного мониторинга (т. е. одновременно в нескольких точках).

Нами были проанализированы некоторые существующие методы мониторинга [11–17], и SNMP (Simple Network Management Protocol) [18, 19] был выбран как основа для моделирования мониторинга КС.

SNMP является стандартным протоколом управления сетью и позволяет обеспечить interoperability систем управления сетью и широко использовать их для управления и мониторинга сетевых узлов в КС. SNMP определяет структуру иерархически организованной базы данных, которая хранится в сетевых узлах КС. В каждом узле содержится информация о предоставляемой услуге, таблице маршрутизации, а также различные статистические данные, отражающие их состояние, которые хранятся в базе данных MIB (Management Information Base) [20]. При этом должны быть измерены такие характеристики, которые описывают различные аспекты поведения сети, и в каждом узле КС могут храниться несколько параметров, отражающих его состояние. Однако, несмотря на то что SNMP является простым протоколом, у него имеются некоторые недостатки по оптимальному использованию пропускной способности сети. Например, опрос узлов сети в виде "один запрос, один ответ", необходимость идентификации объекта (OID — *object identifier*), невозможность фильтрации и т. д. увеличивают время отклика на запросы монитора, что особенно актуально в низкоскоростных сетях.

2. Системы поллинга

Системы поллинга являются разновидностью систем массового обслуживания, состоящих из нескольких очередей, которые обслуживает один сервер. При этом сервер по определенному правилу переключается между очередями и обслуживает

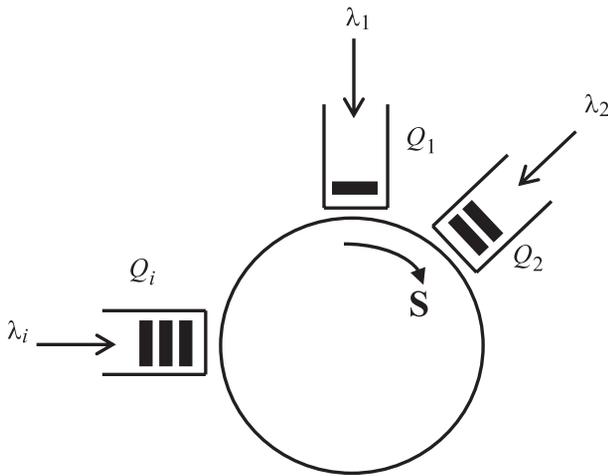


Рис. 1. Модель циклического поллинга

находящиеся в них заявки (рис. 1) [1]. Имеется большое количество литературы по исследованию систем поллинга, и в работах [21—24] приводится анализ систем поллинга с различными дисциплинами обслуживания, например с исчерпывающей (*exhaustive*), шлюзовой (*gated*) и ограниченной (*limited*) дисциплинами.

Термин "поллинг" происходит от так называемой схемы опроса, предназначенной для управления общим каналом передачи данных. В этой схеме опроса центральный компьютер (сервер) опрашивает каждый терминал (очередь) на наличие у него (клиента) какой-либо информации для передачи. При этом получивший запрос терминал передает информацию, а затем сервер переключается на следующий терминал, чтобы уже у него проверить наличие информации для передачи. В итоге сервер опрашивает все терминалы и таким образом управляет процессом передачи информации по общему каналу передачи данных.

Приведенная выше модель может быть рассмотрена как основная стандартная модель поллинга. Предполагается, что модель поллинга состоит из N очередей ($N \geq 1$) Q_1, \dots, Q_N и одного сервера S , циклически опрашивающего очереди по определенной дисциплине $D_i, i = 1, \dots, n$. При этом для переключения сервера S от очереди Q_i к очереди Q_{i+1} затрачивается определенное время, которое называется временем переключения и является независимой случайной величиной. В работе [25] приводится анализ поллинг-моделей без времени переключения, в которой предполагается, что переключения сервера происходят мгновенно.

Модель поллинга аналогична модели синхронного мультиплексирования с временным разделением (СМВР) [26]. Однако модель СМВР отличается тем, что, независимо от того, имеются ли в очередях заявки или нет, для обслуживания каждой очереди сервером выделяется фиксированное время.

Системы поллинга характеризуются процессом поступления в очереди заявок, процессом обслуживания заявок, процессом переключения сервера между очередями, размером буфера и дисциплиной обслуживания.

Процесс поступления заявок происходит в соответствии с их очередностью и независимо друг от друга. Часто в литературе эти процессы считаются пуассоновскими процессами с интенсивностью λ_i , где $i = 1, \dots, N$, т. е. временной интервал между заявками одного и того же типа имеет экспоненциальное распределение. Однако в некоторых реальных системах это утверждение может быть несостоятельным, и поэтому, исходя из сущности сбора данных мониторинга, мы рассматриваем поступление заявок с обновлением.

Процесс обслуживания заключается в том, что каждый раз, когда сервер опрашивает очередь, начинается период обслуживания, в котором обслуживаются заявки текущей очереди. Для обслуживания каждой заявки требуется определенное время. При этом предполагается, что для обслуживания одного и того же типа заявок время обслуживания является независимой, равномерно распределенной случайной величиной и не зависит от состояния системы.

Процесс переключения сервера между очередями заключается в переключении сервера с одной очереди на другую, и обычно для этого требуется некоторое время. Затрачиваемое на переключение между выбранной парой очередей время является независимой случайной распределенной величиной.

Несмотря на то что указанные стохастические временные показатели являются очень важными, они не позволяют полностью идентифицировать систему поллинга. Производительность систем поллинга также зависит от числа очередей, размера буфера, дисциплины обслуживания и дисциплины опроса.

Обычно размер буфера каждой очереди в системе поллинга предполагается неограниченным (бесконечным), в некоторых исследованиях рассматриваются также единичные буферы. Несмотря на то что предположения бесконечности не всегда совпадают с действительностью, считается, что размер буфера не является ограничивающим фактором.

Дисциплины обслуживания. Дисциплина обслуживания определяется числом заявок, которые обслуживаются в течение одного посещения сервером очереди. Внутри очереди заявки обслуживаются в порядке, определяемом дисциплиной обслуживания, причем чаще всего в порядке поступления в очередь. В литературе предлагается множество различных дисциплин обслуживания очередей. Дисциплины обслуживания могут быть классифицированы как дисциплины обслуживания с ограниченным числом заявок, в которых число заявок, обслуживаемых сервером при текущем посещении очереди,

ограничивается, и как дисциплины с ограниченным временем, в которых на посещение сервером очереди налагаются временные ограничения.

Кроме того, дисциплины обслуживания можно разделить на так называемую исчерпывающую, при которой сервер обслуживает заявки до тех пор, пока очередь не опустеет, и шлюзовую [27—29], при которой сервер обслуживает лишь те заявки, которые находились в очереди на момент подключения к ней сервера. В исчерпывающей дисциплине обслуживания вновь прибывающие в очередь заявки могут быть потенциальными кандидатами на обслуживание сервером, а в шлюзовой дисциплине они таковыми не являются. Если сервер обслуживает только те заявки, которые находились в очереди в момент начала цикла, то говорят о глобально-шлюзовой дисциплине, анализ которой проведен в работе [30].

Путем объединения различных дисциплин обслуживания может быть получено множество гибридных дисциплин. Некоторые из этих дисциплин обслуживания изучались в литературе, например, вероятностно-ограниченная дисциплина обслуживания [31]; l_i -ограниченная дисциплина, где число заявок, которое может обслуживать сервер, ограничено числом l_i , $l_i \geq 1$; l_i -уменьшающая дисциплина, при которой сервер обслуживает заявки в очереди до тех пор, пока ее длина не станет на l меньше, чем была в момент подключения сервера, $l \geq 1$.

При вероятностной дисциплине обслуживания число заявок, которое может обслужить сервер в очереди Q_i , определяется значением дискретной случайной величины ξ_i , имеющей закон распределения d_i^j , $j \geq 1$, причем закон распределения может меняться при каждом посещении очереди. Примерами вероятностных дисциплин являются дисциплина Бернулли и биномиальная дисциплина.

При дисциплине Бернулли [32] первая заявка в очереди Q_i обслуживается с вероятностью 1, а каждая последующая — с заданной вероятностью p_i . С вероятностью $1 - p_i$ сервер покидает очередь. Для данной дисциплины закон распределения имеет вид $d_i^j = p^{j-1}$, $j \geq 1$. Биномиальная дисциплина [33], при которой случайная величина ξ_i (число заявок, которое может обслужить сервер в очереди Q_i) имеет биномиальное распределение с параметрами X_i и p_i , где X_i — число заявок в очереди Q_i в момент подключения к ней сервера; p_i — некоторое число, $0 < p_i \leq 1$. Для данной дисциплины закон распределения имеет вид $d_i^j = C_{X_i}^j p_i^j (1 - p_i)^{X_i - j}$, $j = \overline{1, X_i}$, $d_i^j = 0$ для $j > X_i$. Подробная классификация дисциплин обслуживания систем поллинга дана в работе [34], где также приведены соотношения между

числом заявок в системе при различных дисциплинах обслуживания.

Способы переключения. Кроме множества дисциплин обслуживания очередей, которые описывают поведение сервера при обслуживании очереди, также имеется несколько способов переключения (выбора очереди) сервера между очередями. Основной моделью поллинга является циклическая модель, при которой сервер посещает очереди в порядке $Q_1, Q_2, \dots, Q_N, Q_1, Q_2, \dots, Q_N, \dots$. При этом время, затрачиваемое сервером на обход очередей от Q_1 до Q_N , называется циклом. Вместе с тем основная циклическая модель имеет некоторые обобщения, такие как вероятностный, периодический и приоритетный поллинги.

Вероятностный поллинг заключается в том, что сервер с вероятностью p_i , $i = \overline{1, N}$, выбирает очередь Q_i для обслуживания, при этом $\sum_{i=1}^N p_i = 1$. Воз-

можен также и другой вариант выбора очереди с вероятностью p_{ij} , $i, j = \overline{1, N}$, после посещения очереди Q_i сервер переключается к Q_j , $\sum_{j=1}^N p_{ij} = 1$, $i = \overline{1, N}$.

Периодический поллинг заключается в том, что сервер обслуживает очереди в фиксированном

порядке, установленном по таблице поллинга ($T(1), T(2), \dots, T(M)$) длины M ($M \geq N$), $T(i) \in \{1, \dots, N\}$, $i = \overline{1, M}$, в которой каждая очередь обслуживается по крайней мере один раз. При этом предполагается, что таблица поллинга содержит номера всех очередей системы, и время, затрачиваемое сервером на обход очередей от $Q_{T(1)}$ до $Q_{T(M)}$, называется циклом. В работе некоторых систем поллинга выделяют гамильтоновы цикл — это время, за которое сервер посещает все очереди, причем ровно один раз. Частными случаями периодического порядка обхода очередей являются обход типа "звезда", когда очереди обслуживаются в порядке $Q_1, Q_2, Q_1, Q_3, \dots, Q_1, Q_N$, и элеваторный порядок (порядок лифта) обхода очередей, при котором очередь обслуживается в порядке $Q_1, Q_2, \dots, Q_{N-1}, Q_N, Q_{N-1}, \dots, Q_2, Q_1$.

Приоритетный поллинг заключается в том, что система имеет очереди разных приоритетов, и некоторая очередь может быть обслужена сервером, только если более приоритетные очереди не содержат заявки.

Если все очереди системы поллинга имеют дисциплины обслуживания одного типа, то говорят о системе поллинга с данной дисциплиной обслуживания (например, с исчерпывающей, ограниченной и другими дисциплинами обслуживания). Если дисциплины обслуживания очередей различные, то говорят о системе поллинга со смешанной

дисциплиной обслуживания. В зависимости от цели и задачи управления и мониторинга КС может быть выбрана та или иная дисциплина обслуживания.

3. Модель поллинга узлов КС

Поллинг широко используется в КС для определения состояния ее узлов в режиме реального времени. Обычно во многих сетях опрос проводится периодически с фиксированной частотой и для этого используется протокол SNMP.

Исходя из того, что мониторинг КС является процессом сбора данных характеристик узлов, его можно смоделировать как систему поллинга. Для мониторинга КС предлагается простейшая абстрактная модель (рис. 2), которая состоит из центрального узла, так называемого центра мониторинга сети или монитора M и N сетевых узлов, к которым относятся хосты, серверы, маршрутизаторы, коммутаторы, хабы, межсетевые экраны и т. д. Предложенная модель поллинга узлов КС основывается на протоколе SNMP. В этой модели монитор использует SNMP-поллинг и МІВ-базу для получения

данных мониторинга от сетевых узлов. Монитор получает информацию от МІВ-базы путем периодического опроса встроенных в сетевые узлы SNMP-агентов в режиме реального времени. Как правило, опрос проводится синхронно отправлением запросов к SNMP-агентам, находящимся в узлах, и получением информации о текущем состоянии узлов КС. Вместе с тем для точного определения состояния узлов в режиме реального времени опрос узлов должен проводиться часто, и при таком мониторинге небольших КС существенные проблемы не возникают. В то же время постоянное изменение содержания МІВ-таблиц приводит к необходимости частого опроса узлов, чтобы обеспечить актуальность собранных данных мониторинга. Однако с ростом размеров и сложности КС увеличивается и размер МІВ-таблиц (например, таблицы ТСР-соединений и IP-маршрутизации), что приводит к увеличению времени опроса. Но редкий опрос узлов затрудняет понимание их состояния в режиме реального времени, а также в целом состояние КС.

Из описанной выше модели мониторинга КС (см. рис. 2) можно заключить, что процесс опроса узлов сети монитором, т. е. процессы от отправки монитором запроса к сетевым узлам до получения и обработки данных мониторинга, имеют сходство с моделью системы поллинга (см. рис. 1).

Используя это сходство, модель опроса монитором сетевых узлов может быть показана в виде системы поллинга (рис. 3). При этом для оптимизации мониторинга КС следует учитывать порядок опроса узлов сети, число обслуживаемых заявок и порядок обслуживания заявок в каждой очереди.

Математическая модель мониторинга КС (модель поллинга узлов КС) может быть описана следующим образом. Система имеет один сервер (монитор) M и N очередей (узлов сети) Q_1, \dots, Q_N с неограниченным числом мест (с неограниченным буфером) для ожидания (неограниченное число заявок может быть сохранено в каждой очереди без потерь). Предполагается, что единственный монитор M обслуживает по циклу N узлов сети, переключаясь с одного узла на другой. Рассматривается система поллинга с исчерпывающей дисциплиной обслуживания очередей, т. е. при текущем посещении сервером очереди Q_i обслуживаются все заявки (обрабатываются все данные мониторинга), находящиеся в ней, и с вероятностным переключением сервера с очереди на очередь с вероятностью $p_i, i = \overline{1, N}$, что позволит снизить объем трафика поллинга. При этом полагается, что внутри очереди заявки обслуживаются в порядке поступления, и процесс поступления заявок (данных мониторинга) является независимым пуассоновским процессом. Заявки поступают в очередь Q_i с интенсивностью λ_i , где $i = 1, \dots, N$, и общая интенсивность поступления

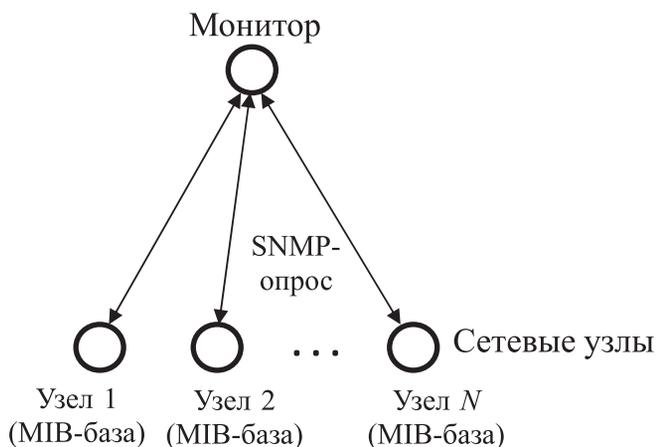


Рис. 2. Абстрактная модель мониторинга КС

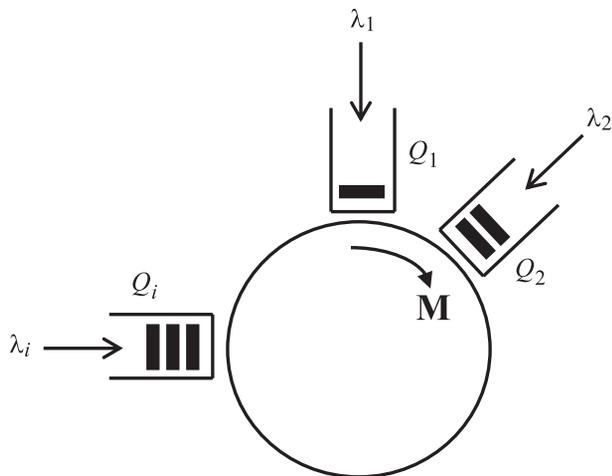


Рис. 3. Модель поллинга монитором сетевых узлов

заявок в систему равна $\lambda = \sum_{i=1}^N \lambda_i$. Заявки, поступающие к очереди Q_i , называются заявками i -класса, и время их обслуживания является независимой случайной величиной со средним значением β_i и вторым моментом $\beta_i^{(2)}$, $i = 1, \dots, N$. Загрузка очереди Q_i обозначается через ρ_i и определяется как $\rho_i := \lambda_i \beta_i$, где $i = 1, \dots, N$ и общая загрузка системы

$\rho := \sum_{i=1}^N \rho_i$. Время обслуживания очереди Q_i является независимой и равномерно распределенной величиной с функцией распределения $B_i(t)$ с моментами $b_i^{(r)} = \int_0^{\infty} t^r dB_i(t)$, $r \geq 1$, и преобразованием Лапласа—Стилтьеса (*Laplace—Stieltjes transform (LST)*)

$\beta_i(s) = \int_0^{\infty} e^{-st} dB_i(t)$, $i = \overline{1, N}$. При этом полагается, что потоки заявок и время обслуживания заявок являются независимыми. По классификации Кендалла такая система называется системой поллинга с очередями типа $M/GI/1$. Необходимым и достаточным условием стабильности (эргодичности) описываемой системы считается $\rho < 1$ [35], и это условие не зависит от времени переключения сервера из очереди в очередь, причем система является незагруженной.

Время ожидания заявки в очереди Q_i определяется временем между поступлением произвольной заявки в систему и началом ее обслуживания. Предполагается, что в очереди заявки обрабатываются по принципу "первым пришел — первым обслужен" (*First Come, First Served (FCFS)*).

Сервер после обслуживания покидает очередь Q_i , используя время переключения типа i , которое является независимой случайной величиной со средним значением t_i , вторым моментом $t_i^{(2)}$, $i = 1, \dots, N$, и преобразованием Лапласа—Стилтьеса $T_i(s)$, $i = \overline{1, N}$. При этом предполагается, что время переключения сервера не зависит от процессов поступления и обслуживания заявок в очереди Q_i .

Среднее время цикла, осуществляемого сервером, как правило, не зависит от очереди Q_i и обозначается как $E[C]$. Среднее время ожидания обслуживания заявки в очереди Q_i обозначается как $E[W_i]$. В работе [36] дифференцированием преобразования Лапласа—Стилтьеса распределения времени ожидания заявок для $E[W_i]$ найдены следующие выражения:

для исчерпывающей дисциплины

$$E[W_i] = (1 - \rho_i) \frac{E[C_i^{*2}]}{2E[C]};$$

для шлюзовой дисциплины

$$E[W_i] = (1 - \rho_i) \frac{E[C_i^2]}{2E[C]}.$$

4. Оптимизация поллинга узлов КС

Исходя из описанной в предыдущем разделе модели поллинга узлов КС, задачу оптимизации процесса мониторинга КС сводим к задаче оптимизации модели поллинга узлов КС. Как было показано выше, системы поллинга в основном характеризуются процессом поступления в очереди заявок, процессом обслуживания заявок, процессом переключения сервера между очередями. Причем каждый процесс характеризуется конкретным параметром, например, процесс поступления характеризуется интенсивностью поступления заявок (данных мониторинга), процесс обслуживания заявок — средним временем ожидания заявок в очереди, а процесс переключения сервера между очередями — временем переключения сервера между очередями. При этом для оптимизации процесса мониторинга КС необходимо динамическое управление поступлением потока данных мониторинга в монитор и дисциплиной обслуживания на основании параметров состояния входного потока, таких как статический приоритет, критичное время жизни заявки, размер штрафа за потерю заявки и т. д. При этом основной задачей является обеспечение в режиме реального времени непрерывности поступления данных мониторинга в монитор и их актуальности. Однако в данной работе вопросы штрафа за потерю заявок не рассматриваются.

Обеспечение непрерывности и актуальности данных мониторинга в режиме реального времени возможно при минимизации среднего времени пребывания заявок в очередях. При этом среднее время пребывания заявки в очереди Q_i может быть определено следующим выражением:

$$E[TS_i] = E[W_i] + \beta_i,$$

где $E[W_i]$ — среднее время ожидания обслуживания заявки в очереди Q_i ; β_i — среднее время обслуживания заявки в очереди Q_i .

Если порог времени пребывания заявки в очереди Q_i (или критичное время жизни заявки) обозначим через TS_i^* , то для минимизации времени пребывания заявки в очереди Q_i необходимо, чтобы выполнялось условие $E[TS_i] \leq TS_i^*$. Учитывая то, что время обслуживания является независимой случайной величиной, минимизация среднего времени пребывания заявки в очереди Q_i может быть достигнута минимизацией среднего времени ожидания обслуживания заявки $E[W_i]$. При этом с минимизацией времени ожидания обслуживания приоритет заявки может быть повышен, поскольку со временем вероятность превышения заявкой порога TS_i^* может увеличиваться. Однако не всегда удастся по-

лучить явные формулы для вычисления среднего времени ожидания заявок в каждой очереди и в работах по исследованию систем поллинга обычно находят приближенные формулы. Иногда задача нахождения средних времен ожидания сводится к нахождению взвешенной суммы этих показателей.

Для оптимизации рассмотренной выше модели поллинга узлов КС используем полученные в работе [37] оптимальные значения вероятностей p_i^* , $i = \overline{1, N}$, для системы поллинга с исчерпывающей или шлюзовой дисциплиной обслуживания очередей и с вероятностной дисциплиной задаваемой вероятности p_i , $i = \overline{1, N}$, минимизирующей взвешенную сумму средних времен ожидания заявок в очереди Q_i :

$$\sum_{i=1}^N \rho_i E[W_i] = \rho \frac{\sum_{i=1}^N \lambda_i \beta_i^{(2)}}{2(1-\rho)} - \frac{t_i}{1-\rho} \sum_{k \in e} \frac{\rho_k^2}{p_k} + \frac{t_i}{1-\rho} \sum_{k=1}^N \frac{\rho_k}{p_k} - \rho t_i + \rho \frac{t_i^{(2)}}{t_i},$$

где e — множество номеров очередей с исчерпывающим обслуживанием.

Задача оптимизации заключается в следующем:

$$\sum_{i=1}^N \rho_i E[W_i] \xrightarrow{p_1, \dots, p_N} \min;$$

$$\sum_{i=1}^N p_i = 1, p_1 \geq 0, \dots, p_N \geq 0.$$

По сути, она является задачей оптимальной маршрутизации сервера при переключении между очередями, и нетрудно видеть, что минимизация этой целевой функции эквивалентна минимизации загрузки системы.

Несмотря на то что неявное выражение для взвешенной суммы средних времен ожидания заявок в очереди Q_i известно, его непосредственная оптимизация невозможна. Эта задача является классической задачей нелинейной оптимизации с линейными ограничениями, которая решается с помощью метода множителей Лагранжа. Решение этой задачи имеет следующий вид [38]:

$$p_k^* = \frac{\sqrt{\rho_k(1-\rho_k)}}{\sum_{j \in e} \sqrt{\rho_j(1-\rho_j)} + \sum_{j \in g} \sqrt{\rho_j}}, k \in e;$$

$$p_k^* = \frac{\sqrt{\rho_k}}{\sum_{j \in e} \sqrt{\rho_j(1-\rho_j)} + \sum_{j \in g} \sqrt{\rho_j}}, k \in g,$$

где g является множеством номеров очередей с шлюзовым обслуживанием.

Современные КС являются настолько сложными и неопределенными системами, что это затрудняет их описание в виде детерминированных моделей.

В работе предлагается методология повышения эффективности мониторинга КС, которая позволит уменьшить время, затрачиваемое на мониторинг, и обеспечит непрерывность обновления и актуальность данных мониторинга. Для этого используется модель системы поллинга и предлагается модель поллинга узлов КС.

Предполагается, что основным показателем оптимальности систем поллинга является минимизация среднего времени ожидания заявок в очереди. Поэтому задача оптимизации процесса мониторинга КС была сведена к задаче оптимизации модели поллинга узлов КС посредством минимизации среднего времени ожидания заявок в очереди.

Список литературы

1. **Levy H., Sidi M.** Polling systems: applications, modeling, and optimization. IEEE Transactions on Communications, 1990, vol. 38, no. 10, pp. 1750–1760.
2. **Network Monitoring and Analysis. A Protocol Approach to Troubleshooting** / Ed. Wilson. New Jersey: Prentice Hall, Inc. Upper Saddle River, 2002. 350 p.
3. **Gao F., Gutierrez J.** A Dynamic Polling Scheme for the Network Monitoring Problem // Proc. of the 7th Asia-Pacific Network Operations and Management Symposium (APNOMS 2003), Fukuoka, Japan, October, 2003. P. 489–500.
4. **Fumio Machida, Masahiro Kawato, Yoshiharu Maeno.** Polling Schedule Optimization for Adaptive Monitoring to Scalable Enterprise Systems // International Journal On Advances in Intelligent Systems. 2008. Vol. 1, no. 1. P. 11–22.
5. **Park S., Park M.** An efficient transmission for large MIB tables in polling-based SNMP // Proc. of the 10th International Conference on Telecommunications. 2003. Vol. 1. P. 246–252.
6. **Скатков А. В., Воронин Д. Ю., Данильчук Д. Н.** Система поддержки принятия решений по выбору структуры системы мониторинга // Технология и конструирование в электронной аппаратуре. 2008. № 2. С. 10–13.
7. **Шыхалиев П. Г.** Об одном методе сокращения размерности анализируемых признаков сетевых трафиков, используемых для мониторинга компьютерных сетей // Телекоммуникации. 2011. № 06. С. 44–48.
8. **Breitbart Y., Dragan F., Gobjuka H.** Effective Network Monitoring // Proc. of the International Conference On Computer Communications and Networks (ICCCN 2004). October 11–13, 2004. Chicago, IL, USA. P. 394–399.
9. **Breitbart Y., Chan C. Y., Garofalakis M., Rastogi R., Silberschatz A.** Efficiently Monitoring Bandwidth and Latency in IP Networks // Proc. of IEEE INFOCOM, 2000.
10. **Breitbart Y., Dragan F., Gobjuka H.** Effective Monitor Placement in Internet Networks // Journal of Networks. 2009. Vol. 4, no. 7. P. 657–666.
11. **Leppinen M., Pulkkinen P., Rautiainen A.** Java- and CORBA-based network management // Computer. 1997. Vol. 30, no. 6. P. 83–87.
12. **Lee J. O.** Enabling network management using Java technologies // IEEE Communications Magazine. 2000. Vol. 38, no. 1. P. 116–123.
13. **Gavalas D., Greenwood D., Ghanbari M., O'Mahony M.** Advanced network monitoring applications based on mobile/intelligent agent // Computer Communications. 2000. Vol. 23, no. 8. P. 720–730.
14. **Romascanu D., Zilbershtein I. E.** Switch monitoring — The new generation of monitoring for local area networks // Bell Labs Technical Journal. 1999. Vol. 4, no. 4. P. 42–54.

15. **Shaffi A. S., Al-Obaidy M.** Managing network components using SNMP // International Journal of Scientific Knowledge. 2013. Vol. 2, no. 3. P. 11—18.
16. **Thottan M. K.** SEQUIN: An SNMP-based MPLS network monitoring system // Bell Labs Technical Journal. 2003. Vol. 8, no. 1. P. 95—111.
17. **Mangaiyarkarasi R., Sivakumar S. R.** Nnetwork monitoring using snmp protocol // International Journal of Power Control Signal and Computation (IJPCSC). Jan-Mar 2012. Vol. 3, no. 1.
18. **Case J., Fedor M., Schoffstall M., Davin J.** Simple Network Management Protocol (SNMP), RFC 1157. Network Working Group, IETF, 1990.
19. **Stallings W.** SNMP, SNMPv2, SNMPv3, and RMON 1 and 2. Addison Wesley, 1996.
20. **McCloghrie K., Rose M.** Management information base for network management of TCP/IP-based internets: MIB-II, RFC 1213, March 1991.
21. **Вишнеvский В. М., Семенова О. В.** Математические методы исследования систем поллинга // Автоматика и телемеханика. 2006. № 2. С. 3—56.
22. **Вишнеvский В. М., Дудин А. Н., Клименок В. И., Семенова О. В., Шпилев С. А.** Система со шлюзовым обслуживанием и отдыхами прибора, зависящими от состояния системы // Четвертая Международная конференция по проблемам управления (МКПУ-IV) 26—30 января 2009: Сб. трудов. М.: Изд. ИПУ РАН, 2009. С. 1767—1772.
23. **Семенова О. В., Петроченков В. В.** Система поллинга с нетерпеливыми заявками и исчерпывающей дисциплиной обслуживания // Distributed Computer and Communication Networks. Theory and Applications (DCCN-2008), М.: ИТПРАС, 2008. С. 45—56.
24. **Takagi H.** Queueing analysis of polling models: an update. // Stochastic Analysis of Computer and Communication Systems / Ed. H. Takagi. Amsterdam: North-Holland Publ. Co., 1990. P. 267—318.
25. **Borst S. C., Boxma O. J.** Polling systems with and without switch-over times // Operations Research. 1997. Vol. 45, no. 4. P. 536—543.
26. **Forouzan I., Behrouz A.** Data communications and networking. — 4th ed. 2007, 1134 p.
27. **Van der Mei R. D., Resing J. A. C.** Analysis of polling models with two-stage gated service: fairness versus efficiency // Mason L., Drwiega T., Yan J. (Eds.). Managing Traffic Performance in Converged Networks — the Interplay between Convergent and Divergent Forces, ITC2007, Lecture Notes in Computer Science 4516, 2007. P. 544—555.
28. **Van der Mei R. D., Resing J. A. C.** Polling systems with two-phase gated service: heavy traffic results for the waiting-time distributions // Probability in the Engineering and Informational Sciences. 2008. Vol. 4, no. 22. P. 623—651.
29. **Van der Mei R. D., Roubos A.** Polling models with multi-phase gated service // Operations Research. 2012. Vol. 1, no. 198. P. 25—56.
30. **Leung K. K.** Cyclic-service systems with probabilistically-limited service // IEEE Journal on Selected Areas in Communications. 1991. Vol. 2, no. 9. P. 185—193.
31. **Boxma O. J., Levy H., Yechiali U.** Cyclic reservation schemes for efficient operation of multiple-queue single-server systems // Annals of Operations Research. 1992. Vol. 3, no. 35. P. 187—208.
32. **Keilson J., Servi L. D.** Oscillating random walk models for GI/G/1 vacation systems with Bernoulli schedules // Journal of Applied Probability. 1986, no. 23. P. 790—802.
33. **Levy H.** Analysis of cyclic polling systems with binomial gated service. In: Hasegawa T., Takagi H., Takahashi Y. (Eds.), Performance of Distributed and Parallel Systems. P. 127—139. Amsterdam: North-Holland, 1989.
34. **Levy H., Sidi M., Boxma O. J.** Dominance relations in polling systems // Queueing Systems. 1990. Vol. 6, is. 1. P. 155—171.
35. **Takagi H.** Analysis of polling systems. (Cambridge, Mass. MIT Press), 1986. 197 p.
36. **Boon M. A. A., Van der Mei R. D., Winands E. M. M.** Applications of polling systems // Surveys in Operations Research and Management Science. 2011. Vol. 16, is. 2. P. 67—82.
37. **Boxma O. J.** Static optimization of queueing systems // Recent Trends in Optimization Theory and Applications / Ed. Agwal R. P. Singapore: World Scientific Publ., 1995. P. 1—16.
38. **Boxma O. J., Levy H., Weststrate J.A.** Optimization of polling systems // Performance '90, eds. P. J. B. King, I. Mitrani, R. J. Pooley. North-Holland, Amsterdam. P. 349—361.

R. H. Shikhaliyev, PhD. on technical sciences, Leading Researcher

Institute for Information Technologies, Azerbaijan National Academy of Sciences, Baku, Azerbaijan
ramiz@science.ab.az

Improving the Monitoring Efficiency of Computer Networks Based on the Polling System Optimization

It is proposed the optimization model for online monitoring of computer networks (CN), which allow minimize of monitoring time. The main goal is to optimize the monitoring of CN nodes at given network resources. For achievement of this purpose it is offered to use model of polling system, by optimizing which it is possible to reach optimization of CN monitoring. The literature contains a number of works, which are devoted to improve the efficiency of network monitoring through the optimization of certain aspects of the polling. The analysis shows, that despite the large contribution to improving the efficiency of network monitoring, these works are based on the architectural solution of the problem of optimization of certain aspects of the polling. And in this works not considered characteristic of the polling system such as the intensity of arriving queries (monitoring data), the mean waiting time in the queue of requests, the server switching time between queues and etc. For this reason, in this paper to improve the efficiency of CN monitoring is proposed to optimize the polling system itself. As one of the main indicators of polling systems optimality is minimal mean waiting time of queries in each queue. Therefore, for optimization of the CN nodes polling, we proposed to minimize of mean waiting time for queries in queues. In our opinion it would be a unified approach to increase the efficiency of CN monitoring based on the polling system.

Keywords: computer networks, network monitoring, network nodes polling, SNMP-polling, polling systems, polling discipline, mean waiting time

References

1. **Levy H., Sidi M.** Polling systems: applications, modeling, and optimization. *IEEE Transactions on Communications*, 1990, vol. 38, no. 10, pp. 1750–1760.
2. **Network Monitoring and Analysis. A Protocol Approach to Troubleshooting.** Ed. Wilson. New Jersey, Prentice Hall, Inc. Upper Saddle River, 2002, 350 p.
3. **Gao F., Gutierrez J.** A Dynamic Polling Scheme for the Network Monitoring Problem. *Proc. of the 7th Asia-Pacific Network Operations and Management Symposium (APNOMS 2003)*, Fukuoka, Japan, October, 2003, pp. 489–500.
4. **Machida F., Kawato M., Maeno Y.** Polling Schedule Optimization for Adaptive Monitoring to Scalable Enterprise Systems. *International Journal On Advances in Intelligent Systems*, 2008, vol. 1, no. 1, pp. 11–22.
5. **Park S., Park M.** An efficient transmission for large MIB tables in polling-based SNMP. *10th International Conference on Telecommunications*, 2003, vol. 1, pp. 246–252.
6. **Skatkov A. V., Voronin D. Yu., Danil'chuk D. N.** Sistema podderzhki prinyatiya reshenii po vyboru struktury sistemy monitoringa. *Tekhnologiya i konstruirovaniye v elektronnoy apparature*, 2008, no. 2, pp. 10–13.
7. **Shikhaliyev R. H.** Ob odnom metode sokrashcheniya razmernosti analiziruemykh priznakov setevykh trafikov, ispol'zuemykh dlya monitoringa komp'yuternykh setei. *Telekommunikatsii*, 2011, no. 6, pp. 44–48.
8. **Breitbart Y., Dragan F., Gobjuka H.** Effective Network Monitoring. *Proc. of the International Conference on Computer Communications and Networks (ICCCN 2004)*, 2004, October 11–13, Chicago, IL, USA, pp. 394–399.
9. **Breitbart Y., Chan C. Y., Garofalakis M., Rastogi R., Silberschatz A.** Efficiently Monitoring Bandwidth and Latency in IP Networks. *Proc. of the Conference on Computer Communications (IEEE INFOCOM 2001)*, Anchorage, Alaska, USA, April 22–26, 2001, vol. 2, pp. 933–942.
10. **Breitbart Y., Dragan F., Gobjuka H.** Effective Monitor Placement in Internet Networks. *Journal of Networks*, 2009, vol. 4, no. 7, pp. 657–666.
11. **Leppinen M., Pulkkinen P., Rautiainen A.** Java- and CORBA-based network management. *Computer*, 1997, vol. 30, no. 6, pp. 83–87.
12. **Lee J. O.** Enabling network management using Java technologies. *IEEE Communications Magazine*, 2000, vol. 38 no. 1, pp. 116–123.
13. **Gavalas D., Greenwood D., Ghanbari M., O'Mahony M.** Advanced network monitoring applications based on mobile/intelligent agent. *Computer Communications*, 2000, vol. 23 no. 8, pp. 720–730.
14. **Romascanu D., Zilbershtein I. E.** Switch monitoring — The new generation of monitoring for local area networks. *Bell Labs Technical Journal*, 1999, vol. 4, no. 4, pp. 42–54.
15. **Shaffi A. S., Al-Obaidy M.** Managing network components using SNMP. *International Journal of Scientific Knowledge*, 2013, vol. 2, no. 3, pp. 11–18.
16. **Thottan M. K.** SEQUIN: An SNMP-based MPLS network monitoring system. *Bell Labs Technical Journal*, 2003, vol. 8 no. 1, pp. 95–111.
17. **Mangaiyarkarasi R., Siyakumar S. R.** Nnetwork monitoring using snmp protocol. *International Journal of Power Control Signal and Computation (UPCSC)*, Jan–Mar 2012, vol. 3, no. 1.
18. **Case J., Fedor M., Schoffstall M., Davin J.** *Simple Network Management Protocol (SNMP)*, RFC 1157, Network Working Group, IETF, 1990.
19. **Stadligns W.** SNMP, SNMPv2, SNMPv3, and RMON 1 and 2, 1996, Addison Wesley.
20. **McCloghrie K., Rose M.** Management information base for network management of TCP/IP-based internets: MIB-II, RFC 1213, March 1991.
21. **Vishnevskii V. M., Semenova Q. V.** Matematicheskie metody issledovaniya sistem pollinga. *Avtomatika i telemekhanika*, 2006, no. 2, pp. 3–56.
22. **Vishnevskii V. M., Dudin A. N., Klimenok V. I., Semenova O. V., Shipilev S. A.** Sistema so shlyuzovym obsluzhivaniem i otdykhami pribora, zavislyashchimi ot sostoyaniya sistemy. *Chetvertaya Mezhdunarodnaya konferentsiya po problemam upravleniya (MKPU-IV): Sbornik trudov.* M.: IPU RAN, January 26–30, 2009, pp. 1767–1772.
23. **Semenova O. V., Petrochenkov V. V.** Sistema pollinga s neterpelivymi zayavkami i ischerpyvayushchei distsiplinoi obsluzhivaniya. *Distributed Computer and Communication Networks. Theory and Applications (DCCN-2008)*, Moscow: IITP RAS, 2008, pp. 45–56.
24. **Takagi H.** Queueing analysis of polling models: an update. In: H. Takagi (ed.), *Stochastic Analysis of Computer and Communication Systems*. North-Holland Publ. Cy., Amsterdam, 1990, pp. 267–318.
25. **Borst S. C., Boxma O. J.** Polling systems with and without switch-over times. *Operations Research*, 1997, vol. 45, no. 4, pp. 536–543.
26. **Forouzan I., Behrouz A.** *Data communications and networking.* — 4th ed., Published by McGraw-Hill, 2007, 1134 p.
27. **Van der Mei R. D., Resing J. A. C.** Analysis of polling models with two-stage gated service: fairness versus efficiency. In: L. Mason, T. Drwiega, and J. Yan (Eds.), *Managing Traffic Performance in Converged Networks — the Interplay between Convergent and Divergent Forces, ITC2007, Lecture Notes in Computer Science 4516*, 2007, Berlin, Springer, pp. 544–555.
28. **Van der Mei R. D., Resing J. A. C.** Polling systems with two-phase gated service: heavy traffic results for the waiting-time distributions. *Probability in the Engineering and Informational Sciences*, 2008, vol. 4, no. 22, pp. 623–651.
29. **Van der Mei R. D., Roubos A.** Polling models with multi-phase gated service. *Operations Research*, 2012, vol. 1, no. 198, pp. 25–56.
30. **Leung K. K.** Cyclic-service systems with probabilistically-limited service. *IEEE Journal on Selected Areas in Communications*, 1991, vol. 2, no. 9, pp. 185–193.
31. **Boxma O. J., Levy H., Yechiali U.** Cyclic reservation schemes for efficient operation of multiple-queue single-server systems. *Annals of Operations Research*, 1992, vol. 3, no. 35, pp. 187–208.
32. **Keilson J., Servi L. D.** Oscillating random walk models for GI/G/I vacation systems with Bernoulli schedules, *Journal of Applied Probability*, 1986, no. 23, pp. 790–802.
33. **Levy H.** Analysis of cyclic polling systems with binomial gated service. In: Hasegawa T., Takagi H., Takahashi Y. (Eds.), *Performance of Distributed and Parallel Systems*. North-Holland, Amsterdam, 1989, pp. 127–139.
34. **Levy H., Sidi M., Boxma O. J.** Dominance relations in polling systems, *Queueing Systems*, 1990, vol. 6, is. 1, pp. 155–171.
35. **Takagi H.** *Analysis of polling systems.* Cambridge, Mass., MIT Press, 1986, 197 p.
36. **Boon M. A. A., Van der Mei R. D., Winands E. M. M.** Applications of polling systems, *Surveys in Operations Research and Management Science*, 2011, vol. 16, issue 2, pp. 67–82.
37. **Boxma O. J.** Static optimization of queueing systems, *Recent Trends in Optimization Theory and Applications.* Ed. Agwal R. P. Singapore: World Scientific Publ., 1995, pp. 1–16.
38. **Boxma O. J., Levy H., Weststrate J. A.** Optimization of polling systems, *Perfomence'90*, eds. P. J. B. King, I. Mitrani, R. J. Pooley, North-Holland, Amsterdam, pp. 349–361.

ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ И ИЗОБРАЖЕНИЙ DIGITAL PROCESSING OF SIGNALS AND IMAGES

УДК 519.246

Б. Г. Кухаренко, канд. физ.-мат. наук, ст. науч. сотр., вед. науч. сотр.,
Институт машиноведения имени А. А. Благонравова РАН, г. Москва, e-mail: kukharenko@imash.ru,
М. О. Солнцева-Чалей, аспирант,
Московский физико-технический институт (ГУ), e-mail: solnceva.chalei@gmail.com

Применение моделей непрерывного профиля для анализа результатов кластеризации многомерных траекторий

Показано применение моделей непрерывного скрытого профиля для определения центральной траектории в достаточно однородных кластерах проекций многомерных траекторий. Модели непрерывного профиля являются скрытыми марковскими моделями с внутренними состояниями времени и масштаба. Определение центральной траектории кластера позволяет, используя оценку близости по мере косинуса, обнаружить потенциально посторонние траектории кластера, которые представляют движения, находящиеся в зоне риска.

Ключевые слова: анализ данных, многомерные траектории, кластеризация, полиномиальная регрессия, модели непрерывного профиля, линейные динамические системы

Введение

Настоящая работа посвящена определению центральных траекторий, представляющих "безопасные коридоры" для самолетов, идущих на посадку. Знание этих центральных траекторий способствует снижению рисков, возникающих из-за расположения взлетно-посадочной полосы в зоне сложного географического ландшафта в горных и прибрежных районах. В работах [1–2] анализируются (дискретные) многомерные траектории самолетов, совершивших посадку в международном аэропорту, находящемся рядом с заливом Сан-Франциско. С помощью метода полиномиальных регрессий, учитывающих сходство геометрической формы траекторий, анализируемые траектории разделяются на пять кластеров [1]. Однако вследствие того, что полиномиальная регрессия дает обобщенное представление о форме траекторий, полученные кластеры являются неоднородными.

Анализ этих кластеров на основе моделей линейных динамических систем определяет компоненты вектора состояний и обнаруживает тонкую структуру кластеров [2]. Компоненты вектора состояний позволяют провести расщепление неоднородного кластера на подкластеры, однако эти компоненты не являются центральными для новых подкластеров и не могут использоваться для обнаружения потенциально посторонних траекторий, которые представляют движения в зоне риска.

Для определения центральных траекторий для новых подкластеров может использоваться модель непрерывного (скрытого) профиля (*Continuous Profile Model* — СРМ) [3–4], основанная на скрытой марковской модели (*Hidden Markov Model* — НММ) [5–6]. Модель непрерывного профиля предполагает, что каждый наблюдаемый временной ряд является зашумленной масштабированной выборкой из некоторого непрерывного скрытого профиля, или скрытой записи (*latent trace*). Наблюдаемые временные ряды генерируются при прохождении последовательности скрытых марковских состояний. СРМ обучается с помощью алгоритма ожидания — максимизации правдоподобия (*Expectation-Maximization algorithm* — EM), который в контексте скрытых марковских моделей называется алгоритмом Баума — Уолша (*Baum — Welch algorithm*) [5–6]. В результате последовательных итераций обучения СРМ определяется наиболее вероятная скрытая запись.

1. Модель непрерывного профиля

В СРМ элементы каждого наблюдаемого временного ряда являются результатом эмиссии, т. е. генерируются непрерывным профилем, или скрытой записью при прохождении через последовательность скрытых состояний с конкретными индексами, как это имеет место в скрытой марковской модели [5–6]. Для учета изменения амплитуды в пределах

конкретного временного ряда и между наблюдаемыми временными рядами к состояниям скрытого времени добавляются состояния скрытого масштаба относительно соответствующих элементов скрытой записи.

1.1. Скрытая марковская модель

Пусть $\mathbf{x}[i] = \{x[i; k], k = \overline{1, N}\}$, $i = \overline{1, K}$, — наблюдаемые временные ряды длиной N . Пусть набор возможных состояний НММ $\mathbf{S} = \{s[j], j = \overline{1, M}\}$ (M — число скрытых состояний НММ). Для наблюдаемого временного ряда $\mathbf{x}[i] = \{x[i; k], k = \overline{1, N}\}$ последовательность скрытых состояний времени $\boldsymbol{\varphi}[i] = \{\varphi[i; k] \in \mathbf{S}, k = \overline{1, N}\}$. В марковской модели вероятности переходов между состояниями $s[m]$ и $s[j]$ — $\mathcal{T}_{s[m], s[j]} \equiv p(\varphi[i; k] = s[j] | \varphi[i; k-1] = s[m])$ не зависят от номера i временного ряда. При условии, что в скрытой последовательности $\boldsymbol{\varphi}[i]$ в момент времени k имеется состояние $s[j]$, вероятность эмиссии $x[i; k]$ равна $\mathcal{A}_{k, s[j]}^i \equiv p(x[i; k] | \varphi[i; k] = s[j])$. В логарифме правдоподобия \mathcal{L} наблюдаемых временных рядов учитывается, что, поскольку состояния $\varphi[i; k]$ скрытые, вероятности переходов между состояниями $\mathcal{T}_{s[m], s[j]}$ и вероятности эмиссии $\mathcal{A}_{k, s[j]}^i$ следует суммировать по всем их возможным значениям:

$$\begin{aligned} \mathcal{L} &= \log \left(\prod_{i=1}^K p(\mathbf{x}[i], \{\mathcal{T}_{s[m], s[j]}\}, \{\mathcal{A}_{k, s[j]}^i\}) \right) = \\ &= \sum_{i=1}^K \log \left(\sum_{\boldsymbol{\varphi}[i]} \mathcal{T}_{\boldsymbol{\varphi}[i], \boldsymbol{\varphi}[i; k]} \left(\prod_{k=1}^N \mathcal{A}_{k, \boldsymbol{\varphi}[i; k]}^i \right) \times \right. \\ &\quad \left. \times \left(\prod_{k=1}^N \mathcal{T}_{\boldsymbol{\varphi}[i; k], \boldsymbol{\varphi}[i; k-1]} \right) \right). \end{aligned} \quad (1)$$

В формуле (1) правдоподобие для каждого наблюдаемого временного ряда факторизуется на три множителя, поскольку условные вероятности состояний в НММ зависят только от предыдущих состояний на одном временном шаге [5–6]. Первый член под логарифмом в формуле (1) — это вероятность начала ряда в конкретном скрытом состоянии, второй — вероятность эмиссии в каждом состоянии и третий — вероятность переходов между состояниями.

Поскольку рассматривается модель со скрытыми переменными, для оценки параметров $\{\mathcal{T}_{s[m], s[j]}\}$, $\{\mathcal{A}_{k, s[j]}^i\}$ в (1) используется EM-алгоритм. Для этого необходимо вычислить предельные апостериорные вероятности каждого состояния $p(\varphi[i; k] | \mathbf{x}[i])$ и определить вероятности переходов. Кроме этого, не-

обходимо вычислить парные предельные апостериорные вероятности $p(\varphi[i; k], \varphi[i; k+1] | \mathbf{x}[i])$. Их вычисление в НММ основано на алгоритме динамического программирования, в контексте этой модели называемого алгоритмом прямой и обратной рекурсии.

1.2. Алгоритм прямой и обратной рекурсии

Для K наблюдаемых временных рядов длиной N и числе скрытых состояний M алгоритм прямой и обратной рекурсии (*Forward-Backward algorithm*) выполняет вычисления с оценкой временной сложности $O(KMN)$ (если матрица переходов разреженная, временная оценка сложности будет меньше). Введем обозначение совместной вероятности $\alpha[i; k; j] \equiv p(x[i; 1], x[i; 2], \dots, x[i; k], \varphi[i; k] = s[j])$. Для всех i и j после инициализации

$$\alpha[i; 1; j] \equiv \mathcal{A}_{1, s[j]}^i \mathcal{T}_{0, s[j]}$$

выполняется прямая рекурсия для $k = \overline{1, N}$

$$\alpha[i; k; j] \equiv \mathcal{A}_{k, s[j]}^i \sum_{m=1}^M \alpha[i; k-1; m] \mathcal{T}_{s[m], s[j]}. \quad (2)$$

Если рекурсия (2) завершена, то можно вычислить

$$\begin{aligned} p(\mathbf{x}[i]) &= \sum_{j=1}^M p(x[i; 1], x[i; 2], \dots, \\ &\quad \dots, x[i; N], \varphi[i; N] = s[j]) = \sum_{j=1}^M \alpha[i; N; j]. \end{aligned} \quad (3)$$

С учетом (3) логарифм правдоподобия \mathcal{L} (1) определяется как

$$\mathcal{L} = \sum_{i=1}^K \log(p(\mathbf{x}[i])) = \sum_{i=1}^K \log \left(\sum_{j=1}^M \alpha[i; N; j] \right). \quad (4)$$

Для вычисления предельных апостериорных вероятностей (*marginals of the posterior*) $p(\varphi[i; k] | \mathbf{x}[i])$ и $p(\varphi[i; k], \varphi[i; k+1] | \mathbf{x}[i])$ используются формулы Байеса для условной вероятности:

$$\begin{aligned} p(\varphi[i; k] = s[j] | \mathbf{x}[i]) &= \\ &= \frac{p(\mathbf{x}[i] | \varphi[i; k] = s[j]) p(\varphi[i; k] = s[j])}{p(\mathbf{x}[i])} \equiv \\ &\equiv \frac{\alpha[i; k; j] p(x[i; k+1], \dots, x[i] | \varphi[i; k] = s[j])}{p(\mathbf{x}[i])} \end{aligned} \quad (5)$$

и

$$\begin{aligned} p(\varphi[i; k-1] = s[j], \varphi[i; k] = s[m] | \mathbf{x}[i]) &= \\ &= \frac{\alpha[i; k-1; j] p(x[i; k], \dots, x[i; N] | \varphi[i; k] = s[m]) \mathcal{T}_{s[j], s[m]}}{p(\mathbf{x}[i])}. \end{aligned} \quad (6)$$

Вычисление совместных вероятностей $\beta[i; k; j] \equiv p(x[i; k+1], x[i; k+2], \dots, x[i; N], \varphi[i; k] = s[j])$

позволяет вычислить вероятности (5), (6). Для всех k и j выполняются инициализация

$$\beta[i; 1; j] = 1$$

и обратная рекурсия для $k = \overline{N-1; 1}$

$$\beta[i; k; j] \equiv \sum_{m=1}^M \mathcal{T}_{s[j], s[m]} \mathcal{A}_{k+1, s[m]}^i \beta[i; k+1; m]. \quad (7)$$

Если $\alpha[i; k; j]$ и $\beta[i; k; j]$ определены в результате прямой (2) и обратной (7) рекурсий, то предельные апостериорные вероятности (5) и (6) вычисляются по формулам

$$p(\mathbf{x}[i]) = \sum_{j=1}^M \alpha[i; N; j] \equiv \sum_{j=1}^M \alpha[i; k; j] \beta[i; k; j]; \quad (8)$$

$$p(\varphi[i; k] = s[j] | \mathbf{x}[i]) = \frac{\alpha[i; k; j] \beta[i; k; j]}{p(\mathbf{x}[i])}; \quad (9)$$

$$\begin{aligned} p(\varphi[i; k-1] = s[j], \varphi[i; k] = s[m] | \mathbf{x}[i]) &= \\ &= \frac{\alpha[i; k-1; j] \mathcal{T}_{s[j], s[m]} \mathcal{A}_{k, s[m]}^i \beta[i; k; m]}{p(\mathbf{x}[i])}. \end{aligned} \quad (10)$$

Эффективное вычисление правдоподобия и предельных апостериорных вероятностей (9), (10) необходимо для оценки параметров в формуле (1) посредством EM-алгоритма. Апостериорная вероятность $p(\varphi[i] | \mathbf{x}[i])$ оценивает, что именно эта последовательность скрытых состояний генерирует наблюдаемый временной ряд. Для оценки наиболее вероятной последовательности скрытых состояний на основе наблюдаемого временного ряда используется алгоритм Витерби [5, 6].

1.3. Модель непрерывного профиля

При обучении СРМ определяется непрерывный профиль, т. е. скрытая запись, вероятности переходов, управляющие марковской эволюцией состояний времени и масштаба, суммарный уровень шума наблюдаемого временного ряда и его глобальный фактор масштаба. После обучения скрытая запись

$$\mathbf{z} = \{z[k], k = \overline{1, M}\} \quad (11)$$

имеет более высокое разрешение по сравнению с ее зашумленными масштабированными копиями — наблюдаемыми временными рядами

$$\mathbf{x}[i] = \{x[i; k], k = \overline{1, N}\}, i = \overline{1, K}.$$

В идеале $M \gg N$, тогда каждый элемент наблюдаемого временного ряда $\mathbf{x}[i]$, $i = \overline{1, K}$, в точности отображается в элемент скрытой записи \mathbf{z} (11). Поскольку разрешение скрытой записи \mathbf{z} (11) выше, чем

у наблюдаемого временного ряда $\mathbf{x}[i]$, $i = \overline{1, K}$, его наблюдаемое время может эффективно ускоряться или замедляться при продвижении вдоль скрытой записи.

Неоднородность выборки элементов скрытой записи (11) и локальный масштаб, используемые при генерации элементов наблюдаемого временного ряда, определяются соответствующей ему последовательностью скрытых состояний СРМ. Каждое скрытое состояние представляет собой пару состояние масштаба/состояние времени

$$\varphi[i; k] = \{\tau[i; k], \chi[k]\}, k = \overline{1, N}. \quad (12)$$

Состояния времени $\tau[i; k]$, $k = \overline{1, N}$, принадлежат последовательности натуральных чисел $\overline{1, M}$, представляющих скрытое время, которое индексирует скрытую запись (11). Состояния масштаба $\chi[k]$, $k = \overline{1, N}$, принадлежат упорядоченному набору $\overline{1, L}$. Распределение вероятности эмиссии элементов наблюдаемого временного ряда $\mathbf{x}[i] = x[i; k]$, $k = \overline{1, N}$, производимых последовательностью скрытых состояний $\varphi[i; k]$, $k = \overline{1, N}$, на основе скрытой записи $\mathbf{z} = z[k]$, $k = \overline{1, M}$, имеет вид

$$\begin{aligned} p(x[i; k] | \mathbf{z}, \varphi[i; k], u[i], \sigma[i]) &\equiv \\ &\equiv \mathcal{N}(x[i; k] | u[i] z[\tau[i; k]] \chi[k], (\sigma[i])^2), \end{aligned} \quad (13)$$

где $\mathcal{N}(x | z, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-z)^2}{2\sigma^2}\right)$ — Гауссово

нормальное распределение; $(\sigma[i])^2$ — это вариация шума в наблюдаемом временном ряду; $u[i]$ — глобальный параметр масштаба, уникальный для наблюдаемого временного ряда, который корректирует (глобальную) разницу масштаба временного ряда $\mathbf{x}[i]$ и скрытой записи \mathbf{z} (11).

Для полного описания СРМ необходимо определить вероятности переходов между скрытыми состояниями. Поскольку вероятности переходов между состояниями времени и вероятности переходов между состояниями масштаба определяются отдельно, совместная вероятность переходов между скрытыми состояниями факторизуется следующим образом:

$$\begin{aligned} \mathcal{T}_{\varphi[i; k], \varphi[i; l]}^i &\equiv p(\varphi[i; k] | \varphi[i; l]) = \\ &= p(\tau[i; k] | \tau[i; l]) p(\chi[k] | \chi[l]). \end{aligned} \quad (14)$$

1.4. Логарифм правдоподобия набора наблюдаемых временных рядов

Логарифм правдоподобия \mathcal{L}^p наблюдаемых временных рядов $\mathbf{x}[i] = \{x[i; k], k = \overline{1, N}\}$, $i = \overline{1, K}$, имеет вид $\mathcal{L}^p = \mathcal{L} + \mathcal{P}$, где \mathcal{L} — это член правдоподобия, происходящий из НММ и состоящий из вероятности

стей эмиссии (13) и вероятностей переходов между скрытыми состояниями (14)

$$\mathcal{L} = \sum_{i=1}^K \log \left(\sum_{\varphi[i]} p(\varphi[i; 1]) \times \left(\prod_{i=1}^N \mathcal{N}(x[i; k] | u[i]z[\tau[i; k]]\chi[k], (\sigma[i])^2) \right) \times \left(\prod_{i=2}^N \mathcal{T}_{\varphi[i; k-1], \varphi[i; k]}^i \right) \right), \quad (15)$$

где $p(\varphi[i; 1])$ — априорные вероятности начальных скрытых состояний, а \mathcal{P} — это логарифм априорной вероятности или штрафной член, обеспечивающий регуляризацию СРМ [4, 5]. Максимум \mathcal{L}^p определяется с помощью EM-алгоритма.

1.5. Оценка предельных апостериорных вероятностей (E-шаг)

На E-шаге (EM-алгоритма) предельные апостериорные вероятности $p(\varphi[i; k] = s | \mathbf{x}[i])$ и $p(\varphi[i; k-1] = s, \varphi[i; k] = s' | \mathbf{x}[i])$ СРМ вычисляются посредством алгоритма прямой и обратной рекурсии, как в НММ (подраздел 1.2). Ввиду разреженности матрицы переходов между скрытыми состояниями (13) вычислительная сложность этого алгоритма линейная по числу скрытых состояний (12) [4, 5].

1.6. Оценка параметров СРМ (M-шаг)

Оценки параметров на M-шаге, которые максимизируют ожидаемый логарифм правдоподобия

$$\langle \mathcal{L}_{\text{comp}}^p \rangle \equiv \langle \mathcal{P} \rangle + \sum_{i=1}^K \langle \log(p(\varphi[i], \mathbf{x}[i])) \rangle, \quad (16)$$

представляются аналитическими формулами [4, 5]. Для оценки элементов скрытой записи (11) используется численная процедура [4, 5]. Производная $\langle \mathcal{L}_{\text{comp}}^p \rangle$ (16) по $z[k]$ (элементу скрытой записи) имеет вид

$$\begin{aligned} \frac{\partial \langle \mathcal{L}_{\text{comp}}^p \rangle}{\partial (z[k'])} &= \frac{\partial}{\partial (z[k'])} \sum_{s=1}^S \sum_{k=1}^N p(\varphi[i; k] = s | \mathbf{x}[i]) \times \\ &\times \log(\mathcal{N}(x[i; k] | u[i]z[\tau[s]]\chi[s], (\sigma[i])^2)) - \lambda \bar{u} \frac{\partial}{\partial (z[k'])} \times \\ &\times \sum_{k=1}^{M-1} (z[k+1] - z[k])^2 = \\ &= - \sum_{\{s | \tau[s] = k'\}} \sum_{k=1}^N p(\varphi[i; k] = s | \mathbf{x}[i]) \frac{\partial}{\partial (z[k'])} \times \\ &\times \frac{(x[i; k] - u[i]z[\tau[s]]\chi[s])^2}{2(\sigma[i])^2} - \lambda \bar{u} (2(z[k'] - z[k' - 1]) - \end{aligned}$$

$$\begin{aligned} &- 2(z[k' + 1] - z[k'])) = \\ &= - \sum_{\{s | \tau[s] = k'\}} \sum_{k=1}^N p(\varphi[i; k] = s | \mathbf{x}[i]) u[i] \chi[s] \times \\ &\times \frac{(x[i; k] - u[i]z[\tau[s]]\chi[s])^2}{2(\sigma[i])^2} - \\ &- \lambda \bar{u} (4z[k'] - 2z[k' - 1] - 2z[k' + 1]), \quad (17) \end{aligned}$$

где λ — коэффициент при штрафном члене, обеспечивающем гладкость скрытой записи (11).

Таким образом, при $\lambda \neq 0$ условие $\frac{\partial \langle \mathcal{L}_{\text{comp}}^p \rangle}{\partial (z[k])} = 0$, $k = \overline{1, M}$, приводит к трехдиагональной системе уравнений (17).

2. Численный эксперимент

В работе [1] представлены результаты кластеризации 112 (трехмерных) посадочных траекторий самолетов на пять кластеров. Эта исходная кластеризация выполняется по методу полиномиальных регрессий и приводит к слишком обобщенному представлению формы траекторий, поэтому в полученных кластерах присутствует неоднородность. Анализ проекций траекторий этих кластеров на основе моделей линейных динамических систем определяет компоненты вектора состояний и обнаруживает тонкую структуру кластеров [2]. В работе [2] эта тонкая структура кластеров показана в проекциях траекторий на координатные оси (ниже рассматриваются только проекции на ось x). На рис. 1 (см. третью сторону обложки) показаны проекции на ось x 16 многомерных траекторий (длина траекторий $N = 156$), входящих в один из исходных кластеров, полученных по методу полиномиальных регрессий (розовый кластер в [1]). Жирные линии 1 и 2 на рис. 1 представляют две компоненты вектора состояний линейной динамической системы, выявляющие тонкую структуру кластера [2]. Приписывание проекций траекторий $[x_i[k], k = \overline{1, 156}]$, $i = \overline{1, 16}$, к одному из подкластеров 1 (красному) или 2 (синему) на рис. 1 выполняется в соответствии с их максимальной близостью к одной из компонент вектора состояний $[y_j[k], k = \overline{1, 156}]$, $j = \overline{1, 2}$, с использованием меры косинуса

$$R = \left(\sum_{k=1}^{156} x_i[k] y_j[k] \right) / \sqrt{\sum_{k=1}^{156} (x_i[k])^2} \sqrt{\sum_{k=1}^{156} (y_j[k])^2},$$

которая учитывает разницу в направлении векторов в пространстве состояний (безотносительно Евклидова расстояния между ними).

В настоящем разделе при анализе тонкой структуры кластера траекторий (рис. 1) в результате применения СРМ в каждом подкластере определяется

(согласующаяся с большинством) обобщенная форма проекций траекторий.

На рис. 2 (см. третью сторону обложки) показаны проекции траекторий красного подкластера и жирной кривой 1 соответствующая компонента вектора состояний (компонента 1 на рис. 1). На рис. 3 (см. третью сторону обложки) показаны проекции траекторий синего подкластера и жирной кривой 1 соответствующая компонента вектора состояний (компонента 2 на рис. 1). Штриховые жирные линии 2 на рис. 2 и 3 показывают скрытые записи (непрерывные скрытые профили) (11) соответствующих СРМ. Непрерывные профили на рис. 2 и 3 отображают проекции траекторий подкластеров по принципу голосования большинства [7]. Непрерывный профиль 2 на рис. 2 отражает ход проекций семи траекторий из 11. Проекция траектории, показанная на рис. 2 зеленым цветом, наиболее удалена (по мере косинуса) от непрерывного профиля 2.

На рис. 3 непрерывный профиль 2 отражает ход проекций четырех траекторий, при этом одна траектория игнорируется. Эта наиболее удаленная (по мере косинуса) от непрерывного профиля 2 проекция траектории показана голубым цветом на рис. 3.

Заключение

В настоящей работе показано применение СРМ для описания обобщенной формы проекций траекторий в однородных кластерах. Непрерывный скрытый профиль достаточно точно отражает фор-

му большинства траекторий в кластере, представляя, таким образом, центральную траекторию этого кластера. В результате определения такой центральной траектории кластера можно улучшить однородность структуры кластера и обнаружить потенциально посторонние траектории кластера, которые представляют движения, находящиеся в зоне риска.

Список литературы

1. Кухаренко Б. Г., Солнцева М. О. Кластеризация управляемых объектов на основе сходства их многомерных траекторий // Информационные технологии. 2014. № 5. С. 3–7.
2. Кухаренко Б. Г., Солнцева М. О. Анализ результатов кластеризации многомерных траекторий посредством моделей линейных динамических систем // Информационные технологии. 2015. № 2. С. 104–109.
3. Listgarten J., Neal R. M., Roweis S. T., Emili A. Multiple alignment of continuous time series / Saul L. K., Weiss Y., Bottou L., eds. // Proceedings of Neural Information Processing Systems (NIPS 2004). December 13–18, 2004, Vancouver, British Columbia, Canada. Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press. 2005. V. 17. P. 5–13.
4. Listgarten J. Analysis of sibling time series data: Alignment and difference detection. PhD thesis. University of Toronto: Graduate Department of Computer Science. 2007.
5. Кухаренко Б. Г. Анализ независимых компонент и скрытая Марковская модель для определения доминантных компонент многомерных временных рядов // Информационные технологии. 2010. № 11. Приложение. С. 1–32.
6. Poritz A. B. Hidden Markov models: A guided tour // Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP). Morgan Kaufmann. 1988. P. 7–13.
7. Dasgupta P., Maskin E. On the robustness of majority rule // Journal of the European Economic Association. 2008. Vol. 6, N 5. P. 949–973.

B. G. Kukharensko¹, Leading Research Scientist, e-mail: kukharenskobg@hotmail.com,
M. O. Solntseva-Chaley², Postgraduate Student, e-mail: solntseva.chalei@gmail.com

¹Blagonravov Institute of Engineering Science of RAS

²Moscow Institute of Physics and Technology (SU)

Applying Continuous Profile Models for Analysis of Multidimensional Trajectory Clustering Results

Problem of clustering trajectories is pre-conditioned by a need to organize motion of objects under control. The polynomial regression method is the best approach to trajectory cluster selection, which estimates a form of general trajectory of each cluster. For a set of sufficiently heterogeneous trajectories, the defined clusters are heterogeneous also. For the inhomogeneous cluster, its polynomial regression is a too strong abstraction. To demonstrate the cluster heterogeneity (and, thus, non full clustering trajectories) a method of dimension reducing is in need. So, linear dynamical system models are applied to object multi dimensional trajectory clustering by the polynomial regression method. Extracted components of system state vector give an opportunity to elicit cluster fine structure in accordance with cosine measure. In present paper applying continuous latent profile models is shown for determining a central trajectory in sufficiently homogeneous clusters of multi-dimensional trajectory projections. Continuous profile models are the hidden Markov models with internal states of time and scale. Defining the cluster central trajectory with affinity estimation by cosine measure in use gives an opportunity to determine the cluster outlying trajectories in potential, which represent motions located in a risk zone

Keywords: data mining, multi-dimensional trajectories, clustering, polynomial regression, continuous profile models, linear dynamical systems

References

1. **Kukhareno B. G., Solntseva M. O.** Klasterizacija upravlyаемых объектов na osnove shodstva ih mnogomernyh trajektoriyi. *Informacionnye tehnologii*. 2014, no. 5, pp. 3–7.
2. **Kukhareno B. G., Solntseva M. O.** Analiz rezultatov klasterizacii mnogomernyh trajektoriyi posredstvom modelei lineinyh dinamicheskikh sistem. *Informacionnye tehnologii*. 2015, no. 2, pp. 104–109.
3. **Listgarten J., Neal R. M., Roweis S. T., Emili A.** Multiple alignment of continuous time series / Saul L. K., Weiss Y., Bottou L., eds. *Proceedings of Neural Information Processing Systems (NIPS 2004). December 13–18, 2004, Vancouver, British Columbia, Canada*. Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press. 2005, vol. 17, pp. 5–13.
4. **Listgarten J.** *Analysis of sibling time series data: Alignment and difference detection*. PhD thesis. University of Toronto: Graduate Department of Computer Science. 2007.
5. **Kukhareno B. G.** Analiz nezavisimyh komponent i skrytaja Markovskaja model dlja opredelenija dominantnyh komponent mnogomernyh vremennyh rjadov. *Informacionnye tehnologii*. 2010, no. 11. Prilozhenie, pp. 1–32.
6. **Poritz A. B.** Hidden Markov models: A guided tour. *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Morgan Kaufmann. 1988, pp. 7–13.
7. **Dasgupta P., Maskin E.** On the robustness of majority rule // *Journal of the European Economic Association*. 2008, vol. 6, no. 5, pp. 949–973.

УДК 621.391

А. В. Вольнская, канд. техн. наук, доц., e-mail: anna-volinskaya@mail.ru,
Уральский государственный университет путей сообщения, г. Екатеринбург,
И. В. Сапожников, аспирант, fbjumper@mail.ru,
Уральский технический институт связи и информатики, г. Екатеринбург

Моделирование линейного мультипараметрического модема для каналов передачи информации с низким отношением сигнал/помеха

Приведены результаты компьютерного моделирования алгоритма мультипараметрической модуляции и демодуляции с оценкой его помехоустойчивости путем корреляционного анализа принятых сигналов.

Ключевые слова: модуляция, помехоустойчивость, свертка, двоичные сигналы, сложные шумоподобные сигналы, коды Баркера, корреляция

Введение

Существуют системы передачи информации, в которых отношение сигнал/помеха (с/п) на входе приемника меньше единицы. К ним относятся, например, системы беспроводной аварийной сигнализации в шахтах, PLC-R каналы (связь с электровозом по контактному проводу), военные системы связи, работающие в условиях радиопротиводействия. В этих системах применение классических видов модуляции затруднено ввиду так называемого порогового эффекта помехоустойчивости, когда отношение сигнал/помеха на входе приемника снижается до единицы, а на выходе — лавинообразно падает [1]. Причина этого явления — нелинейность модуляции, возникающая в результате перемножения сигналов. Кроме того, модуляции подвергается, как правило, один параметр простого гармонического колебания, максимум — два (в комбинированных видах модуляции). Теоретическая возможность применения более сложной шумоподобной несущей обсуждается с середины прошлого века, но только в варианте однопараметрической модуляции, например дисперсии или средней частоты [2]. В начале 80-х годов А. И. Самойловым предложены алгоритм и варианты технической реализации мульт-

типараметрической линейной модуляции шумоподобной несущей путем ее свертки (*convolution*) с передаваемым сигналом и развертки (*deconvolution*) на приемном конце [3]. Назовем такой модем линейным мультипараметрическим модемом (ЛММ). Преимуществом ЛММ является его способность работать при наличии шумов с адаптацией спектра передаваемого сигнала к спектру помех. В упрощенном виде эта технология реализована в стандарте IEEE 802.22 (когнитивное радио) [4] и PLC-каналах передачи сигналов по проводам электрооборудования [5]. Идея общая — энергия сигнала перераспределяется на участки спектра, где минимальна энергия помехи.

Основные положения

На рис. 1 представлены схема и временные диаграммы, поясняющие суть ЛММ, который работает следующим образом. Пусть передаваемое сообщение $s(t)$, имеющее ширину спектра F_s и длительность T_s , разбивается на фрагменты по $2F_s T_s$ отсчетных значений, которые записываются в регистр; шумоподобное несущее колебание $y(t)$ имеет период $2F_s T_s$ отсчетных значений и продвигается по цифровой линии задержки, при этом каждое его

отсчетное значение умножается на соответствующее отсчетное значение модулирующего сигнала и с этим коэффициентом поступает на сумматор. Очевидно, что сигнал $x(t)$ на выходе модулятора является сверткой двух сигналов. При этом каждое его отсчетное значение есть результат линейного взаимодействия всех отсчетных значений несущего колебания со всеми отсчетными значениями передаваемого сообщения. Это обеспечивает как высокую помехоустойчивость, так и высокую скрытность [6, 7].

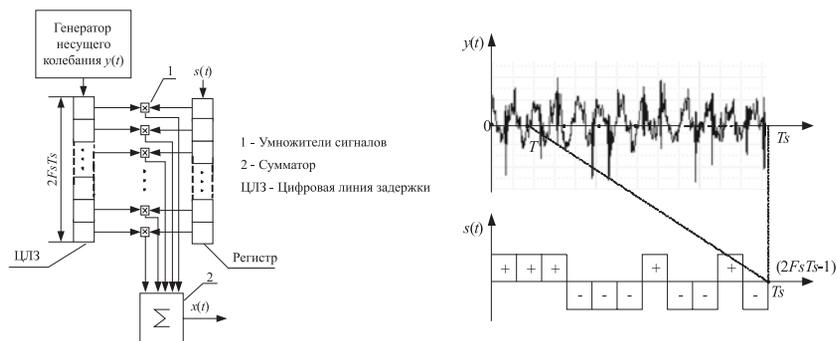


Рис. 1. Схема линейного мультипараметрического модема и временные диаграммы его работы

Оценка помехоустойчивости линейного мультипараметрического модема

Для проверки корректности ЛММ и оценки его помехоустойчивости проведено моделирование в программной среде Visual Studio 2010 [8]. В качестве передаваемых сигналов рассматриваются канонические сигналы Баркера и новые, более сложные, комбинированные сигналы Баркера [9], а в качестве несущего колебания — случайная (шумоподобная) двоичная последовательность. Далее проводится свертка одного из выбранных сигналов с псевдослучайной двоичной последовательностью, в результате чего получается сложное шумоподобное колебание. К этому колебанию добавляется помеха, сформированная с помощью генератора случайных чисел. Для демодуляции вычисляется матрица, обратная матрице несущего колебания и соответствующая демодулирующему колебанию, и выполняется свертка с принятой смесью сигнала и помехи. В результате получаем демодулированный сигнал, для которого вычисляется нормированная функция взаимной корреляции (ВКФ) с переданным сигналом. По главному пику этой функции оценивается возможность надежного обнаружения переданных сигналов.

На рис. 2 приведены результаты моделирования для 11-элементного сигнала Баркера. 11-элементный сигнал Баркера (рис. 2, а) сворачивается с несущим сигналом, представляющим собой случайную двоичную последовательность (рис. 2, б). Модулированный сигнал (полученный в результате свертки) показан на рис. 2, в. На рис. 2, г сверху — сигнал после демодуляции (ниже, для сравнения, — исходный сигнал) и их взаимокорреляционная функция (ВКФ), которая совпадает с автокорреляционной функцией (АКФ), так как сигналы одинаковые.

Теперь к модулированному сигналу добавим помеху. На рис. 3, а снизу показана помеха, а сверху — сумма сигнала и помехи на входе демодулятора. На рис. 3, б — сигнал на выходе демодулятора (снизу — исходный сигнал); пик ВКФ (рис. 3, в) имеет высоту более 0,9, что свидетельствует о хорошем сходстве демодулированного сигнала с переданным.

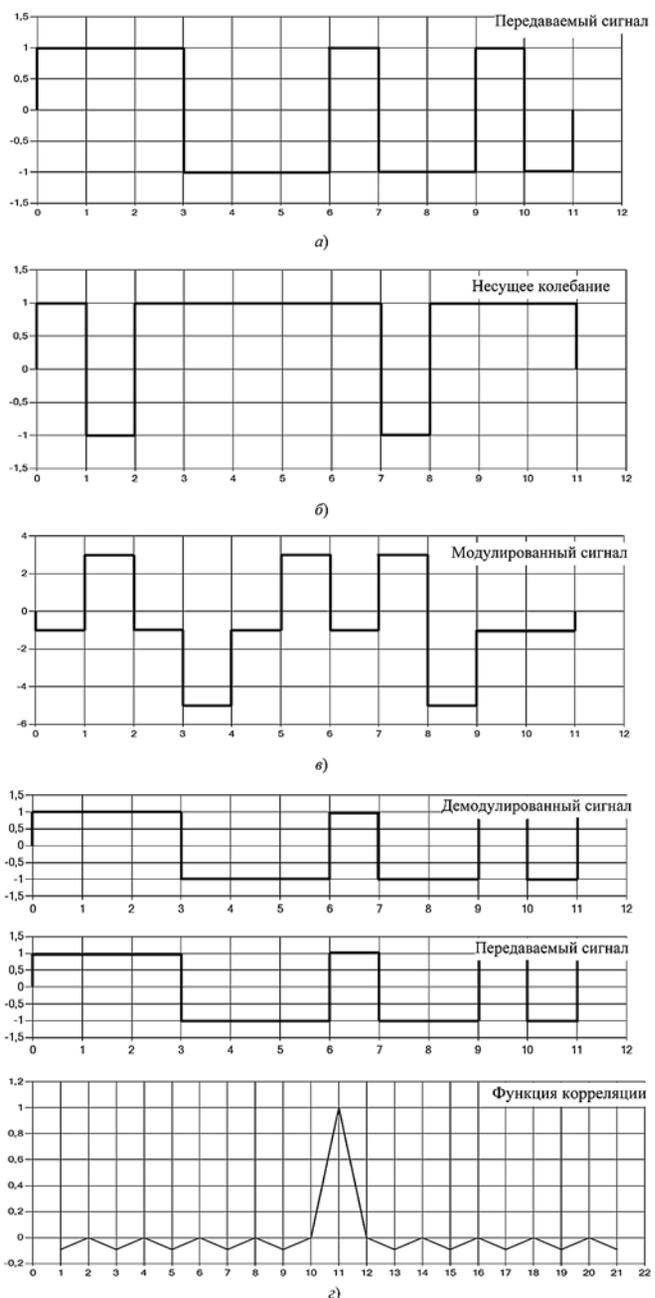


Рис. 2. Результаты моделирования для 11-элементного сигнала Баркера

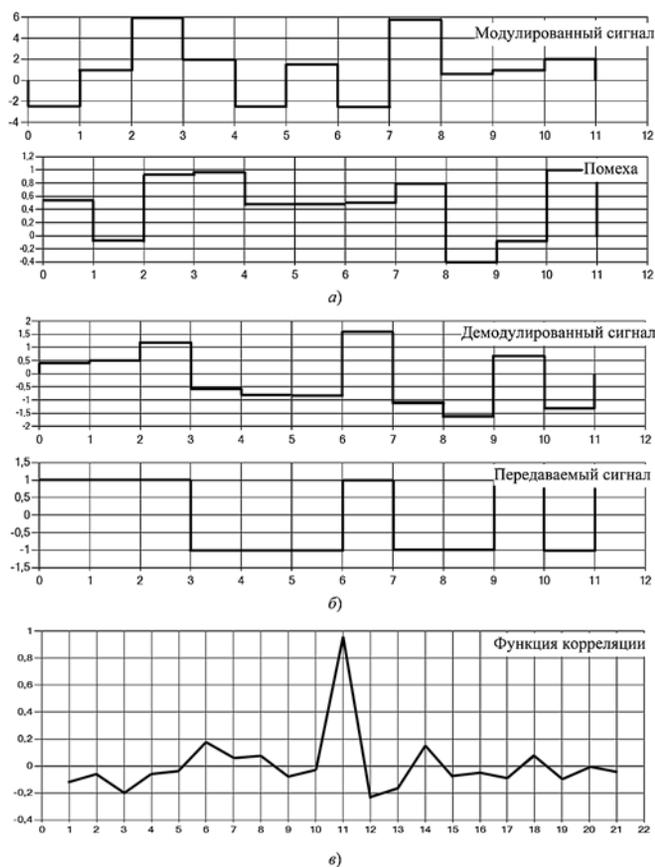


Рис. 3. Результаты моделирования для 11-элементного сигнала Баркера (отношение $c/p = 1$)

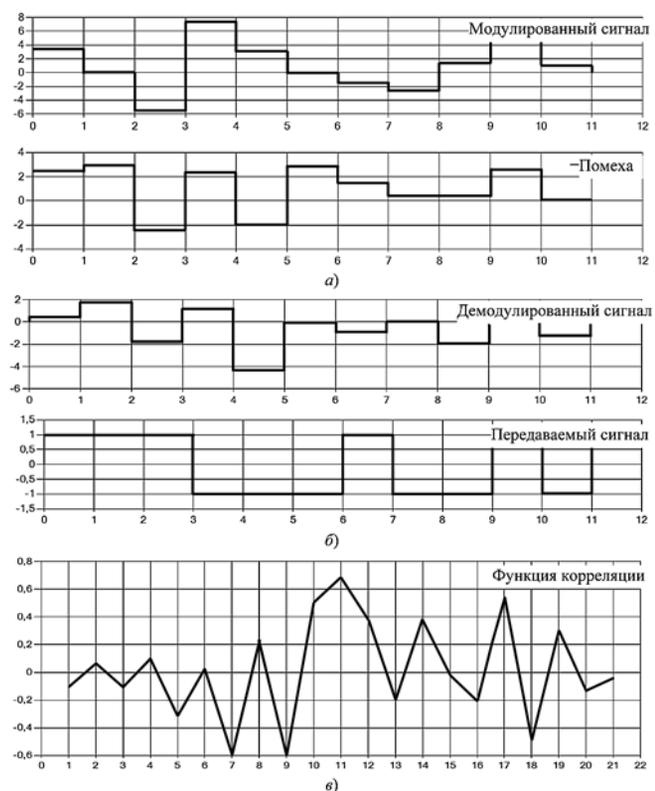


Рис. 4. Результаты моделирования для 11-элементного сигнала Баркера (отношение $c/p = 1/3$)

Увеличим помеху в 3 раза (рис. 4). Видим, что демодулированный сигнал сильно отличается от исходного, поэтому говорить о его надежном обнаружении не приходится.

Теперь проведем эксперимент с более сложным сигналом, например, с одним из найденных нами ранее — комбинационным сигналом Баркера 11—11, содержащим 121 разряд [9]. Видим (рис. 5), что алгоритм модуляции-демодуляции работает корректно, так как демодулированный сигнал в точности совпадает с переданным, что подтверждается значением ВКФ в пиковой точке, равным единице.

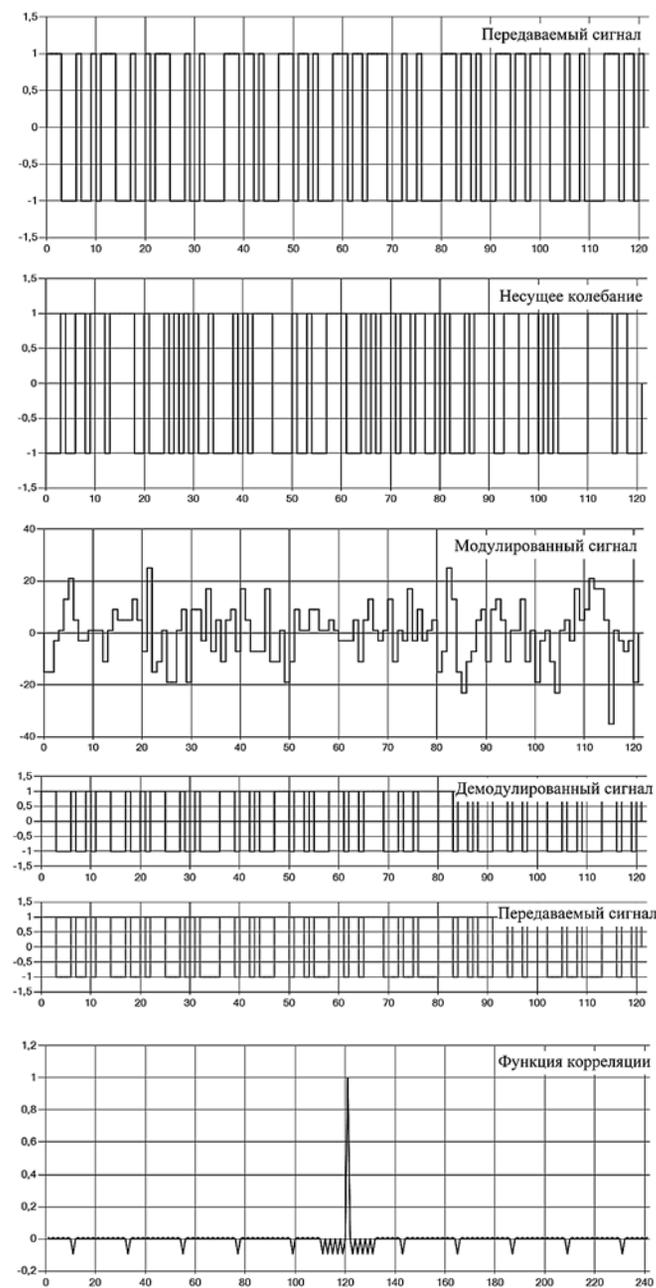


Рис. 5. Результаты моделирования для комбинационного сигнала Баркера 11—11

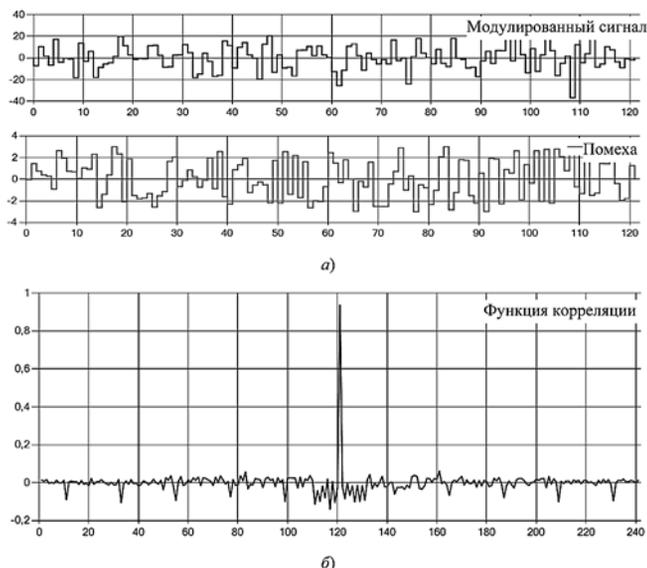


Рис. 6. Результаты моделирования для комбинированного сигнала Баркера 11—11 (отношение $c/p = 1/3$)

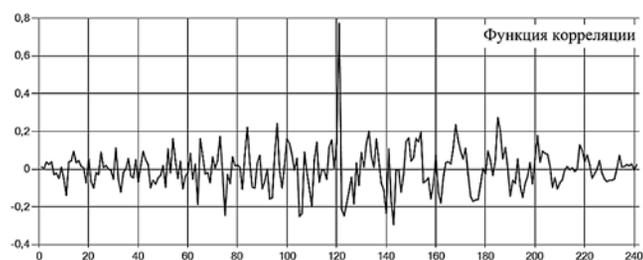


Рис. 7. Результаты моделирования для комбинированного сигнала Баркера 11—11 (отношение $c/p = 1/10$)

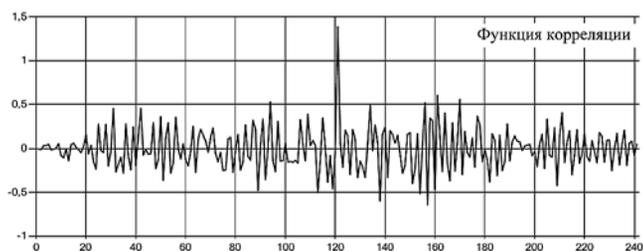


Рис. 8. Результаты моделирования для комбинированного сигнала Баркера 11—11 (отношение $c/p = 1/30$)

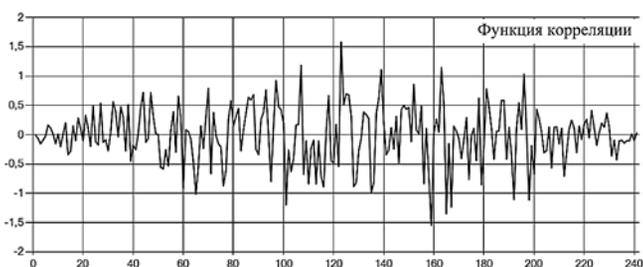


Рис. 9. Результаты моделирования для комбинированного сигнала Баркера 11—11 (отношение $c/p = 1/100$)

Добавим помеху, превышающую сигнал в 3 раза. На рис. 6, *а* показаны сигнал в канале и помеха; на рис. 6, *б* — ВКФ, вид которой говорит о хорошем сходстве переданного и принятого сигналов и возможности надежного обнаружения.

Добавим помеху, превышающую сигнал в 10 раз. На рис. 7 ВКФ имеет главный пик, позволяющий достаточно надежно обнаруживать сигнал.

На рис. 8 представлена ВКФ в случае, когда помеха превышает сигнал в 30 раз. Даже в этом случае можно говорить о высокой вероятности обнаружения.

И лишь помеха, в 100 раз превышающая сигнал, делает обнаружение ненадежным, поскольку главный пик ВКФ соизмерим с боковыми (рис. 9).

Заключение

Итак, применение ЛММ, в котором полезным сигналом является сложный шумоподобный сигнал с наилучшей автокорреляционной функцией (комбинированные сигналы Баркера), а в качестве несущего колебания используется сложное шумоподобное колебание, позволяет, несмотря на сложную помеховую обстановку, организовать надежный канал связи. Кроме того, работа "под шумами" делает такую связь секретной, поскольку сигнал в канале неотличим от помех и обнаружить сам факт сеанса связи трудно. И даже в случае перехвата сигнала злоумышленником осуществлять демодуляцию невозможно, так как для этого нужно знать несущее колебание, которое в данном случае можно рассматривать как ключ шифра.

ЛММ пригоден и для систем связи, и для систем управления. Некоторая небольшая задержка (равная длительности фрагмента сигнала), необходимая для его демодуляции, в большинстве случаев приемлема. Дальнейшие наши исследования направлены на проверку алгоритма для недвоичных сигналов.

Список литературы

1. **Голдман С.** Теория информации. М.: Изд. Иностранной литературы, 1957. 446 с.
2. **Харкевич А. А.** Теория информации. Опознавание образов / Избранные труды в 3 т. М.: Наука, 1973. Т. 3. 524 с.
3. **Самойлов А. И.** Выбор оптимального зондирующего сигнала // Геология и геофизика. 1982. № 7. С. 105—112.
4. **Ray K. J. and Wang B.** Cognitive Radio Networking and Security: A Game-Theoretic View. Cambridge University Press, 2010. 618 p.
5. **Охрименко В.** Широкополосная PLC-технология: проблемы и решения // Электронные компоненты. 2010. № 1. С. 43—49.
6. **Вольнская А. В., Сергеев Б. С.** Предпосылки применения псевдослучайных сигналов-переносчиков в каналах телемеханики железнодорожного транспорта // Транспорт: наука, техника, управление: Научный информационный сборник РАН ВИНТИ. 2011. Вып. 6. С. 39—41.
7. **Вольнская А. В.** Исследование способов адаптации сигналов к помехам в интеллектуальном канале телемеханики же-

лезнодорожного транспорта // Проблемы управления и моделирования в сложных системах: Труды XV Международной конф. Самара: Самарский научный центр РАН, 2013. С. 577—581.

8. **Вольнская А. В., Сапожников И. В.** Результаты компьютерного моделирования мультипараметрической модуляции // Проблемы управления и моделирования в сложных системах:

Труды XVI Международной конф. Самара: Самарский научный центр РАН, 2014. С. 540—548.

9. **Вольнская А. В.** Разработка устройств формирования и оптимальной обработки сложных двоичных сигналов для информационных систем железнодорожного транспорта // Информационные технологии. 2014. № 2. С. 65—70.

A. V. Volynskaya, Ph. D., Associate Professor, e-mail: anna-volinskaya@mail.ru,
The state educational institution of the higher vocational training
"Ural state university of railway transport" (URGUPS),
I. V. Sapozhnikov, Assistant Professor, fbjumper@mail.ru,
The state educational institution of the higher vocational training
"Ural technical institute of communication and informatics" (UrTISI)

Modelling of the Linear Multiparametrical Modem for Channels of Information Transfer with the Low Relation a Signal/Noise

Classical ways of modulation of signals are nonlinear procedures as are based on multiplication of the transmitted signal and the bearing fluctuation. Therefore the so-called threshold effect of a noise stability which consists that at decrease in the relation a signal/hindrance on a receiver entrance to some level close to unit, the relation the signal/hindrance at the exit falls is inherent in them, and communication becomes impossible. For example, in the CDMA standard the minimum relation a signal/hindrance — 6 dB.

If it is required to work at lower relations a signal/hindrance, it is expedient to apply the new type of modulation based on convolution of the transmitted signal from the noise-type bearing. As convolution is linear procedure, the threshold effect is absent that allows to work "under noise". Thus each value of a signal in a communication channel contains information on all transferred message in general, as defines a high noise stability. Besides, such type of communication is protected from interception of information as, first, we won't distinguish a useful signal from hindrances, and secondly, even at its interception it is impossible to carry out demodulation, without knowing the bearing fluctuation.

Results of modeling of a correctness of algorithm of modulation and assessment of its noise stability are given in article. Thus as the transmitted signals Barker's kanonic signals and new — the combined Barker's signals are used.

Modeling showed a correctness of procedures of modulation demodulation for enough difficult signals with base more than 100, and also the high noise stability allowing to work at the relations a signal/hindrance there are a lot of smaller unit.

Keywords: modulation, noise stability, convolution, binary signals, difficult noise-type signals, Barker's codes, computer modeling of a noise stability, correlation

References

1. **Goldman S.** *Teoriya informatsii* (Theory of information). M.: Izd. inostrannoi literatury, 1957. 443 p. (in Russian).
2. **Harkevich A. A.** *Teoriya informatsii. Opoznanie obrazov* (Theory of information. Identification of images). Izbrannye trudy v 3-h t. M.: Nauka, 1973. 524 p. (in Russian).
3. **Samoylov A. I.** Vyibor optimalnogo zondiruyushogo signala. *Geologiya i geofizika*. 1982, no. 7, pp. 105—112 (in Russian).
4. **Ray K. J. Liu and Wang B.** *Cognitive Radio Networking and Security: A Game-Theoretic View*. Cambridge University Press, 2010. 618 p.
5. **Ohrimenko V. S.** Hirokopolosnaya PLC-tehnologiya: problemy i resheniya. *Elektronnyie komponenti*, 2010, no. 1, pp. 43—49 (in Russian).
6. **Volynskaya A. V., Sergeev B. S.** Pregposylki primeneniya psevdosluchainykh signalov-perenoschikov v kanalah telemehaniki

zheleznodorozhnogo transporta. *Transport: nauka, tehnika, upravlenie — nauchnyi informatsionnyi sbornik RAN VINITI*. 2011, vyp. 6, pp. 39—41 (in Russian).

7. **Volynskaya A. V.** Issledovanie sposobov adaptatsii signalov k pomeham v intellektualnom kanale telemehaniki zheleznodorozhnogo transporta. *Problemy upravleniya i modelirovaniya v slozhnykh sistemah: Tr. XV Mezhdunarodnoy konf.*, Samara. 2013, pp. 577—581 (in Russian).

8. **Volynskaya A. V., Sapozhnikov I. V.** Rezultaty komputernogo modelirovaniya multiparametricheskoi modulyatsii. *Problemy upravleniya i modelirovaniya v slozhnykh sistemah: Tr. XVI Mezhdunarodnoy konf.*, Samara, 2014, pp. 540—548 (in Russian).

9. **Volynskaya A. V.** Razrabotka ustroystv formirovaniya i optimalnoi obrabotki slozhnykh dvoichnykh signalov dlya informatsionnykh sistem zheleznodorozhnogo transporta. *Informatsionnye tehnologii*. 2014, no. 2, pp. 65—70 (in Russian).

СИСТЕМЫ СВЯЗИ И ТЕЛЕКОММУНИКАЦИИ CONNECTIONS AND TELECOMMUNICATIONS SYSTEMS

УДК 621.396.99

А. Г. Мильковский, канд. техн. наук, ст. науч. сотр., генеральный директор, corp@tsniimash.ru
А. Ю. Данилюк, канд. техн. наук, доц., первый зам. генерального директора, corp@tsniimash.ru,
С. Н. Карутин, канд. техн. наук, доц., нач. центра, s.karutin@glonass-iac.ru,
П. Б. Глухов, канд. техн. наук, науч. сотр., p.glukhov@glonass-iac.ru,
ФГУП "Центральный научно-исследовательский институт машиностроения", г. Королев, МО

Основы обеспечения целостности системы ГЛОНАСС

Проведен анализ требований к целостности навигационного обеспечения потребителей системы ГЛОНАСС, описаны основные характеристики целостности. Рассмотрены основные сегменты по контролю и мониторингу навигационного поля ГЛОНАСС, а также методы формирования признаков целостности глобальных национальных спутниковых систем. Описаны особенности реализации алгоритма контроля целостности при высокоточных навигационных определениях. Сделаны выводы и даны рекомендации по развитию основных направлений по повышению целостности услуг ГЛОНАСС.

Ключевые слова: глобальные навигационные спутниковые системы, ГЛОНАСС, целостность, точность, RAIM, SBAS, GBAS, СДКМ, precise point positioning, защитный уровень

Введение

Повышение тактико-технических характеристик системы ГЛОНАСС приводит к постоянному расширению круга практических задач в области связи, энергетики, транспорта, имущественных отношений, сельского хозяйства и других, решение которых невозможно уже представить без отечественной глобальной навигационной спутниковой системы (ГНСС). Возрастающая зависимость национальной экономики от состояния и стабильности функционирования ГЛОНАСС показывает необходимость улучшения не только точности и доступности, но и целостности и непрерывности навигационного обеспечения.

Под *целостностью* понимается способность навигационной системы выдавать потребителю своевременное и достоверное предупреждение в тех случаях, когда какие-либо сигналы нельзя использовать по целевому назначению в полном объеме [1]. Контроль целостности ГНСС подразумевает комплекс мероприятий по определению соответствия заданным характеристикам способности этих систем обеспечивать потребителей сигналами тревоги о недостоверности навигационных сигналов ГНСС [2].

Нарушения целостности возникают при сбое в трех сегментах ГНСС: космическом (сегмент навигационного космического аппарата); наземном (сегмент управления); пользовательском, а также в среде распространения радиосигналов. Данные сбои потенциально могут нанести вред потребителю до

того момента, пока он о них не оповещен либо средствами самой системы через скорректированное служебное сообщение, либо через автономный контроль или через вспомогательные наземные службы контроля и мониторинга качества навигационного поля ГНСС. Поэтому задача контроля целостности, являясь критической, должна решаться на разных уровнях:

- на борту навигационного космического аппарата (НКА) с использованием средств бортовой диагностики;
- наземными средствами контроля и управления системы;
- с использованием функциональных дополнений ГНСС;
- в аппаратуре потребителя.

Требования к целостности навигационного обеспечения

В совокупности целостность можно описать такими параметрами, как порог срабатывания тревоги, риск потери целостности, время срабатывания тревоги и защитный уровень [3]:

- *порог срабатывания тревоги (alert limit — AL)* — превышение допустимого отклонения для измеряемого параметра, вызывающее срабатывание тревоги;
- *время срабатывания тревоги (time to alert — TTA)* — максимальное допустимое время между момен-

том выхода навигационной системы за допустимый порог и срабатывания тревоги;

- *риск потери целостности* — вероятность того, что в любой момент времени ошибка определения местоположения (*position error* — *PE*) превысит порог срабатывания тревоги;
- *защитный уровень* (*protection level* — *PL*) — статистическая граничная ошибка, рассчитанная для обеспечения того, чтобы вероятность абсолютной ошибки определения местоположения, превышающей указанную величину, была меньше или равна заданным пределам риска потери целостности;
- *потеря целостности* — событие, длящееся дольше времени срабатывания тревоги, при котором не происходит срабатывания тревоги.

Из рис. 1 [4] можно понять разницу между двумя типами событий в целостности: событий, несущих недостоверную информацию (*misleading information* — *MI*) и опасно недостоверную информацию (*hazardously misleading information* — *HMI*):

- событие *MI* возникает, когда система имеет статус доступной, при этом ошибка определения местоположения превышает безопасный уровень, но не порог срабатывания тревоги;
- событие *HMI* возникает, когда система имеет статус доступной, при котором ошибка определения местоположения превышает порог срабатывания тревоги.

Требуемые значения характеристик контроля целостности изменяются в широких пределах в зависимости от вида потребителей, динамики их движения и оперативности использования уточняемых навигационных параметров при решении потребителями своих целевых задач (подробнее в работе [1]). При этом требования к времени предупреждения о нарушении целостности у морских, речных и на-

земных потребителей ниже, чем у воздушных потребителей, что обусловлено, прежде всего, более высокими скоростями перемещения последних и повышенными рисками, возникающими в условиях, когда решение навигационной задачи по точности ниже требуемого уровня.

Контроль качества параметров радионавигационного поля ГНСС ГЛОНАСС

Решение задачи целостности наземными средствами контроля системы. В настоящее время для организации внутрисистемного контроля качества навигационных услуг силами наземного сегмента в системе ГЛОНАСС используется аппаратура контроля навигационного поля (АКНП), установленная в двух пунктах (в Московской области и г. Комсомольске-на-Амуре). Результаты навигационных определений (а в случае отрицательных данных контроля — и вся соответствующая измерительная информация) передаются в центр управления системой (ЦУС) (Московская обл.), проверяются на достоверность и записываются в центральную базу данных. Последующая обработка информации проводится с помощью комплекса программ контроля навигационного поля.

Оповещение о недостоверности навигационного сигнала предусматривает:

- кратное обнаружение на текущем секундном интервале недостоверности навигационного радиосигнала НКА с номером *n* средствами непрерывного (эпизодического) слежения;

- наискорейшую, без задержек, передачу в ЦУС информации о недостоверности навигационного сигнала;

- принятие в ЦУС решения о недостоверности навигационного сигнала НКА с номером *n* и необходимости выдачи на него соответствующего признака;

- реализацию доставки признака недостоверности навигационного сигнала на НКА с номером *n*;
- оперативное включение в навигационный кадр признака недостоверности;

- возможность передачи потребителю в навигационном сигнале упомянутого признака с необходимой частотой повторения, обеспечивающей наискорейшее от начала приема предупреждение об исключении радиосигнала НКА при уточнении местоположения;

- контроль целостности ГНСС, под которым понимается регистрация обнаружения и прохождения признака недостоверности навигационного сигнала на всех этапах с периодическим предоставлением квитанции в центр управления для апостериорного анализа и в случаях чрезвычайных происшествий.

Формируемый АКНП признак неисправности C_n ($n = 1, \dots, 24$) появляется в неоперативной информации навигационных сообщений (альманахах сис-

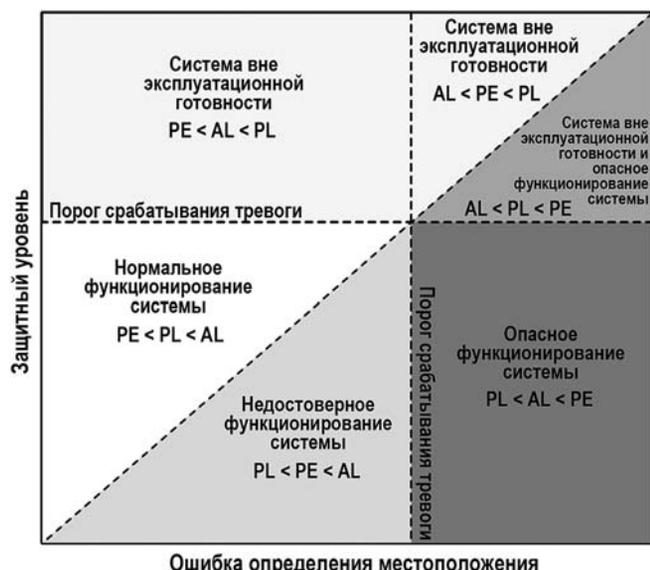


Рис. 1. Стэнфордская диаграмма

темы) всех спутников не позднее чем через 16 ч после появления неисправности. Дискретность передачи данного признака в навигационных сообщениях НКА ГЛОНАСС составляет 2,5 мин.

Решение задачи целостности на борту НКА. Помимо АКНП, на борту всех НКА ГЛОНАСС осуществляется непрерывный автономный контроль функционирования основных бортовых систем. В случае обнаружения нарушений нормального функционирования этих систем, влияющих на качество излучаемого спутником навигационного радиосигнала и достоверность передаваемого навигационного сообщения, на спутнике формируется признак его неисправности B_n ($n = 1, \dots, 24$), который передается потребителю системы в составе оперативной (эфемеридной) информации навигационного сообщения. Дискретность передачи соответствующего признака в навигационных сообщениях НКА ГЛОНАСС составляет 30 с (до 10 с в НКА серии "Глонасс-М"). Максимальная задержка от момента обнаружения неисправности до момента передачи соответствующего признака не превышает 1 мин [5].

В новых навигационных сигналах с кодовым разделением предусмотрена передача двух различных признаков, которые могут как формироваться на борту НКА, так и инициироваться командами наземных средств управления:

- Γ_n — оперативный признак непригодности навигационного радиосигнала n -го НКА;
- l_n — оперативный признак непригодности цифровой информации (ЦИ) n -го НКА в данной строке навигационного кадра.

Признак непригодности ЦИ l_n трактуется как "нестрогий" и говорит о том, что потребитель имеет возможность использовать ранее принятую ЦИ, в составе которой отсутствовал данный признак. Следует отметить, что с точки зрения решения задач навигационных определений признак l_n является

критичным только для строк, содержащих эфемеридно-временную информацию.

Признак непригодности радиосигнала Γ_n формируется на основе автономного контроля в аппаратуре бортового информационно-навигационного комплекса (БИНК) "гладкости" шкалы времени бортового синхронизирующего устройства и формируется, когда фаза навигационного радиосигнала начинает уход от средневзвешенного значения. Для потребителя появление этого признака говорит о необходимости исключения данного НКА из навигационной выборки.

Признаки Γ_n и l_n следуют в служебной части каждой строки навигационного кадра, а время воспроизведения строк в сигнале L3 составляет 3 с, в сигнале L1, L2 — 2 с. Таким образом, индикация результатов автономного контроля БИНК может выполняться не позднее 2—3 с с момента обнаружения.

На текущий момент недостатки бортового канала контроля целостности заключаются в его неполноте. Например, средства самоконтроля рассчитаны на обнаружение не всех возможных нарушений в работе каждой бортовой системы НКА; неисправности самих средств контроля не обнаруживаются и не сопровождаются передачей соответствующего сообщения потребителям; искажение бортовых эфемерид и неправильное прогнозирование частотно-временных поправок не могут быть обнаружены на самом НКА и т. д. [5].

Решение вопросов повышения надежности бортового канала контроля целостности и развития средств бортовой самодиагностики является в настоящее время перспективным направлением, требует дальнейшей проработки в целях максимального парирования неожиданных сбоев ГЛОНАСС и оперативного доведения информации до потребителя. При этом стоит учитывать опыт и планы развития других навигационных систем, разработчики

Требования к бортовой целостности и непрерывности перспективных НКА ГНСС

Аномальная величина	Значение	Вероятность или время
Превышение URE^1	>7 м	$7,6 \cdot 10^{-7}$ в час
Превышение передаваемого порога ошибки псевдодальности	$URE > 4,42 \cdot URA^2$	$7,6 \cdot 10^{-6}$ в час
Вероятность ложной тревоги	$URA > 2,5 \cdot \text{суточная СКП } URE$	$< 1\%$ в год
Запланированный перерыв в работе	$< 1,3$ перерывов в год	1 в год
Средняя продолжительность запланированного перерыва в работе	< 12 ч один перерыв	7,3 ч
Незапланированный перерыв в работе	$< 1,3$ отказов в год	0,55 в год
Средняя продолжительность незапланированного перерыва в работе	< 36 ч один отказ	17 ч
Основной отказ обслуживания	$URE > 4,42 \cdot URA$ или > 30 м	$7,6 \cdot 10^{-7}$ в час
Ошибки за счет скачков фазы	$> 3,6$ м & < 300 м	$7,9 \cdot 10^{-8}$ в час
Погрешность скорости изменения псевдодальности	$> 0,01$ м/с & $< 0,05$ м/с	$7,6 \cdot 10^{-7}$ в час
Погрешность ускорения изменения псевдодальности	$> 0,019$ м/с ²	$1,7 \cdot 10^{-6}$ в час
Деформация сигнала	Опережение/запаздывание ≥ 10 нс Ring $\geq 7,3$ МГц	0,0 в час

¹ URE (User Range Error) — ошибка псевдодальности за счет космического сегмента

² URA (User Range Accuracy) — оценка ожидаемой ошибки псевдодальности, передающаяся в навигационном сообщении

которых уже на этапе создания новых НКА внедряют в процесс производства стандарты и рекомендации по разработке аппаратных и программных средств в авиации, что позволяет повысить показатели бортовой целостности и непрерывности перспективных НКА (см. таблицу) [6] для удовлетворения растущих потребностей военных, коммерческих и гражданских пользователей.

Автономный мониторинг целостности. Благодаря избыточности навигационной информации, получаемой аппаратурой потребителя от рабочего созвездия НКА, возможна организация автономного контроля целостности (Receiver Autonomous Integrity Monitoring — RAIM) в навигационном приемнике.

Существует множество разнообразных способов реализации RAIM, отличающихся статистическим подходом, применяемыми правилами анализа, необходимым числом спутников. Однако в основе всегда лежит анализ внутренней сходимости измерений, выполняющийся путем сравнения координатных решений или их функций, полученных по различным подмножествам наблюдаемой совокупности НКА.

Среди методов RAIM можно выделить фильтрационные методы и методы мгновенной оценки, методы на основе анализа вектора четности или оценки СКО кодовых измерений по результатам фазовых.

В методах мгновенной оценки используется информация, полученная в отдельную эпоху измерений. Предыдущие измерения в построении оценки не участвуют. К данному типу относятся следующие алгоритмы.

Алгоритм оценки максимальной разности решений. При наличии N измерений ($N > 4$) формируются комбинации из $N - 1$ измерений. В качестве тестовой статистики рассматривается максимальная разность решений, полученная по этим комбинациям.

Алгоритм сравнения дальности. Здесь также формируются комбинации из $N - 1$ измерений ($N > 4$), затем дальность до НКА, не участвовавшего в решении, вычисляется с использованием полученных на предыдущем шаге координат. Тестовой статистикой является разность между измеренным значением дальности и ее прогнозом.

Алгоритм сравнения положения. Решение выполняется по комбинации из $N - 1$ измерений ($N > 4$) и по всем N измерениям. В качестве тестовой статистики используются разности полученных решений.

Алгоритм анализа невязок, полученных методом наименьших квадратов (МНК). По результатам измерений выполняется решение по методу МНК. Далее полученные координаты используются для вычисления дальностей до каждого НКА. Эти дальности сравниваются с результатами измерений. Полученная разность и является тестовой статистикой.

Неиспользование полученной ранее измерительной информации является общим недостатком методов мгновенной оценки, что позволяет обнаружить факт отказа только в момент достаточно сильного проявления, а не в момент возникновения.

В фильтрационных методах RAIM используются результаты предыдущих измерений, которые подаются на вход того или иного фильтра, например фильтра Калмана, в целях построения оценки навигационного параметра или положения потребителя на очередную эпоху. Далее эта оценка сравнивается с результатами измерений. Недостатками таких методов являются сложный математический аппарат, который требует значительного времени на выполнение вычислений, и, главное, трудности, возникающие при выявлении плавно нарастающих ошибок навигационных параметров.

Методы, основанные на анализе вектора четности, более корректны в математическом отношении, чем те, что основаны на анализе невязок. Они получили наибольшее распространение, а один из них был рекомендован к использованию в авиационной навигационной аппаратуре. Однако и этот метод сложен в вычислительном отношении, поскольку требует численного интегрирования плотности вероятности функции смещенного χ^2 -распределения для различного числа степеней свободы.

Метод оценки среднего квадратичного отклонения (СКО) кодовых измерений по результатам фазовых предполагает сравнение приращения дальности потребитель—спутник между последовательными эпохами измерений, вычисленного по результатам кодовых измерений с тем же приращением, полученным по фазовым измерениям. По невязкам вычисляется СКО кодовых измерений. Если это значение превышает номинальное для ГНСС значение, то делается вывод о возникновении сбоя. Недостатком метода является отсутствие анализа данных в эфемеридном сообщении. Кроме того, весьма трудно выявляются сбои, связанные с недопустимо высоким дрейфом бортового стандарта частоты.

К безусловным достоинствам методов автономного мониторинга целостности относятся:

- возможность при наличии избыточных навигационных измерений выявления одиночных или групповых сбоев в работе НКА непосредственно в аппаратуре потребителя;
- обеспечение мгновенной выдачи сигналов предупреждения при превышении значений решающей статистики априорно заданных значений уровней предупреждения.

Основные недостатки данных методов следующие:

- ✓ зависимость их работоспособности от числа используемых НКА;
- ✓ необходимость расчета и использования различных значений решающих порогов для конкретных параметров рабочего созвездия и значений геометрического фактора (ГФ). Ошибки,

связанные с выбором порога отбраковки изменений, могут приводить в стандартных RAIM алгоритмах к необоснованному длительному исключению из обработки "хорошего" спутника, вызывая тем самым недопустимое ухудшение ГФ и небезопасное снижение точности.

Вместе с тем стоит отметить, что в последние годы проводятся активные работы по развитию алгоритмов автономного контроля целостности. Наиболее масштабные исследования выполняет Федеральная авиационная администрация США, которая разрабатывает улучшенный алгоритм автономного контроля целостности (Advanced RAIM — ARAIM). Основной целью работ по развитию ARAIM является обеспечение возможности управления по высоте воздушными судами в глобальном масштабе на основании сигналов НКА нескольких ГНСС, предоставляющих услуги посредством навигационных радиосигналов в двух и более частотных диапазонах. Совместное использование сигналов двух и более ГНСС снижает уязвимость навигационного обеспечения потребителя в условиях деградации какой-либо системы.

Информация целостности функциональных дополнений. При решении практических задач, критически важных для безопасности человека (в первую очередь в гражданской авиации), в ГНСС широко используют системы функциональных дополнений наземного (Ground Based Augmentation System — GBAS) и космического базирования (Space Based Augmentation System — SBAS). Эти системы обеспечивают потребителя дополнительной корректирующей информацией и информацией целостности, формируемой на основе измерений текущих навигационных параметров (ИТНП) одной или сети наземных станций сбора измерений.

Дифференциальная система космического базирования SBAS является гораздо более сложной системой, чем GBAS, поскольку разработана для обеспечения потребителя информацией, позволяющей предотвращать использование ГНСС при сбоях в

их работе, не обнаруженных или не исправленных ее собственными средствами в процессе эксплуатации. Если при автономном мониторинге в GBAS решается задача обнаружения сбоя вне зависимости от причины его возникновения, то в основе SBAS лежит принцип раздельной компенсации составляющих погрешности навигационно-временных определений (НВО) в целях уменьшения их значений до приемлемого уровня. Кроме того, SBAS осуществляет оценку потенциальных угроз нарушения целостности и обеспечивает потребителя данными, гарантирующими, что остаточная погрешность НВО не превысит априорно заданного уровня.

В соответствии с работой [7] основными причинами нарушения целостности ГНСС являются сбои в работе НКА ГНСС, космических аппаратов (КА) на геостационарной орбите (ГСО), станций мониторинга, центров управления и средств закладки информации на борт КА на ГСО, а также возмущения среды распространения.

Структурная схема алгоритма контроля целостности SBAS, основанная на проведении анализа корректирующей информации (КИ) в целях обеспечения целостности с учетом перечисленных выше источников угрозы ее нарушения, приведена на рис. 2.

В соответствии с требованиями [8] технические средства мониторинга должны обеспечивать формирование КИ, применение которой совместно с ИЦ обеспечит точность НВО потребителя, удовлетворяющую априорно заданным пороговым значениям. Поскольку для решения задач НВО используется метод наименьших квадратов, то при формировании ИЦ оцениваются значения погрешности эфемеридно-временной информации (ЭВИ) после применения КИ σ_{UDRE} (User Differential Range Error — UDRE) и погрешности компенсации ионосферной погрешности в узлах ионосферной сетки σ_{GIVE} (Grid Ionospheric Vertical Error — GIVE). При этом для формирования КИ и ИЦ используют два подхода — "читай, затем контролируй" и "контролируй, затем считай", которые одинаково приемлемы для



Рис. 2. Структурная схема алгоритма контроля целостности в SBAS

построения системы SBAS. Первый подход предполагает, что на основе вторичной обработки ИТНП формируются КИ и ИЦ, которые затем подвергаются процедуре верификации в целях проверки соответствия достигнутого уровня точности пороговым значениям. Подобная верификация позволяет в случае возникновения сбоя выдать потребителю предупреждение о том, что либо соответствующий НКА нельзя использовать, либо его мониторинг не осуществляется. Второй подход предполагает, что первоначально ИТНП подвергается тщательной проверке на соответствие требованиям к целостности. При таком подходе сформированные КИ и ИЦ больше не подвергаются проверке.

Стоит отметить, что одной из важнейших характеристик услуг функциональных дополнений SBAS является максимально допустимое время предупреждения о нарушении целостности в случае возникновения и обнаружения угрозы такого события, которое определяется временем формирования ИТНП и их доставки в центр обработки, задержками на обработку и формирование КИ и ИЦ, временем формирования сообщения и задержкой на трассе распространения "земная станция — КА на ГСО" и "КА на ГСО — потребитель".

В России к функциональным дополнениям типа SBAS относится система дифференциальной коррекции и мониторинга (СДКМ), которая, как и аналогичные зарубежные системы, предназначена для формирования и доставки потребителям широкозонной корректирующей информации и информации о целостности навигационного поля ГНСС ГЛОНАСС и GPS.

Целостность при высокоточном абсолютном местопределении. Внедрение новых навигационных радиосигналов в нескольких частотных диапазонах, увеличение числа НКА, развертывание новых глобальных и региональных навигационных систем, создание глобальных сетей измерительных станций сделали доступными для большого числа потребителей возможности абсолютных высокоточных навигационных определений.

Метод высокоточного абсолютного местопределения (Precise Point Positioning — PPP) основан на использовании ионосферно-свободных двухчастотных ИТНП НКА, учете детальной физической модели измерений и высокоточной эфемеридно-временной информации.

Добиться высокой точности навигационных определений позволяет однозначная физическая сущность явлений, влияющих на соответствующие составляющие бюджета погрешности ИТНП. Несмотря на то что физическая природа данных явлений хорошо изучена, что позволяет выявлять аномалии в их состоянии, наличие целого ряда внешних условий, необходимых для успешного функционирования алгоритмов PPP, является серьезным огра-

ничением, с которым нельзя не считаться. К этим внешним условиям можно отнести:

- необходимость непрерывного наблюдения и формирования ИТНП по радиосигналам большого числа НКА в течение длительных интервалов времени (5...30 мин);
- необходимость проведения навигационных определений по всем видимым НКА, в том числе входящим в орбитальные группировки различных ГНСС;
- зависимость от точности и оперативности обновления ЭВИ, а также надежности линии передачи данной информации в аппаратуре потребителя.

Опуская детальный анализ технологии PPP, отметим, что методический аппарат оценки целостности навигационного обеспечения, разработанный при создании SBAS, может быть использован для оценки качества высокоточного навигационного обеспечения с целью гарантировать их достоверность.

Поскольку контроль целостности осуществляется в рамках решения конкретных детально описанных задач, то разработка алгоритмов контроля целостности при решении задач высокоточной навигации основывается на тщательном анализе требований, предъявляемых потребителем в каждом конкретном случае. Наиболее сложной задачей является обеспечение целостности при управлении наземным транспортным средством, поскольку остальные типы потребителей (воздушные и морские суда, железнодорожный транспорт) работают в гораздо более благоприятных условиях.

В целом при решении задачи разработки алгоритма контроля целостности при высокоточных навигационных определениях требуется учитывать следующие факторы.

1. Метод PPP определяет абсолютное, а не относительное местоположение аппаратуры потребителя, вследствие чего при наличии ошибок в привязке локальной геодезической системы координат к общеземной могут возникать ошибки в интерпретации полученных результатов навигационных определений.

2. Использование ИТНП, полученных по несущим колебаниям, привносит дополнительные требования к качеству работы схемы слежения за фазой аппаратуры потребителя, особенно при сбоях в работе схемы фазовой автоподстройки частоты в части перескоков из одного устойчивого состояния в другое.

3. Достижение высокой точности возможно только при использовании актуальной высокоточной ЭВИ, что обеспечивается качественной работой центров ее формирования и надежностью каналов передачи данной информации.

4. Соответствующие уровни защиты и тревоги необходимо выбирать с некоторым запасом для снижения вероятности ложной тревоги.

Заключение

Необходимость сохранения за Российской Федерацией лидирующих позиций не только в сфере навигационно-временного обеспечения, но и в сфере обеспечения конкурентоспособности отечественной высокотехнологичной продукции невозможно без развития услуг на базе радиосигналов системы ГЛОНАСС. Активное использование ГЛОНАСС для решения задач, связанных с обеспечением безопасности жизнедеятельности в условиях высоких требований по точности навигационно-временных определений, требует системной увязки подходов по повышению точности и целостности услуг ГЛОНАСС.

Совершенствование бортового канала контроля целостности и средств самодиагностики, модернизация наземного сегмента контроля и управления системы, внедрение улучшенных алгоритмов автономного контроля целостности и их интеграция с возможностями функциональных дополнений ГНСС, а также развитие методического аппарата оценки целостности при высокоточном навигационном обеспечении позволят в ближайшей перспективе повысить устойчивость функционирова-

ния космической системы ГЛОНАСС, расширить ее функциональные возможности, сферы использования, закрепив тем самым участие России на мировом рынке космических товаров и услуг.

Список литературы

1. **Радионавигационный** план Российской Федерации. М.: Интернавигация, 2010.
2. **ГОСТ Р 54460—2011.**
3. **Карутин С. Н., Власов И. Б., Дворкин В. В.** Дифференциальная коррекция и мониторинг глобальных навигационных спутниковых систем. М.: Изд-во Московского университета, 2014. 464 с.
4. <http://waas.stanford.edu/metrics.html> (дата обращения 20.07.2014).
5. **Перов А. И., Харисов В. Н.** ГЛОНАСС. Принципы построения и функционирования. 5-е изд., перераб. и доп. М.: Радиотехника, 2010. 800 с.
6. **Shaw S., Katronick A. J.** GPS III Signal Integrity Improvements // Proc. of the 26th International Technical Meeting of the ION Satellite Division, ION GNSS+ 2013, Nashville, Tennessee, September 16—20, 2013. P. 936—945.
7. **Minimum** Operational Performance Standards for Global Positioning / Wide Area Augmentation System Airborne Equipment, RTCA/DO-229D, Prepared by SC-159. RTCA Inc. Washington, D. C., December 13, 2006.
8. **Software** Consideration in Airborne Systems and Equipment Certification (DO-178C). RTCA. 2011.

A. G. Milkovskiy, PhD., Senior Research Associate, Director General, corp@tsniimash.ru,
A. Y. Danilyuk, PhD., Associate Professor, Senior Deputy Director General, corp@tsniimash.ru,
S. N. Karutin, PhD., Associate Professor, Head of IAC PNT, s.karutin@glonass-iac.ru,
P. B. Glukhov, PhD., Research Associate, p. glukhov@glonass-iac.ru
Central Research Institute for Machine Building, Federal State Unitary Enterprise,
Korolev, Moscow Reg., Russia

Ensuring GLONASS Integrity

Requirements on GLONASS integrity are analyzed; basic integrity parameters are described. Core GLONASS performance monitoring elements are addressed together with the methods for GNSS integrity indicators generation onboard a navigation satellite through the navigation system ground control facilities, in a user receiver and space- and ground based augmentation systems. Trade off analysis on basic techniques for receiver autonomous integrity monitoring (RAIM) is conducted. Structure chart of SBAS integrity monitoring algorithm is provided based on correction information analysis ensuring the GNSS integrity. Main distinctive features of RAIM in high precision navigation solutions are described. Conclusions and recommendations are given as to major improvement directions of GLONASS integrity.

Keywords: global navigation satellite systems, GLONASS, integrity, accuracy, RAIM, ARAIM, SBAS, GBAS, SDCM, precise point positioning, protection level

References

1. **Radionavigatsionnyy plan Rossiyskoy Federatsii.** Moscow, Intemavigatsiya, 2010.
2. **GOST R 54460—2011.**
3. **Karutin S. N., Vlasov I. B., Dvorkin V. V.** *Differentsialnaya korrektsiya i monitoring globalnikh navigatsionnikh sputnikovikh sistem.* Moscow, Izdatelstvo Moskovskogo Universiteta, 2014, 464 p.
4. <http://waas.stanford.edu/metrics.html>
5. **Perov A. I., Kharisov V. N.** *GLONASS. Printsipy postroyeniya i fuktsionirovaniya.* Moscow, Radiotekhnika, 2010, 800 p.
6. **Shaw S., Katronick A. J.** GPS III Signal Integrity Improvements. *Proc. of the 26th International Technical Meeting of the ION Satellite Division, ION GNSS+ 2013, Nashville, Tennessee, September 16—20, 2013.* P. 936—945.
7. **Minimum** Operational Performance Standards for Global Positioning. *Wide Area Augmentation System Airborne Equipment, RTCA/DO-229D, Prepared by SC-159.* RTCA Inc. Washington, D. C., December 13, 2006.
8. **Software** Consideration in Airborne Systems and Equipment Certification (DO-178C) / RTCA. 2011.

Д. А. Перепелкин, канд. техн. наук, доц., e-mail: dmitryperpelkin@mail.ru
ФГБОУ ВПО "Рязанский государственный радиотехнический университет"

Концептуальный подход динамического формирования трафика программно-конфигурируемых телекоммуникационных сетей с балансировкой нагрузки

Предложена математическая модель динамического формирования трафика программно-конфигурируемых телекоммуникационных сетей с балансировкой нагрузки, позволяющая повысить эффективность процессов маршрутизации в условиях динамических изменений параметров линий связи за счет использования дополнительной информации о резервных маршрутах.

Ключевые слова: телекоммуникационные сети, программно-конфигурируемые сети, алгоритмы маршрутизации, адаптивная маршрутизация, алгоритм парных перестановок маршрутов, многопутевая маршрутизация, балансировка нагрузки, качество обслуживания, динамические изменения, резервные маршруты, сетевой трафик

Введение

Внедрение новых сервисов и услуг в современных телекоммуникационных сетях заставляет операторов связи искать новые подходы обеспечения требуемого качества обслуживания (*Quality of Service, QoS*) сетевых запросов и приложений, направленные на повышение быстродействия и надежности сетей, возможностей интегрированной передачи данных, голоса и видеоинформации. Обеспечение высокоскоростного и надежного обмена информацией между узлами телекоммуникационной сети при жестких требованиях к задержкам передачи информации в условиях возможных всплесков трафика в каналах связи является одной из важнейших проблем.

Для повышения *QoS*-параметров телекоммуникационных сетей необходимо использовать эффективные модели и алгоритмы адаптивной маршрутизации [1]. В связи с этим особое внимание уделяется внедрению и поддержке решений многопутевой стратегии маршрутизации (*Multipath routing*), использованию композитных метрик линий связи, максимально учитывающих численные значения различных *QoS*-параметров, а также обеспечению масштабируемости маршрутных решений, т. е. способности сохранить в заданных пределах свою эффективность в условиях роста территориальной распределенности телекоммуникационной сети, числа и типов обслуживаемых трафиков пользователей.

Разработка новых перспективных подходов для решения задачи динамического управления трафиком телекоммуникационных сетей с балансировкой нагрузки позволяет повысить эффективность их функционирования за счет уменьшения трудоемкости построения оптимальных маршрутов и обеспечить надежность и быстродействие передачи данных.

Постановка задачи

Выбор маршрутов в узлах связи телекоммуникационной системы проводится в соответствии с реализуемым алгоритмом маршрутизации.

В современных протоколах маршрутизации (OSPF, IGRP и EIGRP) для построения таблиц маршрутизации используют два известных алгоритма: Беллмана—Форда с трудоемкостью порядка $O(N^3)$ и Дейкстры с трудоемкостью $O(N^2)$, где N — число маршрутизаторов в телекоммуникационной сети [2—4]. Данные алгоритмы обеспечивают автоматическое обновление таблиц маршрутизации и оперативно реагируют на все изменения конфигурации связей. Однако применение этих алгоритмов в условиях динамического формирования трафика при построении дерева оптимальных маршрутов в телекоммуникационных сетях является неэффективным ввиду высокой трудоемкости поиска оптимальных маршрутов и необходимости полного пересчета таблиц маршрутизации.

Развитие и внедрение в телекоммуникационных сетях технологии программно-конфигурируемых сетей (*Software-Defined Networks, SDN*) [5—7] позволяет формулировать различные задачи оптимальной маршрутизации [8, 9] и балансировки сетевого трафика [10, 11]. Активные научные исследования в рамках технологии *SDN* ведутся по развитию моделей и алгоритмов многопутевой маршрутизации. Так, в работе [12—14] предложена потоковая модель многопутевой маршрутизации с балансировкой нагрузки на основе теории *GERT*-сетей и генетических алгоритмов. В работах [15, 16] предложена потоковая модель многопутевой маршрутизации с балансировкой нагрузки и учетом критерия минимальной многопутевой задержки передачи и критерия равенства нулю оценки джиттера.

В работе [17] предложен алгоритм парных переходов, позволяющий за счет сбора дополнительной

информации учесть возможные изменения параметров линий связи телекоммуникационной сети и не выполнять полный пересчет маршрутных таблиц. Это дало возможность снизить трудоемкость расчета таблиц маршрутизации до значения порядка $O(kN)$, где k — число фактически выполненных парных переходов. В работах [18–21] на основе алгоритма парных переходов предложены эффективные алгоритмы адаптивной ускоренной маршрутизации при динамическом подключении узлов и линий связи в сети, которые также позволили снизить трудоемкость построения оптимальных маршрутов передачи данных до значения $O(kN)$.

В работах [22–28] предложен алгоритм парных перестановок маршрутов при динамических изменениях в структуре и нагрузках на линиях связи телекоммуникационной сети, что позволило снизить трудоемкость построения оптимальных маршрутов передачи данных до значения $O(N)$.

Разработка новых, более эффективных моделей и алгоритмов динамического формирования трафика телекоммуникационных сетей с балансировкой нагрузки в условиях динамических изменений параметров линий связи позволяет повысить надежность и быстродействие передачи данных и уменьшить трудоемкость построения таблиц маршрутизации до значения $O(N)$ за счет использования дополнительной информации о резервных маршрутах.

Разработка математической модели

В общем случае для решения задачи динамического формирования трафика применяется графовая модель телекоммуникационной сети, в которой множество вершин графа соответствует множеству узлов связи или маршрутизаторов в сети, а множество ребер — возможным каналам связи между этими узлами. Каждое ребро, соответствующее каналу связи, имеет свой вес. На практике весу ребра могут соответствовать стоимость аренды канала связи, затраты на оплату единицы трафика, передаваемого по каналу связи, соответствующему данному ребру, либо более сложная функция, учитывающая большее число параметров корпоративной сети.

Математическую модель телекоммуникационной сети представим в виде неориентированного взвешенного связного графа $G = (V, E, W, Z)$, где V — множество вершин (узлов связи); $|V| = N$, E — множество ребер (каналов или линий связи); $|E| = M$, W — множество весов ребер (стоимость каналов или линий связи); Z — множество трафиков (пользовательских приложений, служебных данных и т. д.).

Пусть на графе G в некоторый момент времени уже решена задача поиска оптимальных маршрутов до всех узлов множества $V_s = V \setminus \{v_s\}$ из начального узла v_s , т. е. построено дерево оптимальных маршрутов с корнем в узле v_s . Обозначим это дерево как T_g . Рассмотрим множество каналов E графа G .

По признаку вхождения каналов в дерево T_g можно разделить исходное множество E на два подмножества: $E_T \in T_g$ и $E_R \notin T_g$, $E_T \cup E_R = E$.

Множество каналов дерева E_T — это множество каналов дерева T_g для графа G . Для заданного графа G , согласно свойству дерева, мощность множества E_T будет равняться мощности множества V минус единица: $|E_T| = |V| - 1$.

Множество каналов замены для дерева E_R — это множество каналов графа G , не вошедших в дерево T_g . При соответствующих условиях некоторый канал $e_{i,j} \in E_R$, инцидентный узлам v_i и v_j , может перейти в множество каналов дерева E_T , заменив собой некоторый канал $e_{k,p} \in E_T$. При этом инцидентность канала $e_{k,p}$ узлу v_i или v_j является обязательным условием. В свою очередь, канал $e_{i,j}$ перейдет в множество E_R . Будем называть такие переходы парными переходами и обозначать их $e_{i,j} - e_{k,p}$.

В множестве E_R можно выделить два подмножества. Множество каналов замены E_S для дерева — это такое подмножество множества E_R , элементы-каналы которого участвуют, по крайней мере, в одном отношении парного перехода. Множество непарных каналов E_P — это такое подмножество множества E_R , элементы-каналы которого не участвуют ни в одном отношении из множества R . В общем случае множество E_P может быть пустым, $|E_P| = 0$. Множество E_S будет пустым только при условии, что исходный связный граф G является деревом, и тогда задача поиска оптимальных маршрутов лишена смысла.

Обозначим $w_{i,j}$ — маршрутную метрику канала, соединяющего узлы v_i и v_j . Узел v_i располагается относительно v_j ниже по иерархии в дереве оптимальных маршрутов. Множество E_T — множество каналов, каждый элемент которого входит, по крайней мере, в один оптимальный маршрут из начального узла связи; E_R — множество остальных каналов, т. е. $E_R \cup E_T = E$, $E_R \cap E_T = \emptyset$. Обозначим V_T — множество узлов, до которых найден оптимальный маршрут из начального узла связи; V_R — множество остальных узлов, т. е. $V_R \cup V_T = V$, $V_R \cap V_T = \emptyset$.

Для каждого канала связи $e_{i,j} \in E$ на шкале значений весов определены точки вхождения в дерево $w_{i,j}^t$ и в множество замены $w_{i,j}^s$, причем $w_{i,j}^t \leq w_{i,j}^s$. Под точками вхождения понимается максимально возможный вес канала $e_{i,j}$ при его вхождении в множество каналов дерева $E_T \in T_g$ и в множество каналов замены для дерева $E_S \notin T_g$, соответственно.

Будем называть R_k совокупностью подмножества $V^{(Vk)} \subseteq V$ узлов, через которые проходит оптимальный маршрут до узла v_k из исходного узла v_s , и подмножества $E^{(Vk)} \subseteq E$ каналов, составляющих этот маршрут.

Назовем V_k деревом T_k , или совокупностью подмножества $V_T^{(V_k)} \subseteq V$, состоящего из всех узлов, оптимальные маршруты до которых из исходного узла содержат узел v_k , и подмножества $E_T^{(V_k)} \subseteq E$ каналов, составляющих эти маршруты после v_k при движении от узла v_s .

Обозначим множество маршрутов до узла v_i из исходного узла v_s через Π_i , где элемент множества $\pi_{i,k} \in \Pi_i$ будет множеством не повторяющихся каналов $e_{i,j} \in E$, образующих вместе маршрут, соединяющий v_s и v_i . Каждому $\pi_{i,k} \in \Pi_i$ поставим в соответствие число, равное сумме весов входящих в него каналов, т. е. длину или маршрутную метрику маршрута $d_{i,k} \in D_i$, где D_i представляет собой множество оценок оптимальных маршрутов до узла v_i из исходного узла v_s . На множестве Π_i задан селектор H , возвращающий оптимальный маршрут из множества Π_i . В том случае, если существует несколько маршрутов в Π_i с минимальной длиной, то выбирается один из них. Оптимальный маршрут до узла v_i будем обозначать $\pi_i = H(\Pi_i)$, оценку его маршрутной метрики — d_i .

Каждому трафику из множества Z поставим в соответствие ряд параметров: $w_{i,j}^z$ — маршрутная метрика z -го трафика в канале связи $(i, j) \in E$; s_z — маршрутизатор-отправитель; t_z — маршрутизатор-получатель. Для оценки доли z -го трафика, протекающего в канале $(i, j) \in E$, используем управляющую переменную $x_{i,j}^z$. В соответствии с физикой решаемой задачи на переменные $x_{i,j}^z$ накладываем следующие ограничения:

$$0 \leq x_{i,j}^z \leq 1. \quad (1)$$

Маршрутную метрику d_t оптимального маршрута π_t между маршрутизатором-источником и маршрутизатором-получателем пакетов данных z -го трафика определим по выражению

$$d_t = \min_{(i,j) \in E} \sum_{z \in Z} \sum_{(i,j) \in \pi_t} w_{i,j}^z x_{i,j}^z. \quad (2)$$

Тогда маршрутную метрику k -го маршрута из имеющихся маршрутов между маршрутизатором-источником и маршрутизатором-получателем определим по выражению

$$d_{k,t} = \sum_{z \in Z} \sum_{(i,j) \in \pi_{k,t}} w_{i,j}^k x_{i,j}^z \text{ для всех } k \in K, \quad (3)$$

где $w_{i,j}^k$ — маршрутная метрика канала связи, соединяющего узлы с номерами i и j и входящего в маршрут с номером k узлу связи с номером t ; K —

общее число маршрутов между маршрутизатором-источником и маршрутизатором-получателем.

Суммарную маршрутную метрику всех имеющихся маршрутов между маршрутизатором-источником и маршрутизатором-получателем определим следующим образом:

$$D_t = \sum_{k=1}^K d_{k,t} \text{ для всех } k \in K. \quad (4)$$

Для выполнения балансировки нагрузки между имеющимися маршрутами определим долю информации $y_{k,t}$, проходящей по каждому из маршрутов:

$$y_{k,t} = \frac{p_{k,t}}{P_t}, \quad (5)$$

где $p_{k,t} = \frac{D_t}{d_{k,t}}$; $P_t = \sum_{k=1}^K p_{k,t}$ для всех $k \in K$.

На переменные $y_{k,t}$ наложим дополнительные ограничения:

$$\sum_{k=1}^K y_{k,t} = 1; 0 \leq y_{k,t} \leq 1. \quad (6)$$

Для управления процессом балансировки трафика в телекоммуникационной сети необходимо выполнить следующее условие:

$$\sum_{(i,j) \in E} w_{i,j} x_{i,j}^z \leq \alpha c_{i,j}; 0 \leq \alpha \leq 1; (i, j) \in E, \quad (7)$$

где $w_{i,j}$ — маршрутная метрика канала связи; $c_{i,j}$ — максимально-допустимая нагрузка канала связи; α — коэффициент балансировки трафика или динамически управляемый порог максимальной загрузки каналов связи в телекоммуникационной сети.

Для обеспечения оптимальности процесса балансировки нагрузки между разными маршрутами необходимо обеспечить минимизацию максимальной загрузки каналов связи, входящих в маршрут:

$$\min \alpha = \min_{z \in Z} \max_{(i,j) \in E} \frac{\sum_{(i,j) \in E} w_{i,j} x_{i,j}^z}{c_{i,j}}. \quad (8)$$

При оценке качества конечных решений воспользуемся такими показателями QoS , как отклонение от длины оптимального маршрута и средняя многопутевая задержка передачи.

В первом случае сначала определяется допустимое значение отклонения от длины оптимального маршрута, а затем перед балансировкой отбрасываются все маршруты, не соответствующие данному показателю. Для этого используем следующее выражение:

$$Q(M) = \max_{\pi_k \in \Pi} \{d(\pi_k)\} - \min_{\pi_k \in \Pi} \{d(\pi_k)\}. \quad (9)$$

Во втором случае определяется средняя задержка пакетов $\tau(\pi_k)$ вдоль маршрута π_k из множества маршрутов $\Pi = \{\pi_1; \pi_2; \dots; \pi_k; \dots; \pi_K\}$, где K — число маршрутов между маршрутизатором-источником и маршрутизатором-получателем:

$$\tau(\pi_k) = \sum_{(i,j) \in \pi_k} \tau_{i,j}. \quad (10)$$

Далее выполняется расчет межконцевой многопутевой задержки с помощью выражения

$$T(M) = \max_{\pi_k \in \Pi} \{\tau(\pi_k)\}. \quad (11)$$

Оценку джиттера пакетов, вызванного реализацией многопутевой стратегии маршрутизации, т. е. разницы в значениях средних задержек пакетов вдоль разных маршрутов, проведем с помощью формулы

$$J(M) = \max_{\pi_k \in \Pi} \{\tau(\pi_k)\} - \min_{\pi_k \in \Pi} \{\tau(\pi_k)\}. \quad (12)$$

Для разработки алгоритма динамического формирования трафика программно-конфигурируемых телекоммуникационных сетей с балансировкой нагрузки в условиях динамических изменений параметров линий связи сформулируем следующие теоремы.

Теорема 1. Если $nw_{i,j} > w_{i,j}$ и $e_{i,j} \in E_T$, то изменению могут подвергнуться оптимальные маршруты и оценки их длин для узлов $V_T^{(V_j)}$.

Доказательство. Пусть увеличилась маршрутная метрика канала связи $e_{i,j} \in E_T$, который входит, по крайней мере, в один оптимальный маршрут π_k , например в $\pi_{k,p}$. Узлы связи v_k , в оптимальные маршруты до которых канал связи $e_{i,j}$ не входит, будут составлять множество V_T узлов связи, оптимальные маршруты до которых после изменения метрики останутся прежними (не изменятся последовательность каналов связи и значения длин их оптимальных маршрутов). Действительно, пусть существует оптимальный маршрут $\pi_k = \pi_{k,p}$ до узла связи v_k и известно, что канал связи $e_{i,j}$ не входит в этот маршрут. Тогда увеличение метрики этого канала со значения $w_{i,j}$ до $nw_{i,j}$ не изменит маршрут этого пути и не повлияет на его значение $d_{k,p}$, т. е. $\pi_{k,p}$ и $d_{k,p}$, поскольку еще до увеличения метрики рассматриваемого канала включение этого канала в оптимальный маршрут приводило к увеличению длины маршрута. Все узлы связи, не вошедшие в множество V_T , будут составлять множество V_R . Оптимальные маршруты до узлов множества $v \in V_R$ станут "недействительными", т. е. невозможно будет без дополнительного расчета сказать, останутся они такими же или оптимальный маршрут до них не будет включать изменившийся канал связи. Теорема 1 доказана.

Теорема 2. Если $nw_{i,j} < w_{i,j}$ и $e_{i,j} \in E_T$, то без изменения останутся оптимальные маршруты для узлов множества $v \in V_T^{(V_j)} \cup V^{(V_i)}$, а для узлов множества $V^{(V_i)}$ неизменными останутся и оценки длин оптимальных маршрутов.

Доказательство. Пусть уменьшилась маршрутная метрика канала связи $e_{i,j} \in E_T$, входящего в оптимальный маршрут $\pi_k = \pi_{k,p}$ до узла связи $v_k \in V$. Канал связи $e_{i,j}$ после изменения также будет входить в оптимальный маршрут π_k до узла v_k . Поскольку метрика канала связи $w_{i,j}$ изменилась, то измениться должны длины всех маршрутов $\pi_{i,p}$, в которые входит этот канал связи. Действительно, если канал связи $e_{i,j}$ входит в какой-либо оптимальный маршрут и метрика этого канала уменьшается, то это изменение не потребует изменения оптимального маршрута $\pi_{k,p}$ (последовательности каналов) и длина маршрута $d_{k,p}$ изменится (уменьшится) на значение изменения метрики канала. Маршруты $\pi_s, v_s \notin V_T^{(V_j)} \cup V^{(V_i)}$ станут "недействительными", т. е. невозможно будет без дополнительного расчета сказать, останутся они такими же или оптимальный маршрут до них будет включать изменившийся канал связи. Теорема 2 доказана.

Теорема 3. Если $nw_{i,j} > w_{i,j}$ и $e_{i,j} \notin E_T$, то исходное дерево оптимальных маршрутов и оценки длин маршрутов всех узлов связи не изменятся.

Доказательство. Пусть канал связи, не входящий ни в один оптимальный маршрут, увеличивает свою маршрутную метрику $w_{i,j}, e_{i,j} \in E_R$. Никаких изменений дерева оптимальных маршрутов при этом не происходит. Действительно, пусть канал связи $e_{i,j} \in E_T$ входит в маршрут $\pi_{k,p}$ до некоторого узла связи v_k , который не является оптимальным для v_i , т. е. $\pi_{k,p} \neq \pi_k$. В этом случае существует такой маршрут $\pi_{k,t} = \pi_k$, что $d_{k,p} = d_{k,t}$. Тогда после увеличения метрики канала связи $w_{i,j}$ увеличится оценка длины $d_{k,p}$ и неравенство $d_{k,p} > d_{k,t}$ останется справедливым. То есть оптимальный маршрут и его оценка до узла v_k останутся неизменными. Теорема 3 доказана.

Теорема 4. Если $nw_{i,j} < w_{i,j}$ и $e_{i,j} \notin E_T$, то без изменения останутся оптимальные маршруты и оценки их длин для вершин множества $V^{(V_i)}$.

Доказательство. Пусть уменьшилась маршрутная метрика канала связи $e_{i,j} \in E_R$, который не входит ни в один оптимальный маршрут. Допустим, что этот канал связи входит в маршрут $\pi_{i,k} \neq \pi_i$ и $\pi_{i,p} \neq \pi_j$. Если изменившийся канал связи $e_{i,j}$ не уменьшает оценок обоих инцидентных ему узлов связи v_i и v_j , т. е. $d_{i,k} \geq d_i$ и $d_{j,p} \geq d_j$, то дерево оптимальных мар-

шрутов не изменится. Действительно, рассматриваемый канал связи оказывает влияние прежде всего на инцидентные ему узлы множества V . Если включение канала связи $e_{i,j}$ в дерево оптимальных маршрутов не уменьшает оценок маршрута d_i и d_j , то такое включение только увеличит оценки узлов. Так как существовавшие до изменения маршруты до этих узлов имели меньшую длину, то данный канал связи не включается в дерево оптимальных маршрутов. Если включение этого канала связи приводит к уменьшению оценки какого-либо из инцидентных узлов, например v_i , то эта оценка $d_{i,k}$ будет оценкой оптимального маршрута до узла v_i , и канал $e_{i,j}$ войдет в искомое дерево оптимальных маршрутов. Это происходит в силу того, что после изменения не существует иного оптимального маршрута π_i до узла v_i , кроме маршрута $\pi_{i,k}$, содержащего канал $e_{i,j}$. Этот оптимальный маршрут $\pi_{i,k}$ не будет существовать, если не будет оптимальных маршрутов до всех промежуточных узлов $v_p \in V^{(V_i)}$ этого маршрута. Оптимальные маршруты до остальных узлов связи станут "недействительными", т. е. невозможно будет сказать, останутся они такими же или оптимальный маршрут до них будет включать изменившийся канал связи. Теорема 4 доказана.

Теорема 5. Если $nw_{i,j} > w_{i,j}$ и $e_{i,j} \in E_T$ и $nw_{i,j} > nw_{i,j}^t$ (точки вхождения в дерево), то изменению могут подвергнуться оптимальные маршруты и оценки их длин для узлов $V_T^{(V_j)}$ и новые оптимальные маршруты к этим узлам будут проходить через каналы, состоящие в отношении парного перехода к каналам этих узлов.

Доказательство. Пусть увеличилась маршрутная метрика канала связи $e_{i,j} \in E_T$, который входит, по крайней мере, в один оптимальный маршрут π_k , например в $\pi_{k,p}$. Согласно теореме 1, узлы v_k , в оптимальные маршруты до которых канал связи $e_{i,j}$ не входит, будут составлять множество V_T узлов, оптимальные маршруты до которых после изменения метрики останутся прежними (не изменятся последовательность каналов связи и длина оптимальных маршрутов). Для узлов $V_T^{(V_j)}$ среди парных переходов, соответствующих этим узлам, будут находиться каналы, имеющие минимальную длину маршрута к этим узлам. Теорема 5 доказана.

Следствие. При увеличении метрики канала связи, входящего в дерево оптимальных маршрутов для узлов $V_T^{(V_j)}$, маршрутная степень которых больше двух, новые оптимальные маршруты будут проходить через каналы, состоящие в отношении

парного перехода к каналам, входящим в исходный граф.

Теорема 6. Если $nw_{i,j} < w_{i,j}$ и $e_{i,j} \notin E_T$ и новое значение $nw_{i,j} < nw_{i,j}^t$ (точки вхождения в дерево), то новые оптимальные маршруты к узлам множества $v \in V_T^{(V_j)} \cup V^{(V_i)}$ будут проходить через каналы связи, состоящие в отношении парного перехода к каналам этих узлов.

Доказательство. Пусть уменьшилась маршрутная метрика канала связи $e_{i,j} \in E_R$, который не входит ни в один оптимальный маршрут. Согласно теореме 4 без изменения останутся оптимальные маршруты и оценки их длин для узлов множества $V^{(V_i)}$. Так как $nw_{i,j} < nw_{i,j}^t$, включение этого канала приводит к уменьшению оценки какого-либо из инцидентных узлов, например v_i , и эта оценка $d_{i,k}$ будет оценкой оптимального маршрута до узла v_i , и канал связи $e_{i,j}$ войдет в искомое дерево оптимальных маршрутов. Это происходит в силу того, что после изменения не существует иного оптимального маршрута π_i до узла v_i , кроме маршрута $\pi_{i,k}$, содержащего канал связи $e_{i,j}$. Этот оптимальный маршрут $\pi_{i,k}$ не будет существовать, если не будет оптимальных маршрутов до всех промежуточных узлов связи $v_p \in V^{(V_i)}$ этого маршрута. Теорема 6 доказана.

Следствие. При уменьшении метрики канала связи, не входящего в дерево оптимальных маршрутов для узлов $V_T^{(V_j)} \cup V^{(V_i)}$, маршрутная степень которых больше двух, новые оптимальные маршруты будут проходить через каналы связи, состоящие в отношении парного перехода к каналам, входящим в исходный граф.

На основе сформулированных и доказанных выше теорем разработан алгоритм динамического формирования трафика программно-конфигурируемых телекоммуникационных сетей с балансировкой нагрузки в условиях динамических изменений параметров линий связи на основе данных о резервных маршрутах.

Анализ трудоемкости предложенного алгоритма показал его эффективность по сравнению с известными алгоритмами, эта трудоемкость равна $O(N)$.

Таким образом, разработанные математическая модель и алгоритм являются эффективными при динамическом формировании трафика программно-конфигурируемых телекоммуникационных сетей с балансировкой нагрузки в условиях динамических изменений нагрузки на линиях связи за счет использования дополнительной информации о резервных маршрутах.

Пример работы алгоритма

Рассмотрим работу алгоритма динамического формирования трафика программно-конфигурируемых телекоммуникационных сетей с балансировкой нагрузки в условиях возможных динамических изменений параметров линии связи на примере графа G , приведенного на рис. 1 (см. четвертую сторону обложки), в котором уже решена задача поиска оптимальных маршрутов и построено соответствующее дерево.

Для представленного графа G множество каналов дерева равно $E_T = \{e_{1,2}; e_{1,3}; e_{1,4}; e_{2,5}; e_{3,6}; e_{5,7}; e_{5,8}\}$, множество каналов замены — $E_S = \{e_{2,4}; e_{4,5}; e_{4,6}; e_{4,7}; e_{6,8}\}$; множество непарных каналов будет $E_P = \{e_{7,8}\}$. Если рассмотреть канал связи $e_{2,5}$, то для него точка вхождения в дерево будет составлять 70, а точка вхождения в множество замены — 180. При этом данный канал связи находится в отношении парного перехода с каналом $e_{4,5}$, который, в свою очередь, находится в отношении парного перехода множества замены с каналом $e_{5,7}$. После попадания $e_{2,5}$ в множество непарных каналов эта парная перестановка примет вид: $e_{4,5} - e_{5,7}$.

После того как сформирован список парных переходов, дополнительно рассчитываем списки оптимальных и резервных маршрутов из исходного узла связи до каждого узла в сети.

Таким образом, списки оптимальных и резервных маршрутов до всех узлов графа телекоммуникационной сети из исходного узла V_S будут сформированы следующим образом.

Узел связи V_2 .

Маршрут $\pi_2 = \{e_{1,2}\}$ с оценкой $d_2 = 20$ — оптимальный маршрут;

Маршрут $\pi_2^{(1)} = \{e_{1,4}; e_{4,2}\}$ с оценкой $d_2^{(1)} = 50 + 70 = 120$ — резервный маршрут;

Маршрут $\pi_2^{(2)} = \{e_{1,4}; e_{4,5}; e_{5,2}\}$ с оценкой $d_2^{(2)} = 50 + 40 + 50 = 140$ — резервный маршрут.

Узел связи V_3 .

Маршрут $\pi_3 = \{e_{1,3}\}$ с оценкой $d_3 = 20$ — оптимальный маршрут;

Маршрут $\pi_3^{(1)} = \{e_{1,4}; e_{4,6}; e_{6,3}\}$ с оценкой $d_3^{(1)} = 50 + 50 + 60 = 160$ — резервный маршрут.

Узел связи V_4 .

Маршрут $\pi_4 = \{e_{1,4}\}$ с оценкой $d_4 = 50$ — оптимальный маршрут;

Маршрут $\pi_4^{(1)} = \{e_{1,2}; e_{2,4}\}$ с оценкой $d_4^{(1)} = 20 + 70 = 90$ — резервный маршрут;

Маршрут $\pi_4^{(2)} = \{e_{1,3}; e_{3,6}; e_{6,4}\}$ с оценкой $d_4^{(2)} = 20 + 60 + 50 = 130$ — резервный маршрут.

Узел связи V_5 .

Маршрут $\pi_5 = \{e_{1,2}; e_{2,5}\}$ с оценкой $d_5 = 20 + 50 = 70$ — оптимальный маршрут;

Маршрут $\pi_5^{(1)} = \{e_{1,4}; e_{4,5}\}$ с оценкой $d_5^{(1)} = 50 + 40 = 90$ — резервный маршрут;

Маршрут $\pi_5^{(2)} = \{e_{1,4}; e_{4,7}; e_{7,5}\}$ с оценкой $d_5^{(2)} = 50 + 90 + 60 = 200$ — резервный маршрут;

Маршрут $\pi_5^{(3)} = \{e_{1,3}; e_{3,6}; e_{6,8}; e_{8,5}\}$ с оценкой $d_5^{(3)} = 20 + 60 + 60 + 60 = 200$ — резервный маршрут.

Узел связи V_6 .

Маршрут $\pi_6 = \{e_{1,3}; e_{3,6}\}$ с оценкой $d_6 = 20 + 60 = 80$ — оптимальный маршрут;

Маршрут $\pi_6^{(1)} = \{e_{1,4}; e_{4,6}\}$ с оценкой $d_6^{(1)} = 50 + 50 = 100$ — резервный маршрут;

Маршрут $\pi_6^{(2)} = \{e_{1,2}; e_{2,5}; e_{5,8}; e_{8,6}\}$ с оценкой $d_6^{(2)} = 20 + 50 + 60 + 60 = 190$ — резервный маршрут.

Узел связи V_7 .

Маршрут $\pi_7 = \{e_{1,2}; e_{2,5}; e_{5,7}\}$ с оценкой $d_7 = 20 + 50 + 60 = 130$ — оптимальный маршрут;

Маршрут $\pi_7^{(1)} = \{e_{1,4}; e_{4,7}\}$ с оценкой $d_7^{(1)} = 50 + 90 = 140$ — резервный маршрут;

Маршрут $\pi_7^{(2)} = \{e_{1,4}; e_{4,5}; e_{5,7}\}$ с оценкой $d_7^{(2)} = 50 + 40 + 60 = 150$ — резервный маршрут;

Маршрут $\pi_7^{(3)} = \{e_{1,3}; e_{3,6}; e_{6,8}; e_{8,7}\}$ с оценкой $d_7^{(3)} = 20 + 60 + 60 + 20 = 160$ — резервный маршрут.

Узел связи V_8 .

Маршрут $\pi_8 = \{e_{1,2}; e_{2,5}; e_{5,8}\}$ с оценкой $d_8 = 20 + 50 + 60 = 130$ — оптимальный маршрут;

Маршрут $\pi_8^{(1)} = \{e_{1,3}; e_{3,6}; e_{6,8}\}$ с оценкой $d_8^{(1)} = 20 + 60 + 60 = 140$ — резервный маршрут;

Маршрут $\pi_8^{(2)} = \{e_{1,4}; e_{4,7}; e_{7,8}\}$ с оценкой $d_8^{(2)} = 50 + 90 + 20 = 160$ — резервный маршрут.

Работа предложенного алгоритма основывается на том, что при динамическом изменении метрики канала связи, входящего в дерево оптимальных маршрутов, или метрики канала, находящегося в отношении парного перехода к каналу из дерева оптимальных маршрутов, необходимо просмотреть списки оптимальных и резервных маршрутов до каждого узла в сети, куда входит канал, метрика которого изменилась.

Рассмотрим работу алгоритма динамического формирования трафика программно-конфигурируемых телекоммуникационных сетей с балансировкой нагрузки на основе данных о резервных маршрутах при следующих возможных ситуациях в сети.

Для графа G телекоммуникационной сети выполним балансировку трафика между маршрутизатором-отправителем V_1 и маршрутизатором-получателем V_5 в условиях динамических изменений параметров линий связи телекоммуникационной сети. Для упрощения расчетов примем, что в сети имеется один тип трафика ($z = 1$), коэффициент балансировки трафика равен 1 ($\alpha = 1$) и максимально-допустимая нагрузка любого канала связи в сети не может превышать 100 Мбит/с ($c_{i,j} = 100$ Мбит/с). Список возможных маршрутов для выполнения балансировки трафика имеет следующий вид:

- 1) $V_1 - V_2 - V_5$; длина 70; доля информации, проходящей по маршруту, — 0,404;
- 2) $V_1 - V_4 - V_5$; длина 90; доля информации, проходящей по маршруту, — 0,314;
- 3) $V_1 - V_4 - V_7 - V_5$; длина 200; доля информации, проходящей по маршруту, — 0,141;
- 4) $V_1 - V_3 - V_6 - V_8 - V_5$; длина 200; доля информации, проходящей по маршруту, — 0,141.

Результат балансировки трафика между V_1 и V_5 в соответствии сформированному списку маршрутов, представлен на рис. 2 (см. четвертую сторону обложки). При этом канал связи $e_{1,4}$ является общим для маршрутов с номерами 2 и 3.

Рассмотрим возможные случаи динамических изменений параметров линии связи в телекоммуникационной сети и изменения объема передаваемой информации по разным маршрутам при выполнении балансировки трафика.

На рис. 3 (см. четвертую сторону обложки) приведен случай увеличения маршрутной метрики канала связи $e_{1,2}$, входящего в дерево оптимальных маршрутов, так что $nw_{1,2} = 90 > w_{1,2} = 20$. Список маршрутов для выполнения балансировки трафика примет следующий вид:

- 1) $V_1 - V_4 - V_5$; длина 90; доля информации, проходящей по маршруту, — 0,393;
- 2) $V_1 - V_2 - V_5$; длина 140; доля информации, проходящей по маршруту, — 0,253;
- 3) $V_1 - V_4 - V_7 - V_5$; длина 200; доля информации, проходящей по маршруту, — 0,177;
- 4) $V_1 - V_3 - V_6 - V_8 - V_5$; длина 200; доля информации, проходящей по маршруту, — 0,177.

На рис. 4 (см. четвертую сторону обложки) рассмотрен случай уменьшения маршрутной метрики канала связи $e_{4,7}$, входящего в множество замены для дерева оптимальных маршрутов так, что $nw_{4,7} = 50 < w_{4,7} = 90$. Список маршрутов для выполнения балансировки трафика примет следующий вид:

- 1) $V_1 - V_2 - V_5$; длина 70; доля информации, проходящей по маршруту, — 0,389;
- 2) $V_1 - V_4 - V_5$; длина 90; доля информации, проходящей по маршруту, — 0,303;
- 3) $V_1 - V_4 - V_7 - V_5$; длина 160; доля информации, проходящей по маршруту, — 0,172;

- 4) $V_1 - V_3 - V_6 - V_8 - V_5$; длина 200; доля информации, проходящей по маршруту, — 0,136.

Заключение

В работе предложена математическая модель динамического формирования трафика программно-конфигурируемых телекоммуникационных сетей с балансировкой нагрузки, позволяющая повысить эффективность процессов маршрутизации, качество обслуживания пользовательских приложений и различных типов трафика в условиях динамических изменений параметров линий связи за счет использования дополнительной информации о резервных маршрутах. В дальнейшем предполагается усовершенствование предложенной математической модели в целях ее применения для решения задачи многопутевой адаптивной маршрутизации и балансировки трафика по отдельным подсетям телекоммуникационной сети [29].

Работа выполнена при финансовой поддержке гранта Президента РФ для молодых ученых — кандидатов наук МК-819.2014.9.

Список литературы

1. Вишневский В. М. Теоретические основы проектирования компьютерных сетей. М.: Техносфера, 2003. 512 с.
2. Олифер В. Г., Олифер Н. А. Основы компьютерных сетей. СПб.: Питер, 2009. 352 с.
3. Корячко В. П., Перепелкин Д. А. Корпоративные сети: технологии, протоколы, алгоритмы. М.: Горячая линия — Телеком, 2011. 219 с.
4. Корячко В. П., Перепелкин Д. А. Анализ и проектирование маршрутов передачи данных в корпоративных сетях. М.: Горячая линия — Телеком, 2012. 235 с.
5. McKeown N., Anderson T., Balakrishnan H. et al. Openflow: Enabling Innovation in Campus Networks // ACM SIGCOMM Computer Communication Review. 2008. Vol. 38, N. 2. P. 69–74.
6. Смелянский Р. Л. Программно-конфигурируемые сети // Открытые системы. СУБД. 2012. № 9. С. 18.
7. Ефимушкин В. А., Ледовских Т. В., Корабельников Д. М., Язык Д. Н. Сравнительный анализ архитектур и протоколов программно-конфигурируемых сетей // Электросвязь. 2014. № 8. С. 9–14.
8. Корячко В. П., Перепелкин Д. А., Иванчикова М. А. Алгоритм адаптивной маршрутизации в корпоративных сетях нескольких провайдеров связи // Вестник Рязанского государственного радиотехнического университета. 2013. № 2 (44). С. 52–56.
9. Корячко В. П., Перепелкин Д. А., Иванчикова М. А. Алгоритм парных переходов каналов связи при динамическом изменении нагрузки в корпоративных сетях нескольких провайдеров связи с различными зонами покрытия // Вестник Рязанского государственного радиотехнического университета. 2014. № 48. С. 68–76.
10. Ижванов Ю. Л. Динамическая оценка состояния компьютерных сетей на основе метода сетевой томографии и задачи балансировки трафика // Информатизация образования и науки. 2013. № 3 (19). С. 35–40.
11. Никульчев Е. В., Паян С. В., Плужник Е. В. Динамическое управление трафиком программно-конфигурируемых сетей в облачной инфраструктуре // Вестник Рязанского государственного радиотехнического университета. 2013. № 3 (45). С. 54–57.
12. Шибанов А. П., Корячко В. П., Ижванов Ю. Л. Моделирование агрегированного телекоммуникационного канала с технологией открытых потоков // Радиотехника. 2012. № 3. С. 109–112.
13. Ижванов Ю. Л., Корячко В. П., Шибанов А. П. и др. Оптимизация сетей с дозированной балансировкой нагрузки и пи-

ринговыми каналами // Вестник Рязанского государственного радиотехнического университета. 2013. № 1 (43). С. 67–74.

14. **Корячко В. П., Лукьянов О. В., Шибанов А. П.** Нахождение скрытого параллелизма протоколов для улучшения характеристик сети передачи данных полигонного измерительного комплекса // Вестник Рязанского государственного радиотехнического университета. 2014. № 47. С. 68–75.

15. **Лемешко А. В., Вавенко Т. В.** Усовершенствование поточковой модели многопутевой маршрутизации на основе балансировки нагрузки // Проблемы телекоммуникаций. 2012. № 1 (6). С. 12–29.

16. **Лемешко А. В., Вавенко Т. В.** Разработка и исследование поточковой модели адаптивной маршрутизации в программно-конфигурируемых сетях с балансировкой нагрузки // Доклады Томского государственного университета систем управления и радиоэлектроники. 2013. № 3 (29). С. 100–108.

17. **Уваров Д. В., Перепелкин А. И., Корячко В. П.** Построение дерева кратчайших путей в графе на основе данных о парных переходах // Системы управления и информационные технологии. 2004. № 4 (16). С. 93–96.

18. **Перепелкин Д. А., Перепелкин А. И.** Разработка алгоритмов адаптивной маршрутизации в корпоративных вычислительных сетях // Вестник Рязанского государственного радиотехнического университета. 2006. № 19. С. 114–116.

19. **Перепелкин А. И., Перепелкин Д. А.** Разработка алгоритма динамической маршрутизации на базе протокола OSPF в корпоративных вычислительных сетях // Вестник Рязанского государственного радиотехнического университета. 2009. № 28. С. 68–72.

20. **Перепелкин Д. А.** Алгоритм адаптивной ускоренной маршрутизации на базе протокола OSPF при динамическом добавлении элементов корпоративной сети // Вестник Рязанского государственного радиотехнического университета. 2010. № 34. С. 65–71.

21. **Перепелкин Д. А.** Алгоритм адаптивной ускоренной маршрутизации на базе протокола OSPF при динамическом отказе элементов корпоративной сети // Вестник Рязанского государственного радиотехнического университета. 2011. № 37. С. 53–58.

22. **Корячко В. П., Перепелкин Д. А., Перепелкин А. И.** Алгоритм парных перестановок маршрутов в корпоративных сетях // Системы управления и информационные технологии. 2010. Т. 40, № 2. С. 51–56.

23. **Корячко В. П., Перепелкин Д. А., Перепелкин А. И.** Повышение эффективности функционирования корпоративных сетей при динамических изменениях в их структуре и нагрузках на линии связи // Вестник Рязанского государственного радиотехнического университета. 2010. № 33. С. 49–55.

24. **Перепелкин Д. А., Перепелкин А. И.** Повышение качества функционирования корпоративных сетей на базе протокола OSPF // Качество. Инновации. Образование. 2010. № 12. С. 51–56.

25. **Перепелкин Д. А., Перепелкин А. И.** Алгоритм адаптивной ускоренной маршрутизации в условиях динамически изменяющихся нагрузок на линиях связи корпоративной сети // Информационные технологии. 2011. № 3. С. 2–7.

26. **Перепелкин Д. А.** Алгоритм парных перестановок маршрутов на базе протокола OSPF при динамическом подключении узлов и линий связи корпоративной сети // Вестник Рязанского государственного радиотехнического университета. 2013. № 4–1 (46). С. 67–75.

27. **Перепелкин Д. А.** Алгоритм парных перестановок маршрутов на базе протокола OSPF при динамическом отказе узлов и линий связи корпоративной сети // Вестник Рязанского государственного радиотехнического университета. 2014. № 47. С. 84–91.

28. **Перепелкин Д. А.** Динамическое формирование структуры и параметров линий связи корпоративной сети на основе данных о парных перестановках маршрутов // Информационные технологии. 2014. № 4. С. 52–60.

29. **Перепелкин Д. А.** Алгоритм формирования оптимальных структур сегментов корпоративных сетей с учетом данных о структуре базовой сети и связности каналов связи // Вестник Рязанского государственного радиотехнического университета. 2014. № 50–1. С. 56–62.

D. A. Perepelkin, PhD, Associate Professor, Ryazan State Radio Engineering University, Ryazan
E-mail: dmitryperepelkin@mail.ru

Conceptual Approach of Dynamic Traffic Formation of Software-Defined Telecommunication Networks with Load Balancing

At present the actual area of research is the development of models and algorithms of multipath adaptive routing and load balancing in the technology of software-defined network. Providing high-speed and reliable data transmission between the nodes of communication networks under stringent requirements of delay data transmission during possible load changes on the communication links is one of the most important problems.

Model of dynamic traffic formation with load balancing in software-defined network which allows improving the quality of services of networks requests and applications is offered in article. It also ensuring performance and reliability of data transmission based on additional data about the backup routes.

The mathematical model takes into account the character of available heterogeneous traffic and possible changes of metrics in communications network. It performs load balancing in the traffic channels on the basis of data of the replacement backup routes. It ensures compliance with the quality of service requirements of the network application due to the calculation of indicators on the average transmission delay in the channels communication and evaluation of packet jitter.

Software of process modeling adaptive routing and load balancing in communication networks, confirming the validity of the proposed mathematical model and characterize the effectiveness of the proposed approach in comparison with existing analogues is developed.

Keywords: telecommunication networks, software-defined networks, routing algorithms, adaptive routing, routes pairs permutations algorithm, multipath routing, load balancing, quality of service, dynamic changes, backup routes, network traffic

References

1. **Vishnevskij V. M.** *Teoreticheskie osnovy proektirovaniya komp'juternykh setej* (Theoretical Basis of Computer Networks Design), Moscow, Tehnosfera, 2003, 512 p. (in Russian).
2. **Olifer V. G., Olifer N. A.** *Osnovy komp'juternykh setej* (Basis of Computer Networks), Saint Petersburg, Piter, 2009, 352 p. (in Russian).
3. **Korjachko V. P., Perepelkin D. A.** *Korporativnye seti: tehnologii, protokoly, algoritmy* (Corporate Networks: Technologies, Protocols, Algorithms), Moscow, Gorjachaja linija — Telekom, 2011, 219 p. (in Russian).
4. **Korjachko V. P., Perepelkin D. A.** *Analiz i proektirovanie marshrutov peredachi dannykh v korporativnykh setjakh* (Analysis and Design of Data Transmission Routes in Corporate Networks), Moscow, Gorjachaja linija — Telekom, 2012, 235 p. (in Russian).
5. **McKeown N., Anderson T., Balakrishnan H.** et al. Openflow: Enabling Innovation in Campus Networks. *ACM SIGCOMM Computer Communication Review*, 2008, vol. 38, no. 2, pp. 69–74.
6. **Smeljanskij R. L.** Programmno-konfiguriruemye seti. *Otkrytye sistemy*. SUBD, 2012, no. 9. pp. 18 (in Russian).
7. **Efimushkin V. A., Ledovskih T. V., Korabel'nikov D. M., Jazykov D. N.** Sravnitel'nyj analiz arhitektury i protokolov programmno-konfiguriruemykh setej. *Jelektrosvjaz*, 2014, no. 8, pp. 9–14 (in Russian).
8. **Korjachko V. P., Perepelkin D. A., Ivanchikova M. A.** Algoritmi adaptivnoj marshrutizacii v korporativnykh setjakh neskol'kih provajderov svjazi. *Vestnik Rjazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*, 2013, vol. 44, no. 2, pp. 52–56 (in Russian).
9. **Korjachko V. P., Perepelkin D. A., Ivanchikova M. A.** Algoritmi parnykh perehodov kanalov svjazi pri dinamicheskom izmenenii nagruzki v korporativnykh setjakh neskol'kih provajderov svjazi s razlichnymi zonami pokrytija. *Vestnik Rjazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*, 2014, vol. 48, pp. 68–76 (in Russian).
10. **Izhvanov Ju. L.** Dinamicheskaja ocenka sostojanija komp'juternykh setej na osnove metoda setевой tomografii i zadacha balansirovki trafika. *Informatizacija obrazovanija i nauki*, 2013, vol. 19, no. 3, pp. 35–40 (in Russian).
11. **Nikul'chev E. V., Pajain S. V., Pluzhnik E. V.** Dinamicheskoe upravlenie trafikom programmno-konfiguriruemykh setej v oblachnoj infrastrukture. *Vestnik Rjazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*, 2013, vol. 45, no. 3, pp. 54–57 (in Russian).
12. **Shibanov A. P., Korjachko V. P., Izhvanov Ju. L.** Modelirovanie agregirovannogo telekommunikacionnogo kanala s tehnologiej otkrytykh potokov. *Radiotekhnika*, 2012, no. 3, pp. 109–112 (in Russian).
13. **Izhvanov Ju. L., Korjachko V. P., Shibanov A. P.** et al. Optimizacija setej s dozirovannoj balansirovkoj nagruzki i piringovymi kanalami. *Vestnik Rjazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*, 2013, vol. 43, no. 1, pp. 67–74 (in Russian).
14. **Korjachko V. P., Luk'janov O. V., Shibanov A. P.** Nahozhdenie skrytogo parallelizma protokolov dlja uluchshenija harakteristik seti peredachi dannykh poligonno izmeritel'nogo kompleksa. *Vestnik Rjazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*, 2014, vol. 47, pp. 68–75 (in Russian).
15. **Lemeshko A. V., Vavenko T. V.** Uovershenstvovanie potokovoy modeli mnogoputevoj marshrutizacii na osnove balansirovki nagruzki. *Problemy telekommunikacij*, 2012, vol. 6, no. 1, pp. 12–29 (in Russian).
16. **Lemeshko A. V., Vavenko T. V.** Razrabotka i issledovanie potokovoy modeli adaptivnoj marshrutizacii v programmno-konfiguriruemykh setjakh s balansirovkoj nagruzki. *Doklady Tomskogo gosudarstvennogo universiteta sistem upravlenija i radioelektroniki*, 2013, vol. 29, no. 3, pp. 100–108 (in Russian).
17. **Uvarov D. V., Perepelkin A. I., Korjachko V. P.** Postroenie dereva krachajshih putej v grafe na osnove dannykh o parnykh perehodah. *Sistemy upravlenija i informacionnye tehnologii*, 2004, vol. 16, no. 4, pp. 93–96 (in Russian).
18. **Perepelkin D. A., Perepelkin A. I.** Razrabotka algoritmov adaptivnoj marshrutizacii v korporativnykh vychislitel'nykh setjakh. *Vestnik Rjazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*, 2006, vol. 19, pp. 114–116 (in Russian).
19. **Perepelkin A. I., Perepelkin D. A.** Razrabotka algoritma dinamicheskoy marshrutizacii na baze protokola OSPF v korporativnykh vychislitel'nykh setjakh. *Vestnik Rjazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*, 2009, vol. 28, pp. 68–72 (in Russian).
20. **Perepelkin D. A.** Algoritmi adaptivnoj uskorennoj marshrutizacii na baze protokola OSPF pri dinamicheskom dobavlenii jelementov korporativnoj seti. *Vestnik Rjazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*, 2010, vol. 34, pp. 65–71 (in Russian).
21. **Perepelkin D. A.** Algoritmi adaptivnoj uskorennoj marshrutizacii na baze protokola OSPF pri dinamicheskom otkaze jelementov korporativnoj seti. *Vestnik Rjazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*, 2011, vol. 37, pp. 53–58 (in Russian).
22. **Korjachko V. P., Perepelkin D. A., Perepelkin A. I.** Algoritmi parnykh perestanovok marshrutov v korporativnykh setjakh. *Sistemy upravlenija i informacionnye tehnologii*, 2010, vol. 40, no. 2, pp. 51–56 (in Russian).
23. **Korjachko V. P., Perepelkin D. A., Perepelkin A. I.** Povyshenie jeffektivnosti funkcionirovanija korporativnykh setej pri dinamicheskikh izmenenijah v ih strukture i nagruzkah na linii svjazi. *Vestnik Rjazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*, 2010, vol. 33, pp. 49–55 (in Russian).
24. **Perepelkin D. A., Perepelkin A. I.** Povyshenie kachestva funkcionirovanija korporativnykh setej na baze protokola OSPF. *Kachestvo. Innovacii. Obrazovanie*, 2010, no. 12, pp. 51–56 (in Russian).
25. **Perepelkin D. A., Perepelkin A. I.** Algoritmi adaptivnoj uskorennoj marshrutizacii v uslovijah dinamicheski izmenjajushihhsja nagruzok na liniyah svjazi v korporativnoj seti. *Informacionnye tehnologii*, 2011, no. 3, pp. 2–7 (in Russian).
26. **Perepelkin D. A.** Algoritmi parnykh perestanovok marshrutov na baze protokola OSPF pri dinamicheskom podkljuchenii uzlov i linii svjazi korporativnoj seti. *Vestnik Rjazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*, 2013, vol. 46, no. 4–1, pp. 67–75 (in Russian).
27. **Perepelkin D. A.** Algoritmi parnykh perestanovok marshrutov na baze protokola OSPF pri dinamicheskom otkaze uzlov i linii svjazi korporativnoj seti. *Vestnik Rjazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*, 2014, vol. 47, pp. 84–91 (in Russian).
28. **Perepelkin D. A.** Dinamicheskoe formirovanie struktury i parametrov linii svjazi korporativnoj seti na osnove dannykh o parnykh perestanovkah marshrutov. *Informacionnye tehnologii*, 2014, no. 4, pp. 52–60 (in Russian).
29. **Perepelkin D. A.** Algoritmi formirovanija optimal'nykh struktur segmentov korporativnykh setej s uchedom dannykh o strukture bazovoj seti i svjaznosti kanalov svjazi. *Vestnik Rjazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*, 2014, vol. 50, no. 1, pp. 56–62 (in Russian).

БЕЗОПАСНОСТЬ ИНФОРМАЦИИ CRYPTOSAFETY INFORMATION

УДК 004.273:004.056

Д. О. Маркин, соторудник, e-mail: admin@nikitka.net, **В. В. Комашинский**, канд. техн. наук, доц.,
И. Ю. Баранов, канд. техн. наук, доц.,
Академия ФСО России, г. Орел

Модель управления профилем защиты мобильного устройства при доступе к услугам с разным уровнем конфиденциальности

Описана модель процесса управления логической структурой мобильного устройства, дана характеристика логической структуры мобильного устройства, показано ее место в системе управления доступом к услугам с разным уровнем конфиденциальности, описана модель логической структуры взаимосвязей между функциональными блоками мобильного устройства, а также модель управления профилем защиты мобильного устройства при доступе к услугам с разным уровнем конфиденциальности на основе матричных графовых грамматик.

Ключевые слова: мобильное устройство, профиль защиты, реконфигурация логической структуры, динамические графы, графовые грамматики

Введение

Современный мобильный телефон, как и достаточно широкий спектр мобильных устройств (МУ), обладающих и вычислительными, и коммуникационными ресурсами, представляет собой медиаустройство для повседневной работы и развлечений, где функция телефонных переговоров не является первостепенно важной [1]. Для улучшения показателя экономичности основные узлы современных МУ агрегированы в составе микросхемы класса SoC (System-on-Chip, система на чипе), на которую возлагается весь перечень задач сбора, обработки, хранения и обмена пользовательской и служебной информацией [2]. Такая SoC зачастую объединяет на одном кристалле несколько ядер процессора, коммуникационный процессор (baseband processor), графический сопроцессор и др. Добавление при необходимости микроконтроллеров для кодирования речи, высокочастотных блоков для работы в различных стандартах сети сотовой связи, интерфейсных блоков Wi-Fi сети, модулей GPS/ГЛОНАСС, а также набор интерфейсов для взаимодействия с различными типами устройств (USB, SD, MMC, UART и др.) обеспечивает конфигурирование МУ для разных уровней рынка и требований пользователей и обеспечивает многофункциональность МУ.

В то же время на современном этапе развития телекоммуникационных и вычислительных инфраструктур часто возникает задача обеспечения информационного взаимодействия между специализированными информационно-телекоммуникационными ресурсами (ИТ-ресурсами), обладающими различным уровнем конфиденциальности, и вы-

полнения требований по защите информации. На данный момент эта задача в отношении МУ не решена по ряду различных причин, в том числе технологических и нормативно-правовых. Однако, учитывая современный уровень развития технологий проектирования МУ и существующие тенденции, объективно можно построить МУ с управляемой логической структурой взаимосвязей функциональных блоков или модулей на микросхеме класса SoC таким образом, чтобы при необходимости обмена информационными потоками разного уровня конфиденциальности формировалась такая логическая структура МУ, при которой данные потоки, с одной стороны, не пересекались бы, а с другой — были защищены с помощью средств защиты информации на аппаратном уровне МУ. Своеобразным прототипом подобного решения могут служить принципы построения современных программируемых радиостанций, рассмотренных в работах [3, 4].

В данной работе предлагается формальная модель управления профилем защиты (ПЗ) единого многофункционального мобильного устройства (ММУ) при доступе к разнокатегорированным услугам (услугам с разным уровнем конфиденциальности), основанная на изменении логической структуры трактов прохождения информационных потоков. Сущность управления профилем защиты в ММУ заключается в реконфигурации логической структуры взаимосвязей между функциональными блоками ММУ таким образом, чтобы передача информационных потоков с различным уровнем конфиденциальности осуществлялась с выполнением требований по защите информации автоматизиро-

ванно либо автоматически в зависимости от уровня конфиденциальности передаваемой информации. Пример решения подобной задачи в отношении интегрированных объектов информатизации описан в работе К. М. Зорина [5]. Основой предложенного Зориным подхода по реконфигурации логической структуры является обеспечение эффективного функционирования искусственного интеллекта как средства передачи информации, однако вопросы управления логической структурой ММУ как объекта информатизации [6], обеспечивающего защищенное сетевое взаимодействие с помощью информационных потоков с разным уровнем конфиденциальности, в данной работе не рассматриваются. В описываемой модели управления профилем защиты ММУ при доступе к разнокатегорированным услугам предлагается распространить идеи реконфигурации логической структуры искусственного интеллекта на логическую структуру взаимосвязей между функциональными блоками ММУ, чтобы обеспечить безопасный обмен информационными потоками с разным уровнем конфиденциальности между объектами информатизации.

Общий вид модели процесса управления логической структурой мобильного устройства и формальная постановка задачи

Описание изменяемой логической структуры ММУ как информационной системы (ИС) может быть представлено как последовательность задач синтеза на каждом этапе функционирования. В этом случае функционирование ММУ может быть описано динамическими моделями [7, 8]. Таким образом, динамическая модель описания ММУ рассматривает процедуры синтеза его логической структуры в каждый момент времени и с учетом его поведения в предыдущие моменты времени.

В контексте данной работы понятия профиля защиты ММУ и логическая структура ММУ являются синонимичными и характеризуют взаимосвязи между функциональными блоками ММУ, определяющие тракты прохождения информационных потоков с разным уровнем конфиденциальности. В контексте управления профилем защиты при доступе к услугам с разным уровнем конфиденциальности целями изменения состояния ММУ (его логической структуры) является обеспечение безопасного информационного обмена между ММУ и защищенной информационной инфраструктурой. Решение о реконфигурации ММУ принимает некоторая подсистема управления на основе результатов наблюдения за ним, которое заключается в выборе альтернативных управляющих воздействий — альтернативных логических структур ММУ. Очевидно, что каждая логическая структура ММУ отличается от других по целому ряду параметров, таких как быстрдействие, надежность, пропускная способность и, самое главное, защищенность. Данные

параметры могут выступать в качестве критериев оценивания того или иного управляющего воздействия для приведения логической структуры ММУ в оптимальное состояние для текущих условий.

В общем случае целью изменения структуры ММУ является достижение заданных требований надсистемы или, иными словами, выполнение требований по защите информации вследствие возникновения внутренних факторов или ограничений (возникновения необходимости передачи информации с иным уровнем конфиденциальности).

В общем случае процесс управления конфигурацией ММУ может быть представлен в виде следующих выражений.

Состояния $S(t)$ моделируемого ММУ зависят от случайных воздействий $\xi_S(t)$ и процессов управления $x(t)$:

$$S(t) = F\{x(\tau_1), \xi_S(\tau_2), t\}, \tau_1, \tau_2 \in [t_0, t] \cap T. \quad (1)$$

Каналы мониторинга можно описать процессом

$$S_U(t) = F_{S_U}\{S(\tau_1), \xi_{S_U}(\tau_2), t\},$$

где $\xi_{S_U}(t)$ — стохастический порождающий процесс, определяющий случайный характер мониторинга.

Системы принятия решений задаются стохастическими функционалами, вырабатывающими в каждый момент времени t рандомизированные значения $u(t)$ принятых решений по наблюдениям $S_U(t)$ при $\tau \in [t_0, t] \cap T$:

$$u(t) = F_U\{S_U(\tau_1), \xi_U(\tau_2), t\},$$

где $\xi_U(t)$ — стохастический порождающий процесс, определяющий рандомизацию решений $u(t)$.

Канал управления характеризуется процессом управления $x(t)$ в зависимости от принятых решений:

$$x(t) = F_x\{u(\tau_1), \xi_x(\tau_2), t\}, \quad (2)$$

где $\xi_x(t)$ — стохастические порождающие процессы, определяющие случайный характер реконфигурации.

Таким образом, для моделирования необходимо определить вектор анализируемых показателей эффективности и приведенные ранее функционалы, определяющие динамику состояний ММУ при известном начальном состоянии, а также интерпретировать характеристики и параметры функционалов значениями, характеризующими реальное ММУ.

Характеристика логической структуры многофункционального мобильного устройства

Для построения модели управления ПЗ ММУ необходимо конкретизировать характеристику ММУ и его структуру. Суть реконфигурации логической структуры ММУ будет заключаться в воздействии на структурные элементы ММУ в целях повышения эффективности функционирования с учетом обеспечения требуемого уровня безопасности информационного взаимодействия. Пример типичной структуры ММУ, существующей на современном этапе развития, представлен на рис. 1.

Важно отметить, что при работе ММУ в режиме обработки конфиденциальной информации могут предъявляться дополнительные требования, заключающиеся в необходимости отключения ряда функциональных блоков ММУ, создающих предпосылки или возможность утечки информации.

Задачи формирования необходимой логической структуры трактов прохождения информационных потоков в ММУ могут возлагаться на аппаратно-программный модуль доверенной загрузки, а команда на выполнение изменения логической структуры ММУ в целях создания необходимых условий для обработки информации с заданным уровнем конфиденциальности (т. е., по сути, на выполнение реконфигурации ММУ) может поступать по каналу управления, использующего, например, корпоративную защищенную беспроводную локально-вычислительную сеть на базе технологии Wi-Fi, состоящую из доверенных точек доступа и контроллера беспроводной сети. Таким образом, инфраструктура системы управления профилем защиты ММУ может выглядеть так, как представлено на рис. 2.

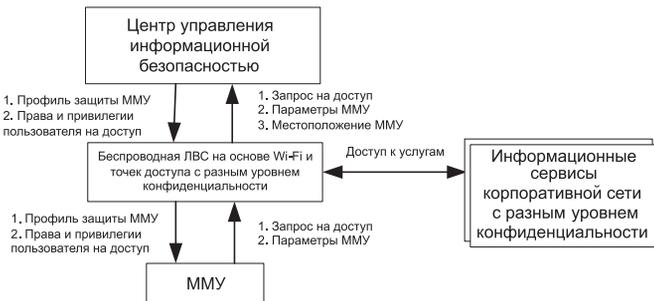


Рис. 2. Обобщенная схема инфраструктуры управления ПЗ ММУ при доступе к услугам с уровнем конфиденциальности

Задача формирования профиля защиты ММУ возлагается на центр управления информационной безопасностью. Опишем формальную модель логической структуры взаимосвязей между функциональными блоками ММУ, которая является объектом управления.

Формальное описание модели логической структуры взаимосвязей между функциональными блоками мобильного устройства

Модель логической структуры ММУ может быть представлена в виде совокупности функциональных блоков, обеспечивающих передачу информационных потоков, которую можно обозначить выражением

$$B = \{b_i^k\},$$

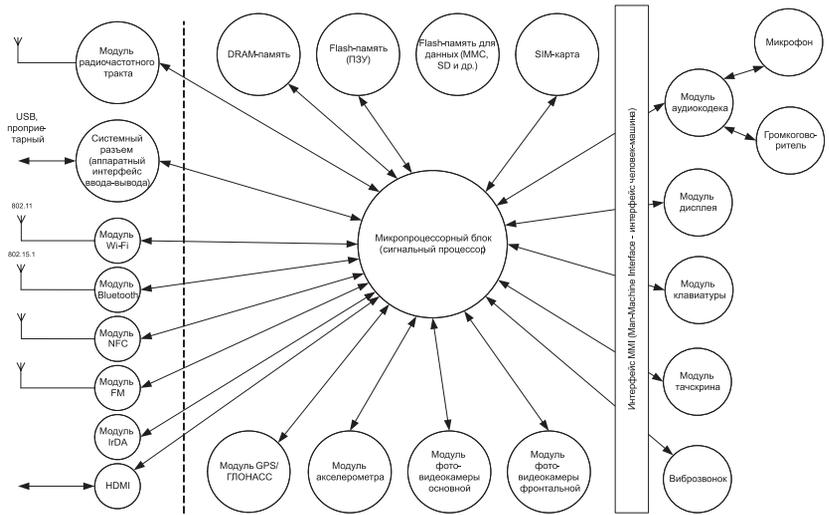


Рис. 1. Схема логической структуры взаимосвязей между функциональными блоками многофункционального мобильного устройства

где $i = \overline{1, N_B}$, N_B — число функциональных блоков (компонентов) ММУ; $k = \overline{1, N_t}$, N_t — число типов интерфейсов функциональных блоков. Для более полного и общего учета возможностей передачи информационных потоков дополнительно может быть определено число режимов работы каждого типа интерфейса $l(k)$.

На основе данных параметров можно определить варианты трактов прохождения информационных потоков через ММУ с использованием индикатора возможности установления связи между i -м и j -м функциональными блоками:

$$I_{i,j}^{k,l} = \{0, 1\}, \quad i, j = \overline{1, N_B}, \quad k = \overline{1, N_t}, \quad l = l(k).$$

В случае $I_{i,j}^{k,l} = 1$ связь между функциональными блоками существует, в противном случае — нет. Данные величины отражают наличие информационных направлений между компонентами ММУ.

Дополнительно, с учетом наличия различных типов интерфейсов, может быть учтен показатель связности $f^{k,l}$, обеспечиваемый k -м типом интерфейса в l -м режиме работы. Показатель связности подразумевает наличие связи один к одному, один ко многим или многие ко многим.

Математическое описание маршрутов передачи информационного потока через компоненты ММУ может быть представлено в виде некоторого множества M , учитывающего наличие информационных направлений между компонентами. Матричное представление маршрута будет иметь вид

$$m = \begin{bmatrix} i_1 & \dots & i_{L_m-1} & i_{L_m} \\ k_1 & \dots & k_{L_m-1} & 0 \\ l_1 & \dots & l_{L_m-1} & 0 \end{bmatrix},$$

где L_m — общее число функциональных блоков, входящих в маршрут m с длиной $L_m - 1$; (i_1, \dots, i_{L_m}) —

последовательность интерфейсов ММУ в маршруте; (k_1, \dots, k_{L_m-1}) — последовательность типов интерфейсов в маршруте; (l_1, \dots, l_{L_m-1}) — последовательность режимов работы соответствующих типов интерфейсов в маршруте. В соответствии с этим множество M определяет допустимые маршруты передачи информационных потоков.

С использованием данной матрицы маршрута информационного потока может быть получен индикатор действующего маршрута в виде выражения

$$I_m = I_{i_1, i_2}^{k_1, l_1}, I_{i_2, i_3}^{k_2, l_2}, \dots, I_{i_n, i_{n+1}}^{k_n, l_n} I_{i_1} I_{i_2} \dots I_{i_{n+1}}, \quad n = L_m - 1. \quad (3)$$

Таким образом, состояние логической структуры ММУ, обслуживающей информационные потоки с разным уровнем конфиденциальности, в соответствии с выражением (1) в каждый момент времени t задается вектором массивов:

$$s^{<0>}(t) = \langle \{I_i(t)\}, \{b_i^k(t)\}, \{I_{i,j}^{k,l}(t)\}, \{f^{k,l}\}, M \rangle,$$

где I_i , $i = \overline{1, N_B}$, — индикатор существования функционального блока (компонента) ММУ; M и $f^{k,l}$ — исходные данные.

Процесс реконфигурации $x^{<0>}(t)$ логической структуры ММУ, представленный в общем виде выражением (2), заключается в управлении параметрами (составом) ММУ. Такое управление можно задать массивом $\{b_{xi}^k(t)\}$, где $b_{xi}^k(t)$ — число интерфейсов k -го типа, введенных на i -й функциональный блок за промежуток времени (t_0, t) , т. е.

$$x^{<0>}(t) = \langle b_{xi}^k(t) \rangle. \quad (4)$$

Процесс управления логической структурой ММУ в общем виде задается выражением (2) и заключается в реконфигурации структуры данной ММУ, определяемой выражением (4). Реальное поведение логической структуры ММУ примет вид

$$b_{i,j}^{k,l}(t) = b_{xi,j}^{k,l}(t) I_{i,j}^{k,l}(t) I_i(t) I_j(t).$$

Связность линий между функциональными блоками ММУ в каждый момент времени t может быть определена в виде

$$f^{k,l} = f^{k,l}(t) b_{i,j}^{k,l}(t),$$

а также может быть представлена в виде матрицы связности функциональных блоков ММУ:

$$F^{k,l}(t) = [f^{k,l}] = [f^{k,l}(t) \cdot b_{i,j}^{k,l}(t)],$$

определяющей допустимые маршруты прохождения информационных потоков в рамках информационной инфраструктуры логической структуры ММУ на момент времени t .

Если перейти от представления логической структуры ММУ в виде наличия в ее составе задан-

ных функциональных блоков с их интерфейсами и режимами работы к представлению в виде связности между данными блоками, то тогда процесс реконфигурации логической структуры ММУ и, соответственно, процесс управления информационными потоками может быть представлен в виде выражения

$$x^{<2>}(t) = \{f_x(m, t)\}.$$

Тогда в общем виде процесс поведения логической структуры в виде динамики изменения связности между функциональными блоками ММУ и с учетом случайных воздействий на нее примет вид

$$f(m, t) = [f_x(m, t) - f_{\xi_S}(m, t)] I(m, t),$$

где $I(m, t)$ — индикатор действующего маршрута m в момент времени t , представленный выражением (3).

На основе представленных выражений можно сделать вывод, что процесс управления профилем защиты ММУ является управлением логической структурой, представленной в виде связности между функциональными блоками ММУ. Таким образом, на центр управления информационной безопасностью возлагается задача выбора оптимальной матрицы связности функциональных блоков ММУ (профиля защиты ММУ), обеспечивающего защищенный доступ к разнокатегорированным услугам корпоративной сети, исходя из текущих условий и параметров доступа.

Модель управления профилем защиты мобильного устройства на основе матричных графовых грамматик

Задачей управления информационными потоками с разным уровнем конфиденциальности является адаптация логической структуры ММУ до такого состояния, при котором будут выполняться требования по обеспечению безопасной передачи информационных потоков через функциональные блоки ММУ, а также ряд других требований, предъявляемых той или иной надсистемой. В общем виде процесс адаптации может быть представлен как последовательность улучшающихся структур:

$$W_i, W_{i+1}, \dots, W_N, i = \overline{1, N}, W = \{W_i\},$$

где W — множество допустимых логических структур ММУ; W_i — логическая структура ММУ на i -й итерации; N — число итераций, при этом на каждой итерации выполняется условие

$$W_N \succ W_{N-1} \succ \dots \succ W_{i+1} \succ W_i$$

т. е. некоторая логическая структура ММУ на $(i+1)$ -й итерации предпочтительней (лучше) структуры на предыдущей i -й итерации процесса адаптации.

Стоит отметить, что наиболее подходящим формальным описанием изменения логической структуры ММУ, которое выражается, в том числе, в виде

изменения маршрутов информационных потоков, является теория графов. Функциональные блоки логической структуры ММУ в этом случае представляются в виде вершин графа, а ребра графа — направления трактов прохождения информационных потоков.

Наиболее подходящим аппаратом для описания процесса генерации (эволюционной адаптации) структуры являются графовые грамматики, которые делятся на порождающие и трансформирующие. Порождающая графовая грамматика содержит конечное множество исходных графов и конечное множество правил допустимых локальных преобразований графов. Если набор исходных графов не задан, грамматика называется трансформирующей. Для решения подобных задач анализа и синтеза динамических структур, описываемых графами, в 70-х—80-х гг. советскими учеными была разработана теория "графодинамики" [9—11].

В 2007 г. для описания динамики графа и описания графовых грамматик был разработан (в основном на булевой алгебре и алгебре матриц) строгий алгебраический подход, названный матричными графовыми грамматиками (Matrix Graph Grammars) [12].

В общем виде формальная грамматика задается в виде

$$G = (V_T, V_N, P, S),$$

где V_T — конечный алфавит нетерминальных символов — множество матриц, описывающих матрицы связности функциональных блоков логической структуры ММУ (исходные маршруты информационных потоков в ММУ); V_N — конечный алфавит терминальных символов — множество матриц, описывающих итоговые матрицы связности логической структуры ММУ (маршруты информационных потоков в ММУ); P — конечное множество правил порождения (допустимых операций над матрицами связности логической структуры ММУ); S — начальный нетерминал грамматики (исходная матрица связности функциональных блоков логической структуры ММУ (например, при включении питания)).

Опишем логическую структуру ММУ в терминах теории графов.

Любой граф можно представить матрицей смежности

$$A_G = \{a_{ij}\},$$

где $a_{ij} = 1$, если существует связь между i -м и j -м функциональными блоками.

Пусть L — исходный граф (перед шагом трансформации, т. е. $L \in V_T$), R — конечный граф (после шага трансформации, т. е. $R \in V_N$). Необходимо определить действие, преобразующее исходный граф L в R :

$$p(L, R): L \rightarrow R,$$

где $p \in P = \Xi = (\xi_1, \xi_2, \dots, \xi_p)$.

Суть преобразования графа заключается в удалении либо добавлении связей между вершинами графа, т. е. по сути в изменении матрицы связности $F^{k,l}$, поэтому введем понятие матрицы удаления связей и матрицы добавления связей.

Тогда матрица удаления связей может быть представлена в виде

$$e = L \wedge (\overline{L \wedge R}) = L \wedge \overline{R}, e = (e)_{ij} = \begin{cases} 1, & \text{удаляется связь;} \\ 0, & \text{связь не удаляется,} \end{cases}$$

а матрица добавления связей в виде

$$r = R \wedge (\overline{L \wedge R}) = R \wedge \overline{L}, r = (r)_{ij} = \begin{cases} 1, & \text{добавляется связь;} \\ 0, & \text{связь не добавляется.} \end{cases}$$

Теперь на основе данных выражений можно представить преобразование $p(L, R): L \rightarrow R$ операций над исходной матрицей связности $L \in V_T \subseteq V$ и матрицами удаления и добавления связей:

$$p: L \rightarrow R \Rightarrow R = r \vee (\overline{e} \wedge L).$$

С помощью этого подхода можно синтезировать любой случайный граф. Чтобы получаемые в результате преобразования $p(L, R): L \rightarrow R$ случайные графы описывали реальную структуру ММУ, необходимо задать ряд ограничений на преобразования графа, т. е. в терминах формальной грамматики правила порождения.

На рис. 3 представлен пример осуществления преобразования $p(L, R): L \rightarrow R$ в случае с операцией полного дублирования вершины.

Матрица связности для исходного и конечного графов выглядит как

$$L = \left(\begin{array}{cccc|c} 0 & 0 & 0 & 1 & k \\ 0 & 0 & 0 & 1 & l \\ 0 & 0 & 0 & 1 & m \\ 1 & 1 & 1 & 0 & n \end{array} \right), R = \left(\begin{array}{cccc|c} 0 & 0 & 0 & 1 & k \\ 0 & 0 & 0 & 1 & l \\ 0 & 0 & 0 & 1 & m \\ 1 & 1 & 1 & 0 & n \\ 0 & 0 & 0 & 0 & z \end{array} \right).$$

Матрицы удаления и добавления связей —

$$e = \left(\begin{array}{cccc|c} 0 & 0 & 0 & 0 & k \\ 0 & 0 & 0 & 0 & l \\ 0 & 0 & 0 & 0 & m \\ 0 & 0 & 0 & 0 & n \\ 0 & 0 & 0 & 0 & z \end{array} \right), r = \left(\begin{array}{cccc|c} 0 & 0 & 0 & 0 & k \\ 0 & 0 & 0 & 0 & l \\ 0 & 0 & 0 & 0 & m \\ 0 & 0 & 0 & 0 & n \\ 1 & 1 & 1 & 0 & z \end{array} \right).$$

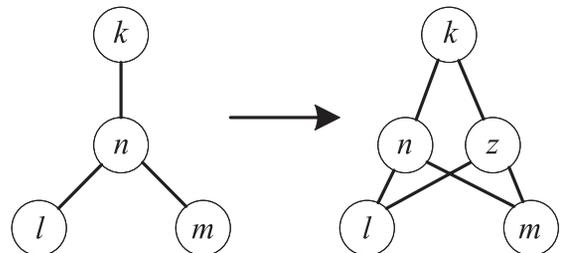


Рис. 3. Полное дублирование вершины

Все преобразования имеют вид

$$L = \left(\begin{array}{cccc|c} 0 & 0 & 0 & 1 & 0 & k \\ 0 & 0 & 0 & 1 & 0 & l \\ 0 & 0 & 0 & 1 & 0 & m \\ 1 & 1 & 1 & 0 & 0 & n \\ 0 & 0 & 0 & 0 & 0 & z \end{array} \right) \vee \left[\begin{array}{cccc} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{array} \right] \wedge$$

$$\wedge \left(\begin{array}{cccc|c} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{array} \right) = \left(\begin{array}{cccc|c} 0 & 0 & 0 & 1 & 1 & k \\ 0 & 0 & 0 & 1 & 1 & l \\ 0 & 0 & 0 & 1 & 1 & m \\ 1 & 1 & 1 & 0 & 0 & n \\ 1 & 1 & 1 & 0 & 0 & z \end{array} \right).$$

Аналогичным образом могут быть описаны и другие операции над графами, такие как понижение степени вершины, повышение степени вершины, понижение степени вершины разбиением, понижение размерности задачи путем разбиения графов и т. д.

Также различные комбинации отдельных подобных правил могут быть объединены в последовательности преобразований [12].

Теперь, на основе описания с помощью матричных графовых грамматик процедуры управления логической структурой ММУ, влияющей на маршруты информационных потоков, можно определить процесс адаптации логической структуры ММУ в соответствии с заданными критериями. В общем виде процесс адаптации может быть представлен в виде рекуррентного соотношения

$$c_{n+1}^i = c_n^i + \gamma_n Q_i(c_n, X_n),$$

где $c_n = \{c_n^i\}$ — параметры, характеризующие качество структуры по заданным критериям; X_n — случайная величина (некоторый граф, представляющий логическую структуру ММУ в момент n); Q_i — детерминированная функция, определяющая качество логической структуры ММУ; γ_n — числовая последовательность, определяющая преобразования из множества $\Xi = (\xi_1, \xi_2, \dots, \xi_p)$.

Для решения задачи адаптации может быть использован адаптивный алгоритм [15]

$$c_n = c_{n-1} - \gamma_n \frac{dQ(c_{n-1}, X_n)}{dc}.$$

Вероятность сходимости данной рекуррентной последовательности, как показано в работах [13, 14], равна единице при выполнении следующих условий:

$$\sum_{n=1}^{\infty} \gamma_n = \infty, \quad \lim_{n \rightarrow \infty} \gamma_n = 0;$$

$$\inf_{\delta < (c - c^*) < \frac{1}{2}} M_x \left\{ (c - c^*) \frac{dQ(X, c)}{dc} \right\} > 0, \quad \forall \varepsilon > 0;$$

$$M_x \left\{ \left[\frac{dQ(X, c)}{dc} \right]^2 \right\} \leq (1 + c^2), \quad d > 0.$$

Ограничения, накладываемые ММУ на возможности по генерации логической структуры, приводят к тому, что траектория процесса аппроксимации не выходит за пределы множества допустимых значений, что делает доказательство сходимости итеративного процесса тривиальным.

Необходимо определить правила остановки для адапционного алгоритма. Способ определения момента остановки — конечного значения номера итерации N — состоит в выборе по достаточно малым разностям между несколькими последовательными значениями $c_{n-i+1}, c_{n-1}, i = 1, \dots, k$. Данное условие может быть записано в виде

$$N = \min \left\{ n: \max_{i=1, k} |c_{n-i+1} - c_{n-i}| < \delta \right\}$$

или

$$N = \min \left\{ n: \sum_{i=1}^k |c_{n-i+1} - c_{n-i}| < \delta \right\},$$

где δ — некоторая величина, равная предельной допустимой разнице между последовательными значениями параметров, характеризующих качество оптимизируемой структуры.

Для обоснования правила остановки необходимо решить вопрос выбора:

$\delta(\varepsilon, \alpha)$, чтобы выполнялось

$$P \left(|c_n - c^*| < \frac{\varepsilon}{|c_n - c_{n-1}|} < \delta \right) \geq 1 - \alpha;$$

δ и k в зависимости от ε и α , чтобы выполнялось

$$P \left(|c_n - c^*| < \frac{\varepsilon}{\max_{i=1, k} |c_{n-i+1} - c_{n-i}|} < \delta \right) \geq 1 - \alpha$$

или

$$P \left(|c_n - c^*| < \frac{\varepsilon}{\sum_{i=1}^k |c_n - c_{n-i}|} < \delta \right) \geq 1 - \alpha,$$

где c^* — решение задачи; k — число итераций; ε — величина, равная разнице между c_n и c^* ; α — уровень значимости.

Конкретные значения параметров δ и k в зависимости от ε и α могут быть получены в результате моделирования процесса функционирования системы управления профилем защиты ММУ.

Заключение

В данной работе обоснована актуальность задачи управления информационными потоками с разным уровнем конфиденциальности при решении задач информационного взаимодействия различных информационных систем. Представлены общий вид

модели процесса управления логической структурой информационной системы и формальная постановка задачи. Дана характеристика логической структуры многофункционального мобильного устройства, обслуживающего информационные потоки с разным уровнем конфиденциальности.

Предложена модель управления информационными потоками с разным уровнем конфиденциальности на основе управления логической структурой информационной системы с использованием аппарата матричных графовых грамматик. Заданы критерии останки адаптационного алгоритма, используемого для адаптации логической структуры информационной системы и поиска оптимальной.

Предложенная модель может быть использована при эксплуатации многофункциональных мобильных устройств в условиях защищенной корпоративной сети, в которой осуществляется обработка информации с разным уровнем конфиденциальности, а также в любых других условиях, когда окружающая обстановка требует изменения конфигурации логической структуры в целях оптимального расходования ресурсов одновременно с выполнением предъявляемых к устройству требований.

Список литературы

1. Хрусталева Д. А. Мобильные телефоны Siemens. Принципы устройства и ремонт. М.: Изумруд, 2004. 256 с.
2. Заяц А. Обзор и тестирование смартфона Caesar A9600, а также знакомство с MT6589 — четырехъядерной SoC Medi-

aTek для бюджетных решений. URL: <http://ixbt.com/md/pda/> (Дата обращения: 03.03.2014 г.).

3. Щербак Н. Программируемые радиостанции — будущее тактической связи // Электроника: Наука, Технология, Бизнес. 2001. № 5. С. 16—19.

4. Uhm M. Adaptivity in Action for SDR and Cognitive Radio // COTS Journal. February. 2006. URL: <http://www.cotsjournalonline.com> (Дата обращения: 14.10.2014 г.).

5. Зорин К. М. Модель и методика реконфигурации логической структуры интегрированного объекта информатизации // Известия СПбГЭТУ "ЛЭТИ". 2010. № 6. С. 20—25.

6. ГОСТ Р 51275—2006. Защита информации. Объект информатизации. Факторы, воздействующие на информацию. Общие положения. Введ. 2006.12.27. М.: Федеральное агентство по техническому регулированию и метрологии, 2007. 8 с. (Национальный стандарт Российской Федерации).

7. Чуднов А. М., Барашков П. Н., Ткаченко А. П., Ткаченко К. А. Модель системы связи с управляемыми структурами в конфликтных условиях. Л.: ВАС, 1986.

8. Советов Б. Я., Яковлев С. А. Моделирование систем: учеб. для вузов. М.: Высшая школа, 2007.

9. Петров С. В. Графовые грамматики и задачи графодинамики // Автоматика и телемеханика. 1977. № 10.

10. Петров С. В. Нормальная форма графовых грамматик // Автоматика и телемеханика. 1977. № 6.

11. Айзерман М. А., Гусев Л. А., Петров С. В., Смирнова И. М., Тененбаум Л. А. Динамический подход к анализу структур, описываемых графами (основы графодинамики) // Сб. науч. трудов. Академия наук СССР "Исследования по теории структур". М.: Наука, 1988.

12. Perez P. P. Matrix Graph Grammars: An Algebraic Approach to Graph Dynamics. VDM Verlag, 2009. С. 5—80.

13. Браверман Э. М., Розоноэр Л. И. Сходимость случайных процессов в теории обучения машин I, II // Автоматика и телемеханика, № 1. 1969.

14. Цыпкин Я. З. Адаптация и обучения в автоматизированных системах. М.: Наука. 1968.

15. Ферманн Х. Х. О некоторых подходах к определению правила останки для алгоритмов адаптации // Проблемы случайного поиска. Кн. 4 / Под общ. ред. Л. А. Растрюгина. Рига: Зинатне, 1975. С. 55—59.

D. O. Markin, Employee, e-mail: admin@nikitka.net, V. V. Komashinskij, Associate Professor, I. Yu. Baranov, Associate Professor, The Academy of the Federal security service of Russia

Mobile Device Security Profile Management Model Using Access to Services With Different Privacy Level

The article contains description of mobile device logical structure management process model. This model contains the analytic model of mobile device logical structure, the logical structure of mobile device functional blocks relations, the mobile device security profile management model based on matrix graph grammars.

The essence of mobile device logical structure management process model is the reconfiguration of the logical structure of the interconnections between the mobile device functional blocks such as CPU, RAM, communication modules, for example, GSM, CDMA, LTE, Wi-Fi, Bluetooth, video and audio subsystems, etc. The main objective of this model is providing the required level of information security.

Mobile device logical structure features are presented. These features allow to present informational traffic routes between mobile device functional blocks. So using these routes the mobile device logical structure can be justified.

The authors show location of this model in the services access control system using access to services with different privacy level. The optimal mobile device logical structure achieve by using the mathematical apparatus of dynamic graphs and optimization iterative algorithms.

Using this model the authors offer mobile device security profile management model using access to services with different privacy level based on matrix graph grammars.

Keywords: mobile device, security profile, logical structure reconfiguration, dynamic graphs, graph grammars

References

1. **Hrustalev D. A.** *Mobil'nye telefony Siemens. Principy ustrojstva i remont.* M.: Izumrud, 2004. 256 p.
2. **Zajac A.** *Obzor i testirovanie smartfona Caesar A9600, a takzhe znakomstvo s MT6589 — chetyrehadernoi SoC MediaTek dlia biudzhetnyh reshenij.* URL: <http://ixbt.com/md/pda/> (Data obrashhenija: 03.03.2014.).
3. **Shherbak N.** Programmiruemye radiostancii — budushhee takticheskoi svyazi / N. Shherbak. *Jelektronika: Nauka, Tehnologija, Biznes.* 2001, no. 5, pp. 16–19.
4. **Uhm M.** Adaptivity in Action for SDR and Cognitive Radio. *COTS Journal.* — February, 2006. URL: Rezhim dostupa: <http://www.cotsjournalonline.com> (Data obrashhenija: 14.10.2014).
5. **Zorin K. M.** Model' i metodika rekonfiguracii logicheskoi struktury integrirovannogo ob'ekta informatizacii. *Izvestija SPbGJeTU "LJeTI".* 2010, no. 6, pp. 20–25.
6. **GOST R 51275—2006.** *Zashhita informacii. Ob'ekt informatizacii. Faktory, vozdeystvujushhie na informaciju. Obshhie polozenija. Vved.* 2006.12.27. — M.: Federal'noe agentstvo po tehničeskomu regulirovaniju i metrologii, 2007. 8 p. (Nacional'nyj standart Rossijskoj Federacii).
7. **Chudnov A. M., Barashkov P. N., Tkachenko A. P., Tkachenko K. A.** *Model' sistemy svyazi s upravljaemymi strukturami v konfliktnyh uslovijah.* L.: VAS, 1986.
8. **Sovetov B. Ja., Jakovlev S. A.** *Modelirovanie sistem: ucheb. dlja vuzov.* M.: Vysshaja shkola, 2007.
9. **Petrov S. V.** Grafovyje grammatiki i zadachi grafodinamiki. *Avtomatika i telemehanika.* 1977, no. 10.
10. **Petrov S. V.** Normal'naja forma grafovych grammatik. *Avtomatika i telemehanika.* 1977, no. 6.
11. **Ajzerman M. A., Gusev L. A., Petrov S. V., Smirnova I. M., Tenenbaum L. A.** *Dinamicheskij podhod k analizu struktur, opisyvaemyh grafami (osnovy grafodinamiki) / Sb. nauch. trudov. Akademija nauk SSSR "Issledovanii po teorii struktur".* M.: Nauka, 1988.
12. **Perez P. P.** *Matrix Graph Grammars: An Algebraic Approach to Graph Dynamics.* VDM Verlag, 2009, pp. 5–80.
13. **Braverman Je. M., Rozonojer L. I.** *Shodimost' sluchajnyh processov v teorii obuchenija mashin I, II / Je. M. Braverman. Avtomatika i telemehanika.* 1969, no. 1.
14. **Cypkin Ja. Z.** *Adaptacija i obuchenija avtomatizirovannyh sistemah.* M.: Nauka. 1968.
15. **Fermann H. H.** *O nekotoryh podhodah k opredeleniju pravila ostanovki dlja algoritmov adaptacii. Problemy sluchajnogo poiska.* Kn. 4. Pod obshh. red. L. A. Rastrigina. Riga: Zinatne, 1975, pp. 55–59.

УДК 004.942; 004.056.55; 004.384

Д. В. Капулин, канд. техн. наук., зав. каф., e-mail: dkapulin@sfu-kras.ru,

О. В. Дрозд, студент, e-mail: olvidrozd@gmail.com

Сибирский федеральный университет, г. Красноярск

Устройство аппаратного шифрования производственных данных

Предложено устройство, обеспечивающее защищенный обмен производственной информацией с использованием алгоритмов ГОСТ 28147—89. Реализация устройства выполнена с применением ПЛИС Xilinx Spartan-6. Подробно рассмотрены этапы проектирования устройства, проведено имитационное моделирование разработанного устройства в среде Simulink.

Ключевые слова: аппаратное шифрование, передача данных, защита информации, беспроводная связь, wi-fi

Введение

Широкое распространение технологий беспроводного *Ethernet* в корпоративном секторе и в секторе электронных устройств для частного пользования закономерным образом приводит к росту внимания к этим технологиям со стороны производителей и интеграторов автоматизированных систем управления технологическими процессами (АСУ ТП). Применительно к АСУ ТП беспроводные сети и устройства передачи данных обладают следующими преимуществами [1]:

- возможность расположения устройств приема-передачи в труднодоступных местах;
- удобство развертывания и обслуживания устройств;
- оперативное добавление устройств в корпоративную сеть или исключение из нее;
- возможность расположения устройств приема-передачи данных на подвижных объектах.

Кроме того, внедрение беспроводных устройств контроля технологических и производственных параметров открывает новые возможности по применению систем автоматизации, такие как обеспечение доступа к объекту, контроль периметра объекта, наблюдение за перемещениями персонала на территории предприятия, автоматизация контроля проведения инспекций и технического обслуживания, контроль экологических параметров окружающей среды и т. д.

Разработку контрольно-измерительных приборов АСУ ТП со встроенными беспроводными интерфейсами ведет компания *Yokogawa* с использованием стандарта промышленной беспроводной связи ISA 100/11a. Номенклатура беспроводных измерительных преобразователей *Yokogawa* включает в себя датчики температуры, абсолютного, избыточного и дифференциального давления [2]. Среди производителей промышленных беспроводных точек дос-

тупа следует также отметить компании *Yokogawa* и *Hirschmann* [1].

При использовании беспроводных технологий для организации связи и управления устройствами, входящими в состав АСУ ТП, следует уделять внимание и защите передаваемой информации, которая зачастую носит конфиденциальный характер. Так, компания *Yokogawa* уделяет особое внимание разработке беспроводных устройств контроля и управления системами нефте- и газопроводов, аппаратуры технологического и производственного учета с применением аппаратного шифрования информации. В качестве алгоритмов шифрования в промышленных сетях, построенных с использованием подобного оборудования, наиболее широкое применение находят алгоритмы *DES* и *AES*. Основными недостатками алгоритма шифрования *DES* являются существование слабых ключей, низкая устойчивость алгоритма к прямому перебору ключей, в том числе с использованием аппаратных средств, низкая устойчивость при атаке с использованием дифференциального криптоанализа. В качестве основного недостатка алгоритма шифрования данных *AES* следует отметить недостаточную изученность математического аппарата шифрования [3].

Указанные недостатки алгоритмов шифрования *DES* и *AES* позволяют сделать вывод о недостаточном уровне защищенности каналов беспроводной связи для их массового применения в АСУ ТП, особенно на объектах энергетики и оборонно-промышленного комплекса. Проблема безопасности беспроводных сетей передачи данных в АСУ ТП усугубляется тем, что в России отсутствуют государственные стандарты по информационной безопасности АСУ ТП [4], аналогичные таким стандартам, как ISA SP99, IEEE 1402, IEC 62351 [5–7]. Вместе с тем существует апробированный алгоритм криптографического преобразования, закрепленный стандартом ГОСТ 28147–89 [8].

Постановка задачи

Цель исследования состоит в разработке методов и средств обеспечения безопасной передачи производственных или технологических данных по промышленным беспроводным сетям передачи данных с использованием алгоритма криптографического преобразования (ГОСТ 28147–89). При этом следует отметить наличие различных подходов к реализации данного алгоритма на базе разнообразных аппаратных средств [9–11].

Для постижения поставленной цели предлагается метод проектирования устройства для организации беспроводного защищенного обмена информацией. Реализацию устройства

предлагается выполнить с помощью криптографического модуля, сформированного на базе программируемой логической интегральной схемы (ПЛИС) *Xilinx Spartan-6 XC6SLX25* с использованием серийного комплекта разработчика на базе данной ПЛИС. Разрабатываемое устройство должно обеспечивать взаимодействие с мобильными устройствами сторонних производителей посредством протоколов *USB 2.0* и *IEEE 802.11*.

Описание устройства обмена данными

Рассмотрим процесс разработки устройства для организации защищенного обмена информацией (данными) по промышленным сетям беспроводной связи. На рис. 1 изображена структурная схема устройства защищенной передачи данных. Для защиты канала передачи данных необходимо использовать как минимум два подобных устройства, одно из которых связано с передатчиком и выполняет шифрование передаваемых данных, второе устройство связано с приемником и предназначено для дешифрования полученных данных. При этом устройства аналогичны и взаимозаменяемы.

На рис. 1 приняты обозначения: 1.1 — устройство защищенной передачи данных; 1.2 — проводной *USB* интерфейс; 1.3 — преобразователь интерфейсов *USB/UART*; 1.4 — криптографический блок на базе ПЛИС; 1.5 — радиointерфейс *IEEE 802.11* со встроенной радиоантенной 1.7; 1.6 — внешняя радиоантенна (опционально); 1.8 — энергонезависимая память; 1.9 — программатор энергонезависимой памяти; 1.10 — проводной *RS-232* интерфейс; 1.11 — аккумуляторная батарея; 1.12 — источник электропитания; 1.13 — переключатель режимов работы; 1.14 — генератор тактовых импульсов; 1.15 — аккумуляторная батарея; 1.16 — автоматизированное рабочее место.

В качестве преобразователя интерфейсов *USB-UART* используется микросхема *FT232R* (производитель *FTDI Chip*, корпус *SSOP-28*). Взаимодействие с внешними устройствами осуществляется посредством проводного интерфейса *USB* (1.15 на рис. 1). Для обеспечения взаимодействия с устройствами по

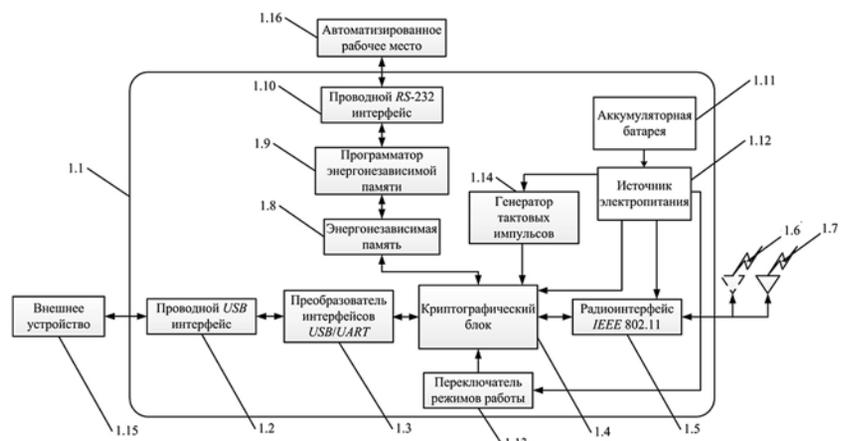


Рис. 1. Структурная схема устройства обеспечения защищенной передачи данных

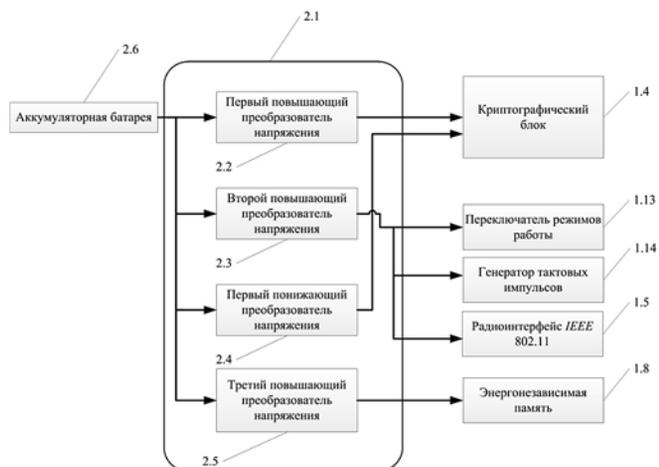


Рис. 2. Структурная схема источника электропитания

беспроводным каналам связи используется модуль *WizFi 220* (производитель *WIZnet Co.*) со встроенной антенной, также возможно подключение внешней антенны через разъем стандарта *U.FL*. Модуль *WizFi 220* поддерживает стандарты беспроводной передачи данных *IEEE 802.11b/g/h*, протоколы безопасности *WEP* и *WPA/WPA2*. Взаимодействие между модулем и внешними устройствами может осуществляться посредством интерфейсов *UART*, *SPI* и *I²C*.

В качестве ключевого запоминающего устройства используется микросхема электрически стираемого перепрограммируемого постоянного запоминающего устройства *24LC02* (производитель *Microchip*, корпус *SOP-8L*) емкостью 2048 бит, что позволяет хранить восемь секретных ключей по 256 бит каждый. Взаимодействие между ключевым запо-

минающим устройством и внешними устройствами осуществляется посредством интерфейса *I²C*. Программатор энергонезависимой памяти построен по схеме, предложенной Клаудио Ланконелли [12], взаимодействие с автоматизированным рабочим местом специалиста по информационной безопасности (1.16 на рис. 1) осуществляется посредством интерфейса *RS-232*. В качестве генератора тактовых импульсов используется генератор тактовых импульсов *KXO-197* (производитель *Geyer*, Германия).

Рассмотрим систему электропитания устройства (рис. 2). В состав источника электропитания 2.1 входят три повышающих преобразователя напряжения 2.2; 2.3; 2.5 и один понижающий преобразователь напряжения 2.4. В качестве повышающих преобразователей напряжения используются повышающие преобразователи напряжения *MAX1675* (производитель *Maxim*, корпус *UMAX10*), в качестве понижающего преобразователя напряжения используется понижающий преобразователь напряжения *LM3674* (производитель *National Semiconductor*, корпус *SOT-23*). Понижающий преобразователь напряжения используется для обеспечения ПЛИС питающим напряжением 1,2 В, для обеспечения питания всех остальных потребителей используются повышающие преобразователи напряжения, кроме преобразователя интерфейсов, питание которого осуществляется через цепи питания и общего провода разъема *USB*. В качестве аккумуляторных батарей 2.6 используются два литий-ионных аккумулятора формфактора 18650 емкостью по 3200 мА · ч, для управления процессом зарядки аккумуляторных батарей используется контроллер заряда *bq24002* (производитель *Texas Instruments*, корпус *R-PDSO-G20*). Процесс заряда аккумуля-

торных батарей осуществляется через цепи питания и общего провода разъема *USB*, таким образом заряжать аккумуляторные батареи можно от персонального компьютера или ноутбука. Заряда двух аккумуляторных батарей достаточно для 48 ч непрерывной работы при максимальном энергопотреблении, притом что модуль *WizFi 220* будет работать только в режиме передачи.

Рассмотрим структуру криптографического блока и назначение модулей, входящих в его состав (рис. 3). Криптографический блок реализован на базе ПЛИС *Xilinx Spartan-6 XC6SLX25* с использованием серийного комплекта разработчика *SK-iMX53-XC6SLX* [13]. В качестве языка реализации алгоритма шифрования выбран язык описания аппаратуры *Verilog*.

Криптографический блок состоит из процессора 3.1, криптографического сопроцессора 3.2, первого универ-

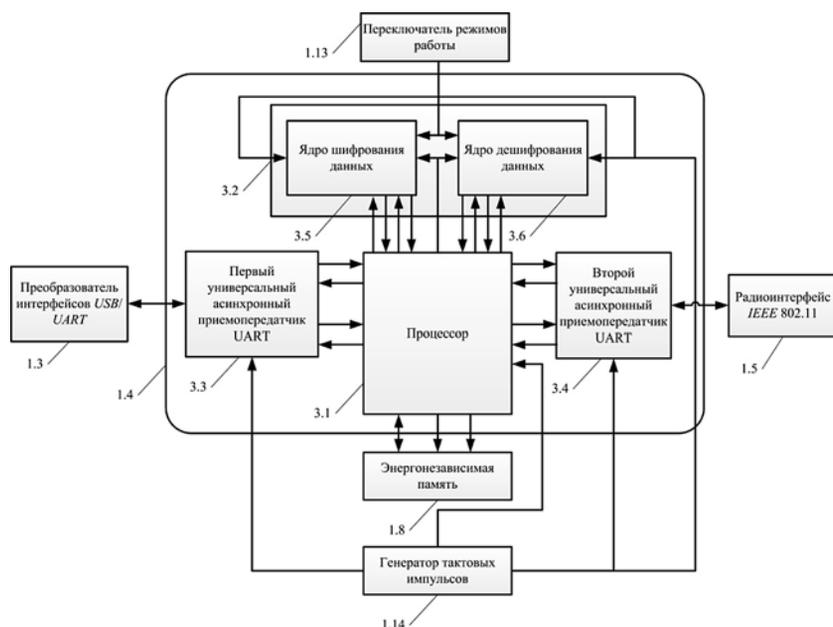


Рис. 3. Структурная схема криптографического блока

сального асинхронного приемопередатчика *UART* 3.3, второго универсального асинхронного приемопередатчика *UART* 3.4. В состав криптографического сопроцессора входит ядро шифрования данных 3.5 и ядро дешифрования данных 3.6.

Основной задачей первого универсального асинхронного приемопередатчика является прием пакетов открытых данных от внешних устройств и передача пакетов открытых данных внешним устройствам, при этом непосредственным приемником и передатчиком данных является преобразователь интерфейсов *USB/UART*, с которым, в свою очередь, взаимодействуют мобильные устройства, подключаемые посредством проводного *USB* интерфейса. Также первый универсальный асинхронный приемопередатчик обеспечивает управление процессом передачи и приема данных.

Задача второго универсального асинхронного приемопередатчика — прием пакетов закрытых данных от внешних устройств и передача пакетов закрытых данных внешним устройствам, при этом непосредственным приемником и передатчиком данных является радиointерфейс *IEEE 802.11* на базе модуля *WizFi 220*, с которым взаимодействуют мобильные устройства посредством беспроводного канала связи стандарта *IEEE 802.11*. Также второй универсальный асинхронный приемопередатчик обеспечивает управление процессом передачи и приема данных.

На рис. 4 представлена структурная схема ядра шифрования данных, при этом структуры ядра шифрования и дешифрования данных аналогичны. Ядра предназначены соответственно для осуществления процедур шифрования и дешифрования последовательностей двоичных данных в соответствии с алгоритмом шифрования согласно ГОСТ 28147—89. Ядра шифрования и дешифрования данных включают в себя блоки выполнения криптографических преобразований в режиме простой замены 4.1, режиме гаммирования 4.2 и режиме гаммирования с обратной связью 4.3. В состав ядер шифрования и дешифрования данных также входят мультиплексоры 4.5, 4.6 и демultipлексоры 4.4, 4.7, предназначенные для коммутации внутренних каналов передачи данных в зависимости от выбранного режима работы криптографического сопроцессора.

Выбор режима шифрования/дешифрования данных осуществляется с помощью переключателя режимов работы. Переключатель режимов работы представляет собой *DIP*-переключатель на три контактные группы, размещаемый на поверхности печатной платы устройства. С помощью переключателя

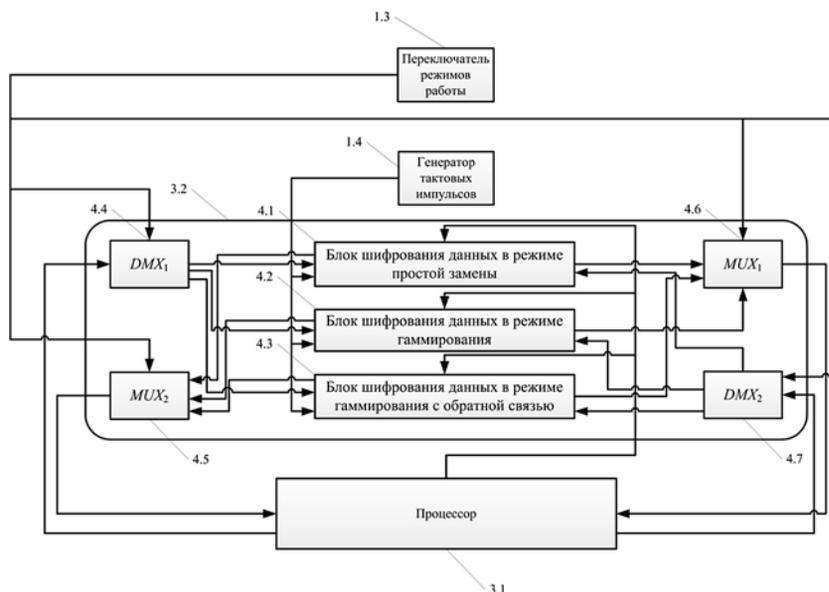


Рис. 4. Структурная схема ядра шифрования данных

режимов работы выполняется переключение криптографического сопроцессора, входящего в состав криптографического блока, на иные режимы работы. Первая контактная группа соответствует режиму простой замены, вторая контактная группа соответствует режиму гаммирования, третья контактная группа соответствует режиму гаммирования с обратной связью.

Основными функциями процессора является обеспечение:

- считывания информации из энергонезависимой памяти, что включает в себя также адресацию памяти и управление процессом чтения информации;
- взаимодействия с первым и вторым универсальными асинхронными приемопередатчиками *UART*, что включает в себя прием данных, передачу данных, управление процессом приема данных, управление процессом передачи данных;
- взаимодействия с криптографическим сопроцессором, что включает в себя передачу и прием как открытых, так и закрытых данных, управление процессами приема и передачи как открытых, так и закрытых данных, передачу криптографическому процессору секретных ключей, предварительно считанных из энергонезависимой памяти.

Выработка секретных ключей осуществляется с использованием программного продукта, реализующего линейный конгруэнтный метод для генерации псевдослучайных чисел. Возможна разработка аппаратного генератора случайных чисел с использованием таких процессов, как тепловой шум полупроводниковых приборов, фотоэлектрический эффект и неравномерность в задержках логических элементов [14].

Параметры реализации алгоритма криптографического преобразования по ГОСТ 28147-89 на базе ПЛИС

Семейство ПЛИС	Модель ПЛИС	Число логических ячеек	Задержка, нс	Потребляемая мощность, Вт	Частота, МГц	Пропускная способность, Мбит/с
<i>Artix-7</i>	<i>XC7A200</i>	3808	131,0860	0,0730	7,6286	61,0286
<i>Spartan-6</i>	<i>XC6SLX25</i>	3808	161,3270	0,0290	6,1986	49,5887

В табл. 1 представлены некоторые параметры реализации шифроалгоритма ГОСТ 28147—89 на базе ПЛИС. Для рассматриваемого случая представлены параметры реализации 32 раундов шифрования данных в режиме простой замены.

Методика проектирования устройства защищенного обмена информацией

Исходя из приведенной структуры разработанного криптографического блока сформулируем методику проектирования подобного рода устройств шифрования на базе ПЛИС. Основная идея предлагаемой методики проектирования состоит в интеграции основных этапов проектирования (математическое моделирование, аппаратная реализация, отладка в составе системы) в единый итерационный цикл проектирования на основе включения в него дополнительных этапов проектирования и автоматизации процесса передачи формализованных описаний проектируемого устройства и данных, получаемых при моделировании и системной интеграции между этапами проектирования.

Предлагаемая методика проектирования состоит из следующих этапов.

1. Формализация технического задания на систему и устройства на базе ПЛИС. Разработка структурной схемы как для системы, так и для устройства.

2. Разработка идеализированной (линеаризованной) математической модели системы с использованием формата чисел с плавающей запятой, выполнение численного моделирования в среде *MATLAB/Simulink* и отладка разработанной модели.

3. Преобразование идеализированного алгоритма, реализованного в формате с плавающей точкой, в алгоритм с представлением чисел с фиксированной точкой. Оптимизация разрядностей по критерию минимизации аппаратных затрат.

4. Замена блоков проектируемого устройства на библиотечные компоненты, ориентированные на дальнейшую реализацию в аппаратуре с использованием пакетов *Altera DSP Builder*, *Xilinx System Generator* или *MATLAB HDL Coder* [15].

5. Аппаратно-программное моделирование созданного устройства.

6. Отладка и анализ параметров разработанного устройства на базе ПЛИС в реальном системном окружении.

Преимуществами предлагаемой методики проектирования устройств на базе ПЛИС по сравнению

с классической методикой проектирования [16] являются:

- снижение влияния человеческого фактора при переходе от математической модели устройства к аппаратной реализации;
- непрерывность процесса проектирования устройства;
- получение в процессе проектирования отлаженной математической модели, которая является прототипом для создаваемого устройства и может быть использована не только в процессе проектирования, но и для дальнейшей оптимизации алгоритма работы устройства на основе данных, полученных из реального системного окружения проектируемого устройства.

Апробация методики проектирования проведена с использованием отладочного комплекса, включающего в себя серийный комплект разработчика *SK-iMX53-XC6SLX* и четыре отладочных модуля, реализующих ключевые узлы разработанного устройства (рис. 5). Узлы комплекса реализованы в соответствии со структурными схемами, приведенными на рис. 1—4. В состав отладочного комплекса входят модуль преобразователя интерфейсов *USB-UART*; модуль радиointерфейса *IEEE 802.11*, в состав которого входит модуль *WizFi 220* со встроенной антенной; модуль, включающий в себя энергонезависимую память и программатор энергонезависимой памяти; модуль, включающий в себя аккумуляторную батарею и источник электропитания.

Моделирование устройства защищенной передачи данных выполнено в среде *MATLAB/Simulink*.

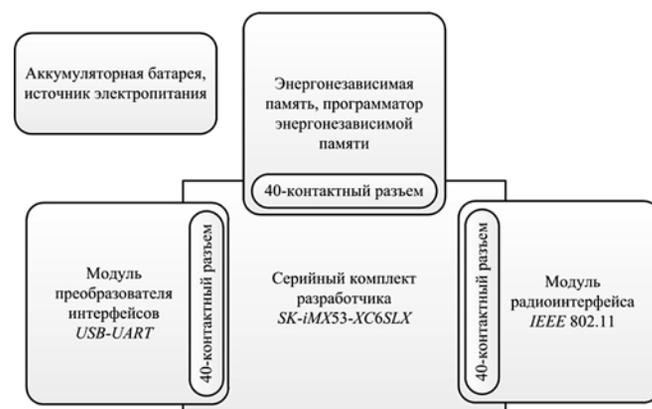


Рис. 5. Структурная схема отладочного комплекса

**Результаты моделирования устройства
для обмена закрытой документальной информацией**

Стандарт	Время работы СМО, мкс	Число поступивших заявок	Число обслуженных заявок	Число необслуженных заявок
802.11g	100	56	56	0
802.11n	100	156	156	0
802.11ac	100	1364	1362	2
Gigabit Ethernet	100	1040	1038	2
Fast Ethernet	100	104	104	0
USB 1.0	100	12	12	0
USB 2.0	100	498	497	1
USB 3.0	100	1915	1872	43

Устройство было представлено в виде трехканальной (три потока шифрования) системы массового обслуживания (СМО) с параметрами, соответствующими различным стандартам обмена информацией. Анализ результатов моделирования, приведенных в табл. 2, показывает, что устройство обеспечивает надежную обработку и передачу данных с использованием большинства стандартных интерфейсов. Применение устройства совместно с интерфейсом *USB 3.0* возможно при организации обработки поступающих пакетов данных в промежуточных накопителях (при одновременном снижении скорости обработки).

Рассматриваемая методика проектирования предлагаемого устройства обмена информацией может быть представлена в виде диаграммы *IDEF3*, приведенной на рис. 6.

Применение методики проектирования для разработки устройства аппаратного шифрования информации заканчивается его тестированием и отладкой. Предварительно должны быть проведены схмотехническое проектирование, размещение и трассировка элементов на печатной плате. Для устройства, рассматриваемого в настоящей работе, такие операции выполнены с использованием системы автоматизированного проектирования *Altium Designer*.



Рис. 6. Алгоритм проектирования устройства защищенного обмена информацией

Заключение

В результате проведенного исследования предложена и апробирована методика проектирования, а также предложена и реализована структура устройства аппаратного шифрования данных по алгоритму криптографического преобразования ГОСТ 28147—89. Разработанное устройство предназначено для безопасного обмена производственной и технологической информацией по промышленным беспроводным сетям передачи данных. Структура устройства спроектирована таким образом, чтобы была обеспечена возможность совместной работы с приборами (в том числе и с мобильными) сторонних производителей, поддерживающими интерфейсы *USB* и *IEEE 802.11*, что имеет особое значение при построении беспроводных сетей передачи данных в АСУ ТП на объектах промышленности, энергетики и оборонно-промышленного комплекса.

Работа выполнена при поддержке Красноярского краевого фонда поддержки научной и научно-технической деятельности.

Список литературы

1. **Беспроводные** точки доступа Hirschmann BAT [Электронный ресурс]. URL: <http://www.hirschmann.ru/industrial/catalog/bat54/bat.series> (дата обращения: 12.01.2015).
2. **Беспроводные** контрольно-измерительные приборы Yokogawa [Электронный ресурс]. URL: <http://yokogawa.kippostavka.ru/wireless.htm> (дата обращения: 12.01.2015).
3. **Синьковский А. В.** Разработка эффективных решений по защите информации с использованием фрактального моделирования в условиях автоматизированного проектирования и производства: Автореф. дис. канд. техн. наук. М., 2007. 28 с.
4. **Лукацкий А. В.** Безопасность АСУ ТП: от слов к делу [Электронный ресурс]. URL: <http://www.gosbook.ru/node/61562> (дата обращения: 12.01.2014).
5. **ISA99**, Industrial Automation and Control Systems Security [Электронный ресурс]. URL: <https://www.isa.org/isa99/> (дата обращения: 12.01.2015).
6. **P1402** — Standard for Physical Security of Electric Power Substations [Электронный ресурс]. URL: <http://standards.ieee.org/develop/project/1402.html> (дата обращения: 12.01.2015).
7. **Core IEC Standards** [Электронный ресурс]. URL: <http://www.iec.ch/smartgrid/standards/> (дата обращения: 12.01.2015).
8. **ГОСТ 28147—89.** Системы обработки информации. Защита криптографическая. Введ. впервые; дата введ. 01.07.90. М.: Изд-во стандартов, 1996. 26 с.
9. **Rabie A. M., Magdy S. A.** Metamorphic-Key-Hopping GOST Cipher and Its FPGA Implementation // The International Journal of Computer Science and Communication Security. 2013. Vol. 3. P. 51—60.
10. **Коробицын В. В., Ильин С. С.** Реализация симметричного шифрования по алгоритму ГОСТ—28147 на графическом процессоре // Информационные технологии. 2008. № 10. С. 46—51.

11. **Коробицын В. В., Ильин С. С.** Реализация симметричного шифрования по алгоритму ГОСТ—28147 на графическом процессоре с использованием технологии CUDA // Информационные технологии. 2011. № 4. С. 41—46.

12. **Lanconelli Open Systems** [Электронный ресурс]. URL: <http://www.lancos.com/index.html> (дата обращения: 12.01.2015).

13. **Отладочная плата SK-iMX53-XC6SLX** [Электронный ресурс]. URL: <http://www.starterkit.ru/html/index.php?name=shop&op=view&id=76> (дата обращения: 12.01.2015).

14. **Саранча С. Н.** Методика определения параметров аппаратного генератора случайных чисел, реализованного в ПЛИС архитектуры FPGA // Автоматизированные системы управления и приборы автоматизации: Всеукраинский межведомственный научно-технический сб. Харьков, 2011. Вып. 157. С. 89—94.

15. **FPGA and ASIC Design with HDL Coder and HDL Verifier** [Электронный ресурс]. URL: <http://www.mathworks.com/fpga-design/solutions.html> (дата обращения 12.01.2015).

16. **Proakis J. G., Salehi M.** Communication systems engineering / New Jersey: Prentice-Hall, 2002. 801 p.

D. V. Kapulin, Head of Chair, e-mail: dkapulin@sfu-kras.ru,

O. V. Drozd, Student, e-mail: olvidrozd@gmail.com, Siberian Federal University, Krasnoyarsk

Hardware-Based Encryption Production Data Device

This Research is devoted to the analysis of methods and techniques of data hardware encryption for production and technological information. The main objective of this paper is developing and applying the automated design method for hardware encryption tools and devices to establish the secure data channels in industrial wireless networks. The paper proposes a method of designing an electronic device, providing secure data transmission in industrial wireless data networks using cryptographic transformation algorithm GOST 28147—89. Using the proposed method, the device for data exchange with protected data is developed. The designing of the device is hold on the FPGA Xilinx Spartan-6 XC6SLX25, hardware description language — Verilog. Simulation of the developed device in a queuing system is run in Matlab/Simulink. This device is able to wireless protected data transmission between any tools, support USB and IEEE 802.11.

Keywords: hardware encryption, data transfer, data protection, wireless, wi-fi

References

1. **Besprovodnye tochki dostupa Hirschmann BAT**, URL: <http://www.hirschmann.ru/industrial/catalog/bat54/bat.series> (accessed 12.01.2015).

2. **Besprovodnye kontrol'no-izmeritel'nye pribory Yokogawa**, URL: <http://yokogawa.kip-postavka.ru/wiriless.htm> (accessed 12.01.2015).

3. **Sin'kovskii A. V.** *Razrabotka effektivnykh reshenii po zashchite informatsii s ispol'zovaniem fraktal'nogo modelirovaniya v usloviyakh avtomatizirovannogo proektirovaniya i proizvodstva* (The development of effective solutions to data protection using fractal simulation in a computer-aided design and manufacturing), Candidate's thesis abstract, Moscow. 2007. 28 p. (in Russian).

4. **Lukatskii A. V.** *Bezopasnost' ASU TP: ot slov k delu*. URL: <http://www.gosbook.ru/node/61562> (accessed 12.01.2014).

5. **ISA99**, *Industrial Automation and Control Systems Security*. URL: <https://www.isa.org/isa99/> (accessed 12.01.2015).

6. **P1402** — *Standard for Physical Security of Electric Power Substations*. URL: <http://standards.ieee.org/develop/project/1402.html> (accessed 12.01.2015).

7. **Core IEC Standards**. URL: <http://www.iec.ch/smartgrid/standards/> (accessed 12.01.2015).

8. **GOST 28147—89**. *Sistemy obrabotki informatsii. Zashchita kriptograficheskaya*. (Information processing systems. Cryptographic protection, State Standart 28147—89). Moscow: Standarty, 1996, 26 p. (in Russian).

9. **Rabie A. M., Magdy S.** A Metamorphic-Key-Hopping GOST Cipher and Its FPGA Implementation. *The International Journal of Computer Science and Communication Security*. 2013, vol. 3, pp. 51—60.

10. **Korobitsyn V. V., Il'in S. S.** Realizatsiya simmetrichnogo shifrovaniya po algoritmu GOST—28147 na graficheskom protsessore (Implementation of the symmetric encryption algorithm GOST—28147 on the GPU). *Informatsionnye tekhnologii*. 2008, no. 10, pp. 46—51 (in Russian).

11. **Korobitsyn V. V., Il'in S. S.** Realizatsiya simmetrichnogo shifrovaniya po algoritmu GOST—28147 na graficheskom protsessore s ispol'zovaniem tekhnologii CUDA (Implementation of the symmetric encryption algorithm GOST—28147 on the GPU with CUDA). *Informatsionnye tekhnologii*. 2011, no. 4, pp. 41—46 (in Russian).

12. **Lanconelli Open Systems**. URL: <http://www.lancos.com/index.html> (accessed 12.01.2015).

13. **Otladochnaya plata SK-iMX53-XC6SLX**. URL: <http://www.starterkit.ru/html/index.php?name=shop&op=view&id=76> (accessed 12.01.2015).

14. **Sarancha S. N.** Metodika opredeleniya parametrov apparatnogo generatora sluchainykh chisel, realizovannogo v PЛИS arkhitektury FPGA (Method of parameter's definition for the hardware random number generator implemented in the FPGA). *Avtomatizirovannye sistemy upravleniya i pribory avtomatiki, Vseukrainskii mezhdostvennyi nauchno-tekhnicheskii sbornik*. Kharkiv. 2011, iss. 157, pp. 89—94 (in Russian).

15. **FPGA and ASIC Design with HDL Coder and HDL Verifier**, URL: <http://www.mathworks.com/fpga-design/solutions.html> (accessed 12.01.2015).

16. **Proakis J. G., Salehi M.** *Communication systems engineering*. NJ.: Prentice-Hall, 2002. 801 p.

ГЕОИНФОРМАЦИОННЫЕ СИСТЕМЫ GEOINFORMATION SYSTEMS

УДК 528.91: 659.2

А. Н. Бешенцев, канд. геогр. наук, ст. науч. сотр., зав. лаб., e-mail: anbesh@gmail.com,
Байкальский институт природопользования СО РАН, г. Улан-Удэ

Геоинформационная система долгосрочного мониторинга природопользования

Представлены результаты создания и использования геоинформационной системы мониторинга хозяйственного освоения территории на примере бассейна озера Байкал. Описана методика создания картографических информационных ресурсов на основе ретроспективных топографических карт для исследования динамики географической среды. Представлен опыт создания картографического хранилища векторных данных и практический пример качественно-количественной оценки хозяйственного преобразования территории.

Ключевые слова: ГИС, динамика природопользования, ретроспективные карты, картографическое хранилище векторных данных, геоинформационный мониторинг

Введение

Современная информатизация территориальной деятельности общества способствует тому, что все большая часть географических исследований реализуется на основе картографических информационных ресурсов. Важное значение в обществе приобретают географические информационные системы [1], телекоммуникационные картографические сервисы, банки и базы геоданных. Современные приборы и методики оценки географической реальности обеспечивают формализацию геосвойств различной природы происхождения и интеграцию наук о Земле на базе геоинформационной технологии.

В настоящее время большинство государственных и научно-исследовательских организаций, связанных с системой территориального управления и планирования, формализовали значительные массивы тематических и топографических карт разных лет издания Корпуса военных топографов, ГУГиК, Генерального штаба СССР и Роскартографии. Эти карты создаются уже около 200 лет в единых картографических проекциях и системах координат, характеризуются единством картографируемых объектов и отображаемых параметров, сходством принятых классификаций, преемственностью методов составления и принципов генерализации. Они представляют собой разновременные модели физико-географического и социально-экономического состояния территории и служат исходными документами для метрической оценки долговременной динамики географической среды и трансформации природных ландшафтов. По этим причинам они являются

важными информационными ресурсами для получения нового геознания о хозяйственном освоении территории и надежным источником геоинформационного прогнозирования [2].

Постановка задачи исследования

Индикатором хозяйственной деятельности является динамика географических объектов и процессов, которая фиксируется разновременными картографическими и аэрокосмическими материалами посредством изменения их планово-высотных и субстанциональных характеристик. Мониторинг и исследование этих изменений имеют большое практическое значение, так как позволяют определить уровень антропогенного воздействия на природную среду, выявить негативные и позитивные стороны природопользования, сформулировать рекомендации по оптимизации для органов управления. Внедрение геоинформационной технологии позволяет автоматизировать процесс использования разновременных картографических ресурсов, повышает точность и оперативность изысканий, снижает уровень субъективизма, обеспечивает возможность манипулирования значительными объемами геоданных [3]. Разработка методологических основ геоинформационного мониторинга долговременных природных и социально-экономических процессов является важным исследовательским направлением. Создание ГИС, позволяющей хранить значительные массивы разновременных пространственных данных, анализировать их и получать новую информацию об изменениях природных ланд-

шафтов, отвечать на запросы пользователя и оперативно выдавать информацию в любой форме, является актуальной задачей.

Материалы и методы

В целях непрерывного отслеживания территориальных параметров хозяйственного освоения бассейна озера Байкал и негативных последствий природопользования в Байкальском институте природопользования (БИП) СО РАН разработана и внедрена геоинформационная система мониторинга природопользования (ГИСМП) на основе пакета ArcGIS (рис. 1).

ГИСМП представляет собой программно-управляемый комплекс периодической картографической регистрации объектов и процессов природопользования, позволяющий в интерактивном режиме оценивать и прогнозировать долговременную динамику хозяйственного использования территории и изменения природной среды региона [4]. Информационной основой ГИСМП является совокупность массивов разновременных топографических растровых и векторных карт 1896—1980 гг. издания. Первый временной срез представлен листами топографической карты, созданной Корпусом военных топографов в 1896—1914 гг. в масштабе 2 версты в 1 дюйме (1:84 000), в поперечно-цилиндрической проекции Гаусса. Для оценки пригодности ретроспективных карт как документов мониторинга выполнен анализ их достоверности и полноты содержания совместно с изучением географической литературы, дополнительных картографических материалов и статистических источников начала XX века. Для определения точности и принципов генерализации осуществлен сравнительный анализ ретроспективной карты с современной, сходной по назначению.

На основании анализа можно сделать вывод, что оцениваемые карты являются высокоинформативными источниками и подробно отображают физико-географическое состояние и систему природополь-

зования территории в начале XX века. Они являются важной метрической базой для мониторинга природопользования Байкальского региона, поскольку представляют результат первой геодезической съемки отображаемой части России.

В процессе автоматизации ретроспективных карт осуществляются послойная векторизация элементов содержания и геометрическая коррекция создаваемых слоев [5], а также наполнение таблиц атрибутов, которые формируются автоматически при экспорте в среду ArcGIS (см. таблицу). Минимизация геометрических искажений ретроспективной топоосновы представляет собой совокупность операций исправления векторных слоев на основе ряда геометрических преобразований сети регистрационных точек, однозначно установленных на современной и ретроспективной топоосновах. Для каждого листа формируется отдельный массив таких точек, в качестве которых поочередно были использованы геодезические пункты, отметки высот, центры населенных пунктов, устья рек, пересечения дорог, культовые объекты. Установлено, что наименьшую ошибку планового положения объектов дает преобразование по устьям рек. На листах с хорошо развитой речной сетью среднеквадратическая ошибка планового положения объектов уменьшалась до 0,2 мм, а вблизи регистрационных точек было полное совмещение.

В результате аффинного преобразования сети регистрационных точек выполняется координатная трансформация ретроспективных векторных слоев. Чем больше регистрационных точек, тем выше точность первичной коррекции. Для последующей коррекции на ретроспективной топооснове вновь указываются координаты положения регистрационных точек и выполняется трансформация посредством алгоритма "резиновый лист". В этом случае исправление геометрических искажений осуществляется вручную векторами смещения.

В результате создаются геоинформационные ресурсы, представляющие совокупность векторных слоев (shp-файлы) и однозначных таблиц атрибу-



Рис. 1. Функциональная структура ГИС

Методика создания и использования картографических информационных ресурсов для мониторинга природопользования

Процедуры мониторинга	Операции мониторинга
Редакционно-подготовительные работы	Оценка картографических материалов Разработка редакционных указаний
Формализация листов ретроспективных карт	Сканирование листов Создание растровых покрытий по параметрам исходной картографической проекции Векторизация элементов содержания
Создание ретроспективных объектных слоев	Экспорт векторных слоев в среду ArcGIS Склейка объектных слоев Редактирование слоев Создание топологии
Геометрическая коррекция ретроспективных объектных слоев	Проецирование современной топоосновы по параметрам исходной картографической проекции Создание сети регистрационных точек Координатная трансформация ретроспективных слоев по сети регистрационных точек Оценка точности преобразования Редактирование слоев с помощью алгоритма "резиновый лист" Обновление топологии ретроспективных слоев
Наполнение таблиц атрибутов ретроспективных слоев	Разработка системы классификации и кодирования объектов природопользования Ввод атрибутивных данных по слоям
Создание объектных слоев динамики природопользования	Совмещение ретроспективных и современных векторных слоев по необходимым временным срезам Устранение осколочных полигонов Обновление топологии Редактирование таблиц атрибутов
Геоинформационное картографирование динамики природопользования	Создание ГИС-проекта Составление карт динамики природопользования Пространственная оценка динамики Ранжирование территории по степени динамики
Геоинформационное моделирование динамики природопользования	Пространственное моделирование динамики (геометрическое, проекционное, масштабное, сетевое анализ, буферизация) Субстанциональное моделирование динамики (математическое, семиотическое) Моделирование посредством запросов
Представление результатов мониторинга	Инвентаризационные карты динамики объектов природопользования Синтетические карты взаимосвязей динамики объектов природопользования Аналитические карты зонирования и районирования территории во взаимосвязи с экологическими последствиями динамики природопользования Прогнозные карты сценариев развития природопользования Системы запросов динамики природопользования Графические и табличные материалы

тов (dbf-таблицы), размещенных в среде ArcGIS, и регистрирующие физико-географические параметры объектов природопользования исследуемой территории в начале XX века. Информационная структура базы данных ресурсов разработана на основе элементов содержания топографических карт: гидрография; населенные пункты; дорожная сеть; растительность; грунты; промышленные, сельскохозяйственные и социальные объекты; рельеф. Ресурсы имеют математическую основу, классификации элементов содержания и точность, аналогичные параметрам современных геоизображений, и могут быть использованы при реализации всех аналитических операций программной среды. Для отображения других временных срезов использованы листы топографической карты масштаба 1:100 000 издания 2002 г. (Роскартография) и издания 1962 г. (ГУГиК).

Обсуждение результатов

В результате совмещения разновременных векторных слоев объектов природопользования и природных ландшафтов создано картографическое хранилище векторных данных [4]. Для каждого атрибута создана avl-легенда, позволяющая моментальное формирование визуального представления геоданных. База данных картографического хранилища содержит атрибуты совмещенных слоев, а также описывает планово-высотные метрические параметры пространственных объектов и является информационной основой для междисциплинарных научных исследований бассейна озера Байкал.

Методика использования хранилища представляет собой совокупность последовательных операций программной среды и заключается в формировании множества картографических представле-



Рис. 2. Схема использования картографического хранилища векторных данных

ний, в которых меняются лишь элементы содержания и способы картографического изображения объектов, а топологическая связность массивов данных сохраняется и не зависит от их комбинирования. Применение указанного подхода обеспечивает топологическую целостность данных и удобство любых преобразований как в интерактивном режиме, так и в автоматическом (рис. 2). Такое управляемое картографирование оптимизирует решение традиционных задач, связанных с выбором математической основы и компоновки карт, позволяет оперативную смену проекций, свободное масштабирование, обеспечено средствами визуализации и алгоритмами автоматической генерализации, а составление и оформление карт, подготовка к изданию реализуются на одном рабочем месте.

Важным механизмом моделирования пространственной информации является интерактивная работа с ГИС посредством геоинформационных запросов — совокупности последовательных операций программной среды, формируемых пользователем и отвечающих заданным субстанциональным и топологическим критериям. С помощью пакета ArcGIS можно выполнять различные запросы к картогра-

фическому хранилищу от простого выбора объектов для их идентификации, определения местоположения по критериям выборки до анализа пространственных взаимосвязей между различными объектами.

Например, на запрос "Создать модель динамики распаханности природных ландшафтов в XX в." первым шагом его реализации является выбор объектов моделирования (пашня и ландшафты). Затем устанавливаются топологические отношения между этими слоями и выполняется автоматизированная метрическая оценка образовавшихся полигонов. В результате этих операций создается новый слой, регистрирующий пространственное и количественное состояние моделируемого явления (в данном случае распаханность ландшафтов). При этом картографическая модель отображает плано-высотную характеристику явления. Графическая модель представляет количественную характеристику ареалов в виде метрических параметров и числа участков. Математическая модель представляет собой таблицу, содержащую метрические параметры распаханности ландшафтов.

Заключение

Разработка методики автоматизации ретроспективных карт позволяет говорить о создании нового геоинформационного ресурса [6], обеспечивающего надежный картографический мониторинг и прогнозирование регионального природопользования. Значительный объем содержательной информации этих ресурсов позволяет их использовать при изучении долговременных природных (опустынивание, заболачивание и др.) и социально-экономических (добыча полезных ископаемых, миграция населения и др.) процессов. Кроме того, они фиксируют состояние земной поверхности в период строительства и ввода в эксплуатацию Транссиба и являются основным материалом при исследованиях воздействия магистрали на географическую среду региона.

Созданная ГИС является программно-техническим комплексом для регистрации и оценки динамики объектов природопользования, обеспечивающим междисциплинарную интеграцию научных исследований хозяйственного освоения бассейна озера Байкал. Она соответствует международным требованиям к аналогичным продуктам, открыта для дополнения пространственными и атрибутивными данными, методически проста и управляема.

Внедрение ГИС в органы управления и планирования соседствующих государств (Российская Федерация и Монголия) будет способствовать формированию единого информационного пространства и способствовать повышению эффективности межгосударственного природоохранного и социально-экономического взаимодействия.

Список литературы:

1. **Бугаевский Л. М., Цветков В. Я.** Геоинформационные системы. М.: Златоуст, 2000. 224 с.

2. **Кудж С. А.** Исследование окружающего мира методами геоинформатики // Вестник МГТУ МИРЭА "MSTU MIREA HERALD". 2013. № 1 (1). С. 95–102.

3. **Цветков В. Я.** Ситуационное моделирование в геоинформатике // Информационные технологии. 2014. № 6. С. 64–69.

4. **Бешенцев А. Н.** Информационная концепция картографического мониторинга геосистем: Дисс. ... д-ра геогр. наук. Иркутск, 2013. 332 с.

5. **Бешенцев А. Н.** Картографический мониторинг природопользования // Геодезия и картография. 2011. № 3. С. 14–18.

6. **Савиных В. П., Цветков В. Я.** Геоданные как системный информационный ресурс // Вестник Российской Академии Наук. 2014. Т. 84, № 9. С. 826–829.

A. N. Beshentsev, Head of the GIS-laboratory, e-mail: anbesh@gmail.com,

Baikal institute of nature management Siberian branch of Russian, Academy of Sciences, g. Ulan-Ude

The Geoinformation System of Long-Term Monitoring of Nature Use

The article contains the results of the development and use of geographic information system for long-term monitoring of nature use of the territory by the example of the Lake Baikal basin. As an information base used arrays of multi-temporal topographic maps. To assess the suitability of these materials for a long-term monitoring of documents made a comprehensive analysis of the topographic map sheets, scale 1:84,000, created in 1896–1914 years. Described the technique of automating, geometric correction and creation cartographic information resources, based on them serving as a reference document for monitoring territory. Submitted by the experience of creating a multi-temporal Cartographical Data Warehouse. The technique of works with Data Warehouse, is the possibility of forming a plurality of cartographic representations, which changes the way of cartographic images of objects, but the integrity and coherence of the topological data sets does not depend on their combination. Presented the mechanism of interactive works with Data Warehouse through the geoinformation requests. Presented a practical example of qualitative and quantitative assessment of the transformation of the territory.

Keywords: GIS, the dynamics of nature use, retrospective maps, cartographical data warehouse, geographic information monitoring

References

1. **Bugaevskij L. M., Cvetkov V. Ja.** *Geoinformacionnye sistemy*. M.: Zlatoust, 2000. 224 p.

2. **Kudzh S. A.** Issledovanie okruzhajushhego mira metodami geo-informatiki. *Vestnik MGTU MORJeA "MSTU MIREA HERALD"*. 2013, no. 1 (1), pp. 95–102.

3. **Cvetkov V. Ja.** Situacionnoe modelirovanie v geoinformatike. *Informacionnye tehnologii*. 2014, no. 6, pp. 64–69.

4. **Beshentsev A. N.** *Informacionnaja koncepcija kartograficheskogo monitoringa geosistem*: Dis. doktora geogr. nauk. Irkutsk. 2013. 332 p.

5. **Beshentsev A. N.** Kartograficheskij monitoring prirodopozovaniya. *Geodezija i kartografija*. 2011, no. 3, pp. 14–18.

6. **Savinyh V. P., Cvetkov V. Ja.** Geodannye kak sistemnyj informacionnyj resurs. *Vestnik Rossijskoj Akademii Nauk*. 2014, vol. 84, no. 9, pp. 826–829.

ЖУРНАЛ В ЖУРНАЛЕ



**НЕЙРОСЕТЕВЫЕ
ТЕХНОЛОГИИ**

№ 8

АВГУСТ

2015

Главный редактор:

ГАЛУШКИН А.И.

Редакционная коллегия:

АВЕДЬЯН Э.Д.
БАЗИЯН Б.Х.
БЕНЕВОЛЕНСКИЙ С.Б.
БОРИСОВ В.В.
ГОРБАЧЕНКО В.И.
ЖДАНОВ А.А.
ЗЕФИРОВ Н.С.
ЗОЗУЛЯ Ю.И.
КРИЖИЖАНОВСКИЙ Б.В.
КУДРЯВЦЕВ В.Б.
КУЛИК С.Д.
КУРАВСКИЙ Л.С.
РЕДЬКО В.Г.
РУДИНСКИЙ А.В.
СИМОРОВ С.Н.
ФЕДУЛОВ А.С.
ЧЕРВЯКОВ Н.И.

**Иностранные
члены редколлегии:**

БОЯНОВ К.
ВЕЛИЧКОВСКИЙ Б.М.
ГРАБАРЧУК В.
РУТКОВСКИЙ Л.

Редакция:

БЕЗМЕНОВА М.Ю.
ГРИГОРИН-РЯБОВА Е.В.
ЛЫСЕНКО А.В.
ЧУГУНОВА А.В.

Осипов В. Ю.

Ассоциативно-пространственная адресация к памяти рекуррентных нейронных сетей 631

Лёзина И. В.

Сравнительный анализ погрешности прогноза биржевых индексов методами экспоненциального сглаживания и нейросетевого моделирования 637

В. Ю. Осипов, доктор техн. наук, проф., e-mail: osipov_vasily@mail.ru,
Санкт-Петербургский институт информатики и автоматизации РАН

Ассоциативно-пространственная адресация к памяти рекуррентных нейронных сетей

В интересах расширения когнитивных возможностей рекуррентных нейронных сетей исследуются подходы к наделению их ассоциативно-пространственной адресацией к памяти. Наличие дополнительной пространственной адресации к памяти в рекуррентных нейронных сетях позволяет им управлять ассоциативными взаимодействиями обрабатываемых сигналов в зависимости от текущих состояний слоев. Предлагается новый метод такой ассоциативно-пространственной адресации, снижающий сложность ее реализации. Он базируется на усовершенствованной модели рекуррентной нейронной сети с управляемыми синапсами. Приводятся результаты моделирования, подтверждающие целесообразность и справедливость предлагаемых решений.

Ключевые слова: рекуррентная нейронная сеть, ассоциативно-пространственная адресация к памяти, когнитивная обработка информации

Введение

Основами всей современной когнитивной обработки информации в рекуррентных нейронных сетях (РНС) выступают ассоциативное запоминание и ассоциативный вызов из памяти сигналов. Однако этого, недостаточно для наделения РНС широкими когнитивными возможностями [1, 2]. Желательно, чтобы РНС имели, помимо ассоциативной, также пространственную адресацию к памяти в зависимости от текущих состояний слоев сети.

В последние годы некоторые результаты в этом направлении совершенствования РНС достигнуты. Они опираются на известные положения по управлению синапсами [3–5] и пространственной обработкой сигналов в нейронных сетях [6–10].

В работах [11–13] предложены методы когнитивной обработки информации в двухслойной РНС с управляемыми синапсами. Нейронам такой сети свойственны три состояния: ожидания, возбуждения и временной невосприимчивости после возбуждения. Каждый нейрон одного слоя в общем случае связан синапсами со всеми нейронами другого слоя. Нейроны одного слоя связаны друг с другом не имеют. Время задержки единичных образов в образуемых контурах этой сети меньше времени невосприимчивости нейронов после возбуждения. Подаваемые в РНС сигналы (в общем случае цветные динамические воздействия) предварительно раскладывают на составляющие в базисе, согласованном с входным слоем. При этом каждую из них преобразуют в последовательность единичных образов (импульсов) с частотой повторения как функцией от амплитуды составляющей. Сигналы в сети представляют в виде последовательных совокупностей единичных образов (СЕО) в соответствии с заданными правилами. При передаче СЕО от слоя к слою осуществляют их пространственные сдвиги. За счет цикличности обработки этих СЕО с простран-

ственными сдвигами каждый слой РНС разделяется на конечное число равных полей. Сеть наделается прозрачной логической структурой. Результаты распознавания запоминают на элементах сети. Причем за счет частичного отражения СЕО от принимающих слоев частично стирают устаревшую информацию с синапсов [14, 15]. В итоге обрабатываемые СЕО продвигаются вдоль слоев сети в заданных направлениях. Они успешно ассоциируются друг с другом и с запомненными сигналами с учетом обратных результатов распознавания. Приоритетность коротких связей между нейронами взаимодействующих слоев обеспечивает однозначное соответствие между входом и выходом РНС. В качестве результатов обработки используют последовательные СЕО на выходном слое после обратного преобразования их в исходные сигналы. Пример такой РНС с простой спиральной структурой слоев показан на рис. 1.

Через воздействие на синапсы, в зависимости от текущих состояний слоев, в работах [16, 17] реко-

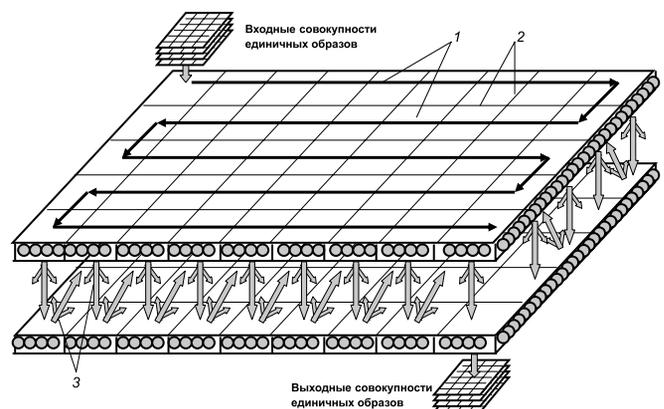


Рис. 1. Рекуррентная нейронная сеть со спиральной структурой слоев:

1, 3 — направления продвижения СЕО вдоль слоев и между ними; 2 — линии разбивки слоев на логические поля

мендовано изменять формы и поворачивать поперечные сечения расходящихся и сходящихся единичных образов в РНС. Это позволило наделить РНС не только ассоциативной, но и пространственной адресацией к ее памяти. На основе ее были предложены решения по формированию в РНС трех сигнальных систем [18, 19], обеспечивающих глубокую обработку информации и взаимодействие с внешним миром.

Однако в известных работах [16—19] зависимости параметров ассоциативно-пространственного обращения к памяти от текущих состояний слоев РНС оставлены без должного внимания. Кроме этого, реализация дополнительной пространственной адресации к памяти в известной РНС осуществляется путем воздействий на синапсы со стороны блока управления. В интересах этого блок управления считывает текущие состояния слоев и вырабатывает соответствующие решения. Однако технически реализовать это управление в РНС больших размеров довольно сложно.

Требуется уточнить модель РНС с рассматриваемыми свойствами, развить методы ассоциативно-пространственной адресации к памяти (АПАП) и взаимодействия сигналов, снизить сложность реализации.

Модели РНС с АПАП

Согласно модели [14—19] различают прямые $w_{ij}(t)$ и обратные $w_{ij}^*(t)$ веса (проводимости) синапсов, которые определяются как

$$\begin{aligned} w_{ij}(t) &= k_{ij}(t) \cdot \beta(r_{ij}(t)) \cdot \eta_{ij}(r_{ij}(t)); \\ w_{ij}^*(t) &= k_{ij}^*(t) \cdot \beta(r_{ij}(t)) \cdot \eta_{ij}(r_{ij}(t)), \end{aligned} \quad (1)$$

где $k_{ij}(t) = 1 - \exp(-\gamma \cdot g_{ij}(t))$, $k_{ij}^*(t) = 1 - \exp(-\gamma(g_{ij}(t) - g_0))$ — прямой и обратный весовые коэффициенты; γ — постоянный коэффициент; $g_{ij}(t)$ — число запомненных воздействий единичных образов на синапс на момент времени t ; $(g_{ij}(t) - g_0)$ — число запомненных воздействий единичных образов, которые могут быть стерты с синапса, $g_{ij}(t) \geq g_0$; $\beta(r_{ij}(t))$ — функция ослабления расходящихся единичных образов; $\eta_{ij}(r_{ij}(t))$ — функция ослабления сходящихся единичных образов; $r_{ij}(t)$ — удаленность связываемых через синапсы нейронов. При формализации запоминания и стирания результатов воздействий на синапсы единичных образов величина $g_{ij}(t)$ равна

$$g_{ij}(t) = g_{ij}(t - \Delta t) \pm \Delta g^{(*)}(I_{ij}^{(*)}(t)), \quad (2)$$

$\Delta g(I_{ij}(t))$, $\Delta g^*(I_{ij}^*(t))$ — прирост и снижение $g_{ij}(t)$ в зависимости от проходящего через синапсы тока $I_{ij}(t)$ в прямом и тока $I_{ij}^*(t)$ в обратном направлениях. В частном случае $\Delta g(I_{ij}(t)) = A I_{ij}(t)$, а $\Delta g^*(I_{ij}^*(t)) = D \cdot I_{ij}^*(t) \leq (g_{ij}(t - \Delta t) - g_0)$, где A , D — константы.

Каждая из связей (синапсов) имеет свои значения функций ослабления $\beta_{ij}(r_{ij})$, $\eta_{ij}(r_{ij})$, соответственно, расходящихся и сходящихся единичных образов.

При возбуждении i -го нейрона на его выходе формируется заряд Q_i . Переносимый часть этого заряда прямой ток $I_{ij}(t)$, протекающий через ij -й синапс, выражается как $I_{ij}(t) = U_i(t)w_{ij}(t)$, где U_i — потенциал на выходе i -го нейрона передающего слоя. Полагается, что на входе возбуждаемого j -го нейрона принимающего слоя потенциал близок к нулю. Это допущение характерно для всех известных импульсных нейронных сетей [3, 4].

Суммарному заряду, переносимому от i -го нейрона к возбуждаемым j -м нейронам за отводимое время T , свойственно условие

$$\sum_{j=1}^n \int_0^T I_{ij}(t) dt = Q_i = \text{const.}$$

Изменение в расходящемся пучке любого $I_{ij}(t)$ через функцию $\beta(r_{ij}(t))$, от которой зависит $w_{ij}(t)$, приводит к пространственному перераспределению переносимых зарядов. При этом относительные токи $\delta I_{ij}(t)$ через синапсы расходящегося пучка равны

$$\delta I_{ij}(t) = I_{ij}(t) / \sum_{j=1}^n I_{ij}(t) = w_{ij}(t) / \sum_{j=1}^n w_{ij}(t). \quad (3)$$

Суммарный ток $I_{j\Sigma}(t)$ на входе возбуждаемого j -го нейрона без учета частичного отражения сигналов равен $I_{j\Sigma}(t) = \sum_{i=1}^n U_i(t)w_{ij}(t)$. Для того, чтобы j -й нейрон возбудился, этот ток должен перенести некоторый конечный заряд

$$Q_{jR} = \sum_{i=1}^n \int_0^T U_i(t)w_{ij}(t) dt.$$

В формировании заряда Q_{jR} участвуют токи $I_{ij}(t)$ от различных нейронов. Изменение любого из них через функцию $\eta_{ij}(r_{ij}(t))$ в составе $w_{ij}(t)$ приводит к перераспределению вкладов переносимых зарядов в Q_{jR} .

Обратный ток $I_{ij}^*(t)$ через ij -й синапс выражается как

$$I_{ij}^*(t) = I_{j\Sigma}(t)K_j(t)w_{ij}^*(t) / \sum_{i=1}^n w_{ij}^*(t),$$

$K_j(t) = I_{j\Sigma}^*(t)/I_{j\Sigma}(t)$ — коэффициент отражения единичных образов от принимающих нейронов; $I_{j\Sigma}^*(t)$ — суммарный ток, отражаемый от j -го нейрона.

Результирующий ток, протекающий через входное сопротивление j -го нейрона, равен $I_{j\text{вх}}(t) = I_{j\Sigma}(t) - I_{j\Sigma}^*(t) = I_{j\Sigma}(t)(1 - K_j(t))$. В случаях, когда ток $I_{j\text{вх}}(t)$ или соответствующий накопленный заряд больше порогового значения, j -й нейрон возбуждается. На его выходе формируется единичный образ (импульс), и нейрон переходит в состояние невосприимчивости. В состоянии невосприимчиво-

сти каждый нейрон находится одинаковое время, которое больше, чем время задержки единичных образов в двухслойных контурах сети.

Изменяя $r_{ij}(t)$ и другие параметры в функциях $\beta(r_{ij}(t))$, $\eta_{ij}(r_{ij}(t))$ в зависимости от текущих состояний слоев РНС, можно осуществлять пространственные сдвиги обрабатываемых сигналов, формировать новые и поворачивать формы поперечных сечений расходящихся и сходящихся единичных образов. В результате обеспечивается возможность ассоциативно-пространственной адресации к памяти и соответствующего взаимодействия сигналов в РНС. С физической точки зрения это достижимо прямым изменением проводимостей синапсов. Такие ассоциативно-пространственная адресация к памяти и взаимодействие сигналов справедливы, если потенциалы на входах возбуждаемых нейронов невелики и практически не зависят от потенциалов на их выходах.

В случаях, когда потенциалы на входах нейронов могут быть сравнимы с потенциалами на их выходах, ситуация существенно меняется. Еще в работе [20] замечено, что в биологических нейронах при их возбуждении токи могут протекать встречно по дендритам и формировать на выходе их значительные потенциалы.

С учетом этого в предлагаемой модели РНС с ассоциативно-пространственной адресацией к памяти (АПАП) прямой ток $I_{ij}(t)$, протекающий через ij -й синапс, должен быть равен $I_{ij}(t) = (U_i(t) - U_{jR}(t))w_{ij}(t)$, где $U_{jR}(t)$ — потенциал на входе j -го нейрона. В этом случае

$$Q_i = \sum_{j=1}^n \int_0^T (U_i(t) - U_{jR}(t))w_{ij}(t)dt = \text{const.} \quad (4)$$

Относительное значение $\delta I_{ij1}(t)$ прямого тока через ij -й синапс в расходящемся пучке единичных образов имеет вид:

$$\begin{aligned} \delta I_{ij1}(t) &= \\ &= (U_i(t) - U_{jR}(t))w_{ij}(t) / \sum_{j=1}^n (U_i(t) - U_{jR}(t))w_{ij}(t). \end{aligned} \quad (5)$$

В отличие от выражения (3), свойственного всем известным моделям импульсных нейронных сетей, отношение (5) учитывает различие потенциалов на синапсах в расходящемся пучке. Аналогично можно получить выражение для относительных токов $\delta I_{ij2}(t)$ применительно к сходящемуся пучку. Оно также зависит от различия потенциалов, приходящихся на отдельные синапсы.

Из анализа выражений (3)—(5) видно, что при изменениях $\Delta U_{ij}(t) = U_{iT}(t) - U_{jR}(t)$ в зависимости от текущих состояний слоев в сети происходит пространственное перераспределение переносимых зарядов в расходящихся и сходящихся пучках. Изменение $\Delta U_{ij}(t)$ равносильно приросту или уменьшению проводимостей синапсов.

Реальные значения $\Delta g(I_{ij}(t))$, $\Delta g^*(I_{ij}^*(t))$ в выражении (2) и изменения проводимостей (весов) (1)

синапсов должны в этом случае определяться в зависимости от перераспределяемых токов в расходящихся и сходящихся пучках единичных образов.

Учет этого эффекта, ранее не принимаемого во внимание, позволяет по-новому взглянуть на пространственную обработку сигналов в РНС. Как в биологических нейронах, так и в их электрических моделях потенциалы на входах могут быть сравнимы с потенциалами на выходах. В результате РНС способна в широких пределах осуществлять пространственное перераспределение переносимых зарядов в расходящихся и сходящихся пучках единичных образов.

Несмотря на это, необходимость в блоке управления, как в работах [16—19], не отпадает. Однако по отношению к синапсам в предлагаемом решении функции этого блока могут быть сведены лишь к неоперативной корректировке через них пространственных параметров расходящихся и сходящихся единичных образов. Формально через блок управления можно существенно изменять структуру РНС в процессе ее функционирования, корректировать базовые функции ослабления расходящихся и сходящихся единичных образов, варьировать порогом возбуждения нейронов и внутренним временем сети.

Базовые функции ослабления сигналов в РНС с АПАП

В качестве базовых функций ослабления сигналов в ряде случаев могут выступать известные радиальные функции Гаусса, Габора, прямая и обратная мультикватратичные и др. [2, 3, 7]. В частности, для расходящихся пучков применима модифицированная функция Габора, определенная только в положительной области. Ее выражение без учета зависимости параметров от времени имеет вид:

$$\beta_{ij}(x_{ij}, y_{ij}) = \exp\left(-\frac{X_{ij}^2 + \gamma^2 Y_{ij}^2}{2\sigma^2}\right) \cos^2\left(\frac{2\pi}{\lambda} X_{ij}\right), \quad (6)$$

где $X_{ij} = x_{ij}\cos\varphi + y_{ij}\sin\varphi + c$; $Y_{ij}(t) = -x_{ij}\sin\varphi + y_{ij}\cos\varphi + f$; φ — угол поворота; c, f — параметры сдвигов по осям X, Y ; $\gamma = \sigma_x/\sigma_y$; σ_x, σ_y — средние квадратические отклонения по осям X, Y ; $\sigma_x = \sigma$; λ — пространственная длина волны. Отличие (6) от классической функции Габора в том, что $\cos(\cdot)$ в правой части возведен в квадрат. Неплохо ведет себя в качестве $\beta(r_{ij}(t))$ также обратная мультикватратичная функция. Эти же функции применимы также в качестве $\eta_{ij}(r_{ij}(t))$.

Задание для РНС базовых функций ослабления единичных образов в расходящихся и сходящихся пучках с конкретными параметрами позволяет разделить сеть не только определенной структурой, но и своеобразной "наследственной" информацией. Именно на основе ее сеть будет обучаться и действовать. Однако однозначных подходов к определению

нию целесообразных базовых функций ослабления для такой РНС пока нет. Формы их могут зависеть от реализуемых в РНС пространственных сдвигов и направлений продвижения СЕО вдоль слоев, количества и типов ассоциируемых в сети сигналов, их реальной взаимосвязи и других факторов.

Опираясь на результаты работ [16–19] и учитывая, что основные ассоциативные взаимодействия единичных образов реализуются не внутри нейросетевых каналов, по которым продвигаются СЕО, а между каналами, можно определиться с исходной ориентацией этих функций в пространстве слоев сети. Целесообразно ориентировать базовые функции ослабления в плоскости слоев перпендикулярно нейросетевым каналам. Причем для функции (6) провалы должны приходиться на каналы с такими же направлениями продвижения СЕО вдоль них, что и у канала, содержащего нейрон, порождающий расходящийся пучок единичных образов. Это позволяет не только обеспечить хорошее ассоциативное взаимодействие сигналов в сети, но и расширить возможности ее памяти, снизить избыточность запоминания информации.

В ряде случаев для определения таких целесообразных функций сначала можно задаться одной из известных радиальных функций. Затем следует приравнять в $w_{ij}(t) = k_{ij}(t) \cdot \beta_{ij}(t) \cdot \eta_{ij}(t)$ все весовые коэффициенты $k_{ij}(t)$ единице. Применяя известные методы [3, 4], можно обучить сеть решению некоторых опорных задач, варьируя только значениями функций ослабления $\beta_{ij}(t)$ расходящихся и $\eta_{ij}(t)$

сходящихся единичных образов. Закрепляя найденные $\beta_{ij}(t)$, $\eta_{ij}(t)$ и снимая ограничения с $k_{ij}(t)$, позволяя этим коэффициентам изменяться по рассмотренным выше правилам, получаем РНС с "наследственной информацией", способную к самообучению.

Результаты ассоциативно-пространственной адресации к памяти РНС

Для того чтобы показать, как могут перераспределяться относительные токи и переносимые заряды в расходящихся пучках единичных образов, сделаем допущения. Пусть $k_{ij}(t)$ и $\eta_{ij}(r_{ij}(t))$ в выражениях (1) принимают значения единицы. Тогда вместо $w_{ij}(t)$ в выражениях (3), (5) можно подставить $\beta(r_{ij}(t))$. Используем в качестве $\beta(r_{ij}(t))$ модифицированную функцию Габора (6). В результате согласно (3) и (5) получим относительные распределения токов в пространстве. Примеры таких сглаженных распределений представлены на рис. 2, а–г.

На рис. 2, а показано распределение согласно (3) для (6) без учета индивидуальной разности потенциалов на синапсах в расходящемся пучке. Рисунки, 2, б–г отражают возможные изменения исходного распределения (рис. 2, а) с таким учетом. Из анализа этих распределений видно, что если потенциалы на входах нейронов принимающего слоя могут быть сравнимы с потенциалами на выходах возбужденных нейронов, то сеть способна в широких пределах изменять адресацию к памяти в зависимости от текущих состояний слоев. Это под-

тверждают существенные изменения (рис. 2, б–г) исходной формы (рис. 2, а) пространственного распределения относительных токов. Причем обеспечивается возможность адресации к памяти не только по направлению, но и по конкретным областям, даже по отдельным элементам. Это справедливо не только для изменения распределения относительной базовой функции (6), но и других. При таком ассоциативно-адресном обращении к памяти сохраняются все возможности селективного запоминания и вызова сигналов, а также стирания устаревшей информации за счет частичного отражения единичных образов от принимающих слоев. Обеспечивается устойчивое функционирование РНС при обработке потоков сигналов в реальном времени. Предельное количество информации C в битах, запоминаемое на синапсах такой сети, равно

$$C = -\frac{n^2}{2} \ln \left(1 - \frac{2W_{\max}}{n^2 \beta \bar{\eta}} \right), \quad (7)$$

где n — число нейронов в каждом из двух слоев сети; W_{\max} — максимальный суммарный вес синапсов сети, при ко-

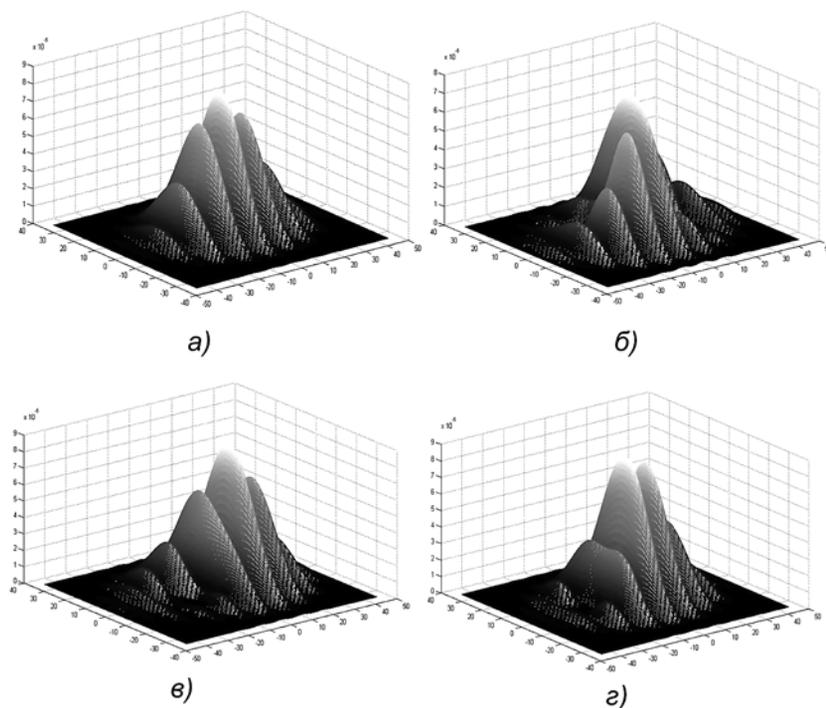


Рис. 2. Примеры пространственного распределения относительных токов в расходящихся пучках единичных образов: а — распределение без учета индивидуальной разности потенциалов на синапсах; б–г — распределения с таким учетом

тором число запоминаемых новых единичных образов равно числу стираемых устаревших образов; $\bar{\beta}$, $\bar{\eta}$ — средние значения функций $\beta(r_{ij}(t))$, $\eta_{ij}(r_{ij}(t))$. Выражение (7) можно получить из анализа выражений (1), (2). Оценка (7) емкости памяти РНС не противоречит результатам работ [2, 21, 22]. В то же время она расширяет взгляды на предельные возможности памяти РНС с учетом стирания устаревшей информации.

Такое обращение к памяти и, соответственно, управление ассоциативным взаимодействием сигналов в РНС позволяет реализовывать условные переходы от обработки одних сигналов к другим.

В случае чистого запоминания входных СЕО они проходят по РНС без искажений, оставляя в ней след о своем появлении в виде изменения весов синапсов с учетом текущих состояний слоев. При запоминании СЕО в РНС со спиральной структурой слоев сети (см. рис. 1) каждая сигнальная конструкция (СЕО, их группа) наиболее прочно связывается с другими, когда при продвижении вдоль слоев расстояние между ними становится минимальным. В итоге устанавливаются прочные связи различных сигнальных конструкций друг с другом. Когда одновременно с запоминанием входных СЕО осуществляется вызов из памяти связанных сигналов, также учитываются текущие состояния слоев. Причем в зависимости от этих состояний, определяющих текущие значения функций ослабления расходящихся и сходящихся единичных образов, возможно формирование различных последовательностей, вызываемых из памяти сигналов. Так можно вызывать из памяти РНС сигнальные конструкции перед или после вызывающего сигнала, дополняя его. Например, пусть РНС помнит сигнальную конструкцию "нейроны связаны синапсами", где каждой букве соответствует одна или несколько СЕО. Тогда в зависимости от текущих состояний ее слоев одним и тем же сигналом "связаны" из памяти РНС могут быть вызваны: "нейроны" или (и) "синапсами", другие связанные сигналы. Также сеть может не отреагировать на входной сигнал. Причем группировка этих конструкций может осуществляться из различных, но связанных между собой, запомненных сигналов.

Допустимы различные варианты структуры РНС, позволяющей переставлять сигнальные конструкции в предложениях с использованием ассоциативно-пространственной адресации к памяти. В частности, помимо структуры, показанной на рис. 1, возможен вариант, приведенный на рис. 3. Основным выходом в такой РНС выступает выход 2, а не выход 1.

В соответствии с этой структурой (рис. 3) сеть может многократно прокручивать внутри себя входную или вызванную из памяти последовательность СЕО и формировать из нее новые последовательности с переставленными сигнальными конструкциями относительно исходного предложения.

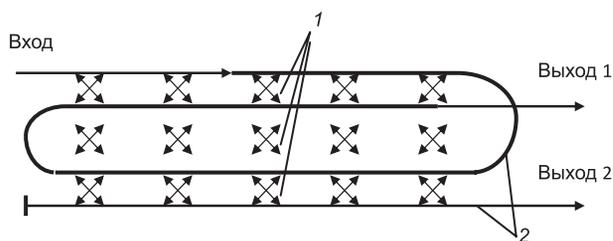


Рис. 3. Структура рекуррентной нейронной сети на уровне нейросетевых каналов, реализующая перестановку сигнальных конструкций в предложениях:

1 — основные ассоциативные взаимодействия сигналов в РНС; 2 — нейросетевые каналы в РНС, по которым продвигаются совокупности единичных образов

Ослабляя одни и усиливая другие связи (1, рис. 3) в зависимости от текущих состояний слоев, РНС может на выходе 2 формировать различные порядки вызываемых из памяти сигналов. Например, пусть РНС держит в своей оперативной памяти на самих нейронах сигнальную конструкцию "веса синапсов это их проводимости". Тогда из нее РНС может сформировать "проводимости синапсов это их веса" и другие конструкции.

Заключение

Таким образом, опираясь на полученные результаты исследования, есть основания утверждать, что не только предлагаемым искусственным, но и биологическим РНС свойственна не просто ассоциативная, а ассоциативно-пространственная адресация к памяти. Рекуррентные сети в зависимости от текущих состояний слоев способны осуществлять пространственное перераспределение переносимых зарядов в расходящихся и сходящихся пучках единичных образов. Имеет место сочетание ассоциативного и пространственного методов обращения к памяти РНС. Такой подход позволяет снять многие ограничения, препятствующие пониманию процессов когнитивной обработки информации в РНС. За счет такой адресации предоставляется возможность оперативно регулировать загрузку памяти второй сигнальной системы РНС входным потоком. Этим обеспечивается глубокая обработка информации в РНС. Также можно оперативно связывать вторую сигнальную систему РНС с третьей и управлять выходом сети, выдавать не все результаты обработки, а только "осознанные".

При реализации синапсов РНС на мемристорах и использовании электрических моделей нейронов с существенной зависимостью потенциалов на входах от выходных уровней можно снизить сложность создания таких сетей. Рекомендуется в известных моделях РНС учитывать текущие потенциалы на входах возбуждаемых нейронов и осуществлять пространственное перераспределение переносимых зарядов в зависимости от состояний слоев.

Предложенные решения развивают методы когнитивной обработки информации в РНС в части

ассоциативно-пространственной адресации к памяти и снижения сложности реализации. Они могут быть использованы при разработке перспективных нейросетевых когнитивных машин с высокоскоростной параллельной обработкой различных сигналов в условиях высокой неопределенности ситуаций и событий.

Список литературы

1. **Haikonen Pentti O. A.** The Role of Associative Processing in Cognitive Computing // *Cognitive Computing*. 2009. N. 1. P. 42–49.
2. **Palm G.** Neural associative memories and sparse coding // *Neural Networks*. 2013. N. 37. P. 165–171.
3. **Haykin S.** *Neural Networks and Learning Machines*. Third Edition. New-York: Prentice Hall, 2008. 936 p.
4. **Galushkin A. I.** *Neural Networks Theory*. Berlin Heidelberg: Springer-Verlag, 2007. 396 p.
5. **Liaw J.-S., Berger T. W.** Dynamic Synapse for Signal Processing in Neural Networks. US Patent 6643627 B2, Nov. 4, 2003.
6. **Cover T. M.** Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern recognition // *IEEE Transactions on Electronic Computers*. 1965. June. P. 326–334.
7. **Jeong S., Lee M.** Adaptive object recognition model using incremental feature representation and hierarchical classification // *Neural Networks*. 2012. N. 25. P. 130–140.
8. **Chartier S., Giguere G., Langlois D.** A new bidirectional heteroassociative memory encompassing correlational, competitive and topological properties // *Neural Networks*. 2009. N. 22. P. 568–578.
9. **Fukushima K.** Artificial vision by multi-layered neural networks: Neocognitron and its advances // *Neural Networks*. 2013. N. 37. P. 103–119.
10. **Kohonen T.** Essentials of the self-organizing map // *Neural Networks*. 2013. N. 37. P. 52–65.
11. **Осипов В. Ю., Осипова М. В.** Способ и устройство интеллектуальной обработки информации в нейронной сети. Патент на изобретение RU 2413304, опубл. 27.02.2011, бюл. № 6.
12. **Осипов В. Ю.** Способ интеллектуальной обработки информации в нейронной сети. Патент на изобретение RU 2427914, опубл. 27.08.2011, бюл. № 24.
13. **Осипов В. Ю.** Рекуррентная нейронная сеть со структурой слоев в виде двойной спирали // *Информационные технологии*. 2014. № 7. С. 56–60.
14. **Осипов В. Ю.** Стирание устаревшей информации в ассоциативных интеллектуальных системах // *Мехатроника, автоматизация, управление*. 2012. № 3. С. 16–20.
15. **Осипов В. Ю.** Пределы памяти рекуррентных нейронных сетей со стиранием устаревшей информации // *Научный вестник Новосибирского государственного технического университета*. 2014. № 3 (56). С. 115–122.
16. **Осипов В. Ю.** Метод управления синапсами в рекуррентной нейронной сети // *Информационные технологии*. 2013. № 7. С. 61–66.
17. **Осипов В. Ю.** Способ интеллектуальной обработки информации в нейронной сети. Патент на изобретение RU 2514931, опубл. 10.05.2014, бюл. № 13.
18. **Осипов В. Ю.** Рекуррентная нейронная сеть с двумя сигнальными системами // *Информационно-управляющие системы*. 2013. № 4. С. 8–15.
19. **Осипов В. Ю.** Ассоциативная интеллектуальная машина с тремя сигнальными системами // *Информационно-управляющие системы*. 2014. № 5. С. 12–17.
20. **Alberts B., Bray P., Lewis J., Raff M., Roberts K., Watson J. D.** (Eds.). *Molecular Biology of The Cell*. Second Edition. New York: Garland Publishing, Inc. 1989. 1219 p.
21. **Palm G.** Information capacity in recurrent McCulloch-Pitts networks with sparsely coded memory states // *Network*. 1992. N. 3. P. 177–186.
22. **Bosch H., Kurfess F.** Information storage capacity of incompletely connected associative memories // *Neural Networks*. 1998. N. 11 (5). P. 869–876.

V. Yu. Osipov, Professor, e-mail: osipov_vasily@mail.ru,
Saint-Petersburg Institute for Informatics and Automation of RAS

Associative and Spatial Addressing to Memory of Recurrent Neural Networks

Approaches to enhance the cognitive capabilities of recurrent neural networks are investigated. In the interests it is proposed to empower such a networks of associative and spatial addressing to memory. The additional spatial addressing to memory of recurrent neural networks allows them to manage the associative interaction of processed signals. The improved model of recurrent neural network with associative and spatial addressing to memory is disclosed. It is shown that a memory addressing can be easily implemented in the electrical models recurrent neural networks. Examples of spatial redistribution of the currents in such networks, depending on the current state of interacting layers are shown. The proposed solutions allow to remove a number of constraints to understanding the processes of cognitive information processing in recurrent neural networks. These solutions can be used to develop advanced cognitive neural machines with high-speed parallel processing of different signals.

Keywords: recurrent neural network, associative and spatial addressing, cognitive information processing

References

1. **Haikonen Pentti O. A.** The Role of Associative Processing in Cognitive Computing, *Cognitive Computing*, 2009, no. 1, pp. 42–49.
2. **Palm G.** Neural associative memories and sparse coding, *Neural Networks*, 2013, no. 37, pp. 165–171.
3. **Haykin S.** *Neural Networks and Learning Machines*. Third Edition. New-York, Prentice Hall, 2008, 936 p.
4. **Galushkin A. I.** *Neural Networks Theory*. Berlin Heidelberg, Springer-Verlag, 2007, 396 p.
5. **Liaw J.-S., Berger T. W.** *Dynamic Synapse for Signal Processing in Neural Networks*. US Patent 6643627 B2, Nov. 4, 2003.
6. **Cover T. M.** Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern recognition, *IEEE Transactions on Electronic Computers*, June, 1965, pp. 326–334.
7. **Jeong S., Lee M.** Adaptive object recognition model using incremental feature representation and hierarchical classification, *Neural Networks*, 2012, no. 25, pp. 130–140.
8. **Chartier S., Giguere G., Langlois D.** A new bidirectional heteroassociative memory encompassing correlational, competitive and topological properties, *Neural Networks*, 2009, no. 22, pp. 568–578.

9. **Fukushima K.** Artificial vision by multi-layered neural networks: Neocognitron and its advances, *Neural Networks*, 2013, no. 37, pp. 103–119.

10. **Kohonen T.** Essentials of the self-organizing map, *Neural Networks*, 2013, no. 37, pp. 52–65.

11. **Osipov V. Yu., Osipova M. V.** Sposob i ustrojstvo intellektual'noj obrabotki informacii v nejronnoj seti [Method and device of intellectual processing of information in neural network]. Patent RU 2413304, 2011-02-27.

12. **Osipov V. Ju.** Sposob intellektual'noj obrabotki informacii v nejronnoj seti [Method of intellectual information processing in neural network]. Patent RU 2427914, 2011-08-27.

13. **Osipov V. Yu.** Rekurrentnaja nejronnaja set' so strukturoj sloev v vide dvojnoj spirali [Recurrent Neural Network with Structure of Layers in the Form of the Double Spiral]. *Informacionnye tehnologii*, no. 7, 2014, pp. 56–60 (In Russian).

14. **Osipov V. Yu.** Stiranje ustarevshej informacii v asociativnyh intellektual'nyh sistemah [Erase Outdated Information in Associative Intelligent Systems], *Mehatronika, aytomatizacia, upravlenie*, 2012, no. 3, pp. 16–20 (In Russian).

15. **Osipov V. Yu.** Predely pamjati rekurrentnyh nejronnyh setej so stiranijem ustarevshej informacii [Limits memory recurrent neural networks with deletion of obsolete information] // *Nauchnyj vestnik NGTU*, 2014, vol. 56, no. 3, pp. 115–122, (In Russian).

16. **Osipov V. Yu.** Metod upravlenija sinapsami v rekurrentnoj nejronnoj seti [The Method of Control Synapses in Recurrent Neural Network], *Informacionnye tehnologii*, no. 7, 2013, pp. 61–66 (In Russian).

17. **Osipov V. Ju.** Sposob intellektual'noj obrabotki informacii v nejronnoj seti [Method for intelligent information processing in neural network]. Patent RU 2514931, 2014-05-10.

18. **Osipov V. Yu.** Rekurrentnaja nejronnaja set' s dvumja signal'nymi sistemami [The Recurrent Neural Network with Two Signal System] // *Informacionno-upravliaiushchie sistemy*, 2013, no. 4, pp. 8–15 (In Russian).

19. **Osipov V. Yu.** Associativnaja intellektual'naja mashina s tremja signal'nymi sistemami [Three Signaling Systems Associative Machine] // *Informacionno-upravliaiushchie sistemy*, 2014, no. 5, pp. 12–17 (In Russian).

20. **Alberts B., Bray D., Lewis J., Raff M., Roberts K., Watson J. D.** (Eds.). *Molecular Biology of The Cell*. Second Edition. Garland Publishing, Inc. New York, 1989. 1219 p.

21. **Palm G.** Information capacity in recurrent McCulloch-Pitts networks with sparsely coded memory states. *Network*, 1992, no. 3, pp. 177–186.

22. **Bosch H., Kurfess F.** Information storage capacity of incompletely connected associative memories. *Neural Networks*, 1998, no. 11 (5), pp. 869–876.

УДК 004.032.26

И. В. Лёзина, канд. техн. наук, доц. кафедры информационных систем и технологий, chuchyck@yandex.ru, Самарский государственный аэрокосмический университет имени академика С. П. Королёва (национальный исследовательский университет) (СГАУ), Самара

Сравнительный анализ погрешности прогноза биржевых индексов методами экспоненциального сглаживания и нейросетевого моделирования

Прогнозирование значений биржевых индексов является актуальной задачей управления экономикой. Получаемые прогнозные значения служат основой для принятия различных управленческих решений. Целью данной работы является сравнение погрешности прогноза реальных данных биржевых индексов Dow Jones и Hang Seng. Для получения прогноза использовали методы экспоненциального сглаживания и нейросетевую модель многослойного перцептрона. Анализ погрешности использованных методов показал более высокую эффективность модели многослойного перцептрона по сравнению с традиционными методами экспоненциального сглаживания.

Ключевые слова: прогнозирование биржевых индексов, экспоненциальное сглаживание, нейросетевое моделирование, многослойный перцептрон

Введение

Выбор метода прогноза зависит в первую очередь от вида имеющейся информации. Прогнозирование на основе временных рядов осуществляется в тех случаях, когда существует ряд значений индекса за предыдущие периоды и предполагается, что в будущем закон его развития сохранится. При этом задача прогнозирования фактически сводится к задаче экстраполяции. К методам экстраполяции относятся метод скользящего среднего, метод наименьших квадратов и метод экспоненциального сглаживания [1–5]. Метод скользящего среднего состоит в замене фактических значений временного ряда расчетными, имеющими значительно меньшие

колебания, чем исходные данные. При этом среднее рассчитывается по группам данных за определенный интервал времени, причем каждая последующая группа образуется со сдвигом на один период. В результате первоначальные колебания временного ряда сглаживаются. Данный метод применяется для краткосрочного прогнозирования. Недостаток метода заключается в том, что требуется много данных для расчета прогнозного значения показателя.

Сущность метода наименьших квадратов состоит в минимизации суммы квадратов отклонений между наблюдаемыми и расчетными значениями. Расчетные значения находят по подобранному уравнению регрессии. Чем меньше расстояние между фактическими значениями и расчетными, тем более

точен прогноз, построенный на основе уравнения регрессии. Недостатком метода наименьших квадратов являются сложность подбора уравнения регрессии и его пересчет по мере поступления новой информации.

Метод экспоненциального сглаживания наиболее эффективен при разработке среднесрочных и долгосрочных прогнозов. Он приемлем при прогнозировании только на один период вперед. Его основные достоинства — простота процедуры вычислений и возможность учета весов исходной информации. При прогнозировании данным методом возникают два затруднения: выбор значения параметра сглаживания и определение начального значения параметра. От значения параметра сглаживания зависит скорость снижения влияния предыдущих наблюдений. Если значение параметра близко к единице, то при прогнозе, в основном, учитывается влияние лишь последних наблюдений. Если значение параметра близко к нулю, то веса значений временного ряда убывают медленно, т. е. учитываются практически все прошлые наблюдения. Точного метода для выбора оптимального значения параметра сглаживания нет, как нет и точного метода расчета экспоненциально взвешенного среднего начального параметра.

Следует также отметить, что при прогнозировании экономических временных рядов метод экспоненциального сглаживания не всегда подходит. Если экономические временные ряды слишком короткие (15—20 наблюдений) или темпы роста показателей достаточно велики, данный метод "не успевает" отразить все изменения [6, 7].

В то же время многие нейросетевые модели выступают в качестве универсального аппроксиматора обучающих данных, поэтому применение нейронных сетей для прогнозирования временных рядов на основе полученной аппроксимирующей функции является весьма перспективным. На вход сети подаются известные значения ряда, а на выходе получают прогноз на необходимое число шагов, т. е. сеть прогнозирует будущие значения ряда на основе его предыдущих значений. К основным преимуществам нейронной сети следует отнести ее универсальность. Сети с одинаковой структурой после обучения могут прогнозировать совершенно различные типы динамики, кроме того, нейронную сеть можно рассматривать как адаптивную модель, поскольку она может дообучаться при поступлении новых сведений.

Основной задачей становится выбор архитектуры нейронной сети и обучение на имеющихся данных. В частности, универсальная теорема об аппроксимации доказана для многослойного персептрона и радиально-базисных сетей [7—12]. Также многослойный персептрон был выбран в качестве исследуемой модели, поскольку его достоинствами являются простота в использовании, хорошо апробированные алгоритмы обучения, способность моделирования функции любой степени сложности.

Модель многослойного персептрона

Для решения задачи сравнения погрешности прогноза, полученного нейронной сетью и методами экспоненциального сглаживания, автором был разработан пакет программ [13]. В модели нейронной сети был использован нейрон сигмоидального типа, в качестве функции активации выбран гиперболический тангенс. Функционирование сигмоидального нейрона можно описать парой уравнений:

$$u_k = \sum_{j=1}^N w_{kj} x_j; \quad (1)$$

$$y_k = \varphi(u_k + b_k), \quad (2)$$

где x_j — входные сигналы; w_{kj} — синаптические веса нейронов; u_k — линейная комбинация входных воздействий; φ — функция активации; y_k — выходной сигнал нейрона. Для обучения сети был использован алгоритм наискорейшего спуска с моментами и метод обратного распространения ошибки, в качестве функции активации — гиперболический тангенс [14]. Для начальной инициализации весовых коэффициентов многослойного персептрона использовали метод роя частиц [15].

Модели экспоненциального сглаживания и двойного экспоненциального сглаживания

Экспоненциальное сглаживание может быть описано формулой следующего вида [16]:

$$S_t = \alpha X_t + (1 - \alpha) S_{t-1}, \quad (3)$$

где S_t — значение индекса на текущем этапе или прогнозируемое значение; S_{t-1} — спрогнозированное значение на предыдущей итерации; X_t — фактическое значение индекса. Когда эта формула применяется рекурсивно, то каждое новое сглаженное значение (которое является также прогнозом) вычисляется как взвешенное среднее текущего наблюдения и сглаженного ряда. Очевидно, результат сглаживания зависит от параметра сглаживания α . Если α равен 1, то предыдущие наблюдения полностью игнорируются. Если α равен 0, то игнорируются текущие наблюдения. Значения α между 0, 1 дают промежуточные результаты.

Также был применен метод двойного экспоненциального сглаживания, где появляется дополнительный параметр сезонного сглаживания γ [16]:

$$S_t = \alpha X_t + (1 - \alpha)(S_{t-1} + b_{t-1}), \quad 0 < \alpha < 1; \quad (4)$$

$$b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1}, \quad 0 < \gamma < 1. \quad (5)$$

За начальные значения S_1 принимается y_1 , а для b_1 существует несколько вариантов (n — число фактических значений индекса):

$$\begin{aligned} b_1 &= X_2 - X_1; \\ b_1 &= [(X_2 - X_1) + (X_3 - X_2) + (X_4 - X_3)]/3; \\ b_1 &= (X_n - X_1)/(n - 1). \end{aligned} \quad (6)$$

Сравнительный анализ погрешностей прогнозирования

Сравнительный анализ погрешностей прогнозирования проводился в среде пакета программ [13] в ходе экспериментальных исследований. В качестве исходных данных использовалась статистика по биржевым индексам с 6.11.2012 по 25.01.2014 [17]. Для сравнения качества прогноза с помощью различных моделей рассчитывали среднее квадратичное отклонение (СКО) по формуле

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - y_{fi})^2}, \quad (7)$$

где y_i — значение индекса, полученное сетью; y_{fi} — фактическое значение индекса. В таблице представлена зависимость СКО прогнозирования на тестовой выборке индексов Dow Jones и Hang Seng для метода экспоненциального сглаживания, двойного экспоненциального сглаживания и многослойного персептрона. Для обучения нейронной сети было взято 150 отсчетов временного ряда и 150 отсчетов — для тестирования. Параметры нейронной сети, подобранные для эксперимента: 5 входов, 9 нейронов в скрытом слое, 4 нейрона в выходном слое, коэффициент момента равен 0,05, коэффициент обучения — 0,6.

Исследование зависимости СКО прогнозирования для метода экспоненциального сглаживания, двойного экспоненциального сглаживания и многослойного персептрона для индексов Dow Jones и Hang Seng

Индекс	Экспоненциальное сглаживание	Двойное экспоненциальное сглаживание	Многослойный персептрон
Dow Jones	0,0303	0,0205	0,0091
Hang Seng	0,0300	0,0213	0,0047

Из приведенной таблицы видно, что для обоих индексов минимальное значение СКО получено при использовании модели многослойного персептрона.

Заключение

Сравнительный анализ погрешностей прогнозирования биржевых индексов показал более высокую эффективность прогноза нейросетевой модели по сравнению с моделями экспоненциального сглаживания: значение СКО прогноза многослойного персептрона составляет 0,0091 и 0,0047 для индексов Dow Jones и Hang Seng соответственно, в то время как значения СКО модели экспоненциального сглаживания равны 0,0303 и 0,0300, а для модели двойного экспоненциального сглаживания —

0,0205 и 0,0213 соответственно. При этом эффективность модели двойного экспоненциального сглаживания незначительно выше модели экспоненциального сглаживания, однако обе модели практически не зависят от вида временного ряда, чего нельзя сказать о модели многослойного персептрона: для временного ряда индекса Hang Seng СКО почти в 2 раза меньше.

Список литературы

1. **Грешилов А. А., Стакун В. А., Стакун А. А.** Математические методы построения прогнозов. М.: Радио и связь, 1997. 112 с.
2. **Владимирова Л. П.** Прогнозирование и планирование в условиях рынка: учеб. пособие. М.: Дашков и Ко, 2005. 400 с.
3. **Новикова Н. В., Поздеева О. Г.** Прогнозирование национальной экономики: учеб.-метод. пособие. Екатеринбург: Изд-во Урал. гос. экон. ун-та, 2007. 137 с.
4. **Слуцкий Л. Н.** Курс МБА по прогнозированию в бизнесе. М.: Альпина Бизнес Букс, 2006. 280 с.
5. **Лукашин Ю. П.** Адаптивные методы краткосрочного прогнозирования временных рядов: учеб. пособие. М.: Финансы и статистика, 2003. 416 с.
6. **Чучуева И. А.** Модель прогнозирования временных рядов по выборке максимального подобия: дисс. канд. техн. наук.: 05.13.01, защищена 20.03.12. М.: МГТУ имени Н. Э. Баумана, 2012. 155 с.
7. **Prajakta S. K.** Time series Forecasting using Holt-Winters Exponential Smoothing // Kanwal Rekhi School of Information Technology Journal, 2004, 13 p. URL: http://www.it.iitb.ac.in/~praj/acads/seminar/04329008_ExponentialSmoothing.pdf.
8. **Pradhan R. P., Kumar R.** Forecasting Exchange Rate in India: An Application of Artificial Neural Network Model [Текст]. Journal of Mathematics Research. 2010. Vol. 2, N. 4. P. 111—117.
9. **Yildiz B., Yalama A., Coskun M.** Forecasting the Istanbul Stock Exchange National 100 Index Using an Artificial Neural Network // International Journal of Science, Engineering and Technology. 2008. Vol. 46. P. 36—39.
10. **Catalao J.** An Artificial Neural Network Approach for Day-Ahead Electricity Prices Forecasting // 6th WSEAS international conference on Neural networks, USA, Stevens Point, 2005. P. 80—83.
11. **Kumar M.** Short-term load forecasting using artificial neural network techniques: Thesis for Master of Science degree in Electrical Engineering. India, Rourkela, National Institute of Technology, 2009. 48 p.
12. **Осовский С.** Нейронные сети для обработки информации / Пер. с польского И. Д. Рудинского. М.: Финансы и статистика, 2002. 344 с.
13. **Свидетельство** о государственной регистрации программы для ЭВМ № 2013614738 "Автоматизированная система прогнозирования финансовых показателей на базе многослойного персептрона" [Текст] / Лезина И. В., Титоренко А. М., Хохлова В. С. — Заявка № 2013612606, дата пост. 02.04.2013, зарегистрировано в Реестре программ для ЭВМ 20.05.2013.
14. **Лёзина И. В., Хохлова В. С.** Прогнозирование финансовых рынков с использованием многослойного персептрона // Перспективные информационные технологии в научных исследованиях, проектировании и обучении (ПИТ 2012): труды научно-технической конференции с международным участием и элементами научной школы для молодежи, посвященной 40-летию кафедры ИСТ СГАУ / Под ред. С. А. Прохорова. Самара: Издательство Самарского научного центра РАН, 2012. С. 168—171.
15. **Лёзина И. В., Хохлова В. С.** Исследование алгоритмов инициализации весов многослойного персептрона для решения задачи прогнозирования // Вестник транспорта Поволжья. № 3 (39) май — июнь 2013 г.
16. **Анализ** временных рядов. URL: <http://www.statsoft.ru/home/textbook/modules/sttimser.html#exponential>.
17. **РосБизнесКонсалтинг** [Электронный ресурс] — <http://expert.rbc.ru/>.

Comparative Analysis of Exchange Index Forecast Error by Using Methods of Exponential Smoothing and Neural Network Modeling

Forecasting the values of stock indices is an actual task of economic management. The resulting values are a basis for making the various management decisions. There are different approaches to forecasting, and the methods choice depends on available information, properties of the researched index and a quality of the forecast. The main goal is to compare the forecast error of real data stock indices Dow Jones and Hang Seng.

The paper uses methods of exponential smoothing and neural network model of multi-layer perceptron for the forecasting. Average-standard deviation is used as a measure of the error. All researches are performed by using software packages developed by the authors.

Error analysis of designed methods and implemented models for check of the training and test samples shows higher efficiency of forecasting the stock indices Dow Jones and Hang Seng when using the multi-layer perceptron model in comparison with traditional methods of exponential smoothing. Analysis of the forecasting error by using the developed software allows you to choose the most efficient model for forecasting the selected stock market indices.

Keywords: forecasting the stock indices, exponential smoothing, neural network modeling, multi-layer perceptron

References

1. Greshilov A. A., Stakun V. A., Stakun A. A. *Matematicheskie metody postroeniya prognozov*. Moscow: Radio i svyaz, 1997, 112 s.
2. Vladimirova L. P. *Prognozirovaniye i planirovaniye v usloviyakh ryinka*, ucheb. posobie. Moscow: Dashkov i Ko, 2005. 400 p.
3. Novikova N. V., Pozdeeva O. G. *Prognozirovaniye natsionalnoy ekonomiki*, ucheb.-metod. posobie. Ekaterinburg: Izd-vo Ural. gos. ekon. un-ta, 2007, 137 s.
4. Sluckin L. N. *Curs MBA po prognozirovaniy v biznese*. Moscow, Alpina Biznes Buks, 280 p.
5. Lukashin Yu. P. *Adaptivnyye metody kratkosrochnogo prognozirovaniya vremennykh ryadov*. Moscow, Finansy i statistika, 2003. 416 p.
6. Chuchueva I. A. *Model prognozirovaniya vremennykh ryadov po vyborke maksimalnogo podobiya*, diss. kand. tehn. nauk.: 05.13.01, zaschislena 20.03.12. Moscow, MGTU imeni N. E. Baumana, 2012. 155 p.
7. Prajakta S. K. Time series Forecasting using Holt-Winters Exponential Smoothing, *Kanwal Rekhi School of Information Technology Journal*, 2004, 13 p. URL: http://www.it.iitb.ac.in/~praj/acads/seminar/04329008_ExponentialSmoothing.pdf.
8. Pradhan R. P., Kumar R. Forecasting Exchange Rate in India: An Application of Artificial Neural Network Model, *Journal of Mathematics Research*. 2010, vol. 2, no. 4, p. 111–117.
9. Yildiz B., Yalama A., Coskun M. Forecasting the Istanbul Stock Exchange National 100 Index Using an Artificial Neural Network, *International Journal of Science, Engineering and Technology*. 2008, vol. 46, p. 36–39.
10. Catalao J. et al. An Artificial Neural Network Approach for Day-Ahead Electricity Prices Forecasting, *6th WSEAS international conference on Neural networks, USA, Stevens Point*, 2005. P. 80–83.
11. Kumar M. *Short-term load forecasting using artificial neural network techniques*: Thesis for Master of Science degree in Electrical Engineering. India, Rourkela, National Institute of Technology, 2009. 48 p.
12. Osovskiy S. *Neyronnyie seti dlya obrabotki informatsii*, per. s polsk. I. D. Rudinskogo, Moscow, Fmansy i statistika, 2002. 344 s.
13. Lyozina I. V., Titorenko A. M., Hohlova V. S. *Svidetelstvo o gosudarstvennoy registratsii programmy dlya EVM № 2013614738 "Avtomatizirovannaya sistema prognozirovaniya dinansovykh pokazateley na baze mnogoslownogo perseptrona"*. Zayavka № 2013612606, data postupleniya 02.04.2013, zaregistrirvano v Reestre programm dlya EVM 20.05.2013.
14. Lyozina I. V., Hohlova V. S. Prognozirovaniye finansovykh ryinkov s ispolzovaniem mnogoslownogo perseptrona, Perspektivnyye informatsionnyie tehnologii v nauchnykh issledovaniyakh, proektirovani i obuchenii (PIT 2012): *trudyi Nauchno-tehnicheskoy konferentsii s mezhdunarodnyim uchastiem i elementami nauchnoy shkoly dlya molodezhi, posvyasch Yonnoy 40-letiyu kafedryi IST SGAU* / ed. S. A. Prohorova, Samara: Izdatelstvo Samarskogo nauchnogo tsentra RAN, 2012, pp. 168–171.
15. Lyozina I. V., Hohlova V. S. Issledovaniye algoritmov initsializatsii vesov mnogoslownogo perseptrona dlya resheniya zadachi prognozirovaniya. *Vestnik transporta Povolzhya*. Nauchnyy zhurnal, 2013. N 3 (39).16. Analiz vremennykh ryadov, URL: <http://www.statsoft.ru/home/textbook/modules/sttimser.html#exponential>
17. RosBiznesKonsalting. URL: <http://export.rbc.ru/>.

Адрес редакции:

107076, Москва, Стромынский пер., 4

Телефон редакции журнала (499) 269-5510

E-mail: it@novtex.ru

Технический редактор *Е. В. Конова*.

Корректор *З. В. Наумова*.

Сдано в набор 06.06.2015. Подписано в печать 20.07.2015. Формат 60×88 1/8. Бумага офсетная.

Усл. печ. л. 8,86. Заказ IT815. Цена договорная.

Журнал зарегистрирован в Министерстве Российской Федерации по делам печати, телерадиовещания и средств массовых коммуникаций.

Свидетельство о регистрации ПИ № 77-15565 от 02 июня 2003 г.

Оригинал-макет ООО "Авансед солюшнз". Отпечатано в ООО "Авансед солюшнз".

119071, г. Москва, Ленинский пр-т, д. 19, стр. 1.