Запорізький національний технічний університет

Радіоелектроніка
Інформатика
Управління

1(26)'2012

Науковий журнал

Виходить двічі на рік

Видається з березня 1999 року

Зареєстрований ***29 січня 2003 року***
Державним комітетом інформаційної політики,
телебачення та радіомовлення України.
Свідоцтво – серія ***КВ № 6904***

Засновник і видавник – Запорізький національний технічний університет

Запоріжжя, ЗНТУ
2012

ISSN 1607-3274

Постановою президії ВАК України № 1-05/4 від 26.05.2010 р. журнал «Радіоелектроніка, інформатика, управління» (скорочена назва – РІУ), який видається з 1999 року, включений до переліку наукових фахових видань України, в яких можуть публікуватися результати дисертаційних робіт на здобуття наукових ступенів доктора і кандидата технічних наук та фізико-математичних наук (радіофізика).

Журнал є донором журналу «Telecommunications and Radio Engineering», який видається в США.

Інтернет-сторінка журналу: http://journal.zntu.edu.ua/ric/index.php?page=index.

Статті, що публікуються в журналі, реферуються в базах даних та РЖ ВІНІТІ (Росія) і «Джерело» (Україна). Журнал РІУ міститься у міжнародній базі наукових видань Index Copernicus (http://journals.indexcopernicus.com/index.php), електронна копія журналу розміщена на сайті Національної бібліотеки України імені В. І. Вернадського НАН України у розділі «Наукова періодика України» за адресою: http://nbuv.gov.ua/portal/.

Журнал розповсюджується за Каталогом періодичних видань України (передплатний індекс – 22914).

РЕДАКЦІЙНА КОЛЕГІЯ

Рекомендовано до видання вченою радою Запорізького національного технічного університету (ЗНТУ), протокол № 06 від 27.02.2012 р.

Рукописи проходять незалежне рецензування з залученням провідних фахівців, за результатами якого редакційна колегія приймає рішення про опублікування.

Журнал зверстаний редакційно-видавничим відділом ЗНТУ.

Адреса редакції: 69063, м. Запоріжжя, вул. Жуковського, 64, ЗНТУ, редакція журналу «РІУ».
Тел: (061) 769-82-96 – редакційно-видавничий відділ
Факс: (061) 764-46-62
E-mail: rvv@zntu.edu.ua

ЗМІСТ

CONTENTS

РАДІОФІЗИКА

РАДИОФИЗИКА

RADIOPHYSICS

УДК 621.315.5:544.03

Gorbenko V. I.[1], Gorban A. N.[2]

[1]Ph.D. in physics, associate professor, Classical Private University
[2]D.Sc. in physics, professor, first vice-rector, Classical Private University

THERMAL DECOMPOSITION OF INDIUM PHOSPHIDE IN VACUUM AND ATOMIC HYDROGEN ENVIRONMENT

The thermal decomposition of indium phosphide has been investigated by Auger-electron spectroscopy and mass-spectroscopy. Scanning electron microscopy has been used for study of indium islands growth on surface of the compound semiconductor. The role of atomic hydrogen in processes of decomposition and growth of metallic islands was determined by comparing with these processes under vacuum.

Key words: indium phosphide, atomic hydrogen, thermal decomposition, scanning electron microscopy.

INTRODUCTION

It is now well established that the interaction of atomic hydrogen with clean InP surface leads to a decomposition of the substrate [1–3]. There are two successive stages of the interaction. During first interaction stage H-atoms binds covalently to the substrate and saturates surface unit cells [1, 4]. The second interaction stage leads to a decomposition of the substrate [2]. Auger Electron Spectroscopy (AES) measurements have shown that the ratio of the intensities of the P(120eV) and In(410eV) peaks decrease during the exposure of indium phosphide in atomic hydrogen. The confirmation of a metal presence on the surface was given by Photoemission Yield Spectroscopy (PYS), too. The adsorption stage of the interaction and the decomposition stage are contiguous at doses of hydrogen exposition about 5×10^3–10^4 L. In accordance with estimations in [1] the number of hydrogen atoms reaching the sample during an exposition 10^4 L is 10^{15} atoms/cm^2. The techniques based on high frequency discharge in wet hydrogen allows to obtain 10^{14}–10^{15} H-atoms per cm^3 and its flow to sample surface about 10^{19}–10^{20} atoms$\times$s$^{-1}\times$cm^{-2}

The aim of this report is to present the investigation results of influence of high concentration of atomic hydrogen on the decomposition and the metallization process of indium phosphide.

EXPERIMENTAL

The vacuum equipment and experimental conditions have been described in details elsewhere[5]. We recall that the expositions of InP samples have been carried out in a specially designed vacuum chamber (reactor). To this chamber via diaphragm the monopole mass spectrometer MX7304A (produced by «SELMI», Ukraine) was connected. Such construction gives an opportunity to record a real-time mass spectra of the gas components. The wet hydrogen fed a discharge vessel, which was connected to the reactor. The hydrogen was excited by high-frequency discharge. During experiments the normal working pressure of gases in the reactor was at the level of 10…25 Pa and a base pressure in the spectrometer chamber was of about 10^{-5} Pa. The maximal concentration of atomic hydrogen was at the level of 10^{15} cm^{-3}.

The samples were cut from *n*-type InP single crystals (n=1,1$\times10^{17}$ cm^{-3} (111)). The standard surface preparations before exposition to the gas mixture in reactor were chemical polishing in a bromine-methanol etchant and successive rinsing in bidistilled and deionised water. During exposition the distance between the discharge and semiconductor sample was about 20 cm that produced conditions for termalization and deionization of gas particles moving from the discharge to the sample.

RESULT AND DISCUSSION

Decomposition. The effect of the high-intensity flow of hydrogen atoms on the decomposition of indium phosphide has been investigated by mass spectrometric method.

The experiments showed that the heating of InP surface up to 800K in hydrogen environment without discharge did

not change the composition of the system gas phase. At higher temperature the diphosphorus molecules were detected. None of the gas hydrogen-phosphorus species were discovered up to 1000 K.

The dependence of diphosphorus partial pressure as function of temperature was similar to that observed during a dissociation of indium phosphide in vacuum. The dependence depicted as Arrenius plots is shown on fig.1 curve (a). From the slope of the curve (a) we have found that the enthalpy of reaction $InP_{(sol)} \rightarrow In_{(sol)}+1/2P_{2(gas)}$ at 298 K is about 36,9 kcal/mole. It is close to the standard value for this reaction [6]. Probably, the decomposition of indium phosphide in such system is caused by a simple dissociation of the compound.

The exciting of hydrogen by high-frequency discharge added the atomic hydrogen to the system. The exposition of indium phosphide at maximum concentration of the atomic hydrogen caused a drastic change of the process of indium phosphide decomposition. The temperature of the beginning of the decomposition was lower than that for vacuum or molecular hydrogen medium. In our experiments this temperature for both vacuum and unexciting hydrogen was close to 800K and it was reduced by 230 K in the presence of atomic hydrogen. Moreover, in mass spectra both the phosphine and the diphosphorus were observed simultaneously. It is really nothing new to find the PH_3 molecules in the systems similar to H/InP. But an appearance of diphosphorus in gas phase at such a low temperature as 570 K was detected for the first time. Finally, the figure 1(b) shows the dependence of diphosphorus partial pressure in the system with atomic hydrogen. Both a shift of the curve to low temperatures and a change of the curve slope are evident. In this case the enthalpy of reaction $InP_{(sol)} \rightarrow In_{(sol)}+1/2P_{2(gas)}$ at 298K was estimated as 9,32 kcal/mole that is strongly differed from the standard value.

It has been experimentally established that during thermal dissociation of indium phosphide the P_2 and P_4 species are initially forming because they are thermodynamic preferable and P-atoms have enough high surface diffusivity [6]. For treatments of InP with atomic hydrogen HREELS measurements have shown the binding covalently of atomic hydrogen to both P-atom and In-atom on the surface [3, 4]. The saturation of surface by atomic hydrogen leads to a breaking of the bonds between In and neighbouring P-atoms. Probably such interaction of atomic hydrogen with surface atoms is able to cause a releasing of phosphorus atoms similar to a thermal treatment in vacuum. We are thinking that the weakening of In-P bonds by hydrogen interaction with surface is a reason of decreasing both a temperature of InP decomposition and an enthalpy of InP dissociation reaction.

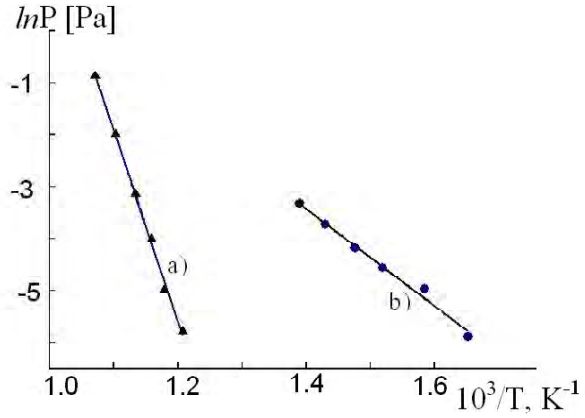


Fig. 1. Dependence of diphosphorus partial pressure as function of temperature:

a) during InP dissociation in vacuum or molecular hydrogen;
b) during InP exposition in hydrogen environment with concentration of H-atoms about 10^{15} cm^{-3}

Growth of Indium Islands. The influence of hydrogen atoms on the process of the indium islands growth has been investigated by a comparing with In-islands growth during InP dissociation in vacuum. The scanning electron microscopy method has been used.

The fig. 2 shows the main features of InP surface (111) after the dissociation in vacuum: hexahedral shaped indium islands (label 1); a simple drop with a spherical form (label 2); a drop which has been formed from hexahedral islands (label 3); and the label 4 marks a hexahedral area which was emptied after transition of an indium island from hexahedral to sphere-like shape. On surface of some samples the islands with triangular shape have been observed too. Probably the growth on the (111) surface of triangular and hexahedral islands is conditioned by dislocations of the semiconductor crystal. At a dislocation core the atoms are weaker bound than atoms in the crystal. Therefore the phosphorus atoms of the dislocation core are able to desorb at the lowest temperature that causes an initial nucleation of indium islands at dislocation. That explains a forming of shapes of the islands similar to shapes of etching pits.

Different from that the decomposition of indium phosphide in hydrogen with high concentration of atomic component leads to a forming only spherical islands (fig. 3, *a–d*).

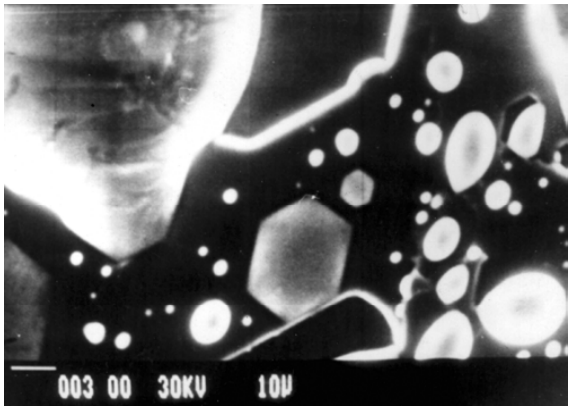


Fig. 2. InP (111) surface after dissociation in vacuum. The temperature of treatment is about 800 K

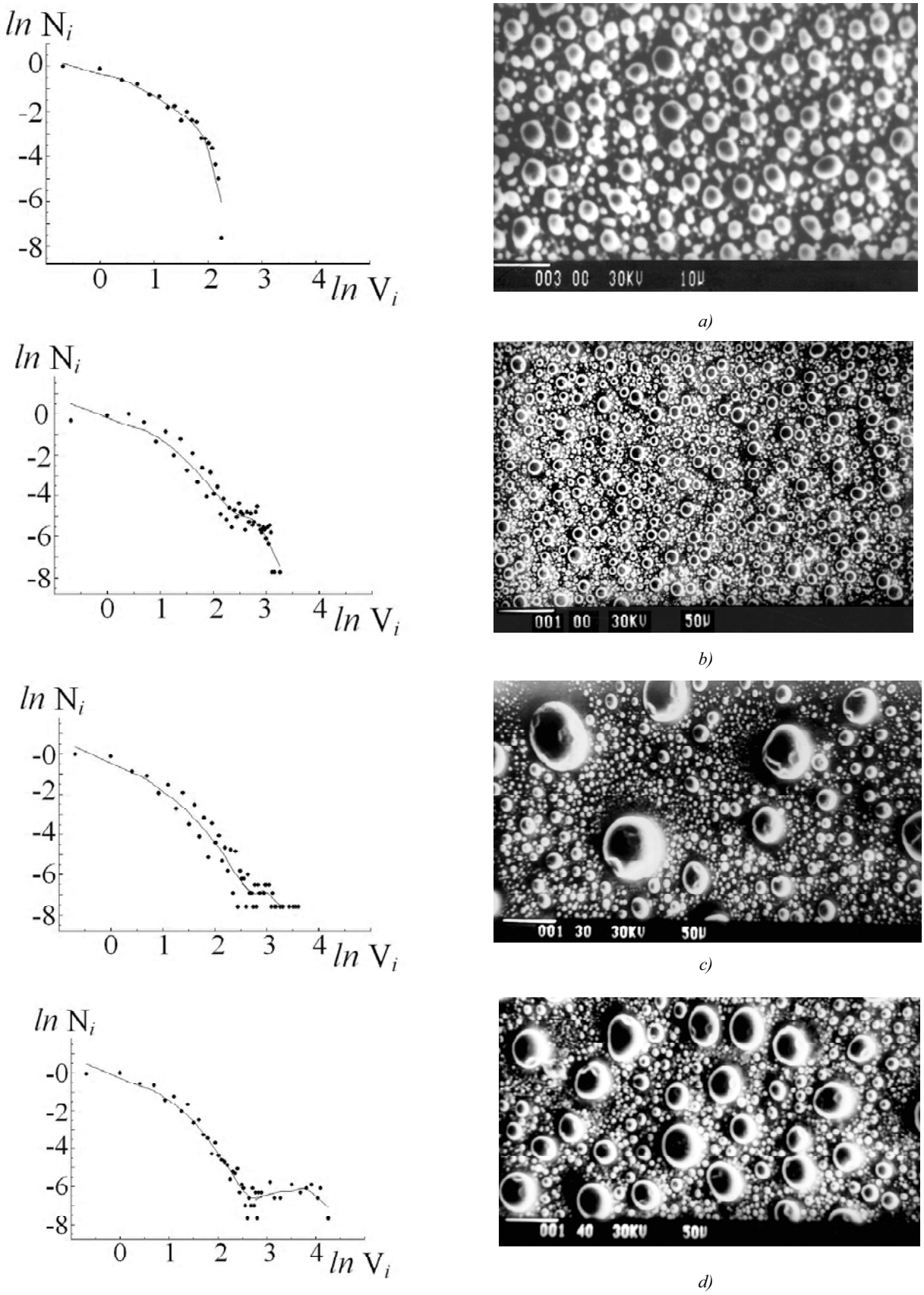


Fig. 3. InP (111) surface after exposition in atomic hydrogen. At left from them the experimental islands size distributions are shown with correspondence to micrographs

Never on such samples the straight-sided islands have not been observed. It is clear that the dislocations do not play a visible role in the process of an interaction of H-atoms with InP surface. The micrographs on fig. 3 are highlighting steps of islands evolution during atomic hydrogen exposition of InP sample.

The first step is incubation not shown on micrographs here. The step duration is depended on temperature of sample and environment. The incubation step comes to an end by formation of small indium drops (for example fig. 3, *a*). At high density of atomic hydrogen flow to a surface the beginning of islands appearance is very difficult to determine exactly. On the left from fig. 3, *a* the typical distribution of islands sizes at start of growth is shown. The sizes of islands form a short range with clear restriction on the right.

The subsequent InP expositions in atomic hydrogen are leading to a spreading of this restriction. Some casual chosen islands grow much faster others. On the distribution in the range of greatest sizes there is a small group which is gradually separating from its basic part. As well as in a case of InP dissociation in vacuum a coalescence is one of the reasons of appearance of large islands. The other reason is non-uniformity of a flow of hydrogen atoms to various areas of a surface. The homogeneity of H-atoms flow is upsetting after forming of metal subsystem on the sample surface. Really the metallic islands catalyse the recombination reaction of hydrogen atoms to molecule. Therefore the forming of indium islands leads to increasing of gradient of the atomic hydrogen concentration near to the sample surface. During recombination act the energy about 4,3 eV is produced. With a combination of an intensive flow of H-atoms it can give the additional heating of islands and cause the enhancing process of decomposition. Really the metallic islands catalyze the recombination reaction of hydrogen atoms to molecule. Therefore the forming of indium islands leads to increasing of gradient of the atomic hydrogen concentration near to the sample surface. During recombination act the energy about 4,3 eV is produced. With a combination of an intensive flow of H-atoms it can give the additional heating of islands and cause the enhancing process of decomposition. Any island that has larger size than nearest neighbours is able to change the gradient of H-atoms concentration and more intensive flow of H-atoms is forming to this island. It is clear such island grows faster than its neighbours at the same time the growth of neighbours is depressed more and more. Fig. 3 *c*, *d* and enclosed distributions illustrate such behavior of islands.

CONCLUSION

The role of atomic hydrogen in the processes of indium phosphide decomposition and growth of indium islands on InP(111) surface has been studied. The increasing of concentration of atomic hydrogen in gas phase causes a decreasing of minimal temperature of the indium phosphide decomposition. During decomposition the species P_2, P_4 and PH_3 are forming. The enthalpy of reaction $InP_{(sol)} \rightarrow In_{(sol)} + 1/2P_{2(gas)}$ is strongly decreased in the presence of atomic hydrogen. The catalytic properties of indium to the reaction of recombination of hydrogen atoms are influencing on process of island growth. The steps of island evolution during atomic hydrogen exposition of InP sample have been established.

REFERENCES

1. *M'hamedi, O.* Interaction of atomic hydrogen with cleaved InP. I. The adsorption stage / M'hamedi O., Proix F., Sebenne C. A. // Journal of Vacuum Science and Technology A. – 1988. – Vol. 6, № 2. – P. 193–198.
2. *Proix, F.* Interaction of atomic hydrogen with cleaved InP. II. The decomposition stage / Proix F., M'hamedi O., Sebenne C. A. // Journal of Vacuum Science and Technology A. – 1988.– Vol. 6, № 2. – P. 199–203.
3. *Proix, F.* Dissociation effects of H and H_2^+ on clean III–V compounds / Proix F. // Physica B. – 1991. – № 170. – P. 457–468.
4. *Schaefer, J. A.* Atomic hydrogen – a local probe for interface characterization / Schaefer J. A. // Surface Science. – 1987. – № 189/190. – P. 127–136.
5. *Gorbenko, V.* Oxidation and metallization in H_2/H_2O/InP system / Gorbenko V., Shvets J., and Gorban A. // Proceedings of the Twenty-Seventh State-of-The-Art Program On Compound Semiconductors (SOTAPOCS XXVII) by editors S. N. G.Chu, D. N. Buckley, K. Wada et. al. Vol. 97 21. – The Electrochemical Society, Pennington, 1997. – P. 375–381.
6. *Panish, M. B.* Phase equilibria and vapor pressures of the system In+P / Panish M. B., Arthur J. R. // J. Chem. Thermodynamics. – 1970. – vol. 2, № 3. – P. 299–318.

Стаття надійшла до редакції 16.11.2011.

Горбенко В. И., Горбань А. Н.

ТЕРМИЧЕСКАЯ ДЕКОМПОЗИЦИЯ ФОСФИДА ИНДИЯ В ВАКУУМЕ И В СРЕДЕ С АТОМАРНЫМ ВОДОРОДОМ

Исследование термической декомпозиции фосфида индия и изменения морфологии поверхности выполнено при помощи Оже-электронной спектроскопии, масс-спектроскопии и сканирующего электронного микроскопа. Влияние атомарного водорода на процесс декомпозиции фосфида индия, возникновение и рост островков индия определено благодаря сравнению с подобными процессами в условиях вакуума.

Ключевые слова: фосфид индия, атомарный водород, термическая декомпозиция, сканирующая электронная микроскопия.

Горбенко В. І., Горбань О. М.

ТЕРМІЧНА ДЕКОМПОЗИЦІЯ ФОСФІДУ ІНДІЮ У ВАКУУМІ ТА В СЕРЕДОВИЩІ З АТОМАРНИМ ВОДНЕМ

Термічна декомпозиція фосфіду індію у вакуумі та в середовищі з атомарним воднем

Дослідження термічної декомпозиції фосфіду індію та морфологічних змін поверхні проводилось за допомогою Оже-електронної спектроскопії, мас-спектроскопії та скануючого електронного мікроскопа. Вплив атомарного водню на процес декомпозиції фосфіду індію, виникнення та ріст індієвих островків визначено завдяки порівнянню з подібними процесами в умовах вакууму.

Ключові слова: фосфід індію, атомарний водень, термічна декомпозиція, скануюча електрона мікроскопія.

УДК 534.222.2

Никонов А. Ю.[1], Небеснюк О. Ю.[2], Шмалий С. Л.[3], Никонова З. А.[4]

[1]Инженер Запорожской государственной инженерной академии
[2, 3]Канд. техн. наук, доцент Запорожской государственной инженерной академии
[4]Канд. техн. наук, профессор Запорожской государственной инженерной академии

ИССЛЕДОВАНИЕ ОСОБЕННОСТЕЙ ДЕФЕКТООБРАЗОВАНИЯ В ВЫСОКОЛЕГИРОВАННОМ КРЕМНИИ ПРИ ОБЛУЧЕНИИ

В статье приведены результаты экспериментального исследования механизмов взаимодействия высокоэнергетического излучения с кремнием.

Ключевые слова: излучение, легирующие примеси, концентрация.

ПОСТАНОВКА ПРОБЛЕМЫ

Применение высокоэнергетических излучений для локального легирования кремния заключается в возможности модифицировать тонкие поверхностные слои без изменения свойств объема (матрицы), нагревать и даже плавить локальные участки, формировать различные метастабильные состояния.

АНАЛИЗ ПОСЛЕДНИХ ДОСТИЖЕНИЙ И ПУБЛИКАЦИЙ

Известно, что наличие значительных концентраций неконтролируемых примесей в кремнии (кислород, углерод и др.) приводит к состоянию, когда влияние легирующей примеси в образовании стабильных радиационных дефектов трудно определить [1].

ЦЕЛЬ СТАТЬИ

Экспериментальное исследование механизмов взаимодействия высокоэнергетического излучения с кремнием, чтобы выбором его оптимальных параметров снизить или вообще исключить структурные несовершенства, влияющие на работу создаваемых приборов.

МАТЕРИАЛЫ И РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Как правило, кристаллы с концентрациями легирующих примесей $10^{13} \div 10^{18}$ см$^{-3}$ содержат $10^{17} \div 10^{18}$ см$^{-3}$ кислорода, углерода и др. Последние могут являться эффективными стоками для первичных дефектов. С ростом уровня легирования кристаллов взаимодействие первичных радиационных дефектов с неконтролируемыми примесями может существенно изменяться как с точки зрения образования стабильных радиационных нарушений, так и рекомбинации первичных дефектов на неконтролируемых примесях. К тому же механизмы взаимодействия первичных дефектов с легирующими примесями могут существенно проявляться лишь при больших концентрациях последних.

При концентрациях легирующей примеси, больших 10^{15} см$^{-3}$, наблюдается падение активности введения дефектов с уровнями E_C–0,17, E_C–0,24, E_V+0,34, E_V+0,21 эВ, что хорошо видно из табл. 1 и рис. 1.

Авторами подтверждено экспериментально, что по мере увеличения уровня легирования кристаллов значительно возрастает начальная скорость удаления носителей δ_n, δ_p.

Как видно из рис. 2, при концентрациях носителей n_0, $p_0 \le 10^{15}$см$^{-3}$ δ_n и δ_p определяются введением указанных центров. При концентрациях носителей, существенно больших 10^{15}см$^{-3}$, введение рационных дефектов не определяет скорость удаления носителей.

Рост скорости удаления носителей с увеличением концентрации легирующих примесей не определяется изменением зарядового состояния первичных радиационных дефектов с $E=E_V+0{,}05$ эВ.

При концентрации носителей больше 10^{15}см$^{-3}$ в кремнии процессы аннигиляции первичных нарушений не определяются температурой облучения, т.к. они не зависят от $T_{\text{обл.}}$ в области $78 \div 300$ К.

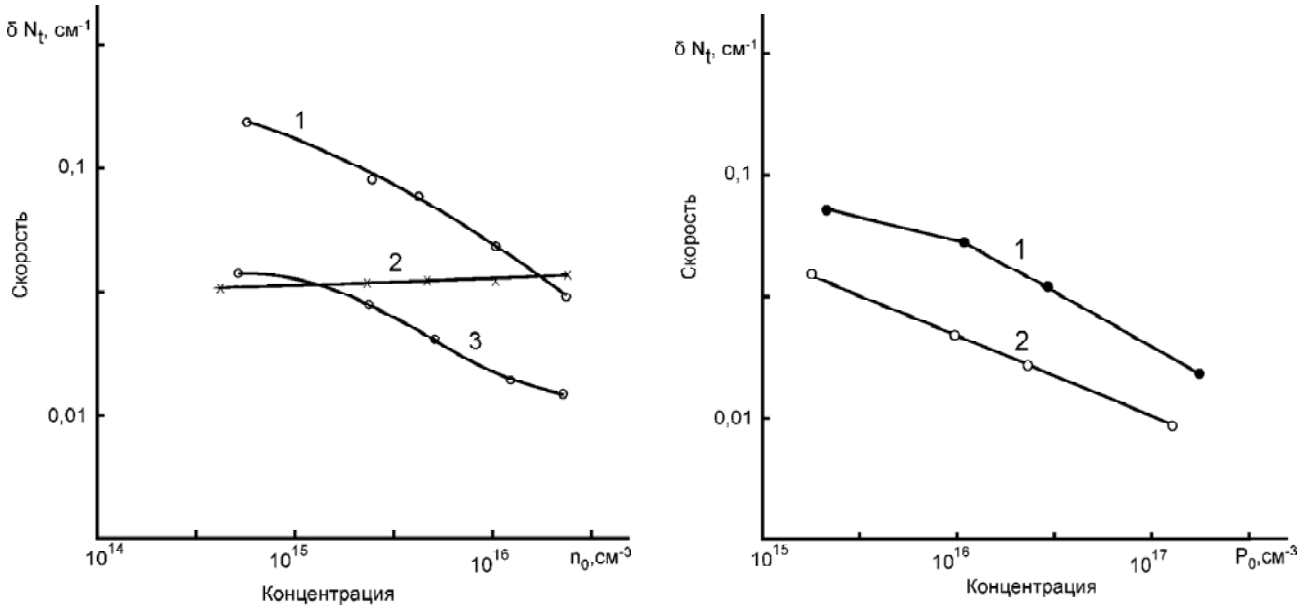
Анализ сведений о природе радиационных дефектов [2–6] показывает, что в состав радиационных дефектов входят вакансии. При сравнивании концентрированных зависимостей скорости удаления носителей и введения радиационных нарушений (рис. 1, 2) экспериментально установлено, что уменьшение потока вакансий на образование дефектов с указанными уровнями не определяет рост скорости удаления носителей при $n_0, p_0 > 10^{15}$см$^{-3}$.

Авторы отмечают, что компенсация высоколегированного кремния обусловлена введением нового типа радиационных дефектов с глубоким уровнем. Причём зависимости скорости удалении носителей от концентрации легирующей примеси при $n_0, p_0 > 3 \cdot 10^{15}$см$^{-3}$ или положение уровня Ферми, измеренное при 78 К, дают рост эффективности введения новых компенсирующих центров от уровня легирования кристаллов.

С ростом концентрации легирующих примесей наряду с взаимодействием первичных радиационных нарушений с неконтролируемыми примесями начинает конкурировать процесс захвата первичных дефектов атомами легирующей примеси. Возможность такого захвата может быть обусловлена наличием зарядов у взаимодействующих компонентов. Междоузельные атомы в n- и p-кремнии выступают как акцепторы и доноры, соот-

Таблица 1. Энергетический спектр и сечения захвата основных носителейв *n*- и *p*- высоколегированном кремнии

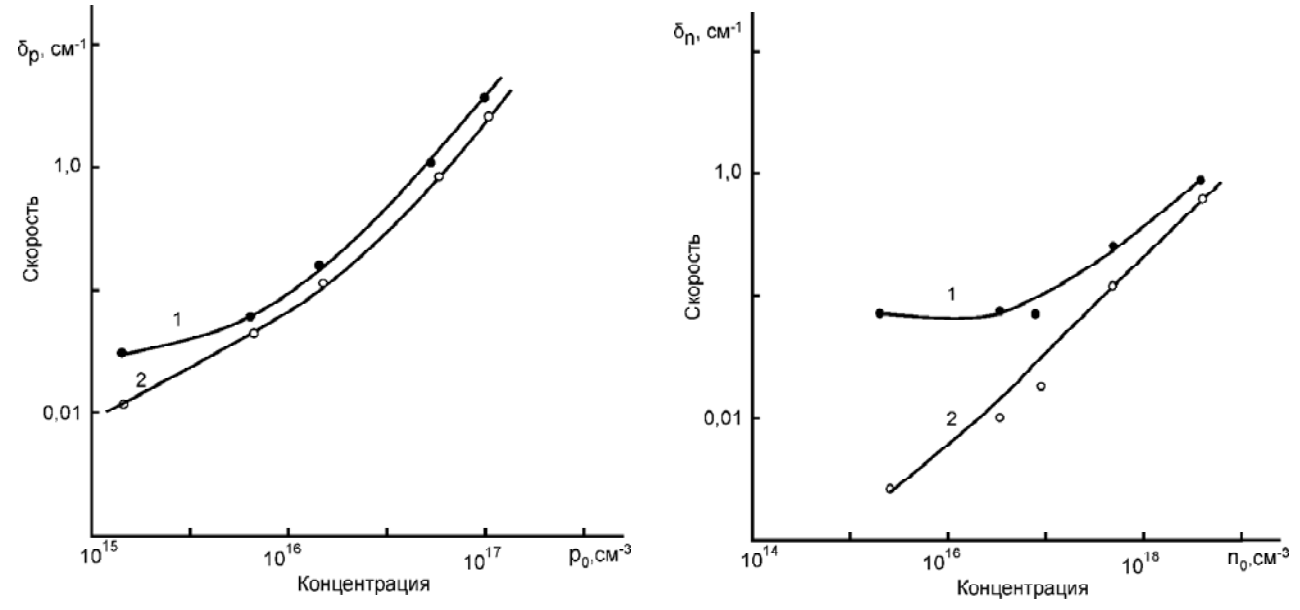
Материал базы	$E_C(_V)$ ± E эВ	$\delta_{n(p)} \cdot 10^2$ см^2
n-кремний	E_C–0,17	4,3
	E_C–0,24	3,6
	E_C–0,41	–
p-кремний	E_V+0,21	1,8
	E_V+0,34	1,1



а) *б)*

Рис. 1. Зависимость скорости введения дефектов в *n*-кремнии (*а*) и *p*-кремнии (*б*) от концентрации носителей при облучении электронами (E_0 = 7 МэВ, $T_{\text{обл.}}$ = 300К):

а) 1 – E_C –0,17эВ; 2 – E_C –0,41эВ; 3 – E_C –0,24эВ *б)* 1 – E_V +0,34эВ; 2 – E_V +0,21эВ



а) *б)*

Рис. 2. Экспериментальная зависимость начальной скорости удаления носителей в *p*-кремнии (*а*) и *n*- кремнии (*б*) от концентрации носителей при облучении электронами (E_0 = 7 МэВ, $T_{\text{обл.}}$ = 300К) $T_{\text{изм}}$.: 1 – 78К; 2 – 300К

ветственно. Если участники реакции обладают зарядами, то радиус взаимодействия, определяемый электростатическим потенциалом, составляет $r_{EH} \approx 50$ Å.

В этом случае с ростом уровня легирования кристаллов возрастает вероятность захвата подвижного междоузельного атома заряженным атомом легирующей примеси. В тоже время сходство концентрационных зависимостей скоростей удаления носителей в *n*- и *p*-кремнии дает возможность предположить, что вакансии в *p*-кремнии в условиях облучения подвижны при 78 К. В таком случае, при отсутствии аннигиляции первичных радиационных дефектов при $n_0, p_0 > 10^{18}$см$^{-3}$ и уменьшении скорости введения дефектов вакансионного типа, возможно образование комплексов (примесный атом-междоузлие) +вакансия. Первоначальным актом, в силу большей подвижности междоузельных атомов, является их взаимодействие с примесью, с последующим захватом вакансии.

ВЫВОДЫ

Таким образом, дефектообразование в высоколегированном кремнии характеризуется следующим:

– аннигиляция первичных радиационных дефектов практически отсутствует при $n_0, p_0 < 10^{17}$см$^{-3}$ и $T \leq 300$ К;

– введение известных вакансионных комплексов не определяет скорость удаления носителей при концентрациях $n_0, p_0 >> 10^{15}$см$^{-3}$;

– компенсация кремния при больших концентрациях легирующих примесей обусловлена введением комплексов: легирующая примесь-междоузлие, либо (легирующая примесь-междоузлие)+вакансия, дающих глубокие уровни в запрещенной зоне ($E_t > 1/2\ E$);

– совершенствование способов локального легирования с целью снижения структурных несовершенств высоколегированного кремния позволит разрабатывать полупроводниковые приборы без ущерба для их структурных и электрофизических характеристик.

СПИСОК ЛИТЕРАТУРЫ

1. *Емцев, В. В.* Примеси и точечные дефекты в полупроводниках / В. В. Емцев, Т. В. Машовец ; под ред. профессора С. М. Рывкина. – М. : Радио и связь, 1981. – 248 с.
2. *Щербачев, К. Д.* Особенности образования радиационных дефектов в слое кремния структур «кремний на изоляторе» / К. Д. Щербачев, В. Т. Бублик, В. Н. Мордкович // Физика и техника полупроводников. – 2011. – Т. 45, вып. 6. – С. 754–758.
3. *Вавилов, В. С.* Радиационные эффекты в полупроводниках и полупроводниковых приборах / В. С. Вавилов, Н. А. Ухин. – М. : Атомиздат, 1969. – 312 с.
4. *Томпсон, М.* Дефекты и радиационные повреждения в металлах, пер. с англ. / М. Томпсон. – М. : Мир, 1971. – 368 с.
5. *Пагава, Т. А.* Два канала отжига дивакансий в облученных кристаллах кремния *n*-типа. / Т. А. Пагава, Н. Т. Бжалава, Н. И. Майсурадзе [та ін.] // Український фізичний журнал. – 2010. – Т. 55, № 11. – С. 1195–1200.
6. *Стась, В. Ф.* Термоакцепторы в облученном кремнии / В. Ф. Стась, И. В. Антонова, Е. П. Неустроев [и др.] // Физика и техника полупроводников. – 2000. – Т. 34, вып. 2. – С. 162–167.

Стаття надійшла до редакції 29.06.2011.
Після доробки 16.11.2011.

Ніконов А. Ю., Небеснюк О. Ю., Шмалій С. Л., Ніконова З. А.

ДОСЛІДЖЕННЯ ОСОБЛИВОСТЕЙ ДЕФЕКТОУТВОРЕННЯ В ВИСОКОЛЕГОВАНОМУ КРЕМНІЇ ПРИ ОПРОМІНЕННІ

У статті наведені результати експериментального дослідження механізмів взаємодії високоенергетичного випромінювання з кремнієм.

Ключові слова: випромінювання, легуючі домішки, концентрація.

Nikonov A. Y., Nebesnjuk O. J., Shmaly S. L., Nikonova Z. A.

RESEARCH OF FEATURES THE FORMATION OF DEFECTS IN HIGHLY ALLOYING SILICON DURING IRRADIATED

In the article the results of experimental research mechanisms interaction high-energy radiation on silicon.

Key words: radiation, doping impurity, concentration.

РАДІОЕЛЕКТРОНІКА ТА ТЕЛЕКОМУНІКАЦІЇ

РАДІОЕЛЕКТРОНИКА ТА ТЕЛЕКОММУНИКАЦИИ

RADIO ELECTRONICS AND TELECOMMUNICATIONS

УДК 621.372.852.001.11

Петрова Е. В.[1], Фурманова Н. И.[2], Фарафонов А. Ю.[3]

[1]Магистр Запорожского национального технического университета
[2]Ассистент, аспирант Запорожского национального технического университета
[3]Канд. техн. наук, доцент Запорожского национального технического университета

РАЗРАБОТКА УПРОЩЕННОГО АЛГОРИТМА ПРОЕКТИРОВАНИЯ МИКРОПОЛОСКОВЫХ ППФ НА ШПИЛЕЧНЫХ РЕЗОНАТОРАХ С ОТВЕРСТИЯМИ В ЭКРАНИРУЮЩЕМ СЛОЕ НА ОСНОВЕ ЭЛЕКТРОДИНАМИЧЕСКОГО АНАЛИЗА В ПРОГРАММЕ ANSOFT HFSS

Проведен анализ микрополосковых полоснопропускающих фильтров (ППФ) на шпилечных резонаторах меандровой формы со щелью в экранирующем слое. Проведены расчеты геометрических параметров ППФ на шпилечных резонаторах на основе фильтров-прототипов нижних частот. Построена модель шпилечного ППФ с отверстиями в экранирующем слое и проведена ее оптимизация. В ходе исследования разработан упрощенный алгоритм проектирования конструкции данного типа фильтров.

Ключевые слова: полоснопропускающий фильтр, связанная линия, щель в экранирующем слое, шпилечные резонаторы, центральная частота.

Развитие систем радиолокации, радионавигации и телекоммуникаций требует создания сверхвысокочастотных (СВЧ) устройств в короткие сроки и с постоянным увеличением требований к их характеристикам. В данных устройствах часто используются традиционные фильтры на параллельных связанных микрополосковых линиях, которые имеют ряд недостатков. Данные фильтры имеют паразитную полосу пропускания на частоте $2f_0$ и из-за больших габаритных размеров конструкции фильтра становится невозможным рациональное использование площади подложки. В случаях, когда имеются ограничения по размерам фильтра, целесообразнее использовать конструкцию фильтров на шпилечных резонаторах. Существует множество конструкций шпилечных фильтров.

Современными исследователями были предложены следующие конструкции фильтров, позволяющие получить желательные характеристики фильтра: повышенную селективность фильтра на шпилечных резонаторах с помощью введения дополнительной емкостной связи [1]; широкую полосу пропускания (до 1 ГГц) с помощью изменения угла наклона между плечами звеньев классического фильтра на шпилечных резонаторах на 75° [2]; миниатюризацию телекоммуникационных устройств приема-передачи и создание изоляции между соседними проводниками фильтров посредством использования многослойной конструкции микрополосковых шпилечных фильтров [3]; уменьшение габаритных размеров фильтра и увеличение ширины микрополосковых линий в конструкции микрополоскового узкополосного шпилечного фильтра с использованием заземления в виде сквозных отверстий [4]. В работе [5] приведено сравнение фильтров на основе шпилечных резонаторов с перекрестными связями в виде каскада из четырех элементов. Рассмотренные фильтры имеют одну или две щели в экранирующем слое с различными видами связей между звеньями шпильки. В работе [6] был исследован микрополосковый фильтр на шпилечных резонаторах, а в статье [7] предлагается конструкция микроплоскового фильтра на связанных линиях с отверстиями в экранирующем слое.

В данной работе предлагается новая конструкция микрополоскового шпилечного фильтра с отверстиями в экранирующем слое. Данный тип конструкции фильтра имеет ряд преимуществ по сравнению с традиционными шпилечными фильтрами. Во-первых, отверстие в экранирующем слое способствует подавлению паразитной полосы пропускания и увеличению затухания в полосе заграждения, а также увеличению ширины полосы пропускания. Во-вторых, такая конструкция позволяет установить скоростное соответствие мод. В-третьих, использование шпилечных фильтров с отверстиями в экранирующем слое позволяет усилить электрическую связь между соседними резонаторами. В-четвертых, увеличивается ширина микрополосковых линий и расстояние между ними, что упрощает технологию изготовления и снижает требования к допускам. Основной проблемой при проектировании фильтров с отверстием в экранирующем слое является то, что не существует алгоритма пересчета его из традиционного фильтра.

Целью данной работы является создание упрощенного алгоритма для проектирования конструкции микрополосковых фильтров на шпилечных резонаторах со щелью в экранирующем слое на основе электромагнитного анализа фильтра, проведенного в системе Ansoft High Frequency Structure Simulator (HFSS).

1. РАСЧЕТЫ И МОДЕЛИРОВАНИЕ

В данной работе для расчета параметров полоснопропускающего фильтра на шпилечных резонаторах используется метод проектирования фильтров на связанных линях на основе фильтров-прототипов нижних частот по методикам, предложенным в [8, 9, 10].

Геометрические параметры шпилечного ППФ, полученные на основе расчета фильтра-прототипа нижних частот, представлены в табл. 1.

Таблица 1. Геометрические параметры ППФ на шпилечных резонаторах

Параметры	Номер связанной линии		
	1 и 5	2 и 4	3
Ширина участка связанных линий W_i, мм	0,4	0,49	0,5
Расстояние между участками связанных линий S_i, мм	0,3	0,89	0,97
Длина участка связанных линий, L_i, мм	10,33	10,23	10,22

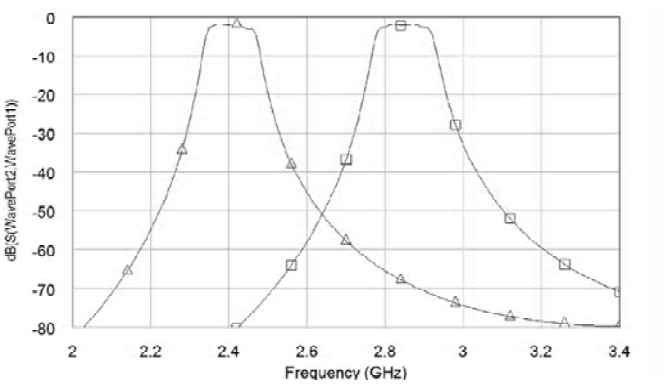


Рис. 1. АЧХ микрополоскового ППФ на шпилечных резонаторах:

– АЧХ схемы с исходными размерами; – АЧХ модифицированной схемы

Фильтр реализуется на подложке из поликора марки ВК–100 с диэлектрической проницаемостью ε=9,6, толщиной h=0,5 мм и слоем металлизации толщиной t=5 мкм. Данный фильтр предназначен для работы на центральной частоте 2,85 ГГц, ширина полосы пропускания 122 МГц.

Для получения амлитудно-частотной характеристики микрополоскового ППФ на шпилечных резонаторах была спроектирована электрическая схема фильтра с помощью группы схемотехнических модулей Schematics в системе Microwave Office от Applied Wave Research (AWR MWO). Из полученной АЧХ (рис. 1) видно, что центральная частота смещена по сравнению с центральной частотой исходной структуры, предложенной в [6], на 440 МГц и составляет 2,41 ГГц. Для смещения центральной частоты до необходимого значения была уменьшена длина участка связанных линий до 8,33 мм с помощью функции Tune системы AWR MWO.

Перерасчет геометрических размеров топологии традиционного микрополоскового шпилечного ППФ для конструкции с отверстиями в экранирующем слое был проведен с помощью программы MaxFCT, которая является собственной разработкой кафедры КТПР Запорожского национального технического университета. Даная программа базируется на квазистатическом анализе поперечного сечения микрополосковой топологии и генетическом алгоритме поиска решений. Результаты расчетов представлены в табл. 2.

Таблица 2. Параметры микрополоскового шпилечного ППФ с отверстиями в экранирующем слое

Параметры	Номер связанной линии		
	1 и 5	2 и 4	3
Волновое сопротивление парных полуволн Z_{oe}, Ом	67,31	52,65	51,46
Волновое сопротивление непарных полуволн Z_{oo}, Ом	40,87	44,10	43,54
Ширина участка связанных линий W_i, мм	0,608	0,793	0,806
Расстояние между участками связанных линий S_1, мм	0,410	1,810	1,818
Ширина отверстия в экранирующем шаре S_2, мм	1,17	2,52	2,40

Для проведения электромагнитного анализа, была построена модель микрополоскового шпилечного ППФ с отверстиями в экранирующем слое, с использованием системы Ansoft HFSS (рис. 2). Длина щелей в экранирующем слое равна длине участков связанных линий.

Из полученной АЧХ (рис. 3) видно, что введение щели в экранирующий слой привело к изменению центральной частоты полосы пропускания, которая составляет 2,967 ГГц, поэтому необходимо провести ее корректировку.

Для получения необходимого значения центральной частоты, была проведена оптимизация конструкции данного фильтра с использованием метода генетического алгоритма, с помощью программы Optimetrics, которая входит в состав системы HFSS. В данной работе переменной для оптимизации была выбрана длина участков связанных линий и длина отверстий в экранирующем слое *L*.

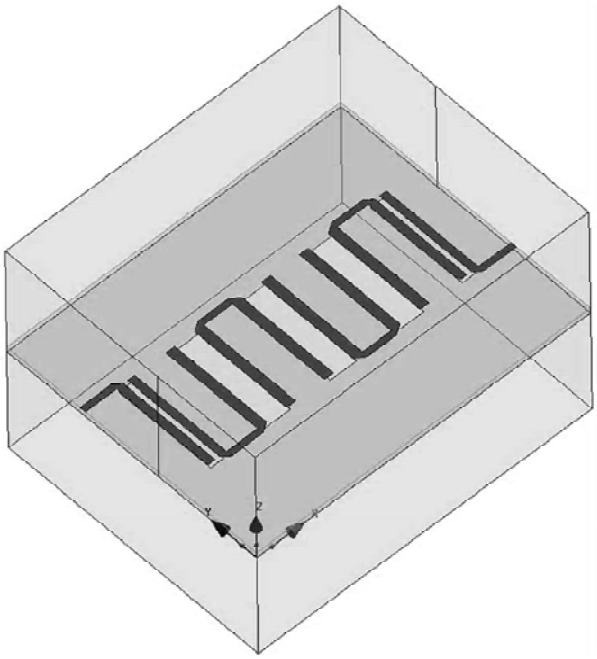


Рис. 2. Модель микрополоскового шпилечного фильтра с отверстиями в экранирующем слое

Установление изменяемой переменной осуществлялось путем замены значения длины участков связанных линий и длины отверстий в экранирующем слое на *L*. Для того, чтобы уменьшить диапазон принимаемых значений переменной и, тем самым, сократить время проведения процесса оптимизации, предварительно был проведен параметрический анализ. Также, для решения параметрических задач была создана целевая функция, которая описывает расположение частотных точек при построении АЧХ. После проведения параметрического анализа был запущен процесс оптимизации. Из полученного графика значений целевой функции относительно номера итерации было получено оптимальное значение длины участков связанных линий и щелей в экранирующем слое. Минимальному значению целевой функции соответствовало оптимальное значение L = 8,83 мм, которое было получено на второй итерации из проведенных шестидесяти шести. Исходя из полученных результатов оптимизации, была перестроена модель исследуемого фильтра (изменена длина участков связанных линий) и проведен перерасчет частотной характеристики (рис. 4).

Из полученной АЧХ видно, что центральная частота составила 2,85 ГГц, что соответствует исходным данным. Также видно, что при введении щели в экранирующий слой увеличилась ширина полосы пропускания исследуемого фильтра, которая составляет 160 МГц.

2. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Для упрощения проектирования микрополосковых ППФ на шпилечных резонаторах с отверстиями в экранирующем слое, которые базируются на основе традиционных микрополосковых фильтров, предложен следующий алгоритм:

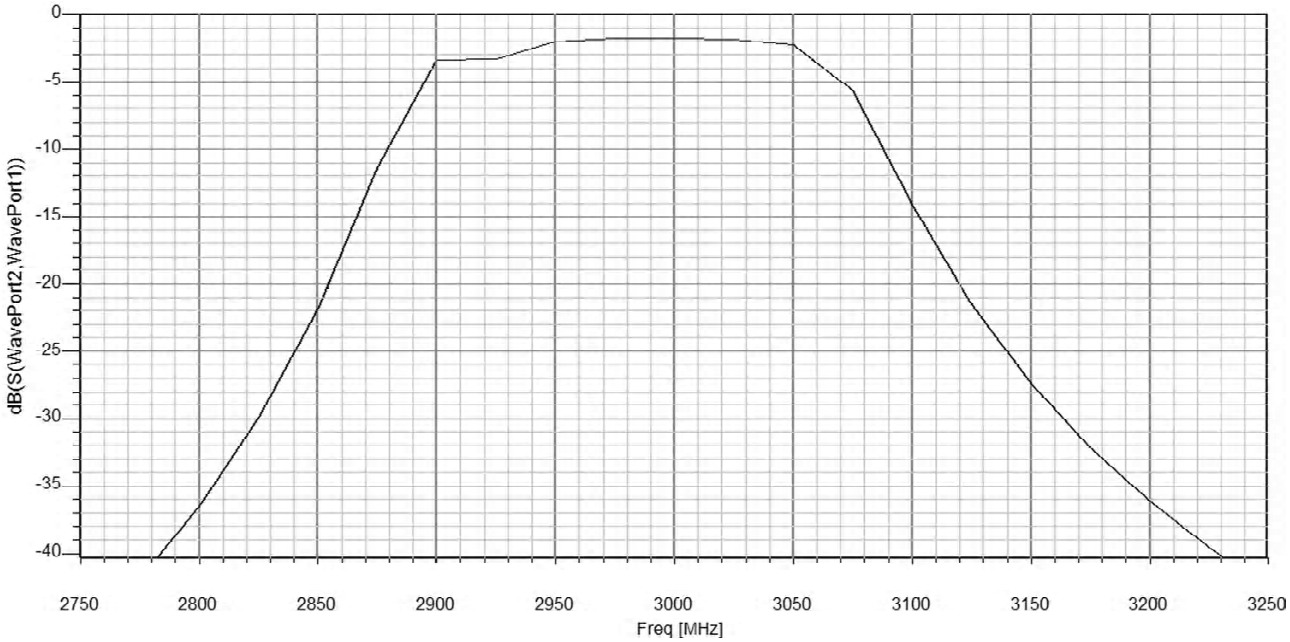


Рис. 3. АЧХ микрополоскового шпилечного фильтра с отверстиями в экранирующем слое

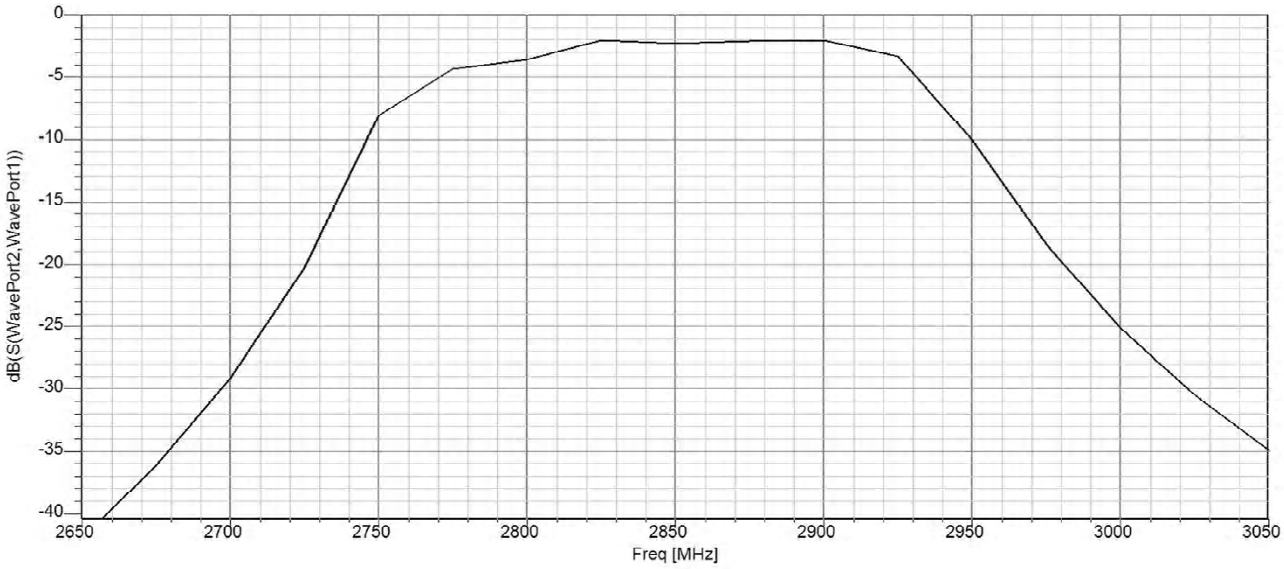


Рис. 4. АЧХ оптимизированного микрополоскового шпилечного фильтра с отверстиями в экранирующем слое

1. Расчет геометрических размеров топологии фильтров на связанных полуволновых резонаторах на основе фильтров-прототипов нижних частот по методикам, предложенным в [8, 9, 10].

2. Построение электрической схемы микрополоскового шпилечного фильтра с помощью группы схемотехнических модулей Schematics системы AWR MWO. Получение АЧХ микрополоскового шпилечного фильтра. Изменение длины связанных линий для смещения центральной частоты с помощью функции Tune системы AWR MWO.

3. Проведение расчетов геометрических размеров поперечного сечения топологии микрополоскового ППФ на шпилечных резонаторах с отверстиями в экранирующем слое с помощью программы MaxFCT.

4. Построение и электромагнитный анализ конструкции шпилечного ППФ с отверстиями в экранирующем слое в системе HFSS на основе результатов расчетов в программе AWR MWO (длина участка связанных линий) и программы MaxFCT (ширина микрополосковых линий, расстояние между ними и ширина отверстий в экранирующем слое).

5. Проведение оптимизации конструкции микрополоскового ППФ на шпилечных резонаторах с отверстиями в экранирующем слое с помощью программы Optimetrics в системе HFSS для получения необходимой АЧХ.

6. Построение микрополоскового шпилечного фильтра с отверстиями в экранирующем слое, проверка результатов.

Целью дальнейших исследований является исследование микрополосковых ППФ на шпилечных резонаторах с варьирующимися углами наклона плечей шпилечных резонаторов [2] с щелями в экранирующем слое.

ВЫВОДЫ

Проведены расчеты геометрических параметров топологии микрополоскового шпилечного ППФ со щелью в экранирующем слое. По полученным результатам спроектирована конструкция данного фильтра. Проведена оптимизация длины участков связанных линий и щелей в экранирующем слое. В ходе работы был создан упрощенный алгоритм для проектирования конструкции микрополосковых фильтров на шпилечных резонаторах со щелью в экранирующем слое, на основе электромагнитного анализа фильтра.

СПИСОК ЛИТЕРАТУРЫ

1. *Николаев, М.* Компактные микрополосковые фильтры с повышенной селективностью / М. Николаев // Современная электроника. – 2008. – № 1. – С. 28–30.
2. *Lotfi Neyestanak, A. A.,* Enhanced Wide Band Microstrip Hairpin Filter / *A. A. Lotfi Neyestanak* // Journal of Mobile Communication. – 2009. – Vol. 3, № 3. – P. 59–61.
3. *Sulaiman, A.* Simple Multilayer Hairpin Bandpass Filter / A. Sulaiman, H. Mokhtar, H. Jusoh, H. Baba, A. Awang, F. Ain // European Journal of Scientific Research. – 2010. – Vol. 42, № 4. – P. 604–613.
4. *Hasan, A.* Novel Microstrip Hairpinline Narrowband Bandpass Filter Using Via Ground Holes / A. Hasan, A. E. Nadeem // Progress In Electromagnetics Research. – 2008. – № 78. – P. 393–419.
5. *Militaru, N.* Enhanced Couplings in Broadband Planar Filters with Defected Ground Structures / N. Militaru, M. G. Banciu, G. Lojewski // Romanian Journal of Information. – 2007. – Vol. 10, № 2. – P. 199–212.
6. *Гипсман, А. И.* Современные методы и результаты квазистатического анализа полосковых линий и устройств / А. И. Гипсман, В. М. Красноперкин, Г. С. Самохин, Р. А. Силин // Обзоры по электронной технике. Серия 1. Электроника СВЧ. Выпуск 1 (1602) – М. : ЦНИИ «Электроника», 1991. – 94 с.
7. *Velazquez-Ahumada, M.-C.* Parallel coupled microstrip filters with ground-plane aperture for spurious band suppression and enhanced coupling / M.-C. Velazquez-Ahumada, J. Martel, F. Medina // IEEE trans. on microwave theory and techniques. – 2004. – Vol. 52, № 3. – P. 1082 – 1086.
8. *Маттей, Д. Л.* Фильтры СВЧ, согласующие цепи и цепи связи, т. I / Д. Л. Маттей, Л. Янг, Е. М. Т. Джонс : пер с

англ. ; под общ. ред. Л. В. Алексеева и Ф. В. Кушнира. – М. : Связь, 1971. – 440 с.
9. *Малорацкий, Л. Г.* Проектирование и расчет СВЧ элементов на полосковых линиях / Л. Г. Малорацкий, Л. Р. Явич. – М. : Сов. радио, 1972. – 232 с.
10. Синтез микрополосковых полосовых фильтров на связанных линиях с отверстиями в экране / А. Ю. Фарафонов, А. Ю. Воропай, Л. М. Карпуков [та ін.] // Радіоелектроніка, інформатика, управління. – 2009. – № 1 (20). – С. 41–44.

Стаття надійшла до редакції 19.01.2012.
Після доробки 26.01.2012.

Петрова К. В., Фурманова Н. І., Фарафонов О. Ю.

РОЗРОБКА СПРОЩЕНОГО АЛГОРИТМУ ПРОЕКТУВАННЯ МІКРОСМУЖКОВИХ СПФ НА ШПИЛЬКОВИХ РЕЗОНАТОРАХ З ОТВОРАМИ В ЕКРАНУЮЧОМУ ШАРІ НА ОСНОВІ ЕЛЕКТРОДИНАМІЧНОГО АНАЛІЗУ В ПРОГРАМІ ANSOFT HFSS

Проведений аналіз мікросмужкових смугопропускаючих фільтрів (СПФ) на шпилькових резонаторах меандрової форми зі щілиною в екрануючому шарі. Проведено розрахунки геометричних параметрів СПФ на шпилькових резонаторах на основі фільтрів-прототипів нижніх частот. Побудована модель шпилькового СПФ з отворами в екрануючому шарі та проведена її оптимізація. В ході дослідження розроблений спрощений алгоритм проектування конструкції даного типу фільтрів.

Ключові слова: смугопропускаючий фільтр, зв'язана лінія, щілина в екрануючому шарі, шпилькові резонатори, центральна частота.

Petrova K. V., Furmanova N. I., Farafonov A. Y.

DEVELOPMENT OF SIMPLIFIED ALGORITHM FOR THE DESIGN OF MICROSTRIP BAND-PASS ON HAIRPIN RESONATORS FILTERS WITH SLOTS IN THE GROUND PLANE ON THE ELECTRODYNAMICS ANALYSIS IN ANSOFT HFSS

Analysis of microstrip band-pass filters hairpin resonators meander shape with slots in the ground plane is presented. The calculations of the geometric parameters of the band-pass on hairpin resonator filters on the basis of prototype filters of reduce frequencies are presented. A model of the band-pass on hairpin resonator filters with slots in the ground plane has been designed and was used optimization. In research a simplified algorithm for the design process of this type of filters is proposed.

Key words: pass-band filter, coupled line, slots in the ground plane, hairpin resonators, center frequency.

МАТЕМАТИЧНЕ ТА КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ

МАТЕМАТИЧЕСКОЕ И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

MATHEMATICAL AND COMPUTER MODELLING

УДК 62-55:681.515

Гостев В. И.

Д-р техн. наук, заведующий кафедрой Государственного университета информационно-коммуникационных технологий г. Киев

ПРОЕКТИРОВАНИЕ ТРЕХРЕЖИМНОГО НЕЧЕТКОГО РЕГУЛЯТОРА ДЛЯ СИСТЕМ АКТИВНОГО УПРАВЛЕНИЯ ОЧЕРЕДЬЮ В TCP/IP СЕТЯХ

Изложены вопросы проектирования трехрежимного нечеткого регулятора при идентичных треугольных функциях принадлежности с тремя термами для систем активного управления очередью в TCP/IP сетях и представлена принципиальная схема регулятора в системе MATLAB.

Ключевые слова: проектирование, нечеткий регулятор, функции принадлежности, активное управление очередью, TCP/IP сети, MATLAB.

ВВЕДЕНИЕ

В последние годы непредсказуемый рост Интернета все более и более указывает на проблему борьбы с перегрузками. Явление перегрузок сети происходит, когда количество данных, введенных в сеть, больше чем количество данных, которые доставляются по назначению. Широко используется активное управление очередью пакетов. При пакетной передаче источники данных должны уменьшать скорость передачи при наличии потери пакетов. Для эффективной борьбы с перегрузками предложены фаззи-регуляторы (регуляторы, работающие на базе нечеткой логики) активного управления очередью. Проектирование таких регуляторов является сложным из-за эвристически вовлеченных в управление правил и функций принадлежности, отсутствия автоматических методов проектирования нечеткой базы знаний и настройки параметров регулятора. Поэтому параметры нечеткой системы управления обычно настраиваются посредством проб и ошибок с использованием эвристических методов и моделирования. Особые сложности возникают из-за широкого диапазона эксплуатационных условий, например, числа соединений, емкости связей, задержек распространения. При синтезе нечетких регуляторов наиболее часто используются треугольные функции принадлежности (ФП) для лингвистических величин. При расчете управляющих воздействий на выходе нечеткого регулятора абсциссу «центра тяжести результирующей фигуры» определяют обычно приближенным методом численного интегрирования. В работах [1–8] по системам активного управления очередью с применением нечетких регуляторов не приводятся сведения по проектированию регуляторов, а в работе [1] изложены неправильные результаты выполнения и настройки регуляторов, хотя в этих работах отмечается, что нечеткий регулятор является одной из главных составляющих таких систем. В предлагаемой ниже работе на основе нового метода проектирования нечетких регуляторов, изложенного в работах [9–11], получены строгие аналитические выражения для управляющих воздействий на выходе трехрежимного нечеткого регулятора при идентичных треугольных функциях принадлежности с тремя термами. В качестве входных воздействий на регулятор, кроме ошибки системы, рассматриваются первая и вторая производные ошибки. Представлена принципиальная схема нечеткого регулятора и изложены вопросы настройки регулятора с проверкой правильности его работы.

РЕШЕНИЕ ЗАДАЧИ

В данной работе рассмотрим нечеткий регулятор, структурную схему которого в интерактивной системе MATLAB можно представить в виде последовательного соединения трех блоков (см. рис. 1): **формирователя ве-**

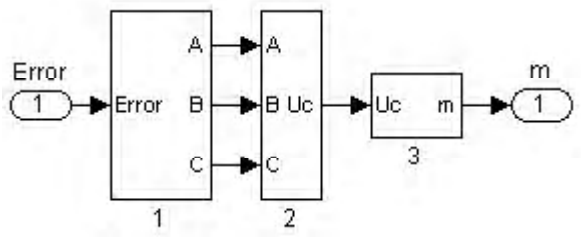


Рис. 1.

личин A(t), B(t) и **C(t)** (блок 1), **блока сравнения величин A(t), B(t)** и **C(t) и расчета** u_c (блок 2) и **блока нормировки выходной переменной** (блок 3) [10]. На вход регулятора поступают ошибка системы θ, скорость изменения (первая производная) ошибки $\dot{\theta}$, ускорение (вторая производная) ошибки $\ddot{\theta}$, m – выходная величина регулятора.

Блоки оценки первой (**1-st drv**) и второй (**2-nd drv**) производных от ошибки реализуют уравнения:

$$\begin{aligned} \dot{\theta}(t) &\approx \{\theta(kh) - \theta[(k-1)h]\}/h, \\ \ddot{\theta}(t) &\approx \{\dot{\theta}(kh) - \dot{\theta}[(k-1)h]\}/h, \end{aligned} \tag{1}$$

где h – шаг квантования (шаг поступления информации на вход регулятора).

На универсальном множестве $U = [-1,\ 1]$ заданы три нечетких подмножества, ФП которых для каждой лингвистической величины определяются по формулам:

$$\mu_1(u) = \begin{cases} 1, & -1 \le u \le -a; \\ (a-u)/(2a), & -a \le u \le a; \\ 0, & a \le u \le 1; \end{cases}$$

$$\mu_2(u) = \begin{cases} 0, & -1 \le u \le -a; \\ (a+u)/(2a), & -a \le u \le a; \\ 1, & a \le u \le 1; \end{cases}$$

$$\mu_3(u) = \begin{cases} (a+u)/a, & -a \le u \le 0; \\ (a-u)/a, & 0 \le u \le a. \end{cases} \tag{2}$$

При поступлении на нечеткий регулятор в какой-то момент времени значений входных переменных, θ^*, $\dot{\theta}^*$ и $\ddot{\theta}^*$ с шагом квантования h осуществляется пересчет входных переменных в переменные u_1^*, u_2^*, u_3^* на универсальное множество $U = [-1,\ 1]$. Пересчет фиксированного значения любой переменной $x^* \in [x_{\min}, x_{\max}]$ в соответствующий элемент $u^* \in [-1,\ 1]$ единого универсального множества (см. рис. 2) при фаззификации и дефаззификации определяется пропорцией:

$$(x_{\max} - x_{\min})/2 = (x^* - x_{\min})/(u^* + 1),$$

откуда

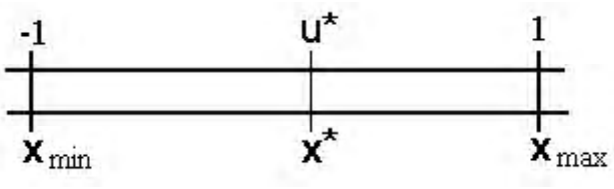


Рис. 2.

$$\left.\begin{aligned} u^* &= 2(x^* - x_{\min})/(x_{\max} - x_{\min}) - 1 \\ x^* &= (x_{\max} - x_{\min})(u^* + 1)/2 + x_{\min} \end{aligned}\right\}. \tag{3}$$

Затем производится расчет значений ФП для этих переменных (см. рис. 3).

Точками на универсальном множестве отмечены возможные для какого-то момента времени значения переменных u_1^*, u_2^*, u_3^*. Для упрощения нормировки (пересчета значений сигналов в значения элементов единого универсального множества) диапазоны изменения входных и выходного сигналов (параметров нечеткого регулятора) принимаем симметричными:

$$A_m = \theta_{\max} = -\theta_{\min};\ \ B_m = \dot{\theta}_{\max} = -\dot{\theta}_{\min};$$
$$C_m = \ddot{\theta}_{\max} = -\ddot{\theta}_{\min};\ \ D_m = m_{\max} = -m_{\min}.$$

Тогда формулы для нормировки (пересчета) принимают вид:

$$\left.\begin{aligned} u_1^* &= (\theta^* + A_m)/A_m - 1; \\ u_2^* &= (\dot{\theta}^* + B_m)/B_m - 1; \\ u_3^* &= (\ddot{\theta}^* + C_m)/C_m - 1. \end{aligned}\right\} \tag{4}$$

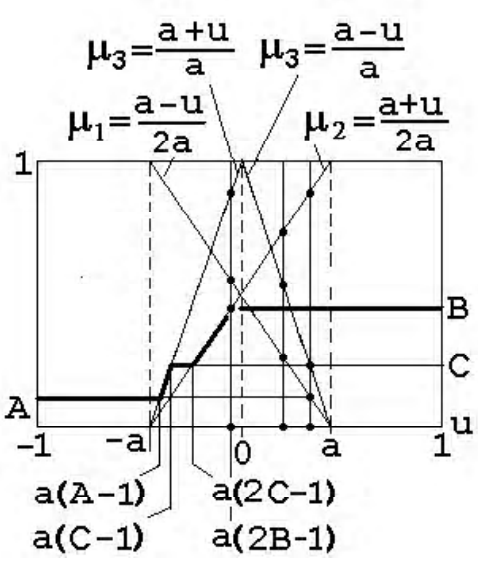
Лингвистическое правило управления нечеткого регулятора:

$$\text{если } (\theta^* = a_1^j) \text{ и } (\dot{\theta}^* = a_2^j) \text{ и } (\ddot{\theta}^* = a_3^j), \text{ то}$$
$$(m^* = a_c^j),\ j = \overline{1,3}, \tag{5}$$

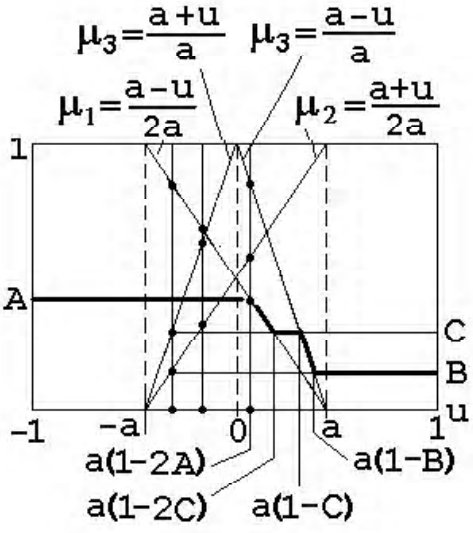
где a_1^j, a_2^j и a_3^j – лингвистические оценки ошибки, первой производной ошибки и второй производной ошибки, рассматриваемые как нечеткие терм-множества, определенные на универсальном множестве, $j = \overline{1,3}$; a_c^j – лингвистические оценки управляющего воздействия на объект, выбираемые из терм-множества переменной m. Лингвистические оценки выбираются из терм-множеств лингвистических переменных θ^*, $\dot{\theta}^*$ и $\ddot{\theta}^*$ и m^*:

$$a_i^j \in \{\textit{отрицательная}\ (1),\ \ \textit{положительная}\ (2),$$
$$\textit{близкая к нулю – нулевая}\ (3)\}.$$

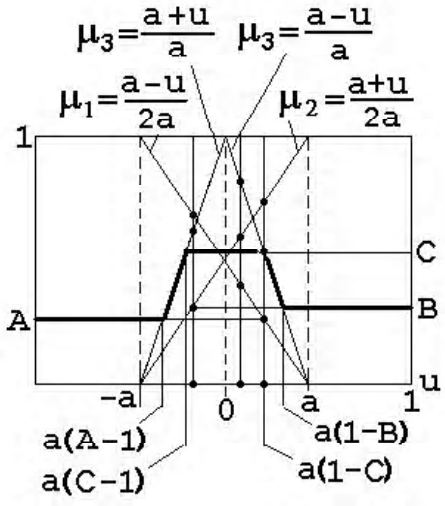
Другими словами, все сигналы (определенные выше лингвистические переменные) характеризуются как отрицательные (j=1), положительные (j=2) или близкие к нулю (j=3).



а)



б)



в)

Рис. 3.

Функция принадлежности управляющего воздействия нечеткому множеству «отрицательная» определяется из системы нечетких логических уравнений:

$$\mu_{1c}(u) = \mu_1(u_1) \wedge \mu_1(u_2) \wedge \mu_1(u_3). \quad (6)$$

Функция принадлежности управляющего воздействия нечеткому множеству «положительная» определяется из системы нечетких логических уравнений:

$$\mu_{2c}(u) = \mu_2(u_1) \wedge \mu_2(u_2) \wedge \mu_2(u_3). \quad (7)$$

Функция принадлежности управляющего воздействия нечеткому множеству «близкая к нулю» определяется из системы нечетких логических уравнений:

$$\mu_{3c}(u) = \mu_3(u_1) \wedge \mu_3(u_2) \wedge \mu_3(u_3). \quad (8)$$

Результирующая функция принадлежности для управляющего воздействия в соответствии с рабочим правилом НР записывается в виде:

$$\mu_c(u) = \mu_{1c}(u) \vee \mu_{2c}(u) \vee \mu_{3c}(u). \quad (9)$$

В выражениях (6)–(9) $\wedge$ – логическое «и», $\vee$ – логическое «или».

В соответствии с лингвистическими правилами управления функция принадлежности управляющего воздействия $\mu_{1c}(u)$ нечеткому множеству «отрицательная» ограничена сверху значением:

$$A = \min[\mu_1(u_1^*), \mu_1(u_2^*), \mu_1(u_3^*)], \quad (10)$$

функция принадлежности управляющего воздействия $\mu_{2c}(u)$ нечеткому множеству «положительная» ограничена сверху значением:

$$B = \min[\mu_2(u_1^*), \mu_2(u_2^*), \mu_2(u_3^*)], \quad (11)$$

функция принадлежности управляющего воздействия $\mu_{3c}(u)$ нечеткому множеству «близкая к нулю» ограничена сверху значением:

$$C = \min[\mu_3(u_1^*), \mu_3(u_2^*), \mu_3(u_3^*)]. \quad (12)$$

Результирующая функция принадлежности для управляющего воздействия на основании выражения (9) получается путем формированием максимума:

$$\mu_c(u) = \max[\mu_{1c}(u), \mu_{2c}(u), \mu_{3c}(u)]. \quad (13)$$

Для определения конкретного значения управляющего воздействия m^* формируется «результирующая фигура», ограниченная результирующей ФП, и производится поиск абсциссы «центра тяжести результирующей фигуры» u_c.

Отметим весьма существенный факт. Какие бы значения не принимали переменные u_1^*, u_2^*, u_3^* на универсальном множестве $U = [-1, 1]$, в зависимости от соотношений величин A и B, «результирующая фигура» может принимать только три конфигурации: при $A \le C \le B$ первая конфигурация показана на рис.3, *а*; при

$A \ge C \ge B$ вторая конфигурация показана на рис. 3, *б*; при $\begin{cases} A \le B \le C \\ B \le A \le C \end{cases}$ третья конфигурация показана на рис. 3, *в*.

Общая формула для определения абсциссы «центра тяжести результирующей фигуры» записывается в виде:

$$u_c = \int_{-1}^{1} u\mu(u)du \Big/ \int_{-1}^{1} \mu(u)du. \qquad (14)$$

Абсцисса «центра тяжести результирующей фигуры» при $A \le C \le B$ определяется по формуле:

$$u_c = \frac{A\int\limits_{-1}^{a(A-1)} udu + \int\limits_{a(A-1)}^{a(C-1)} \frac{a+u}{a} udu + C\int\limits_{a(C-1)}^{a(2C-1)} udu + \int\limits_{a(2C-1)}^{a(2B-1)} \frac{a+u}{2a} udu + B\int\limits_{a(2B-1)}^{1} udu}{A\int\limits_{-1}^{a(A-1)} du + \int\limits_{a(A-1)}^{a(C-1)} \frac{a+u}{a} du + C\int\limits_{a(C-1)}^{a(2C-1)} du + \int\limits_{a(2C-1)}^{a(2B-1)} \frac{a+u}{2a} du + B\int\limits_{a(2B-1)}^{1} du} \qquad (15)$$

при $A \le C \le B$.

После несложных вычислений находим:

$$u_c = \frac{(a^2-1)(A-B)/2 + a^2B^2 - a^2(A^2+C^2)/2 + a^2(A^3-4B^3+3C^3)/6}{(1-a)A + (1+a)B - aB^2 + a(A^2+C^2)/2} \qquad (16)$$

при $A \le C \le B$.

Абсцисса «центра тяжести результирующей фигуры» при $A \ge C \ge B$ определяется по формуле:

$$u_c = \frac{A\int\limits_{-1}^{a(2A-1)} udu + \int\limits_{a(2A-1)}^{a(2C-1)} \frac{a+u}{2a} udu + C\int\limits_{a(2C-1)}^{a(1-C)} udu + \int\limits_{a(1-C)}^{a(1-B)} \frac{a-u}{a} udu + B\int\limits_{a(1-B)}^{1} udu}{A\int\limits_{-1}^{a(2A-1)} du + \int\limits_{a(2A-1)}^{a(2C-1)} \frac{a+u}{2a} du + C\int\limits_{a(2C-1)}^{a(1-C)} du + \int\limits_{a(1-C)}^{a(1-B)} \frac{a+u}{a} du + B\int\limits_{a(1-B)}^{1} du} \qquad (17)$$

при $A \ge C \ge B$.

После несложных вычислений находим:

$$u_c = \frac{(a^2-1)(A-B)/2 - a^2A^2 + a^2(B^2+C^2)/2 + a^2(4A^3-B^3-3C^3)/6}{(1+a)A + (1-a)B - aA^2 + a(B^2+C^2)/2} \qquad (18)$$

при $A \ge C \ge B$.

В качестве примера приведем следующие результаты расчетов при a=0,2:

– при A=0,1, B=0,4, C=0,2 получаем $u_c = 0{,}2774$;

– при A=0,4, B=0,1, C=0,2 получаем $u_c = -0{,}2774$.

Абсцисса «центра тяжести результирующей фигуры» при $\begin{cases} A \le B \le C \\ B \le A \le C \end{cases}$ определяется по формуле:

$$u_c = \frac{A\int\limits_{-1}^{a(A-1)} udu + \int\limits_{a(A-1)}^{a(C-1)} \frac{a+u}{a} udu + C\int\limits_{a(C-1)}^{a(1-C)} udu + \int\limits_{a(1-C)}^{a(1-B)} \frac{a-u}{a} udu + B\int\limits_{a(1-B)}^{1} udu}{A\int\limits_{-1}^{a(A-1)} du + \int\limits_{a(A-1)}^{a(C-1)} \frac{a+u}{a} du + C\int\limits_{a(1-C)}^{a(1-C)} du + \int\limits_{a(1-C)}^{a(1-B)} \frac{a-u}{a} du + B\int\limits_{a(1-B)}^{1} du} \qquad (19)$$

при $\begin{cases} A \le B \le C \\ B \le A \le C \end{cases}$.

После несложных вычислений находим:

$$u_c = \frac{(a^2-1)(A-B)/2 + a^2(B^2-A^2)/2 + a^2(A^3-B^3)/6}{(1-a)A + (1-a)B + 2aC - aC^2 + a(A^2+B^2)/2} \text{ при} \begin{cases} A \le B \le C \\ B \le A \le C \end{cases}. \quad (20)$$

В качестве примера приведем следующие результаты расчетов при a=0,2:

– при A=0,2, B=0,3, C=0,4 получаем $u_c = 0{,}0903$;

– при A=0,3, B=0,2, C=0,4 получаем $u_c = -0{,}0903$.

Полученное значение u_c затем преобразуется в значение управляющего воздействия на объект управления (согласно формуле (3)):

$$m^* = D_m u_c^*. \quad (21)$$

В качестве примера приведем следующие результаты расчетов при a=0,2.

– при A=0,2, B=0,3, C=0,4 получаем $u_c = 0{,}0903$;

– при A=0,3, B=0,2, C=0,4 получаем $u_c = -0{,}0903$.

Полученное значение u_c затем преобразуется в значение управляющего воздействия на объект управления (согласно формуле (3)):

$$m^* = D_m u_c^*. \quad (21)$$

Отметим, что при фиксированных A и B величина C имеет строго определенное значение. Если $A \le B$, то величина C определяется из следующих соотношений:

$$\mu_1 = (a-u^*)/(2a) = A; \Rightarrow u^* = a(1-2A); \Rightarrow \\ \Rightarrow \mu_3 = C = (a-u^*)/a = 2A. \quad (22)$$

Если $A \ge B$, то величина C определяется из следующих соотношений:

$$\mu_2 = (a+u^*)/(2a) = B; \Rightarrow u^* = a(2B-1); \Rightarrow \\ \Rightarrow \mu_3 = C = (a+u^*)/a = 2B. \quad (23)$$

Формирователь величин A(t), B(t) и **C(t)** (блок 1 на рис. 1) проектируется на основании формул (1), (2), (4), (10), (11), (22) и (23). Этот блок показан на рис. 4.

В формирователе ошибка рассогласования квантуется аналого-цифровым преобразователем (АЦП) **(Zero-Order Hold)** с шагом квантования (шагом поступления данных в нечеткий регулятор) h. Ошибка $\theta(k)$ с выхода АЦП, ее первая $\dot{\theta}(k) = [\theta(k) - \theta(k-1)]/h$ и вторая $\ddot{\theta}(k) = [\dot{\theta}(k) - \dot{\theta}(k-1)]/h$ разности (формула (1)) подаются на вход блока нормировки входных переменных, который построен по формулам (4). На выходе блоков **Product, Product1, Product2** структурной схемы формирователя величин **A(t), B(t)** и **C(t)** с учетом «минус единицы» получаем переменные u_i (соответственно u_1, u_2, u_3). Элементами ограничения **(Saturation)** моделируем подмножество ($-a \le u \le a$) универсального множества $U = [-1, 1]$, на которое поступают переменные $u_i, i = 1,2,3$. В блоках **Fcn, Fcn1, Fcn2** записываем

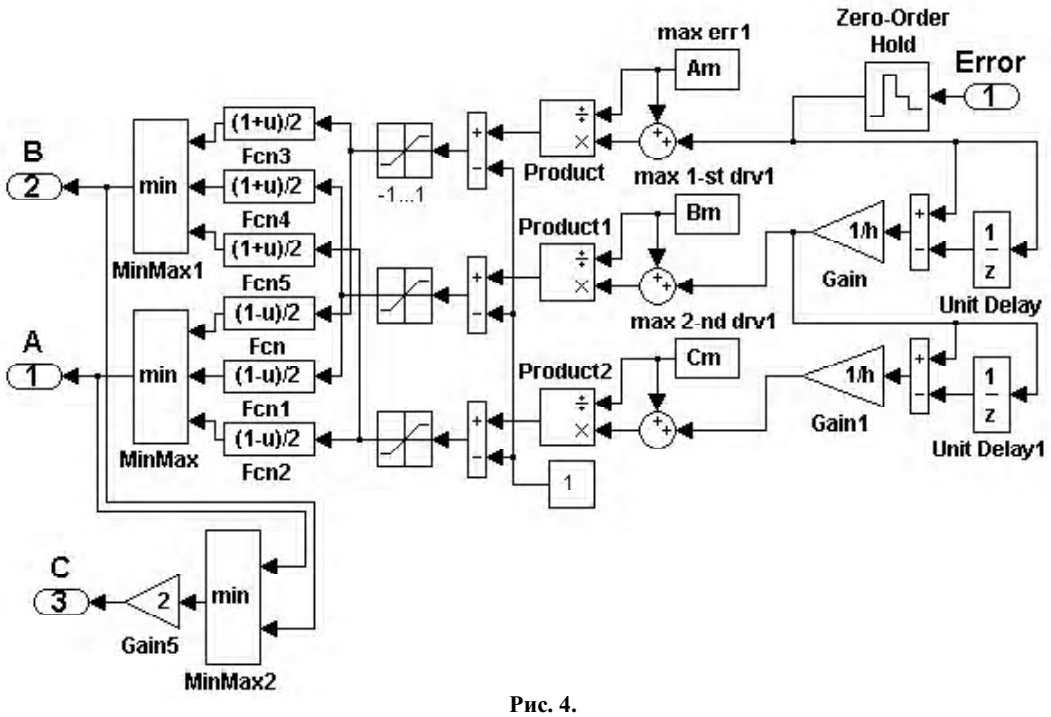


Рис. 4.

аналитические выражения для функций принадлежности $\mu_1(u)$, а в блоках **Fcn3, Fcn4, Fcn5** – аналитические выражения для функций принадлежности $\mu_2(u)$ (формулы (2), (22), (23)). На выходе блоков **Fcn, Fcn1, Fcn2** получаем переменные $\mu_1(u_i)$(соответственно $\mu_1(u_1)$, $\mu_1(u_2)$, $\mu_1(u_3)$), а на выходе блоков **Fcn3, Fcn4, Fcn5** получаем переменные $\mu_1(u_i)$ (соответственно $\mu_2(u_1)$, $\mu_2(u_2)$, $\mu_2(u_3)$). Выражения (10) и (11) вычисляются в блоках **MinMax** и **MinMax1**, на выходе которых получаем значения переменных **A(t)** и **B(t).** Вычисление величины **C(t)** осуществляется таким образом: поскольку при **A(t)** $\leq$ **B(t)** величина **C(t) = 2 A(t),** а при **A(t)** $\geq$ **B(t)** величина **C(t) = 2 B(t),** то достаточно определить меньшую величину и увеличить ее значение в два раза (см. соотношения (22) и (23)). Для этого служат блоки **MinMax2 и Gain5**.

Значения диапазонов $A_m = \theta_{\max} = -\theta_{\min}$; $B_m = \dot{\theta}_{\max} = -\dot{\theta}_{\min}$; $C_m = \ddot{\theta}_{\max} = -\ddot{\theta}_{\min}$ при настройке нечеткого регулятора подбираются либо вручную, либо автоматически путем решения оптимизационной задачи.

Блок сравнения величин A(t), B(t) и C(t) и расчета u_c (блок 2 на рис. 1) проектируется на основании формул (16), (18) и (20). Этот блок показан на рис. 5.

На выходе делителя **Product** формируется величина u_c на основании формулы (16) при $A \leq C \leq B$. На выходе делителя **Product1** формируется величина u_c на основании формулы (18) при $A \geq C \geq B$. На выходе делителя **Product2** формируется величина u_c на основании формулы (20) при $\begin{cases} A \leq B \leq C \\ B \leq A \leq C \end{cases}$.

Переключатели **Switch** и **Switch1** замыкают верхние контакты при условии $A \leq C \leq B$, когда на средних контактах этих переключателей сигналы положительные (в блоках **Switch** и **Switch1** параметр **Threshold**=0,000001). При условии $A \geq C \geq B$, когда на средних контактах переключателей **Switch** и **Switch1** сигналы отрицательные, переключатели замыкают нижние контакты.

Переключатели **Switch2** и **Switch3** замыкают верхние контакты при условии $\begin{cases} A \leq B \leq C \\ B \leq A \leq C \end{cases}$, когда на средних контактах этих переключателей сигналы положительные (в блоках **Switch2** и **Switch3** параметр **Threshold**=0,000001).

При условии $A \leq C \leq B$, когда на среднем контакте переключателя **Switch2** сигнал положительный, а на среднем контакте переключателя **Switch3** сигнал отрицательный, то в переключателе **Switch2** замкнут верхний контакт, а в переключателе **Switch3** замкнут нижний контакт.

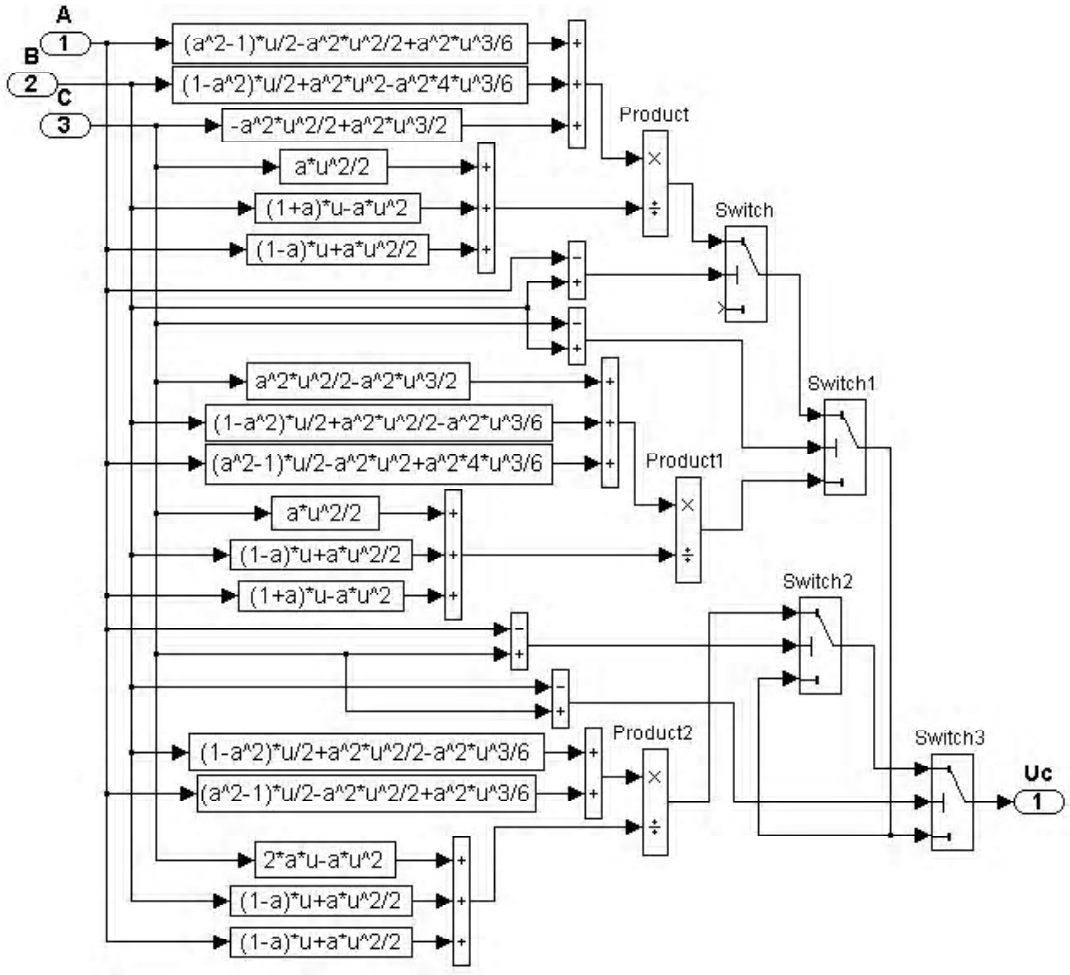


Рис. 5.

При условии $A \ge C \ge B$, когда на среднем контакте переключателя **Switch3** сигнал положительный, а на среднем контакте переключателя **Switch2** сигнал отрицательный, то в переключателе **Switch3** замкнут верхний контакт, а в переключателе **Switch2** замкнут нижний контакт.

Таким образом, при условии $A \le C \le B$ сигнал на выход схемы поступает с выхода делителя **Product**, при условии $A \ge C \ge B$ сигнал на выход схемы поступает с выхода делителя **Product1** и при условии $\begin{cases} A \le B \le C \\ B \le A \le C \end{cases}$ сигнал на выход схемы поступает с выхода делителя **Product2.**

Блок нормировки выходной переменной (блок 3 на рис. 1) с цифро-аналоговым преобразователем (ЦАП) (Zero-Order Hold1 – фиксатором нулевого порядка с передаточной функцией $H(s) = (1 - e^{-hs})/s$), полученный на основании формулы (21), показан на рис. 6. Граничное значение диапазона $D_m = m_{\max} = -m_{\min}$ является параметром, который перестраивается при настройке нечеткого регулятора.

Таким образом, разработана полная принципиальная схема нечеткого регулятора и определены параметры регулятора, необходимые для его настройки. Предложенная схема может практически использоваться как в системах активного управления очередью пакетов в TCP/IP, так и любых других системах автоматического управления.

Логика работы нечеткого регулятора приведена на рис. 7. В регуляторе *входные* функции принадлежности (в формирователе) идентичны *выходным* (по которым рассчитаны формулы в блоке сравнения).

В **формирователе величин A(t), B(t) и C(t)** на входе нечеткого регулятора переменные θ^*, $\dot{\theta}^*$ и $\ddot{\theta}^*$, поступающие в регулятор с шагом квантования h, пересчитываются в переменные u_1^*, u_2^*, u_3^* по формулам (4) и производится расчет значений *входных* функций принадлежности (см. рис. 3 и 7, *а*) для переменных u_1^*, u_2^*, u_3^*. На основе алгоритма Мамдани определяются уровни отсечения A, B и C по формулам (10–12).

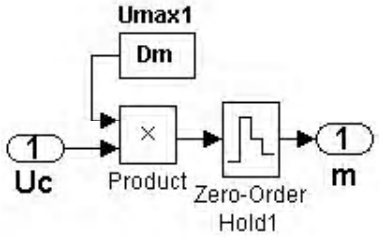


Рис. 6.

В **блоке сравнения величин A(t), B(t) и C(t) и расчета** u_c значения A, B и C, полученные в **формирователе**, откладываются соответствующим образом на *выходные* функции принадлежности, которые в этом регуляторе идентичны входным функциям принадлежности (см. рис. 7, *а*, *б*), определяется результирующая функция принадлежности (жирная линия на рис. 7, *б*) и производится расчет ненормированного выхода регулятора u_c по формулам (14–20).

Далее полученное значение u_c в **блоке нормировки выходной переменной** пересчитывается в выходное напряжение регулятора по формуле (21).

Рассмотрим *режимы работы* спроектированного нечеткого регулятора в системе управления.

Если одна или две из переменных u_i, $i = 1,2,3$, больше a, а две или одна из остальных расположены на универсальном множестве в диапазоне $-a \le u \le a$, то $A = 0$. Если одна или две из переменных меньше $-a$, а две или одна из остальных расположены на универсальном множе-

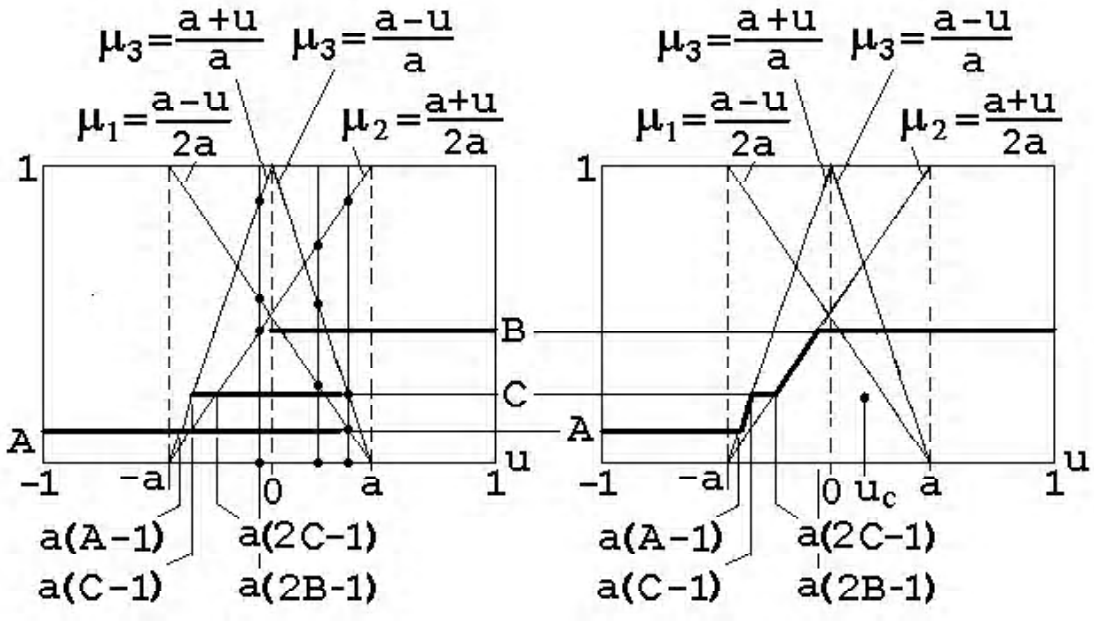


Рис. 7.

стве в диапазоне $-a \le u \le a$, то $B = 0$. Если три переменных больше a ($u_i \ge a, i = 1,2,3,$), то $B \ne 0$ и если три переменных меньше $-a$ ($u_i \le -a, i = 1,2,3,$), то $A \ne 0$. В этих случаях регулятор работает в *первом режиме,* когда входные сигналы (входные воздействия) большие и регулятор обеспечивает высокое быстродействие (быструю отработку входных сигналов).

Если все нормированные входные сигналы $u_i, i = 1,2,3,$ находятся в диапазоне $-a \le u_i \le a$, то регулятор работает во *втором режиме,* когда входные сигналы (входные воздействия) малые и регулятор обеспечивает динамическую точность системы (малые динамические ошибки).

Третий режим работы заключается в следующем. Если одна из переменных $u_i, i = 1,2,3,$ больше a, а другая переменная меньше $-a$, то $A = B = 0$ и на выходе нечеткого регулятора сигнал равен нулю. В этом случае нечеткий регулятор ведет себя как нелинейное корректирующее устройство со случайным прерыванием управляющего воздействия на объект управления.

Нечеткий регулятор переходит из одного режима в другой режим автоматически в зависимости от входных сигналов.

Теперь отметим существенную ошибку в работе [1], в которой приведены *входные* функции принадлежности с двумя термами и *выходные* функции принадлежности с тремя термами. Очевидно, в *выходных* функциях принадлежности будут рабочими только две, идентичные *входным* функциям, поскольку *выходные* функции принадлежности с тремя термами не получают значения уровня отсечения **C(t)**. Поэтому принятые в работе [1] *выходные* функции использовать нецелесообразно, поскольку регулятор будет работать только при двух *выходных* функциях принадлежности, которые идентичные *входным* функциям. Исследование замкнутой системы управления с предложенным регулятором изложено в работе [12].

ВЫВОД

В работе представлена принципиальная схема трехрежимного нечеткого регулятора и изложены вопросы настройки регулятора с проверкой правильности его работы. На основе нового метода проектирования нечетких регуляторов, изложенного в работах [9–11], получены строгие аналитические выражения для управляющих воздействий на выходе нечеткого регулятора при идентичных треугольных функциях принадлежности с тремя термами. В качестве входных воздействий на регулятор кроме ошибки системы рассматриваются первая и вторая производные ошибки, что резко уменьшает динамические ошибки в системах автоматического управления, использующих такой нечеткий регулятор.

СПИСОК ЛИТЕРАТУРЫ

1. *Ming, Liu* Design of a Multi-model Fuzzy Controller for AQM / Liu Ming, Dou Wen-hua, Xiao Rui // Fuzzy Systems and Knowledge Discovery–FSKD : Third International Conference (September 24–28, 2006, Xi'an), Lecture Notes in Computer Science. – China, Xi'an. – 2006, Volume 4223. – P. 739–742.
2. *Weiyan, Liu* A fuzzy-logic control algorithm for active queue management in IP networks / Liu Liu, Shunyi Zhang, Mu Zhang, Tao Liu // Journal of Electronics. – 2008. – No. 1, Vol. 25. – P. 102–107.
3. *Hadjadjaoul, Y.* FAFC: Fast Adaptive Fuzzy AQM Controller For TCP/IP Networks // Y. Hadjadjaoul, A. Nafaa, D. Negru, A. Mehaoua // IEEE GLOBAL Telecommunication Conference (29.11–3.12, 2004). – Dallas, Texas, USA, 2004. – Pp. 95–98
4. *Chrysostomou, C.* Fuzzy Logic Based Congestion Control in TCP/IP Networks for Quality of Service Provisioning / C. Chrysostomou, A. Pitsillides, G. Hadjipollas, A. Polycarpou, M. Sekercioglu // Next Generation Teletraffic and Wired / Wireless Advanced Networking (NEW2AN'04) : Proceedings of the International Conference (2–6 February 2004, St. Petersburg). – St. Petersburg, Russia, 2004. – Pp. 235–243.
5. *Chrysostomou, C.* Fuzzy logic congestion control in TCP/IP tandem networks / C. Chrysostomou, A. Pitsillides // Computers and Communications (IEEE ISCC 2006) : Proceedings of the 11th IEEE Symposium (June 26–29, 2006, Cagliari). – Cagliari, Italy, 2006. – Pp. 123–129.
6. *Chrysostomou, C.* Using Fuzzy Logic Control to Address Challenges in AQM Congestion Control in TCP/IP Networks / C. Chrysostomou, A. Pitsillides // Workshop on Modeling and Control of Complex Systems (MCCS'05) : CD ROM Proceedings (June 30–July 1, 2005, Ayia Napa). – Ayia Napa, Cyprus, 2005. – 22 p.
7. *Di Fatta, G.* A genetic algorithm for the design of a fuzzy controller for active queue management / G. Di Fatta, F. Hoffmann, Lo Re G, A. Urso // IEEE Transactions on Systems, Man, and Cybernetics, Part C : Applications and Reviews. – 2003. – Volume 33, Issue 3. – Pp. 129–134.
8. *Yaghmaee, M. H.* A Fuzzy Based Active Queue Management Algorithm / M. H. Yaghmaee, H. A. Toosi // SPECTS'2003 : Proc. International Symposium on Performance Evaluation of Computer and Telecommunication Systems (July 20–24). – 2003. – Montreal, Canada. – Pp. 458–462.
9. *Гостев, В. И.* Нечеткие регуляторы в системах автоматического управления / Гостев В. И. – К. : Радіоаматор, 2008. – 972 с.
10. *Гостев, В. И.* Новый метод проектирования одного класса нечетких цифровых регуляторов / Гостев В. И. // Проблемы управления и информатики. – 2007. – № 6. – С. 73–84
11. *Гостев, В. И.* Новый метод проектирования одного класса нечетких цифровых регуляторов / Гостев В. И. // Автоматика–2007 : Матеріали XIV міжнародної конференції з автоматичного управління (10–14 вересня 2007 року, м. Севастополь). – Ч.1. – Севастополь : СНУЯЄ та П, 2007. – С. 122–132.
12. *Гостев, В. И., Скуртов С. Н.* Фаззи–системы активного управления очередью в сетях TCP/IP: монография // Гостев В. И, С. Н. Скуртов. – Нежин : ООО Видавництво «Аспект-Поліграф», 2011. – 464 с.

Стаття надійшла до редакції 14.03.2011.
Після доробки 16.01.2012.

Гостєв В. І.

ПРОЕКТУВАННЯ ТРИРЕЖИМНОГО НЕЧІТКОГО РЕГУЛЯТОРА ДЛЯ СИСТЕМ АКТИВНОГО КЕРУВАННЯ ЧЕРГОЮ В TCP/IP МЕРЕЖАХ

Викладені питання проектування трирежимного нечіткого регулятора при ідентичних трикутних функціях приналежності з трьома термами для систем активного керування чергою в TCP/IP мережах та представлена принципова схема регулятора в системі MATLAB.

Ключеві слова: проектування, нечіткий регулятор, функції приналежності, активне керування чергою, TCP/IP мережі, MATLAB.

Gostev V. I.
DESIGNING OF AN THREE-REGIME FUZZY CONTROLLER FOR SYSTEMS OF ACTIVE QUEUE MANAGEMENT IN TCP/IP NETWORKS

Questions of designing of an three-regime fuzzy controller are stated at identical triangular membership functions with three terms for systems of active queue management in TCP/IP networks and the basic scheme of controller in system MATLAB is presented

Key words: designing, fuzzy controller, membership functions, active queue management, TCP/IP networks, MATLAB.

УДК 004.93

Гофман Є. О.[1], Олійник А. О.[2], Субботін С. О.[3]

[1]Аспірант Запорізького національного технічного університету
[2]Канд. техн. наук, доцент Запорізького національного технічного університету
[3]Канд. техн. наук, доцент Запорізького національного технічного університету

СКОРОЧЕННЯ БАЗ ЛІНГВІСТИЧНИХ ПРАВИЛ НА ОСНОВІ ДЕРЕВ РОЗВ'ЯЗКІВ

Розглянуто завдання індукції лінгвістичних правил. Розроблено метод ідентифікації дерев розв'язків для індукції лінгвістичних правил. Створено програмне забезпечення на основі запропонованого методу. Проведено експерименти по розв'язанню практичних задач, що дозволило дослідити ефективність запропонованого методу.

Ключові слова: дерево розв'язків, індукція правил, лінгвістичне правило.

ВСТУП

В наш час широке застосування отримали експертні системи, засновані на лінгвістичних правилах [1, 2], які успішно використовуються в різних прикладних областях, зокрема в технічному та медичному діагностуванні, фінансовому менеджменті, розпізнаванні образів, геологічній розвідці, керуванні комп'ютерними мережами, технологічними процесами, аналізі веб-контенту в інтернет та ін. Широке застосування таких систем обумовлене в першу чергу тим, що вони є прозорими й відносно дешевими в реалізації.

Оскільки бази правил в експертних системах часто характеризуються великим обсягом, актуальним є завдання індукції правил, суть якого полягає в тому, що на основі початкового набору правил необхідно сформувати нову базу правил меншого обсягу, яка в достатній мері представляла б початкову базу правил і була б менш надлишковою.

Існують різні методи індукції правил [3], однак ці методи при обробці правил аналізують їхню якість окремо, не розглядаючи та не враховуючи якість усієї бази в цілому, що приводить до одержання неоптимальних баз нечітких правил. Тому актуальною є розробка нових методів індукції правил, які враховували б якість усієї бази знань, а не тільки окремих правил. Для розв'язання даного завдання пропонується створювати дерева розв'язків [3-5], які б після їхньої побудови переводилися в лінгвістичні правила. Вибір дерев розв'язків обгрунтовується їхньою можливістю виявляти неспостережувані зв'язки всередині досліджуваних об'єктів, процесів і систем.

Метою даної роботи є розробка методу індукції лінгвістичних правил з використанням математичного апарата дерев розв'язків.

Для досягнення поставленої мети необхідно розв'язати такі завдання:

– огляд математичного апарату дерев розв'язків;

– приведення основних етапів ідентифікації дерев розв'язків відповідно до розв'язуваного завдання;

– створення правил перетворення дерев розв'язків у лінгвістичні правила;

– порівняння розробленого підходу з існуючими методами індукції лінгвістичних правил.

ПОСТАНОВА ЗАВДАННЯ

Нехай задана база лінгвістичних правил $RB = \{R^1, R^2 ..., R^{RN}\}$, що описує об'єкти навчальної вибірки $O = \{O^1, O^2, ..., O^N\}$. Тоді на основі навчальної вибірки об'єктів O, необхідно сформувати таку базу лінгвістичних правил $RB^* = \{R^1, R^2, ..., R^{RN^*}\}$, $RN^* << RN$, яка забезпечувала б прийнятну якість прогнозування експертної системи, побудованої на основі отриманої бази лінгвістичних правил RB^*:

$$Q(RB^*) \geq Q_{threshold},$$

де $Q(RB^*)$ – точність прогнозування або класифікації по базі правил RB^*; $Q_{threshold}$ – мінімально припустима точність прогнозування або класифікації.

ДЕРЕВА РОЗВ'ЯЗКІВ

Дерева розв'язків являють собою графові інтелектуальні моделі, у внутрішніх вузлах яких розташовані функції прийняття рішень на основі значень вхідних змінних, а в зовнішніх вузлах (термінальних вузлах, листах) знаходяться значення вихідної змінної, відповідні до умов у внутрішніх вузлах [2, 6, 7].

Завдяки своїй деревоподібній структурі такі моделі дозволяють наочно представляти результати обчислень. Тому вони добре інтерпретуються людьми-фахівцями в прикладних областях, які, як правило, не мають спеціальної математичної підготовки та не знайомі з методами й моделями штучного інтелекту. Дерева розв'язків дозволяють ефективно вирішувати завдання класифікації та прогнозування, забезпечуючи при цьому високу точність.

Для застосування дерев розв'язків на практиці з метою класифікації або прогнозування значень вихідних параметрів досліджуваних об'єктів по наборах значень вхідних характеристик необхідно за допомогою даних навчальної вибірки сформувати дерево розв'язків таким чином, щоб воно щонайкраще описувало досліджуваний об'єкт.

Побудова дерев розв'язків пов'язана з витягом правил з навчальних вибірок. Кожний шлях від кореня дерева до одного з його листів може бути перетворений до логічного висловлення – правилу типу «якщо A, то B», де його антецедент виходить шляхом використання всіх умов, представлених у внутрішніх вузлах від кореня до вихідного листа, а права частина правила виходить із відповідного листа дерева.

Процес побудови дерева розв'язків, як правило, містить такі етапи: розростання, розгалуження, обчислення значення вихідного параметра для листа, скорочення.

У результаті етапу розростання (збільшення, growing) деяка вершина заміняється піддеревом, отриманим шляхом розгалуження цієї вершини. На даному етапі відбувається поділ обраної вершини на деякі нові (у випадку дихотомічного дерева вершина розбивається на дві нові). При цьому перебираються всі ознаки й усі можливі варіанти розгалуження по кожній з ознак. У результаті залишається варіант розбиття, при якому значення критерію якості розбиття є найкращим. Якщо нові вершини є перспективними для наступного поділу (критерії завершення розростання не задоволені), то виконується їхнє розгалуження. У випадку неможливості подальшого поділу вершини вона стає листом, і для неї виконується процедура обчислення значення вихідного параметра. Якщо розгалуження вершини приводить до погіршення якості дерева, то вершина також оголошується листом.

Процедура розгалуження (поділу, splitting) дерева викликається рекурсивно при виконанні етапу розростання. Розгалуження призначено для створення для обраної вершини заданої кількості (для дихотомічних дерев – дві) вершин-нащадків.

Обчислення значення вихідного параметра відбувається шляхом пересування по синтезованому дереву розв'язків від кореневого вузла до листа в залежності від значень вхідних параметрів.

Етап скорочення (усікання, pruning) використовується для спрощення побудованого дерева шляхом відсікання нащадків у обраній вершини, яка в наслідку стає листом з певним значенням. Усікання вузла виконується у випадку, якщо воно не приведе до істотного погіршення апроксимаційних і узагальнюючих характеристик дерева розв'язків.

Таким чином, етап усікання дерева виконується знизу нагору: рух починається від листів дерева та відбувається нагору доти, доки апроксимаційні здатності дерева розв'язків залишаються прийнятними.

ІНДУКЦІЯ ЛІНГВІСТИЧНИХ ПРАВИЛ НА ОСНОВІ ПОБУДОВИ ДЕРЕВ РОЗВ'ЯЗКІВ

Існуючі методи побудови дерев розв'язків [2–7] не враховують особливостей завдання індукції лінгвістичних правил. У зв'язку з цим розробляється новий метод побудови дерев розв'язків для індукції правил. Подібно відомим методам побудови дерев розв'язків, пропонований метод складається з основних фаз: ріст дерева і його згладжування (скорочення), після чого виконується перетворення дерева розв'язків у лінгвістичні правила. Найбільш важливими аспектами пропонованого методу є наступні: використання модифікованої ентропії як оцінної міри і використання згладжування для відсікання.

Таким чином, пропонований метод складається з таких етапів:

– ріст дерева;

– згладжування дерева;

– перетворення дерева розв'язків у лінгвістичні правила.

На етапі росту дерева пропонується використовувати жадібний підхід. У кожному вузлі, що відповідає підмножині T навчальної вибірки, вибирається ознака f і значення v таким чином, що дані з T розділяються на дві підмножини $T^1_{f,v}$ та $T^2_{f,v}$ виходячи з умов $x_{i,f} \le v : T^1_{f,v} = \{x_i \in T : x_{i,f} \le v\}$ і $T^2_{f,v} = \{x_i \in T : x_{i,f} > v\}$. Таке розбиття розділяє множину об'єктів навчальної вибірки на такі, для яких значення ознаки f менше значення v, і на ті, для яких значення ознаки f більше значення v.

З метою розбиття дерева розв'язків для кожного можливого розбиття (f, v) розраховується оціночна функція:

$$Q(f,v) = p_{f,v} g(p^1_{f,v}) + (1 - p_{f,v}) g(p^2_{f,v}),$$

де $p^1_{f,v} = P(y_i = 1 \mid x_i \in T^1_{f,v})$, $p^2_{f,v} = P(y_i = 1 \mid x_i \in T^2_{f,v})$ і $p_{f,v} = P(x_i \in T^1_{f,v} \mid x_i \in T)$; $g(p)$ – модифікована ентропія для ймовірності віднесення вихідної змінної y до розглянутого класу за умови, що x більше або менше значення v ($p^1_{f,v}$ і $p^2_{f,v}$, відповідно):

$$g(p) = -r(p)\ln(r(p)) - (1 - r(p))\ln(1 - r(p)),$$

де $r(p)$ перетворить оцінку ймовірності:

$$r(p) = \begin{cases} \frac{1}{2}(1 + \sqrt{2p - 1}), \text{ якщо } p > 0{,}5; \\ \frac{1}{2}(1 - \sqrt{1 - 2p}), \text{ якщо } p > 0{,}5. \end{cases}$$

Таким чином, чим ближче значення ймовірності до 0,5, тим вище модифіковане значення, а чим далі від 0,5, тим значення нижче.

Функція оцінки розраховується для всіх можливих розбиттів і вибирається розбиття з найменшим значенням оціночної функції. Розбиття починається від кореневого вузла та триває доти, поки не виникне ситуація, коли неможливо зробити нове розбиття.

Після виконання першого етапу може виникнути ситуація «перенавчання» дерева, що може привести до не зовсім коректної роботи дерева на тестових вибірках. У зв'язку із цим на другому етапі проводиться усікання великого дерева, щоб дерево менших розмірів давало більш стабільні оцінки ймовірності й було більш інтерпретабельним.

Далі описується підхід, який замість урізання повного дерева, буде робити переоцінку ймовірності кожного листового вузла шляхом усереднення оцінки ймовірності по шляху проходження від кореневого вузла до листового вузла. Для досягнення даної мети була взята ідея «обважнення дерева» [8]. Якщо використовується дерево для стиснення бінарної класової ознаки y_i, заснованого на x_i, то в такому випадку метод обважнення дерева гарантує, що коефіцієнт стиснення переоціненої ймовірності не буде гірше, чим в успішно усіченому дереві. Оскільки пропонований метод застосовується більшою мірою до трансформованої оцінки ймовірності $r(p)$, ніж безпосередньо до p, то теоретично, результат може бути наступним: шляхом використання переоціненої ймовірності, можна досягнути очікуваної класифікації навчальної множини з не гіршим результатом, ніж у правильно усіченого дерева.

Слід зазначити, що даний підхід також є стисненням, оскільки за допомогою такого підходу оцінка стискується від далеких вузлів дерева в напрямку до оцінок вузлів, які перебувають ближче до кореня дерева.

Нехай вузли T_1 та T_2 є елементами одного рівня із загальним батьківським вузлом T. Нехай $p(T_1)$, $p(T_2)$ і $p(T)$ будуть відповідними оцінками ймовірності. Локальна переоцінена ймовірність може бути обчислена за формулами: $w_T p(T) + (1 - w_T) p(T_1)$ для T_1 та $w_T p(T) + (1 - w_T) p(T_2)$ для T_2. Локальна значимість w_T і супутня функція $G(T)$ розраховуються рекурсивно, ґрунтуючись на таких формулах:

$$\frac{w_T}{1 - w_t} = \frac{c \cdot \exp(-|T| g(p(T)))}{\exp(-|T_1| G(T_1) - |T_2| G(T_2))},$$

$$G(T) = \begin{cases} g(p(T)) + \dfrac{1}{|T|}\log((1 + \dfrac{1}{c})w_T), \text{ якщо } w_T > 0{,}5, \\ \dfrac{|T_1|}{|T|} G(T_1) + \dfrac{|T_2|}{|T|} G(T_2) + \dfrac{1}{|T|}\log((1 + c)(1 - i w_T)), \text{ в іншому випадку.} \end{cases}$$

Параметр c установлюється апріорно та показує Байєсову «оцінку» розбиття. Для листового вузла T установлюється: $G(T) = g(p(T))$ та $w_T = 1$.

Після обчислення значимостей w_T для кожного вузла рекурсивним методом (використовується спадна рекурсія), необхідно розрахувати глобальну оцінку ймовірності для кожного вузла дерева зверху вниз. Даний етап усереднює усі оцінки $r(p)$ від кореневого вузла T_0 до вузла T_h по шляху $T_0, \ldots, T_h$, ґрунтуючись на значимості w_T. Слід зазначити, що значимість w_T є лише локально важливою. Це означає те, що глобальна значимість вузла T_h є $w_T^* = \prod_{i<k}(1 - w_i) w_k$ на всьому шляху. За визначенням, $\sum_{i=1}^{h} w_i = 1$ для будь-якого напрямку, що веде до листа. За наступними рекурсивними формулами обчислюється глобальна переоцінка підлеглих вузлів T_1 і T_2 в батьківському вузлі T:

$$\hat{w}_{T_i} = \hat{w}_T (1 - w_T),$$

$$r^*(T_i) = r^*(T) + \hat{w}_{T_i} w_{T_i} r(p(T_i)),$$

де $r(p(T))$ – перетворення з оцінки ймовірності $p(T)$ у вузлі T. У кореневому вузлі встановлюється: $\hat{w} = 1$. Після обчислення $r^*(T_h)$ для листового вузла T_h в якості оцінки ймовірності можна використовувати $r^{-1}(r(T_h))$. Мітка класу для T_i буде дорівнювати одиниці, якщо $r(T_i) > 0{,}5$, в іншому випадку – нулю. Усікання дерева виконується, починаючи з основи за напрямом вгору шляхом перевірки ідентичності вузлів одного рівня. Якщо ідентичність вузлів виявлена, то вони видаляються й використовується значення батьківського вузла. Дана процедура буде тривати доти, поки вона не стане неможливою. Метод згладжування послідовно поліпшує роботу дерева. Оцінка часової складності – $O(M)$, де M – кількість вузлів неусіченого дерева.

Важливою частиною запропонованого методу є етап перетворення дерева розв'язків в еквівалентний набір лінгвістичних правил, що легко піддаються тлумаченню. Важливість такого перетворення пояснюється двома причинами:

1. Будь-якій людині легше зрозуміти й змінити набір правил, ніж зрозуміти й змінити дерево розв'язків. Потреба в такій зміні очевидна. Наприклад, може виникнути ситуація, коли є деяка невідповідність між навчальною вибіркою й реальною системою, що вимагає ручної модифікації автоматично створеної системи, і, таким

чином, у системі, заснованій на правилах, таку модифікацію можна виконати шляхом простої зміни відповідних правил.

2. Той факт, що набір правил є логічно еквівалентним відповідному дереву розв'язків для даної навчальної вибірки, гарантує, що будь-який математичний аналіз ефективності роботи дерева розв'язків відноситься не тільки до дерева розв'язків, але й до відповідного набору правил.

Найпростіший спосіб перетворення дерева в еквівалентний набір правил полягає в тому, щоб створити набір правил із правил, кожне з яких відповідає окремому листу дерева шляхом формування логічного об'єднання умов на шляху від кореня дерева до листа.

Пропонується підхід, що перетворить дерево розв'язків у набір логічно еквівалентних правил. Метою пропонованого підходу не є одержання доказово мінімального набору правил. Замість цього за допомогою пропонованого підходу проводиться логічне усікання правил.

1. Перевірка умов «>» та «<» у всіх правилах з метою усунення надмірності в описі умов правил. Таким чином, виконується, наприклад таке перетворення: $(x < 3) \cap (x < 5)$ заміняється на $(x < 3)$.

2. Видалення умов, які є логічно надлишковими в контексті всього набору правил, тобто видалення умов, які ідентифікуються виходячи зі структури отриманого дерева розв'язків. Таке спрощення змінює правило, що пов'язано з конкретним листом дерева, при цьому зберігаючи повну адекватність усього набору правил.

Для кожного листа, що віднесений до класу X, створюється правило про те, що об'єкт відноситься до класу X шляхом кон'юнкції умов, що знаходяться на шляху проходження від кореня до X, але використовуючи тільки ті умови, які відповідають наступному правилу: для кожного вузла N на шляху від кореня до листа з міткою X умова, що відповідає батьку N, є частиною кон'юнкції тільки в тому випадку, якщо спрацьовує умова сусідства для вузла N. Умова сусідства для N вважається успішною, якщо: вузол N не є коренем і сусідній вузол відносно N не є листом з міткою X.

Таким чином результуючий набір правил є логічно еквівалентним базовому дереву розв'язків.

Запропонований метод індукції лінгвістичних правил з використанням дерев розв'язків був програмно реалізований за допомогою мови програмування C#.

Для експериментів використовувалися тестові дані, які були взяті із загальнодоступних репозиторіїв [9]. Експериментальні дослідження проводилися на підставі вибірки, яка містила інформацію про ехокардіограми пацієнтів із серцевими приступами. Вибірка містила інформацію про 132 пацієнтів, кожний з яких характеризувався 12 ознаками. Крім того, для кожного пацієнта вказувалося живий він або помер.

Запропонований метод індукції нечітких правил порівнювався з мультиагентним методом і канонічним методом еволюційного пошуку. Виходячи із проведених експериментів, були отримані бази лінгвістичних правил, що характеризуються наступною якістю класифікації пацієнтів: 81,3 %, 79,1 % і 92,7 % для мультиагентного, еволюційного та запропонованого методів, відповідно.

Таким чином, можна відзначити, що запропонований метод побудови дерев розв'язків для індукції лінгвістичних правил забезпечує більш точні результати прогнозування в порівнянні з іншими відомими методами індукції лінгвістичних правил.

ВИСНОВКИ

У роботі вирішено актуальне завдання автоматизації індукції лінгвістичних правил.

Наукова новизна роботи полягає в тому, що розроблено новий метод побудови дерев розв'язків, який дозволяє виконувати індукцію лінгвістичних правил, що досягається за рахунок введення додаткових функцій перетворення при рості дерева, шляхом згладжування дерева розв'язків для його усікання та за рахунок введення критерію сусідства при перетворенні дерева розв'язків.

Розроблений метод ідентифікації дерев розв'язків для індукції лінгвістичних правил дозволяє виконувати перетворення й об'єднання правил, що забезпечує можливість розробки експертних систем на підставі більш логічно прозорих і простих баз лінгвістичних правил.

Практична цінність отриманих результатів полягає в тому, що на основі запропонованого методу розроблено програмне забезпечення, яке дозволяє виконувати індукцію баз правил для одержання баз лінгвістичних правил, на підставі яких можна створювати експертні системи з меншою помилкою класифікації.

СПИСОК ЛІТЕРАТУРИ

1. Encyclopedia of artificial intelligence / Eds.: J. R. Dopico, J. D. de la Calle, A. P. Sierra. – New York : Information Science Reference, 2009. – Vol. 1–3. – 1677 p.
2. *Барсегян, А. А.* Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP : учебное пособие / А. А. Барсегян. – С. Пб. : BHV, 2007. – 384 с.
3. *Quinlan, J. R.* Decision trees and decision making / J. R. Quinlan // IEEE Transactions on Systems, Man and Cybernetics. – 1990. – № 2 (20). – P. 339–346.
4. Quinlan J. R. Induction of decision trees / J. R. Quinlan // Machine Learning. – 1986. – № 1. – P. 81–106.
5. Gelfand S. B. An Iterative Growing and Pruning Algorithm for Classification Tree Design / S. B. Gelfand, C. S. Ravishankar, E. J. Delp // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1991. – № 13 (2). – P. 163–174.
6. *Liu, X.* A decision tree solution considering the decision maker's attitude / X. Liu, Q. Da // Fuzzy Sets and Systems. – 2005. – № 152 (3). – P. 437–454.
7. Classification and regression trees / L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone. – California : Wadsworth & Brooks, 1984. – 368 p.
8. Willems F. M. J. The Context Tree Weighting Method: Basic Properties / F. M. J. Willems, Y. M. Shtarkov, T. J. Tjalkens / / IEEE Transactions on Information Theory. – 1995. – № 3. – P. 653–664.
9. UCI Machine Learning Repository [electronic resource] / Center for Machine Learning and Intelligent Systems. – Access mode : http://archive.ics.uci.edu/ml/datasets.html.

Стаття надійшла до редакції 28.12.2011.

Гофман Е. А., Олейник А. А., Субботин С. А.
СОКРАЩЕНИЕ БАЗ ЛИНГВИСТИЧЕСКИХ ПРАВИЛ НА ОСНОВЕ ДЕРЕВЬЕВ РЕШЕНИЙ

Рассмотрена задача индукции лингвистических правил. Разработан метод идентификации деревьев решений для индукции лингвистических правил. Создано программное обеспечение на основе предложенного метода. Проведены эксперименты по решению практических задач, что позволило исследовать эффективность предложенного метода.

Ключевые слова: дерево решений, индукция правил, лингвистическое правило.

Gofman Ye., Oliinyk A., Subbotin S.
LINGUISTIC RULES BASES REDUCTION BASED ON DECISION TREES

The problem of linguistic rules induction is considered. A method of decision trees identification for linguistic rules induction is developed. The software based on the proposed method is created. Experiments on the solution of practical problems, which allowed to investigate the effectiveness of the proposed method are made.

Key words: decision tree, rules induction, linguistic rule.

УДК 004.78; 004.021; 004.046

Ильяшенко М. Б.[1], Голдобин А. А.[2]

[1]Канд. техн. наук, доцент Запорожского национального технического университета
[2]Ассистент Запорожского национального технического университета

РЕШЕНИЕ ЗАДАЧИ ПОИСКА ИЗОМОРФИЗМА ГРАФОВ ДЛЯ ПРОЕКТИРОВАНИЯ СПЕЦИАЛИЗИРОВАННЫХ ВЫЧИСЛИТЕЛЕЙ

Предлагается усовершенствованный алгоритм поиска изоморфизма графов и результаты исследования его эффективности. Объектом исследования является множество граф-схем алгоритмов достижения цели, полученная после обхода заданной семантической сети абстрактной машиной Уоррена.

Ключевые слова: декларативная логика, предикат, дерево вывода, пролог, рекурсивный обход с возвратом, граф-подграф изоморфизм.

ВВЕДЕНИЕ

В составе программно-лингвистических средств автоматизации проектирования цифровых устройств на программируемых логических интегральных схемах (ПЛИС) широко используются методики формального описания структурно-функциональной организации проектируемого объекта. Методология проектирования на основе математического аппарата теории ориентированных гиперграфов [1] позволяет использовать унифицированные алгоритмы выполнения основных этапов создания цифровых устройств, в том числе специализированных (проблемно-ориентированных) вычислителей.

Ориентированный гиперграф, допускающий петли и кратные дуги, является наиболее общим типом графовых моделей и называется ориентированным псевдогиперграфом. В дальнейшем, для краткости, ориентированный, помеченный псевдогиперграф будем именовать «графом».

Для формального описания свойств узлов и дуг графа удобно применять операторное пространство, образующее многозначную логику. Примером такого пространства является трехзначная логика модальных операторов Лукасевича [2].

Применение многозначного операторного пространства для описания актов синтеза цифрового устройства, которое формально задано графом, и исследование способов преобразования проблемно-ориентированных описаний является актуальной задачей. Ее решение позволяет разрабатывать эффективные инструменты, место применения которых – системы автоматизированного проектирования (САПР) цифровых и микропроцессорных устройств различного назначения.

Абстрактная машина Уоррена (англ. Warren's abstract machine, WAM) [3] представляет собой формальную модель устройства, реализующего основные операции исчисления предикатов первого порядка, представленных дизъюнктами Хорна. Известное приложение WAM – японский проект вычислителей пятого поколения (1982–1992 гг.).

В абстрактной машине Уоррена одновременно выполняется семантический анализ программы на входном языке L_{pro} и формирование таблиц лексико-синтаксического анализа. Результаты анализа используются для генерирования программы на выходном языке L_{wam}. Язык L_{wam} представляет собой набор типовых, функционально-ориентированных операций линейной резолюции в терминах WAM – конечного множества команд типа `unify_variable, put_list` и др.

Формальная грамматика является абстрактной структурой, описывающей множество правил образования строк языка из заданного алфавита терминальных и нетерминальных символов. Генерирующая и анализирующая формальные грамматики используются для решения противоположных задач, в зависимости от семантики грамматического разбора. Генерирующая грамматика образуется конечным множеством порождающих правил (продукций) формирования строк формального

языка. Анализирующая грамматика предназначена для грамматического разбора строк формального языка, поступающих на вход.

Графически правила продукций формальной грамматики представляются в виде ориентированных помеченных псевдогиперграфов. Например, продукции регулярной грамматики описываются в виде *граф-схем алгоритма* (ГСА).

Иерархия Хомского [2] представляет собой иерархию классов последовательно вложенных формальных грамматик, описывающих языки разного типа. В исходном виде иерархия Хомского образована четырьмя уровнями, которые задают законы образования основных классов формальных языков. Каждому уровню иерархии Хомского соответствует своя модель распознающего автомата, начиная с машины Тьюринга с бесконечной лентой, которая предназначена для распознавания продукций грамматик общего вида (уровень 0 иерархии Хомского), и заканчивая конечным автоматом, порядок переключения состояний которого описывается регулярной грамматикой (уровень 3 иерархии Хомского).

Абстрактная машина Уоррена способна распознавать грамматики уровня 2. Это контекстно-свободные грамматики, с помощью которых определяются контекстно-свободные языки. Языки, образованные такими грамматиками, могут быть распознаны с помощью *стекового автомата*. Таким образом, с точки зрения анализа входной программы, поведение WAM эквивалентна работе стекового автомата.

В процессе выполнения запроса WAM способна генерировать описание ориентированного графа, представляющего ГСА выходной программы на регулярном языке L_{wam}.

Так как абстрактная машина Уоррена занимает в иерархии Хомского более высокий уровень по отношению к конечному автомату и способна преобразовывать строки контекстно-свободного языка L_{pro} в конструкции регулярного языка L_{wam}, она может использоваться в следующих приложениях:

– ввод задания на проектирование специализированного вычислителя на ПЛИС;

– конфигурирование микропроцессорных систем с программируемой архитектурой;

– реализация технологий логического, концептуального, функционального программирования.

В данной работе описан алгоритм поиска граф-подграф изоморфизма и приведены результаты исследования его эффективности на примере помеченных ориентированных гиперграфов, которые представляют результат работы абстрактной машины Уоррена.

1. ПОСТАНОВКА ЗАДАЧИ

Графически концептуальная область входной программы для абстрактной машины Уоррена может быть задана *семантической сетью*. Например, на рис. 1 показан концептуальный граф вычислительной модели прямоугольного треугольника.

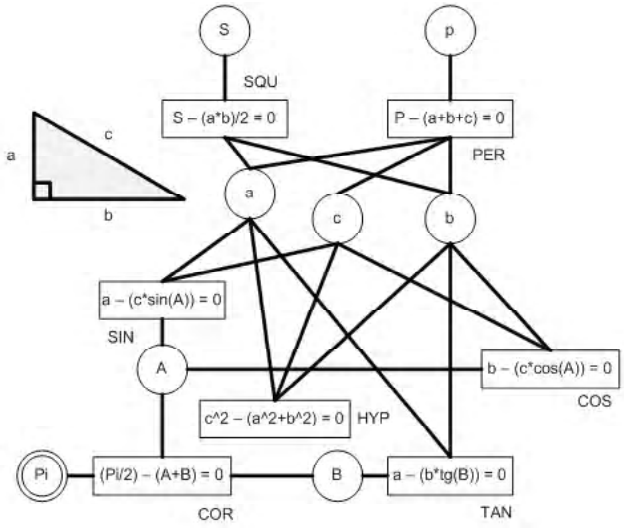


Рис. 1. Концептуальный граф

Программа на входном языке L_{pro} абстрактной машины Уоррена образована объектами двух типов: термом программы p и термом запроса $?-q$. Цель работы машины можно сформулировать так: определив программу p, необходимо составить запрос $?-q$, выполнение которого или закончится неудачно, если p и q нельзя унифицировать, или будет завершено удачно, если получилась связка переменных из q с переменными из p.

При доказательстве теорем методом резолюции, проверка невыполнимости запроса наталкивается на препятствия, связанные с бесконечным числом областей интерпретации запроса. В общем случае, если выбранная область бесконечна, то запрос допускает бесконечно много конкретизаций, т. е. существует бесконечно много интерпретаций относительно языка L_{pro}. Для обхода данной проблемы в WAM клаузальная форма концептуального пространства, заданная программой p, ограничивается эрбановой областью [2].

По этой причине пролог-процессор, реализованный на WAM, реализует стратегию разбора И/ИЛИ-дерева запроса «слева направо и вглубь с возвратом при неудаче». На рис. 2 показан случай поиска конкретизации переменной b в вычислительной модели прямоугольного треугольника, при условии, что грамматический разбор начинается с фразы SQU.

Можно сделать выводы относительно порядка работы машины Уоррена.

1. Концептуальное пространство, в котором выполняется грамматический разбор запроса $?-q$ машиной Уоррена ограничивается множеством конкретизированных во входной программе p переменных и функционалов.

2. Время выполнения запроса $?-q$ зависит от выбора начальной языковой конструкции – хорновского дизъюнкта, определяющего правило в программе на входном языке L_{pro}.

3. Если $H_M^k(G)$ – множество всех ГСА, сформированных машиной Уоррена при выполнении k-го запроса на максимально определенном пространстве, то лю-

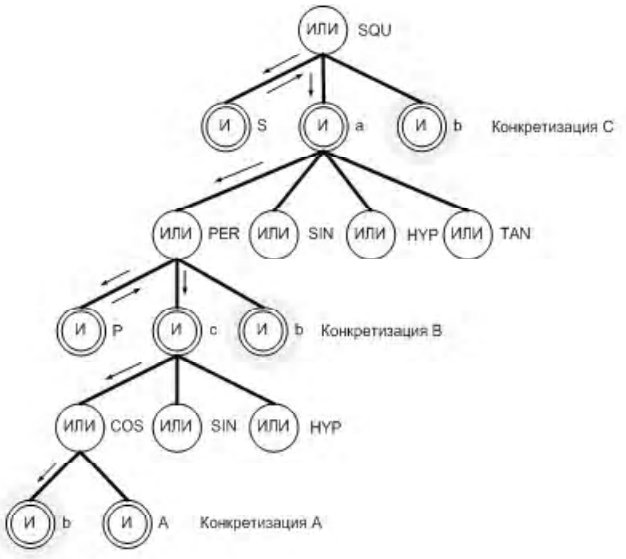


Рис. 2. Пример грамматического разбора

бое множество $H_m^k(G)$ не полностью определенных концептуальных пространств сюръективно отображается на множество $H_M^k(G)$: $H_M^k(G) \subseteq \forall\ H_m^k(G)$.

Не полностью определенным концептуальным пространствам соответствуют различные семантические сети, представляющие их графически. Поиск изоморфного подграфа в ориентированном графе семантической сети концептуального пространства программы для WAM является важной задачей, обеспечивающей значительное повышение эффективности работы генератора программно-лингвистических описаний, созданного на основе машины Уоррена. Для решения этой задачи был разработан усовершенствованный алгоритм определения изоморфизма двух заданных графов.

2. МЕТОД РЕШЕНИЯ ЗАДАЧИ

Пусть даны графы $G_1 = (V_1, E_1, L_1)$ и $G_2 = (V_2, E_2, L_2)$, ... V – множество вершин графа, E – множество ребер графа и L метки вершин графа. Граф G_1 изоморфен подграфу графа G_2 (обозначается, как $G_1 \cong S_2 \subseteq G_2$), если существует подстановка $\varphi : V_2 \to V_1$, такая, что для каждой пары вершин $v_i, v_j \in V_2$, если $(v_i, v_j) \in E_2$, то $(\varphi(v_i), \varphi(v_j)) \in E_1$ и для всех $l_i \in L_2$ выполняется $l_i \in L_2 = \phi(r_i) \in L_1$.

Алгоритм установления граф-подграф изоморфизма для помеченных графов является развитием и продолжением алгоритма установления изоморфности [4].

Алгоритм установления изоморфизма удобно описывать в терминах поиска в пространстве состояний. Каждое состояние s процесса совмещения вершин соответствует частичной подстановке $\varphi(s)$, которая содержит часть вершин полной подстановки. Каждому состоянию так же соответствуют подграфы $G_1(s)$ и $G_2(s)$, полученные из вершин графов G_1 и G_2, вошедших в частичную подстановку $\varphi(s)$, и ребер, соединяющих эти вершины. В дальнейшем обозначим через $\varphi_1(s)$ и $\varphi_2(s)$ проекции подстановки $\varphi(s)$ на V_1 и V_2.

Алгоритм состоит из предварительной и основной части. В предварительной части выполняются операции упорядочивания вершин графов и выполнения однократных, по ходу алгоритма, операций, призванных сократить область поиска основной, переборной части алгоритма.

Предварительная часть алгоритма. Основные действия, выполняемые в предварительной части алгоритма – сортировка вершин графов и формирование матрицы возможных совмещений.

Матрица возможных совмещений $M_{i,j}$ – это бинарная таблица размером $|V_1| \times |V_2|$. Каждому элементы таблицы соответствует пара вершин исходных графов $V_{1,i}$ и $V_{2,j}$. Значения матрицы формируются следующим образом:

– $M_{i,j} = 0$, если на основании предварительных проверок вершины $V_{1,i}$ и $V_{2,j}$ совместить нельзя;

– $M_{i,j} = 1$, в противном случае.

Смысл матрицы возможных совмещений в том, чтобы выполнить однократно в рамках предварительной части алгоритма все проверки, не основанные на информации, полученной в процессе совмещения вершин, тем самым, ускорить обработку соответствующих ограничений, сведя ее к одной операции сравнения.

В программе реализованы следующие предварительные проверки:

1. $M_{i,j} = 0$, если $|V_{1,i}| < |V_{2,j}|$, где $|V_{X,Y}|$ – степень вершины Y графа X;

2. $M_{i,j} = 0$, если $|V_{1,i}^{in}| < |V_{2,j}^{in}|$, где $|V_{X,Y}^{in}|$ – число входящих ребер вершины Y графа X;

3. $M_{i,j} = 0$, если $|V_{1,i}^{out}| < |V_{2,j}^{out}|$, где $|V_{X,Y}^{out}|$ – число исходящих ребер вершины Y графа X;

4. $M_{i,j} = 0$, если $W_{1,i}^{vertex} < W_{2,j}^{vertex}$, где $W_{X,Y}^{vertex}$ – число вершин в волновом разложении подграфа окружения вершины Y графа X;

5. $M_{i,j} = 0$, если $\sum_{l=1}^{k} W_{1,i,l}^{vertex} < \sum_{l=1}^{k} W_{2,j,l}^{vertex}$, $k = 1..|W_{2,j}^{vertex}|, l = 1..4$, где $W_{X,Y,l}^{vertex}$ – число вершин в l-ой волне волнового разложения графа X, начиная с вершины Y;

6. $M_{i,j} = 0$, если $W_{1,i}^{ribes} < W_{2,j}^{ribes}$, где $W_{X,Y}^{ribes}$ – число ребер в волновом разложении подграфа окружения вершины Y графа X;

7. $M_{i,j} = 0$, если $\sum_{l=1}^{k} W_{1,i,l}^{ribes} < \sum_{l=1}^{k} W_{2,j,l}^{ribes}$, $k = 1..|W_{2,j}^{ribes}|, l = 1..4$, где $W_{X,Y,l}^{ribes}$ – число ребер в l-ой волне волнового разложения графа X, начиная с вершины Y;

8. $M_{i,j} = 0$, если $L_{1,i} \neq L_{2,j}$, где $L_{X,Y}$ – метка вершины Y графа X.

Возможно использование и других критериев для оценки возможности совмещения вершин графов. Метод разработки таких критериев основан на волновом разложении графов, начиная с заданной вершины [5]. По мере распространения волны получаются подграфы окружения вершин. Сравнивая параметры соответствующих подграфов окружения вершин графов, которые предполагается совмещать, делается вывод о потенциальной возможности или принципиальной невозможности такого совмещения. В приведенных критериях для этого использовались сумма вершин и ребер в подграфах окружения сравниваемых вершин для всех этапов распространения волны.

Сортировка вершин графов производится с целью ускорения нахождения изоморфной подстановки, в случае, если такая подстановка существует. В переборной части алгоритма переставляются только вершины большего графа, в то время, как порядок вершин меньшего графа не меняется. Порядок следования вершин меньшего графа определяется в предварительной части алгоритма.

Пусть $T_{2,i}$ – количество ребер инцидентных вершинам с меньшими номерами и $P_{2,i} = \sum_{j=1}^{|V_1|} M_{j,i}$ – суммарное количество вариантов совмещения вершины i графа G_2 с вершинами графа G_1. Тогда порядок сортировки вершин графа G_2 следующий:

$$V_{2,i} = V_{2,k}, \text{ где } T_{2,k} = \min_{j=i+1}^{|V_2|}\left(T_{2,j}\right).$$

$$\text{Если } T_{2,i} = T_{2,j}, \text{ то } V_{2,i} = V_{2,k}, \text{ где } P_{2,k} = \min\left(P_{2,i}, P_{2,j}\right).$$

Т. е. вершины графа G_2 сортируются в порядке убывания количества связей с вершинами имеющими меньшие номера или в порядке убывания количества вариантов совмещения вершин, если количество связей одинаково. Такой порядок следования вершин обусловлен тем, что чем больше связей с уже совмещенными имеет вершина, тем жестче будет ограничивающее условие, включающее эту вершину, и, соответственно, меньше общее количество совмещений, которые необходимо перебрать.

Основная часть алгоритма. Эта часть алгоритма представляет собой последовательное наложение вершин с возвратом, описывать которое удобно в терминах метода поиска в пространстве состояний.

Вершины графа G_2 остаются нетронутыми и каждой из них ставится в соответствие одна из вершин графа G_1. При этом проверяется допустимость такого совмещения. Если удается найти соответствие всем вершинам графа G_2, при этом выполнено условие изоморфизма, то найденное состояние возвращается как искомая подстановка.

Пусть $T_{1,i}$ – количество связей вершины i графа G_1 с вершинами $V_{1,j} \in \varphi_1(s)$, а $T_{2,i}$ – количество связей вершины i графа G_2 с вершинами $V_{2,j} \in \varphi_2(s)$.

Начальному состоянию $\varphi(s)_0 = 0$ соответствует состояние, при котором не совмещено еще ни одной пары вершин.

Для получения i-го состояния для вершины $V_{2,i}$ ищется соответствие среди вершин $V_{1,j}$, таких что:

1. $M_{i,j} = 1$, т. е. вершины совместимы на основании предварительных проверок;

2. $T_{1,i} \geq T_{2,j}$;

3. Для $k = 1..i$, если $(v_i, v_k) \in E_1$, то $(\varphi(v_i), \varphi(v_k)) \in E_2$.

Если выполнены все три условия, из которых третье является прямым следствие определения граф-подграф изоморфизма, то соответствующая пара вершин входит в частичную подстановку и формируется новое состояние $\varphi(s)_i$.

Перебор состояний производится методом поиска в глубину.

3. РЕЗУЛЬТАТЫ РАБОТЫ

Результат работы машины Уоррена может быть представлен в виде ГСА – графовой модели, в которой используются следующие типы вершин (блоков) (рис. 3).

Для тестирования производительности разработанного алгоритма формировались случайные графы, состоящие из блоков P и D типа. При формировании графов использовались следующие параметры:

nv – число вершин в генерируемых графах;

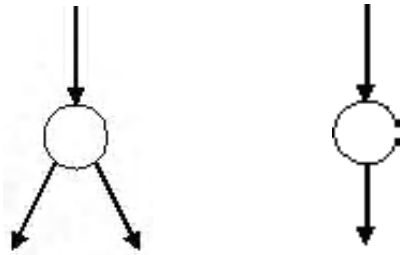
ns – нижняя граница число вершин в генерируемых подграфах;

p – процент вхождения P-блоков (процент вхождения D-блоков составлял 100 %–Np);

l – количество различных меток, используемых для маркировки вершин графов.

ГСА формировался из отдельных блоков P и D типа, на основании параметра p в соотношении p/(100–p). Дуги блоков случайным образом соединялись между собой. Общее число вершин ГСА определялось параметром nv. Всем вершинам случайным образом приписывались метки в диапазоне от 0 до l–1.

Подграф формировался путем удаления части вершин из ГСА. Начиная со случайной вершины, пускалась волна, проходящая как по входящим, так и по исходя-



P блок – условная вершина графа

D блок – операторная вершина графа

Рис. 3. Типы вершин графа

щим дугам, для формирования связного подграфа. Как только на очередном шаге размер подграфа вошедшего в волновое разложение превышал параметр ns, все вершины, которые накрыла волна, выделялись в виде подграфа.

Приводятся результаты численного исследования производительности алгоритма на основании набора сгенерированных графов, описанного выше. Каждое значение формировалось как суммарное время поиска подграфа для 100 пар графов (рис. 4).

Для графов с числом вершин до 1500 включительно, алгоритм в среднем тратит на поиск изоморфного подграфа не более 0,35 секунды машинного времени, что достаточно для решения реальных задач проектирования схемных устройств управления. Ввиду комбинаторной природы алгоритма, наблюдаются пики производительности, при которых время поиска подграфа в отдельных случаях может заметно отличаться от среднего (рис. 5).

Влияние размера искомого подграфа на общую производительность алгоритма значительно менее весомое. При изменении размера подграфа по отношению к размерам ГСА от 1 % до 50 % (т. е. в 50 раз) время работы алгоритма изменилось лишь с 28 секунд до 100 (т. е. в 3 раза). С учетом того, что большинство ГСА функционально-ориентированных конечных автоматов имеют размер порядка 20–40 вершин, фактор размера подграфа не является решающим для производительности алгоритма (рис. 6).

При изменении процента блоков *P*-типа использованных для генерации графов конечного автомата в диапазоне от 10 % до 50 % (в 5 раз), время вычислений изменилось с 33 до 50 секунд (т. е. в 1,5 раза). Следовательно, как и фактор размера искомого подграфа, соотношение числа блоков *P* и *D* типов незначительно влияет на производительность алгоритма, однако с ростом процента *P* блоков производительность алгоритма все же незначительно падает (рис. 7).

Из приведенного графика следует, что фактор числа различных меток приписываемых вершинам графов (т. е. числа различных вычислительных узлов, применяемых при формировании конечного автомата) меньше всего влияет на производительность алгоритма. Вне зависимости от значения параметра *l*, время сравнения одной пары графов остается в очень узких пределах от 0,35 до 0,36 секунды машинного времени. Такое малозначительное влияние параметра *l* объясняется значительным вкладом то-

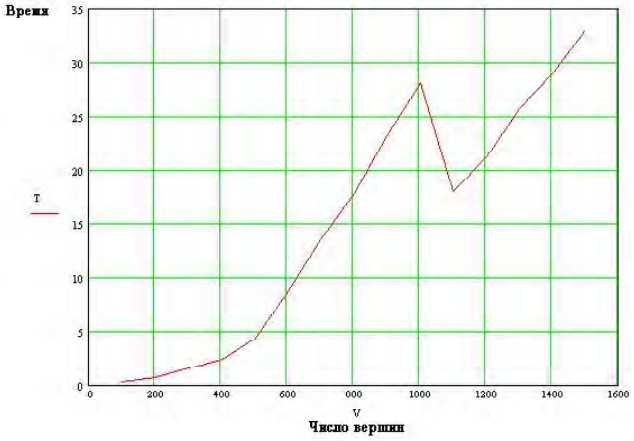


Рис. 4. Зависимость времени поиска подграфа от числа вершин в графе конечного автомата (параметры *ns*=20, *p*=20, *l*=5, *nv*=100…1500)

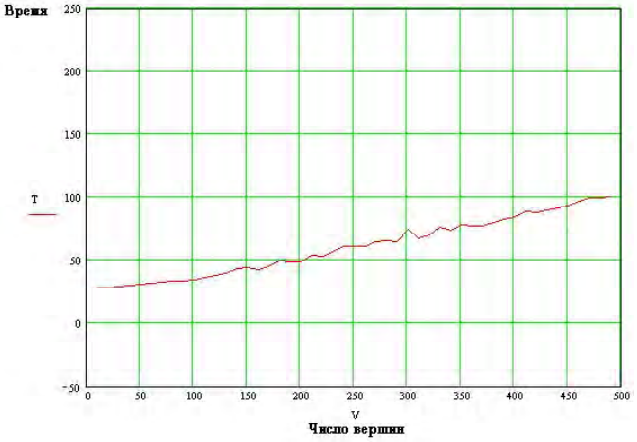


Рис. 5. Зависимость времени поиска подграфа от числа вершин в подграфе типового ГСА (параметры *nv*=1000, *p*=20, *l*=5, *ns*=10…500)

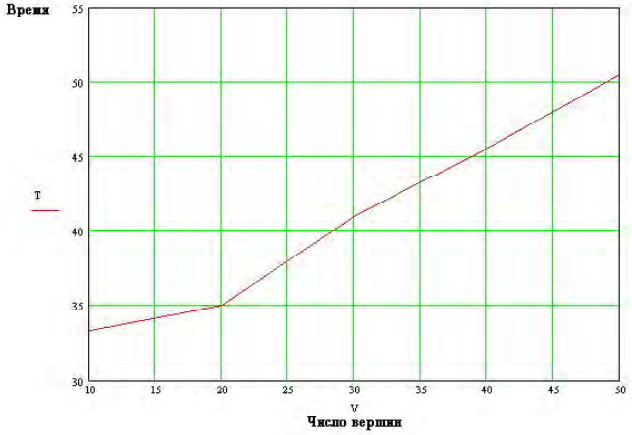


Рис. 6. Зависимость времени поиска подграфа от соотношения *P* и *D* блоков в графе конечного автомата (параметры *nv*=1000, *ns*=100, *p*=10..50, *l*=5)

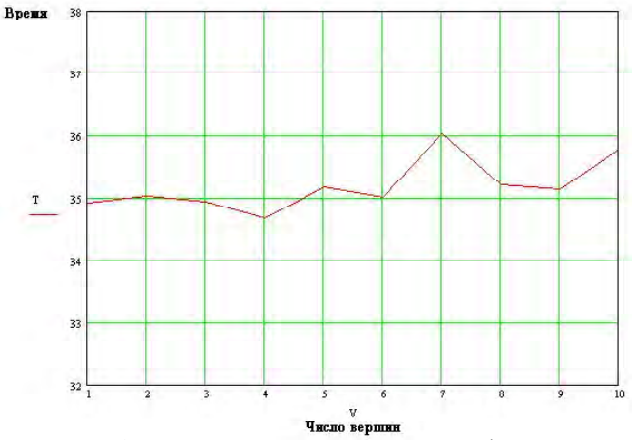


Рис. 7. Зависимость времени поиска подграфа от числа различных меток приписываемых вершинам графов (параметры *nv*=1000, *ns*=100, *p*=20, *l*=1..5)

пологии графов в условие, ограничивающее дерево возможных решений – для случайных графов и графов, не включающих в себя регулярные структуры, большая часть информации используемой для ограничения дерева поиска получается из анализа топологии графов. Но чем больше регулярных структур будут содержать графы, тем более значительным становиться влияние фактора меток приписанных вершинам графов.

ВЫВОДЫ

В работе представлено решение задачи поиска граф-подграф изоморфизма для помеченных графов. Приводится детальное описание разработанного алгоритма и результатов исследования его производительности для графов, характерных для конечных автоматов с программируемой процедурой. Применение описанного алгоритма позволяет расширить класс решаемых задач на автоматы с программируемой структурой, что является актуальным [6, 7, 8].

При проектировании проблемно-ориентированных вычислителей на ПЛИС, данное решение обеспечивает оптимизацию аппаратурных затрат на этапах ввода задания на проектирование, конфигурирования микропроцессорных систем с программируемой архитектурой, реализации технологий логического, концептуального, функционального программирования и др.

Исследовано влияние различных параметров ГСА на производительность алгоритма поиска подграфа. Результаты исследования показали, что разработанный алгоритм может быть эффективно применен в современных САПР, например, для создания функционально-ориентированных конечных автоматов, полученных в результате анализа концептуальной области проблемно-ориентированным пролог-процессором, одним из способов реализации которого является абстрактная машина Уоррена.

СПИСОК ЛІТЕРАТУРИ

1. *Голдобин, А. А.* Квазигомоморфное преобразование гиперграфов в автоматизации проектирования устройств управления / А. А. Голдобин // Радиоэлектроника, информатика, управление. – 2006. – № 1. – С. 41.
2. *Тейз, А.* Логический подход к искусственному интеллекту: от классической логики к логическому программированию : пер. с франц. / Тейз А., Грибомон П., Луи Ж. и др. – М. : Мир, 1990. – 432 с.
3. *Hassan Ait-Kasi Warren's* Abstract Machine : a tutorial reconstruction / Hassan Ait-Kasi Warren's. – MIT Press, 1999. – 144 p.
4. *Ильяшенко, М. Б.* Разработка и исследование параллельного алгоритма проверки граф-подграф изоморфизма / М. Б. Ильяшенко // Радиоэлектроника, информатика, управление. – 2006. – № 1. – С. 63–69.
5. *Пинчук, В. П.* Основанная на волновом разложении система инвариантов для простых графов и алгоритм распознавания изоморфности / В. П. Пинчук. – К., 1995. – Деп. в ГНТБ Украины 10.05.95, N 1002 – Ук95.
6. *Каляев, А. В.* Многопроцессорные системы с программируемой архитектурой / Каляев А. В. – М. : Радио и связь, 1984. – 240 с.
7. *Баркалов, А. А.* Синтез устройств управления на программируемых логических устройствах / Баркалов А. А. – Донецк : РВА ДонНТУ, 2002. – 262 с.
8. *Соловьев, В. В.* Проектирование цифровых систем на основе программируемых логических интегральных схем / Соловьев В. В. – М. : Горячая линия-Телеком, 2001. – 636 с.

Стаття надійшла до редакції 10.10.2011.

Ільяшенко М. Б., Голдобін О. О.

ВИРІШЕННЯ ЗАДАЧІ ПОШУКУ ІЗОМОРФІЗМУ ГРАФІВ ДЛЯ ПРОЕКТУВАННЯ СПЕЦІАЛІЗОВАНИХ ОБЧИСЛЮВАТЕЛІВ

Пропонується вдосконалений алгоритм пошуку ізоморфізму графів та результати дослідження його ефективності. Об'єктом дослідження є множина граф-схем алгоритмів досягнення мети, що була отримана після обходу семантичної мережі, яка задана, за допомогою абстрактної машини Уоррена.

Ключові слова: декларативна логіка, предикат, дерево виводу, пролог, рекурсивний обхід із поверненням, граф-підграф ізоморфізма.

Il'yashenko M. B., Goldobin A. A.

GRAPH-SUBGRAPH ISOMORPHISM PROBLEM SOLVING FOR DESIGNING SPECIAL COMPUTERS

An advanced algorithm for solving graphs isomorphism problem is proposed and experimental results of its efficiency are presented. Object of investigation is set of control flow graphs of solutions achieved, that were received after circumvent of the semantic network by Warren abstract machine.

Key words: declarative logic, predicate, O-Tree, Prolog, recursively returning, graph-subgraph isomorphism.

УДК 621. 391:519.2:519.7

Лисицкая И. В.

Канд. техн. наук, доцент Харьковского национального университета радиоэлектроники

СРАВНЕНИЕ ПО ЭФФЕКТИВНОСТИ СУПЕРБЛОКОВ НЕКОТОРЫХ СОВРЕМЕННЫХ ШИФРОВ

Излагается новая методика оценки показателей доказуемой безопасности блочных симметричных шифров. С применением этой методики выполняется анализ дифференциальных свойств суперблоков трех шифров: AES-а, уменьшенной версии шифра Мухомор и шифра MISTY1. Излагается оригинальная методика оценки максимального значения дифференциала уменьшенной модели двухциклового AES суперблока, и уточняется действительное значение этого максимума. Демонстрируется, что стойкость больших шифров и, в частности шифра Rijndael (AES-а) не зависит от дифференциальных показателей *S*-блоков, используемых в шифрах. Представляется как одно из перспективных решений по построению суперблоков преобразование FI шифра MISTY1, которое примечательно тем, что реализует (за один цикл) дифференциальные свойства случайной подстановки соответствующей степени.

Ключевые слова: доказуемая безопасность, дифференциал, суперблок, случайная подстановка.

ВВЕДЕНИЕ

В этой работе под суперблоком мы будем понимать функционально законченный узел шифра, включающий в себя композицию нескольких преобразований цикловой функции. В частности, в работах [1, 2] AES суперблоком названо отображение 4-х байтового массива $a = [a_0, a_1, a_2, a_3]$ в 4-х байтовый массив $e = [e_0, e_1, e_2, e_3]$, принимающее 4-байтовый ключ k $[k_0, k_1, k_2, k_3]$. Оно состоит из последовательности четырех преобразований:

SubBytes $b_i = S[a_i]$, с S являющимся AES S-блоком;

MixColumns $c = \mathrm{M}_c b$, с M_c являющейся 4 Ч 4 матрицей;

AddRoundKey $d = c \oplus k$, с k являющимся цикловым ключом;

SubBytes $e_i = S[d_i]$.

Авторами отмечается, что дифференциальные вероятности над этой структурой эквивалентны двум AES циклам и доказывается с использованием достаточно громоздких и сложных для понимания теоретических построений с привязкой к дифференциальным характеристикам S-блока AES, что точным значением максимальной ожидаемой дифференциальной вероятности ($MEDP$[1]) для AES супер блока является значение $12{,}34\times2^{-32}$ (есть и варианты значения $MEDP_{32} \approx 13{,}25\times2^{-32}$ [3]).

В итоге формируется граница для дифференциалов над AES, уменьшеному до четырех циклов, следующая из применения границы Хонга и др. [3]:

$$MEDP_{32} \le \left(\max_{x\neq 0, y} DP(x, y) \right)^4$$

к мега блоку, что приводит к результату:

$$MEDP_{128} \le (MEDP_{32})^4 \approx 1{,}881\times 2^{-114}.$$

Имеются работы, где подобным же образом (с привязкой к свойствам S-блоков) выполняется оценка линейных показателей SPN шифров [4]. Этот подход к определению доказуемой стойкости блочных симметричных шифров (БСШ) уже давно вызывает у нас сомнение, так как полученные результаты привязываются к дифференциальным и линейным свойствам S-блоков, используемых в шифрах, что, как показывают наши эксперименты, методически оказывается не верным. Не вызывает удовлетворения и сама методика определения показателей доказуемой стойкости БСШ в виде максимумов средних значений дифференциальных и линейных вероятностей ($MADP$ и соответственно $MALP$).

Мы далее обоснуем свою позицию к определению показателей доказуемой стойкости БСШ, и, в частности, дифференциальных показателей AES суперблока, приведем сравнение для него значений $MADP$ и оценок, полученных с использованием предлагаемого подхода, и заодно обсудим дифференциальные свойства суперблоков еще двух конструкций, где под суперблоками, как уже отмечено выше, будут пониматься функционально обособленные элементы цикловых преобразований других шифров.

1. ПОНЯТИЙНЫЙ АППАРАТ ДИФФЕРЕНЦИАЛЬНОГО И ЛИНЕЙНОГО КРИПТОАНАЛИЗА

Напомним сначала, следуя работе [4], понятийный аппарат линейного и дифференциального криптоанализа.

[1] *В ряде работ наряду с аббревиатурой MEDP (максимум ожидаемой дифференциальной вероятности) используется обозначение MADP (максимум среднего значения дифференциальной вероятности)*

Определение 1. (Дифференциальная и Линейная вероятность): *Дифференциальная вероятность DP^f и линейная вероятность LP^f соответственно для ключезависимой функции f с n-битным входом x и n-битным выходом y* ($x, y \in GF(2^n)$) *есть*:

$$DP^f(\Delta x \to \Delta y) = \frac{\#\{x \in GF(2)^n \mid f(x) \oplus f(x \oplus \Delta x) = \Delta y}{2^n};$$

$$LP^f(\Gamma x \to \Gamma y) = \left(\frac{\#\{x \in GF(2)^n \mid x \cdot \Gamma x) = f(x) \cdot \Gamma y}{2^{n-1}} - 1 \right)^2,$$

где Δx и Δy является входной и выходной разностями, а Γx и Γy является входной и выходной масками; $x \cdot \Gamma x$ обозначает результат скалярного произведения x и Γx.

Определение 2. ($DP^f_{\max}$ и $LP^f_{\max}$): *Максимальным значением дифференциальной и линейной вероятности для ключезависимой функции f называется соответственно*:

$$DP^f_{\max} = \max_{\Delta x \neq 0, \Delta y} DP^f(\Delta x \to \Delta y),$$

$$LP^f_{\max} = \max_{\Gamma x,\, \Gamma x \neq 0,} LP^{f[k]}(\Gamma x \to \Gamma y).$$

Напомним теперь выражения для средних вероятностей *ADP*, *ALHP*, *MADP* и *MALHP* ключезависимой функции $f = f[k](x)$ с *n*-битным входом *x* и *n*-битным выходом $y \in \mathrm{GF}(2^n)$, которая параметризована ключом *k*, используемые во многих публикациях по обоснованию показателей стойкости блочных шифров.

Определение 3. Средним значением дифференциальной вероятности (*ADP*) функции $f[k](x)$ является:

$$ADP^f = \underset{k}{ave}\, DP^{f[k]}(\Delta x \to \Delta y).$$

Определение 4. Средним значением вероятности линейной оболочки (*ALHP*) функции $f = f[k](x)$ является:

$$ALHP^f = \underset{k}{ave}\, LP^{f[k]}(\Gamma x \to \Gamma y).$$

Определение 5. Максимумом среднего значения дифференциальной вероятности (*MADP*) и максимумом среднего значения вероятности линейной оболочки (*MALHP*) функции $f[k](x)$ есть:

$$MADP^f = \max_{\Delta x \neq 0, \Delta y} ADP^f(\Delta x \to \Delta y),$$

$$MALHP^f = \max_{\Gamma x,\, \Gamma y \neq 0} ALHP^f(\Gamma x \to \Gamma y).$$

Из приведенных определений видно, что приведенные показатели определяются максимумом среднего значения дифференциальной вероятности для некоторого фиксированного перехода входной разности Δx в выходную разность Δy, и максимумом среднего значения смещения для маски входа Γx и маски выхода Γy. Эти показатели представляют собой далеко не максимально возможные значения дифференциальных и линейных вероятностей, которые по идее и должны рассматриваться как показатели доказуемой безопасности.

Новая точка зрения к формированию оценок стойкости БСШ к атакам дифференциального и линейного криптоанализа, которая формализуется как два новых метода, состоит в следующем.

Предлагается для оценки стойкости БСШ к атакам дифференциального и линейного криптоанализа пользоваться не *MADP* и *MALHP*, а средними (по множеству ключей) значениями максимумов дифференциальных и линейных вероятностей ключезависимой функции $f[k](x)$, а именно *AMDP* и *AMLHP*.

Определение 6. (*AMDP*). *Среднее* (*по множеству из* 2^h *ключей*) *значение максимальных дифференциальных вероятностей ключезависимой функции* $f[k](x)$ *есть*:

$$AMDP^f = \underset{k}{ave}\, DP^{f[k]}_{\max} = \frac{1}{2^h} \sum_{k=1}^{2^h} DP^{f[k]}_{\max}.$$

Определение 7. (*AMPLH*). *Среднее* (*по ключам*) *значение максимальных вероятностей линейных оболочек функции* $f[k](x)$ *есть*:

$$AMLHP^f = \underset{k}{ave}\, LP^f_{\max}(\Gamma x \to \Gamma y) = \frac{1}{2^h} \sum_{k=1}^{2^h} LP^{f[k]}_{\max}.$$

В обоих случаях 2^h – мощность множества ключей зашифрования, использованных при вычислениях.

Здесь можно отметить сразу, что очевидны неравенства: $MADP^f < AMDP^f$, $MALHP < AMLHP$.

Помимо большей адекватности формируемых оценок (значение оценок для шифров совпадают с соответствующими дифференциальными и линейными показателями случайных подстановок и характеризуют максимально достижимые значения дифференциальных и линейных вероятностей), в последнем случае обеспечиваются и значительные вычислительные преимущества (нет необходимости запоминать полностью все таблицы, а достаточно только определять и помнить их максимальные значения).

2. ОБ УЧАСТИИ *S*-БЛОКОВ В ФОРМИРОВАНИИ МАКСИМАЛЬНЫХ ЗНАЧЕНИЙ ДИФФЕРЕНЦИАЛЬНЫХ И ЛИНЕЙНЫХ ВЕРОЯТНОСТЕЙ ШИФРОВ

Наши исследования с уменьшенными версиями многих шифров показали, что значения максимумов полных дифференциалов и линейных корпусов, которыми оцениваются показатели стойкости шифров к атакам дифференциального и линейного криптоанализа, зави-

сят не от показателей *S*-блоков, используемых в шифрах, а от дифференциальных и линейных показателей случайных подстановок соответствующей степени, к которым асимптотически приходят шифры после определенного начального числа циклов шифрования.

Для иллюстрации этого положения ниже предлагаются результаты исследований дифференциальных свойств 16-битной модели шифра Rijndael [5]. Для таких размеров входных блоков данных вычислительных ресурсов вполне достаточно, чтобы построить целиком таблицу XOR переходов (полных дифференциалов) сразу для всего шифра.

В табл. 1 представлены зависимости средних значений максимумов полных дифференциалов ($AMDP\times 2^{16}$) шифров, использующих *S*-блоки с различными значениями $DP^{s}_{\max} = p$ (δ-равномерности), от числа циклов *r* алгоритма Baby-Rijndael с операцией MixColumns на весь текст (как раз преобразование, являющееся основой структуры названной выше AES суперблоком).

Результаты, представленные в табл. 1, ярко иллюстрируют, что показатели стойкости шифров не зависят от применяемых в них *S*-блоков. Они определяются, как показано и в ряде других наших работ [6–8 и др.], значениями максимумов таблиц XOR разностей и смещений таблиц линейных аппроксимаций случайных подстановок соответствующей степени. У нас, правда, сразу нашлось много оппонентов, которые нас стали убеждать, что то, что свойственно малым шифрам, может не выполняться для их больших прототипов. Однако, последние наши исследования с большими шифрами [9–10] свидетельствуют о том, что и большие шифры также ведут себя как случайные подстановки, т.е. наша позиция является правильной.

В результате свойства AES суперблока не являются решающими для определения показателей доказуемой безопасности шифра Rijndael. Мы, тем не менее, далее рассмотрим дифференциальные свойства этого и других, близких к нему преобразований, с целью совершенствования и развития самой методики определения введенных выше новых показателей к оценке стойкости шифрующих преобразований (шифров) и сравнения их со старыми подходами.

2. ОЦЕНКА ЗНАЧЕНИЙ ДВУХЦИКЛОВОГО ДИФФЕРЕНЦИАЛА AES СУПЕРБЛОКА

Ниже предлагаются результаты вычислительных экспериментов по определению *AMDP* и *MADP* AES суперблока.

Конечно же, построить всю таблицу дифференциальных разностей для AES суперблока, также как и подстановки степени 2^{32}, не удается (не хватает вычислительных возможностей), но вполне достаточно имеющихся вычислительных ресурсов для построения закона распределения переходов отдельной строки таблицы XOR разностей.

Результаты решения этой задачи и представляются в табл. 2.

В правой колонке таблицы мы для сравнения представили закон распределения переходов в строке случайной подстановки степени 2^{32} (строка AES суперблока не «дотягивает» до строки случайной подстановки степени 2^{32}). Отметим здесь, что результаты для отдельных ключей практически не зависят от ключевого материала, т.е. в качестве оценки может выступать дифференциал, рассчитанный для отдельного ключа (это еще одно из достоинств развиваемого подхода).

Как следует из представленных данных значение максимума строки дифференциальной таблицы AES суперблока для выбранного входа равно 32. Это значит, что для всей дифференциальной таблицы ожидаемое значение максимума будет не менее 64-х, что существенно больше значения 12, 34, пропагандируемого в отмеченных ранее зарубежных публикациях (в последующих экспериментах нам удалось найти переход со значением 40).

Полученные результаты позволяют сделать вывод, что ожидаемое максимальное значение двухциклового

Таблица 2. Распределение переходов одной строки таблицы XOR-разностей AES суперблока ($AMDP\times 2^{16}$) для входа (в строку) 010101, безключевой вариант

Значение перехода	Число переходов в строке AES супер блока	Число переходов в строке подстановке степени 2^{32}
0	2605143438	2605070418
2	1302455376	1302484861
4	325637706	325626184
6	54254936	54271858
8	6794838	6784085
10	679254	678418
12	61352	56535
14	4021	4038
16	1543	252
18	13	14
20	291	1
24	52	
28	8	
32	4	

Таблица 1. Значения полного дифференциала ($AMDP\times 2^{16}$) для различных *S*-блоков и количества циклов алгоритма Rijndael с операцией MixColumns на весь текст

Sbox / *r*	Sbox, Сл *p*4F2	Sbox. *p*4 Лабир.	Sbox AES *p*4	Sbox *p*6F0	Sbox *p*6 F2	Sbox DES *p*8	Sbox *p*8 F0	Sbox *p*12 F0
1	16384,00	16384,00	16384,00	24576,00	24576,00	32768,00	32768,00	49152,00
2	83,87	132,00	132,00	490,87	230,40	1152,00	1536,00	5184,00
3	20,73	**19,47**	**18,80**	25,53	35,27	70,87	139,13	146,13
4	**19,60**	18,73	19,00	**19,20**	**18,93**	**19,27**	**23,93**	**19,07**
5	19,13	19,47	19,47	18,93	19,40	19,00	23,87	19,00

дифференциала большого AES супер блока ($AMDP \times 2^{32}$) должно быть больше максимального значения дифференциальной таблицы случайной подстановки степени 2^{32} равного 34.

В другом эксперименте мы для входа 010101 нашли максимальное значение строки дифференциальной таблицы AES суперблока. Оно оказалось равным 32. А затем для найденного максимального перехода $\Delta x = 10101 \rightarrow \Delta y = 661E0000$ были вычислены значения переходов при других значениях ключа. В табл. 3 представлены числа переходов, полученные для 30-ти случайно выбранных ключей зашифрования.

В соответствии с этими данными получено значение *MADP* для строки AES супер блока:

$$MADP(010101, 661E0000) \times 2^{16} = 7{,}86.$$

Для всей дифференциальной таблицы AES суперблока следует ожидать значение максимума в районе 15,72, что несколько больше значения 12,34, используемого в работах по оценке доказуемой безопасности шифра Rijndael (AES). Другое среднее значение $MADP \times 2^{16}$ максимального перехода для входа 010101 в AES получилось равным 8,76. Но дело не в этих небольших разницах в оценках значений *MADP* для AES суперблока. Получается, что приведенные в публикациях результаты являются, говоря мягко, не совсем точными. Самое главное это то, что эти значения не связаны с действительными значениями доказуемой стойкости шифров, как это считается в затронутых публикациях.

В табл. 4 мы приводим результаты исследования 16-битной версии SL преобразования шифра Мухомор [11] (практически изучаются показатели уменьшенной модели самого AES суперблока). В таблице представлены результаты экспериментов по построению законов распределения переходов дифференциальной таблицы SL преобразования.

Видно, что второй цикл является достаточно сложным для рассматриваемого преобразования. Приведем свои соображения по подсчету максимума дифференциальной таблицы для AES суперблока (для двух циклов шифрования AES).

Таблица 3. Числа переходов ($AMDP \times 2^{16}$) входа 010101 в один и тот же выход 661E0000 для 30-ти случайно выбранных ключей зашифрования

Ключ зашифрования	Значение перехода	Ключ зашифрования	Значение перехода
	32	7E0ACA68	8
10C9AA38	4	42DC38BF	0
EF34B4F6	16	F21B574C	8
522F3364	8	6F00601B	16
73B2CD8B	16	59ED7EE9	4
3BB11EC5	0	9F86B693	4
6C1A60C7	0	72569299	4
21F6E7E9	12	358F25B0	4
2FB26869	8	4848E2BE	4
3888589D	4	A0A2430D	16
27A47122	4	437C58F9	8
178448CA	16	3C30A2B6	4
C0D90AAE	4	ECA05DB8	12
7CCD5C0D	4	AEA79D44	8
B202E14F	0	DFE9D423	12
A5C7A90C	4		

При прохождении разностей пар входных блоков через первый цикл (*S*-блоки и преобразование MixColumn) наибольшая вероятность перехода $\Delta X \rightarrow \Delta Y$ обеспечивается при одном активном *S*-блоке. На его входе при прохождении по всем 2^{16} возможным значениям входных разностей каждая фиксированная разность повторяется $2^{16} / 2^4 = 2^{12}$ раз. Линейное преобразование тиражирует каждую разность на все четыре входа *S*-блоков следующего цикла. Если *S*-блок имеет значение δ-равномерности $\delta = 4$ (для *S*-блока AES), то на выходе первого цикла будет зафиксирован переход с максимальным значением $\delta \cdot 2^{12}$ (для AES это будет 2^{14}), причем если *S*-блок имеет 15 максимальных переходов $\delta = 4$, то на выходе первого цикла будет 60 значений $2^{14} = 16384$ (см. табл. 2).

На следующем цикле в прохождении разностей будут участвовать уже все четыре *S*-блока (второго цикла). Одна и та же разность на входе линейного преобразования сформирует повторяющиеся 16384 раза значения разностей (различных) на входах *S*-блоков второго цикла. А это значит, что, проходя *S*-блоки второго цикла, разные пары входов (после сложения с ключевыми битами) для разных *S*-блоков дадут разные значения выходных разностей со своими показателями прохождения (для AES *S*-блока это будут в подавляющем большинстве двойки). Заметим теперь, что ненулевые входные разности будут давать только ненулевые выходные разности. По статистике *S*-блоки имеют около 40 процентов ненулевых переходов (AES полубайтовый *S*-блок имеет 120 нулей в дифференциальной таблице без учета нулевой строки и нулевого столбца), т.е. из всего множества $2^4 - 1 = 15$ возможных ненулевых значений выходов для полубайтового *S*-блока в строке таблицы может быть реализовано только $15 \cdot 0{,}47 = 7$ различных ненулевых выходов.

Итак, нам нужно подсчитать число ситуаций, когда на выходах четырех *S*-блоков второго цикла будет одинаковое число совпадающих выходных разностей. Пусть мы зафиксировали одну (любую) из выходных (ненулевых) разностей первого *S*-блока. К этой разности мы можем выбрать одну из 6 возможных ненулевых разностей с выхода второго *S*-блока, так что вероятность получить набор из двух фиксированных значений разностей будет равна $\frac{1}{7}$. К этим двум разностям можно добавить еще одну из разностей с выхода третьего *S*-блока. Вероятность такой тройки (композиции) будет $\left(\frac{1}{7}\right)^2$. Наконец, к выбранной тройке можно добавить разность с выхода четвертого *S*-блока, и вероятность выбора этой четверки будет, по аналогии с предыдущим, равна $\left(\frac{1}{7}\right)^3$.

Таблица 4. Распределение числа переходов дифференциальной таблицы SL преобразования шифра Мухомор в зависимости от числа циклов шифрования

Количество переходов для ячейки	Число ячеек S_2(Baby$R$), 1 цикл	Количество переходов для ячейки	Число ячеек S_2(Baby$R$), 2 цикла	Количество переходов для ячейки	Число ячеек S_2(Baby$R$), 3 цикла
0	4168654065	**0**	2632290711	**0**	2605108264
16	65610000	**2**	1263628451	**2**	1302316781
32	43740000	**4**	329970420	**4**	325638830
64	10935000	**6**	56464541	**6**	54301413
128	4131000	**8**	10545896	**8**	6796804
256	1508625	**10**	1165323	**10**	678225
512	243000	**12**	324071	**12**	56663
1024	62100	**14**	20706	**14**	4390
2048	16200	**16**	379822	**16**	364
4096	1350	**18**	28307	**18**	21
8192	360	**20**	39641	**20**	5
16384	60	**22**	1296		
		24	7016		
		26	215		
		28	449		
		32	25940		
		34	3133		
		36	4032		
		38	176		
		40	442		
		44	36		
		48	4		
		64	790		
		66	144		
		68	166		
		72	20		
		128	8		
		132	4		
Время, с	192	**Время, с**	350	**Время, с**	509

Итак, с помощью четырех S-блоков мы можем получить фиксированный набор из четырех разностей (одинаковых или разных) с вероятностью $p=\left(\frac{1}{7}\right)^3$ и любой другой отличающийся от выбранной четверки набор из четырех разностей с вероятностью $p-1=1-\left(\frac{1}{7}\right)^3$. В результате можно считать, что мы имеем дело с двумя событиями, подчиняющимися биномиальному закону: одно событие – появление четверки разностей совпадающей с выбранной (вероятность такого события), другое – появление четверки разностей не совпадающей с выбранной (вероятность этого события). Для выборки $2^{14} = 16384$ таких независимых исходов среднее число совпадающих четверок выходных разностей в таком случайном эксперименте будет равно математическому ожиданию биномиального распределения, т. е.

$$2^{14} \cdot \left(\frac{1}{7}\right)^3 = 47{,}7.$$

Но этот результат является, конечно, оценкой снизу. Реальные значения (с учетом особенностей S-блоков, значений максимумов переходов, их распределения по таблице и других показателей) будут в общем случае существенно более высокими (в нашем примере имеется минимальное значение 83,87 и максимальное 5184).

3. СУПЕР БЛОК MISTY1

Мы выделили еще одну криптографическую функцию, заслуживающую внимания. Это FI подстановка в шифре MISTY1, являющегося еще одним из финалистов конкурса NESSIE.

Алгоритм MISTY1 разработан в 1995–1996 гг. командой специалистов под руководством известного криптолога Мицуру Мацуи (Mitsuru Matsui) из компании Mitsubishi Electric (Япония) [12]. Он имеет весьма необычную структуру – основан на «вложенных» сетях Фейстеля. Сначала 64-битный шифруемый блок данных разбивается на два 32-битных субблока, после чего выполняется r-циклов преобразований, имеющих ярко выраженную трехуровневую вложенную структуру. Рекомендуемым количеством раундов алгоритма является 8, но количество раундов алгоритма может быть любым, превышающим 8 и кратным четырем.

Мы не будем здесь приводить описание этой оригинальной конструкции, а интересующихся отправим к

оригинальной разработке [12]. Нас будет интересовать «кирпичики» – преобразования FI этой достаточно сложной конструкции, из которых строится цикловая функция. Ее структуру иллюстрирует рис. 1.

FI также (как и основная конструкция шифра) представляет собой сеть Фейстеля, но в шифре MISTY1 это преобразование осуществляет уже третий уровень вложенности. В отличие от сетей Фейстеля на двух верхних уровнях, данная сеть является несбалансированной: обрабатываемый 16-битный фрагмент делится на две части: 9-битную левую и 7-битную правую. Затем выполняются 3 раунда преобразований, которые состоят из следующих действий:

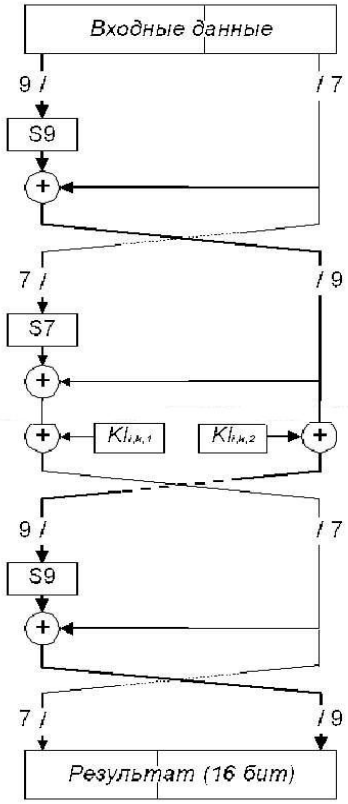


Рис. 1. Структура функции FI

1. Левая часть «прогоняется» через таблицу замен. 9-битная часть (в раундах 1 и 3) обрабатывается таблицей подстановки S_9, а 7-битная (в раунде 2) – таблицей подстановки S_7. Сами таблицы приведены в описании шифра [12].

2. На левую часть операцией XOR накладывается текущее значение правой части. При этом, если справа 7-битная часть, она дополняется нулями слева, а у 9-битной части удаляются слева два бита.

3. Во втором раунде на левую часть операцией XOR накладывается фрагмент ключа раунда $KI_{i,k,1}$, а на правую – фрагмент $KI_{i,k,2}$. В остальных раундах эти действия не выполняются.

4. Левая и правая части меняются местами.

Будем рассматривать функцию FI как суперблок. Сравним свойства этого суперблока с суперблоком AES рассмотренным ранее.

Как видно из представленных данных в рассматриваемом случае мы сразу (за один цикл) получаем закон распределения близкий к закону распределения переходов дифференциальной таблицы случайной подстановки 16-ой степени. Это преобразование на выходе сразу реализует асимптотическое значение максимума полного дифференциала (правда, это достигается внутренней трехцикловой структурой преобразования).

Отметим здесь, однако, что наши эксперименты с совершенными *S*-блоками, так мы назвали *S*-блоки, обладающими показателями случайных *S*-блоков (имеющих законы распределения числа инверсий, возрастаний и циклов, а также законы распределения переходов XOR таблиц и смещений таблиц линейных аппроксимаций, повторяющие соответствующие теоретические законы), показывают, что они ведут себя также как и другие применяемые в шифрах случайные или не случайные *S*-блоки, т.е. их применение не приводит к заметному улучшению характеристик сходимости шифров к асимптотическому значению максимума полного дифференциала. Поэтому вопрос об эффективности использования при

Таблица 5. Значения переходов дифференциальной таблицы супер блока MISTY1 для различных значений ключа зашифрования

Кол-во переходов в ячейке	Число ячеек Ключ 0x0000:	Число ячеек Ключ 0xFF00:	Число ячеек Ключ 0x0F0F:	Число ячеек Ключ 0x1234:	Число ячеек Ключ 0x1111:	Число ячеек Ключ 0xAAAA:
0	2605549364	2605492361	2605539766	2605527119	2605516029	2605520863
2	1300024352	1300109653	1300032540	1300064477	1300064802	1300060159
4	328372588	328366110	328380339	328359715	328384633	328378643
6	53631298	53614855	53627998	53624362	53618333	53624945
8	6564176	6560111	6562427	6566955	6558540	6557969
10	639805	638633	638937	639359	639510	639068
12	50994	50888	50481	50617	50780	50904
14	3430	3414	3517	3406	3374	3447
16	208	193	210	209	212	212
18	9	7	10	5	11	15
20	1	0	0	0	1	0
22				10	0	0

построении шифров преобразований, обладающих показателями более близкими к показателям случайных подстановок, остается пока открытым. Очевидно, что основная компонента обеспечения случайности преобразования все-таки связана с реализацией механизма достаточно глубокого перемешивания обрабатываемых блоков данных внутри «тела» всего шифра – достижения статистической инвариантности распределения разностей на выходе преобразования от ключевых и текстовых битов.

ВЫВОДЫ

Результатами работы следует считать выполненный анализ дифференциальных свойств суперблоков трех шифров: AES, мини Мухомора (SL преобразования этого шифра, как варианта уменьшенного AES суперблока) и шифра MISTY1.

И все же основным результатом является положение, в соответствии с которым свойства AES суперблока не являются решающими для определения показателей доказуемой безопасности шифра Rijndael.

Предложено вместо оценок максимумов средних значений дифференциальных и линейных вероятностей (*MADP* и *MALHP*) суперблоков и шифров рассматривать средние значения максимумов этих вероятностей (*AMDP* и *AMLHP*), как более адекватно характеризующих потенциальные возможности в реализации максимумов дифференциальных и линейных показателей шифрующих преобразований. Эти оценки в несколько раз превышают значения *MADP* и *MALHP* и позволяют получить более точные результаты.

В процессе этого анализа разработана уточненная методика оценки максимального значения дифференциала ($AMDP \times 2^{-32}$) двухциклового AES суперблока. В качестве более точной оценки вероятности максимального значения двухциклового дифференциала (*AMDP*) обосновано значение $48/2^{32}$ (сегодня эксперименты уже дали результат $80/2^{32}$). Показано, что стойкость больших шифров и, в частности, шифра Rijndael (AES-a) не зависит от дифференциальных (и линейных) показателей *S*-блоков, используемых в шифрах. В соответствии с нашими результатами она определяется соответствующими характеристиками случайных подстановок, к которым приходит каждый шифр при увеличении числа циклов шифрования [13].

Представлено как одно из перспективных решений по построению суперблоков (криптографических примитивов) преобразование FI шифра MISTY1. Это преобразование реализует за один цикл (состоящий из последовательности трех простых преобразований) дифференциальные показатели, характерные для случайной подстановки соответствующей степени.

СПИСОК ЛИТЕРАТУРЫ

1. Computational aspects of the expected differential probability of 4-round AES and AES-like ciphers / M. Lamberger, J. Daemen, N. Pramstaller at al // Abstract – 8th Central European Conference on Cryptography 2008. Computing – 2009. – Pp. 85–104. DOI 10.1007/s00607-009-0034-y.
2. *Daemen, J.* 'Understanding two-round differentials in AES' / J. Daemen, V. Rijmen // Proc. Security and Cryptography for Networks (SCN 2006), LNCS, 4116, edited by De Prisco, R., and Yung, M., (Springer). – 2006. – Pp. 78–94.
3. *Keliher, L.* Exast maximum expected differential and linear probability for 2-round advanced encryption standard (AES) / L. Keliher, J. Sui // Cryptology ePrint archive Report 2005/321. –2005. – http://eprint.iacr.org.
4. *Sano, F.* On the Security of Nested SPN Cipher against the Differential and Linear Cryptanalysis / Sano, K. Ohkuma, H. Shimizu, S. Kawamura / IEICE Trans. Fundamentals, January 2003. – vol. E86-a, NO.1. – Pp. 37–46.
5. *Долгов, В. И.* Вариации на тему шифра Rijndael / В. И. Долгов, И. В. Лисицкая, А. В. Казимиров // Прикладная радиоэлектроника. – 2010. – Т.9, №3. – С. 321–325.
6. Криптографические свойства уменьшенной версии шифра «Мухомор» / И. В. Лисицкая, О. И. Олешко, С. Н. Руденко [та ін.] // Спеціальні телекомунікаційні системи та захист інформації. Збірник наукових праць. – Київ. – 2010. – Вип. 2(18). – С. 33–42.
7. *Кузнецов, А. А.* Линейные свойства блочных симметричных шифров, представленных на украинский конкурс / А. А. Кузнецов, И. В. Лисицкая, С. А. Исаев // Прикладная радиоэлектроника. – 2011. – Т. 10, № 2. – С. 135–140.
8. *Лисицкая, И. В.* Об участии *S*-блоков в формировании максимальных значений линейных вероятностей блочных симметричных шифров / И. В. Лисицкая, В. В. Ковтун // Межведомственный научн. технический сборник «Радиотехника». – 2011. – Вып. 166. – С. 17–25.
9. *Лисицкая, И. В.* Большие шифры - случайные подстановки / И. В. Лисицкая, А. А. Настенко // Межведомственный научн. технический сборник «Радиотехника». – 2011. – Вып. 166. –С. 50–55.
10. *Лисицкая, И. В.* Дифференциальные свойства шифра FOX / И. В. Лисицкая, Д. С. Кайдалов // Прикладная радиоэлектроника. – 2011. – Т. 10, № 2. – С. 122–126.
11. Перспективний блоковий симетричний шифр «Мухомор» – основні положення та специфікація / І. Д. Горбенко, М. Ф. Бондаренко, В. І. Долгов [та ін.] // Прикладная радиоэлектроника. – 2007. – Том. 6, №2. – С. 147–157.
12. *M. Matsui,* «New block encryption algorithm Misty», Fast Software Encryption '97, LNCS 1267, E. Biham, Ed., Springer-Verlag. – 1997. – Pp. 64–74.
13. Новая идеология оценки стойкости блочных симметричных шифров к атакам дифференциального и линейного криптоанализа / И. Д. Горбенко, В. И. Долгов, И. В. Лисицкая, Р. В. Олейников // Прикладная радиоэлектроника. – 2010. – Т. 9, № 3. – С. 212–320.

Стаття надійшла до редакції 23.02.2011.
Після доробки 22.02.2012.

Лисицька І. В.

ПОРІВНЯННЯ ЗА ЕФЕКТИВНІСТЮ СУПЕРБЛОКІВ ДЕЯКИХ СУЧАСНИХ ШИФРІВ

Викладається нова методика оцінки показників доказової безпеки блокових симетричних шифрів. Із застосуванням цієї методики виконується аналіз диференціальних властивостей суперблоків трьох шифрів: шифру AES, зменшеної версії шифру Мухомор і шифру MISTY1. Викладаються результати обчислювальних експериментів по визначенню значень AMDP і MADP AES суперблоку. Демонструється, що стійкість великих шифрів і, зокрема шифру Rijndael (AES) не залежить від диференціальних показників S-блоків, використ-

товуваних у шифрах. Представляється як одне з перспективних рішень з побудови суперблоку перетворення FI шифру MISTY1, яке примітно тим, що реалізує за один цикл диференціальні показники випадкової підстановки відповідного степеня.

Ключові слова: доказова безпека, диференціал, суперблок, випадкова підстановка.

Lysytska I. V.

COMPARING oN EFFECTIVENESS OF SUPERBOXES some MODERN SIPHERS

New method of assessment indicators provable security block symmetric ciphers sets out. With application of this method are analyzed for differential properties superblock three ciphers: cipher AES, the reduced version cipher Muhomor and cipher MISTY1. The results of computational experiments to determine the values of AMDP and MADP AES superblock are presented. Demonstrated that the resistance of large ciphers and, in particular cipher Rijndael (AES) is independent of the differential properties of S-blocks used in the ciphers. It seems like one of the promising solutions for building superblocks transformation FI cipher MISTY1, which is noteworthy that sells for one cycle of differential performance random permutation corresponding degree.

Key words: of provable security, differential, superblock, random permutation.

УДК 004.3

Баркалов А. А.[1], Мальчева Р. В.[2], Солдатов К. А.[3]

[1]Д-р техн. наук, проф. Университета Зеленогурского (Польша)
[2]Канд. техн. наук, доцент Донецкого национального технического университета
[3]Аспирант Донецкого национального технического университета

ОПТИМИЗАЦИЯ СХЕМЫ АВТОМАТА МУРА, РЕАЛИЗУЕМОЙ В БАЗИСЕ ПЛИС

В статье предлагается метод, предназначенный для уменьшения числа входных переменных и промежуточных термов в реализуемых системах булевых функций. Предложенный метод основан на расширении кодов состояний перехода и замене логических условий. Применение предложенного метода позволяет до 20 % уменьшить общее число макроячеек в блоках БЛУ и БФП.

Ключевые слова: автомат Мура, ПЛИС, ГСА, псевдоэквивалентные состояния, замена логических условий.

ВВЕДЕНИЕ

Практически любая цифровая система включает в свой состав устройство управления (УУ) [1]. При реализации схем УУ часто используется модель микропрограммного автомата Мура [2]. В настоящее время программируемые логические интегральные схемы (ПЛИС) [3] широко применяются для реализации схем УУ. Существуют два основных класса ПЛИС: CPLD (Complex Programmable Logic Devices) и FPGA (Field-Programmable Gate Arrays) [4, 5]. Для уменьшения числа макроячеек ПЛИС в схеме УУ необходимо уменьшать число входных переменных и промежуточных термов в реализуемых системах булевских функций (СБФ) [6]. В настоящей работе предлагается метод решения этой задачи для микропрограммного автомата (МПА) Мура. Метод основан на расширении кодов состояний перехода и замене логических условий.

Целью исследований является оптимизация схемы МПА Мура за счет расширения кодов состояний перехода и замены логических условий.

Задачей исследований является разработка метода синтеза МПА Мура, позволяющего уменьшить число макроячеек ПЛИС в схеме автомата. При этом алгоритм управления представляется в виде граф-схемы алгоритма (ГСА) [1].

ОБЩИЕ ПОЛОЖЕНИЯ И ОСНОВНАЯ ИДЕЯ ПРЕДЛОЖЕННОГО МЕТОДА

Пусть автомат Мура задан прямой структурной таблицей (ПСТ) со столбцами [1]: a_m, $K(a_m)$, a_S, $K(a_S)$, X_h, Φ_h, h. Здесь a_m – исходное состояние МПА; $K(a_m)$ – код состояния $a_m \in A$ разрядности $R_A = \lceil \log_2 M \rceil$, для кодирования состояний используются внутренние переменные из множества $T = \{T_1,...,T_{R_A}\}$; a_S, $K(a_S)$ – соответственно состояние перехода и его код; X_h – входной сигнал, определяющий переход $< a_m, a_S >$, и равный конъюнкции некоторых элементов (или их отрицаний) множества логических условий $X = \{x_1,...,x_L\}$; Φ_h – набор функций возбуждения триггеров памяти МПА, принимающих единичное значение для переключения памяти из $K(a_m)$ в $K(a_S)$, $\Phi_h \subseteq \Phi = \{\varphi_1,...,\varphi_{R_A}\}$; $h = 1,...,H$ – номер перехода. В столбце a_m записывается набор микроопераций Y_q, формируемых в состоянии $a_m \in A$, где $Y_q \subseteq Y = \{y_1,...,y_N\}$, $q = 1,...,Q$. Эта таблица является основой для формирования систем функций:

$$\Phi = \Phi(T, X), \quad (1)$$

$$Y = Y(T), \quad (2)$$

задающих логическую схему МПА. Системы (1)–(2) являются основой для реализации схемы МПА Мура, структура которой показана на рис. 1. Условимся обозначать этот МПА символом U_1.

В МПА U_1 блок формирования функций возбуждения памяти (БФП) реализует систему (1). Блок формирования микроопераций (БФМ) реализует систему (2). Память состояний МПА реализуется на регистре (Рг), состоящем из *D*-триггеров [5]. По сигналу Start в Рг записывается нулевой код начального состояния $a_1 \in A$. По сигналу Clock содержимое Рг меняется в зависимости от функций (2).

При реализации схемы U_1 в базисе FPGA схема БФП реализуется на элементах табличного типа (LUT, look-up table). Для реализации БФП используются встроенные блоки памяти (EMB, embedded memory block) [7, 8]. При реализации схемы U_1 в базисе CPLD схема БФП реализуется на макроячейках программируемой матричной логики (PAL, programmable array logic). Для реализации схемы БФМ могут использоваться макроячейки PAL, либо внешние программируемые ПЗУ. Отметим, что существуют микросхемы CPLD, в которых имеются встроенные EMB. К таким CPLD относятся, например, микросхемы Delta 3K [9]. В данной статье мы рассматриваем случай реализации БФМ на программируемых ПЗУ, которые могут быть как встроенными, так и внешними.

Для оптимизации числа термов в системе (1) предлагается использовать наличие классов псевдоэквивалентных состояний автомата Мура [10]. Число блоков ППЗУ можно уменьшить, если входными переменными БФМ будут адресные разряды наборов микроопераций (НМО) [11]. Число входов схемы БФП можно уменьшить за счет замены логических условий $X_l \in X$ некоторыми переменными $P_g \in P$, где $|P| << |L|$[2]. Все эти идеи положены в основу предлагаемого метода. Обозначим предлагаемый МПА символом U_2.

МЕТОД СИНТЕЗА АВТОМАТА U_2

Одной из особенностей МПА Мура является наличие псевдоэквивалентных состояний [2], то есть состояний с одинаковыми переходами под воздействием одинаковых входных сигналов. Такие состояния соответствуют операторным вершинам [1] алгоритма управления, выходы которых связаны со входом одной и той же вершины алгоритма.

Пусть $\Pi_A = \{B_1,...,B_I\}$ – разбиение множества A на классы псевдоэквивалентных состояний. Закодируем классы $B_i \in \Pi_A$ двоичными кодами $K(B_i)$ разрядности:

$$R_B = \lceil \log_2 I \rceil. \quad (3)$$

Пусть исходная ГСА Г включает Q попарно различных наборов микроопераций (НМО) $Y_q \subseteq Y$. Закодируем набор Y_q двоичным кодом $K(Y_q)$ разрядности:

$$R_Y = \lceil \log_2 Q \rceil. \quad (4)$$

Пусть операторная вершина b_t ГСА Г соответствует состоянию $a_m \in B_i$ и пусть в ней записан набор микроопераций Y_q. Тогда код состояния $a_m \in A$ можно представить в виде конкатенации кодов:

$$K(a_m) = K(B_i) * K(Y_q), \quad (5)$$

где * – знак конкатенации.

Пусть $X(a_m) \subseteq X$ – множество логических условий, определяющих переходы из состояния $a_m \in A$. Пусть $L_m = |X(a_m)|$ и $G = \max(L_1,...,L_m)$. Тогда логические условия $x_l \in X$ можно заменить некоторыми переменными $P_g \in P$, где $|P| = G$ [12].

Пусть $X(B_i) \subseteq X$ – множество логических условий, определяющих переходы из состояний $a_m \in B_i$, где $B_i \in \Pi_A$. В силу определения псевдоэквивалентных состояний справедливо равенство $X(B_i) = X(a_m)$, где $a_m \in B_i$. Таким образом, логические условия $x_l \in X$ можно заменить переменными $P_g \in P$ для классов состояний.

Представление кодов состояний в виде (5) и замена логических условий позволяет получить МПА Мура U_2 (рис. 2), предлагаемый в данной работе. Как видно из рис. 2, автомат U_2 включает блок замены логических условий (БЛУ) и блоки БФП и БФМ. Рассмотрим особенности модели U_2.

Блок БЛУ осуществляет замену логических условий $x_l \in X$. Для этого формируется система функций:

$$P = P(X, \tau). \quad (6)$$

Переменные $\tau_R \in \tau$, где $|\tau| = R_B$, используются для кодирования классов $B_i \in \Pi_A$.

Блок БФП реализует систему функций:

$$\Phi = \Phi(P, \tau). \quad (7)$$

Число функций системы (7) определяется как $R_B + R_Y$. Отметим, что в общем случае $R_A < R_B + R_Y$. Однако блок

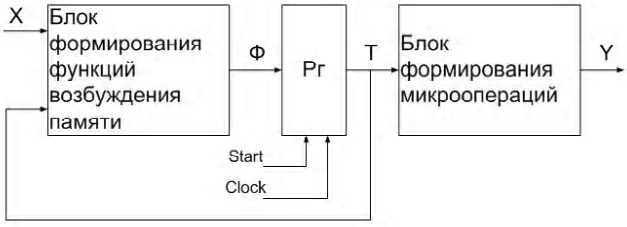


Рис. 1. Структурная схема МПА Мура U_1

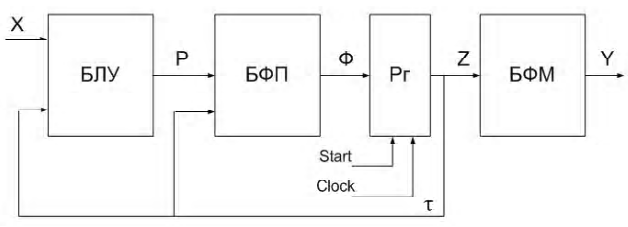


Рис. 2. Структурная схема МПА Мура U_2

БФП в автомате U_1 имеет $R_A + L$ входов, а в автомате U_2 – только $R_B + G$. Кроме того, переход к классам $B_i \in \Pi_A$ позволяет значительно уменьшить число термов в функциях (7) по сравнению с (1).

Блок БФМ реализует систему функций:

$$Y = Y(Z), \tag{8}$$

где переменные $z_R \in Z$ используются для кодирования наборов микроопераций. При этом $|Z| = R_Y \le R_A$. Если выполняется условие:

$$R_Y < R_A, \tag{9}$$

то число блоков ППЗУ в схеме БФМ автомата U_2 уменьшается в m раз, где:

$$m = 2^{R_A - R_Y}. \tag{10}$$

Сравнение автоматов U_1 и U_2 показывает, что автомат U_2 обладает меньшим быстродействием. Это связано с наличием блока БЛУ. Таким образом, предлагаемый метод применим, если он обеспечивает заданное быстродействие управляемой цифровой системы.

Предлагаемый метод синтеза МПА Мура U_2 по отмеченной ГСА Г включает следующие этапы:

1. Формирование разбиения Π_A и кодирование классов $B_i \in \Pi_A$.
2. Формирование таблицы переходов МПА по системе обобщенных формул перехода.
3. Кодирование НМО и определение расширенных кодов состояний $a_m \in A$.
4. Формирование таблицы блока замены логических условий.
5. Формирование прямой структурной таблицы автомата U_2.
6. Формирование таблицы блока БФМ.
7. Реализация схемы автомата в заданном элементном базисе.

Рассмотрим пример применения предложенного метода.

ПРИМЕР ПРИМЕНЕНИЯ ПРЕДЛОЖЕННОГО МЕТОДА

Пусть для некоторой ГСА Γ_1 получено разбиение Π_A, где $\Pi_A = \{B_1,...,B_7\}$. Пусть $A = \{a_1,...,a_{18}\}$ и $B_1 = \{a_1\}$, $B_2 = \{a_2,...,a_6\}$, $B_3 = \{a_7, a_8\}$, $B_4 = \{a_9,...,a_{12}\}$, $B_5 = \{a_{13}, a_{14}\}$, $B_6 = \{a_{15}\}$, $B_7 = \{a_{16}, a_{17}, a_{18}\}$. Таким образом, M=18, R_A=5, I=7, R_B=3. Пусть в ГСА Γ_1 имеется Q=12 попарно различных НМО, тогда R_Y=4. Закодируем наборы $Y_q \subseteq Y$ тривиальным образом: $K(Y_1)$=0000, $K(Y_2)$=0001, … , $K(Y_{12})$=1011. Закодируем классы $B_i \in \Pi_A$ следующим образом: $K(B_1)$=000, … , $K(B_7)$=110. Пусть фрагмент системы обобщенных формул перехода [2] имеет следующий вид:

$$B_2 \to x_3 a_7 \vee \overline{x_3} x_4 a_9 \vee \overline{x_3 x_4} a_{14},$$

$$B_3 \to x_5 a_9 \vee \overline{x_5} x_6 a_{16} \vee \overline{\overline{x_5} x_6} a_{14}. \tag{11}$$

Система типа (11) является основой для построения таблицы переходов МПА, имеющей столбцы B_i, a_m, X_h, h. Здесь X_h – конъюнкция логических условий, определяющая переход из состояний $a_S \in B_i$ в состояние $a_m \in A$, $h = \overline{1, H}$ – номер перехода. Для фрагмента (11) таблица переходов имеет 6 строк (табл. 1).

Связь табл. 1 и фрагмента (11) очевидна. Пусть в состоянии a_7 формируется НМО Y_3, в состоянии a_9 – Y_5, в состоянии a_{14} – Y_8, в состоянии a_{16} – Y_6. Это позволяет определить расширенные коды данных состояний: $K(a_7)$=0100010, $K(a_9)$=0110100, $K(a_{14})$=1000111, $K(a_{16})$=1100101. В этих кодах первые три разряда совпадают с кодом $K(B_i)$, где $a_m \in B_i$, а последние четыре разряда определяются кодом НМО.

Пусть для ГСА Γ_1 L=14, при этом $X(B_1)=\{x_1,x_2\}$, $X(B_2)=\{x_3, x_4\}$, $X(B_3)=\{x_5, x_6\}$, $X(B_4)=\{x_1, x_5, x_7\}$, $X(B_5)=\{x_8, x_9, x_{10}\}$, $X(B_6)=\{x_3, x_{11}, x_{12}\}$, $X(B_7)=\{x_{13}, x_{14}\}$. Как следует из анализа этих множеств G=3 и $P=\{p_1, p_2, p_3\}$. Таблица блоков БЛУ имеет 7 строк и 4 столбца (табл. 2).

Из табл. 2 следует система (6). Например, из анализа столбца P_1 можно получить функцию:

$$P_1 = x_1(B_1 \vee B_4) \vee x_3(B_2 \vee B_6) \vee x_8 B_5.$$

Используя коды классов $B_i \in \Pi_A$ можно получить окончательное выражение:

$$P_1 = \overline{x_1}\tau_1\tau_2\tau_3 \vee x_1\overline{\tau_1}\tau_2\tau_3 \vee x_3\overline{\tau_1\tau_2}\tau_3 \vee x_3\tau_1\overline{\tau_2}\tau_3 \vee x_8\tau_1\tau_2\tau_3.$$

Отметим, что подобные функции тривиально реализуются на мультиплексоре [12]. Как известно, мультиплексор является стандартным библиотечным элементом САПР [7–9].

Прямая структурная таблица МПА U_2 строится, как расширение таблицы переходов столбцами $K(B_i)$, $K(a_m)$, Φ_h и заменой столбца X_h столбцом P_h. В столбце Φ_h записываются функции $D_r \in \Phi$, принимающие единичные значения на h-м переходе МПА. Для нашего примера табл. 1 преобразовывается в табл. 3 тривиальным образом.

Таблица 1. Фрагмент таблицы переходов автомата Мура

B_i	a_m	X_h	h	B_i	a_m	X_h	h
B_2	a_7	x_3	1	B_3	a_9	x_5	4
	a_9	$\overline{x_3}x_4$	2		a_{16}	$\overline{x_5}x_6$	5
	a_{14}	$\overline{x_3 x_4}$	3		a_{14}	$\overline{x_5 x_6}$	6

Таблица 2. Таблица блока замены логических условий

B_i	P_1	P_2	P_3	i	B_i	P_1	P_2	P_3	i
B_1	x_1	x_2	–	1	B_5	x_8	x_9	x_{10}	5
B_2	x_3	x_4	–	2	B_6	x_3	x_{11}	x_{12}	6
B_3	–	x_5	x_6	3	B_7	–	x_{13}	x_{14}	7
B_4	x_1	x_5	x_7	4	–	–	–	–	

Таблица 3. Фрагмент прямой структурной таблицы МПА U_2

B_i	$K(B_i)$	a_m	$K(a_m)$	P_h	Φ_h	h
B_2	001	a_7	0100010	P_1	D_2D_6	1
		a_9	0110100	$\overline{P_1}P_2$	$D_2D_3D_5$	2
		a_{14}	1000111	$\overline{P_1}\,\overline{P_2}$	$D_1D_5D_6D_7$	3
B_3	010	a_9	0110100	P_2	$D_2D_3D_5$	4
		a_{16}	1100101	$\overline{P_2}P_3$	$D_1D_2D_5D_7$	5
		a_{14}	1000111	$\overline{P_2}\,\overline{P_3}$	$D_1D_5D_6D_7$	6

Эта таблица является основой для формирования системы (7). Так, с учетом минимизации, из табл. 3 можно получить функцию:

$$D_1 = \overline{\tau_1}\,\overline{\tau_2}\tau_3\,\overline{P_1}\,\overline{P_2} \vee \overline{\tau_1}\tau_2\,\overline{\tau_3}\,\overline{P_2}\,.$$

Таблица блока БФМ строится тривиальным образом, и этот этап в данной статье не рассматривается. Последний этап метода связан с применением промышленных САПР фирм-производителей ПЛИС [7–9]. Этот этап мы также здесь не рассматриваем.

ЗАКЛЮЧЕНИЕ

Предложенный метод позволяет гарантировано уменьшить число переходов МПА Мура до величины этого параметра эквивалентного автомата Мили. При этом соответственно уменьшается число термов в функциях возбуждения памяти МПА.

Использование метода замены логических условий позволяет уменьшить число входных переменных в функциях возбуждения памяти. Это особо важно для минимизации числа LUT-элементов при реализации схемы на ПЛИС типа FPGA.

Представление кода состояния в виде конкатенации кодов классов псевдоэквивалентных состояний и наборов микроопераций может привести к увеличению разрядности кода состояния. Однако при выполнении условия (9) такое представление позволяет уменьшить число блоков памяти в схеме формирования микроопераций. Кроме того, предлагаемое представление позволяет закодировать классы так, чтобы оптимизировать число макроячеек в блоке замены логических условий.

Проведенные авторами исследования показали, что применение предложенного метода позволяет до 20 % уменьшить общее число макроячеек в блоках БЛУ и БФП по сравнению с этим параметром для блока БВП автомата U_1. Кроме того, число блоков памяти в БФМ практически всегда уменьшалось в два раза. Отметим, что время цикла МПА U_2 в 1,5 раза больше, чем для эквивалентного автомата U_1.

Научная новизна предложенного подхода заключается в использовании расширенного представления кодов состояний и замены логических условий для уменьшения числа макроячеек ПЛИС и блоков ППЗУ в схеме автомата Мура.

Практическая значимость метода заключается в уменьшении стоимости схемы МПА Мура по сравнению с известными в литературе аналогами.

СПИСОК ЛИТЕРАТУРЫ

1. *Baranov, S.* Logic and System Design of Digital Systems / Baranov S. – Tallinn : TUT Press, 2008. – 328 pp.
2. *Barkalov, A.* Logic Synthesis for FSM-based Control Units / A. Barkalov, L. Titarenko. – Berlin : Springer, 2009. – 233 pp.
3. *Грушницкий, Р. И.* Проектирование систем с использованием микросхем программируемой логики / Р. И. Грушницкий, А. Х. Мурзаев, Е. П. Угрюмов. – С. Пб. : БХВ.-Петербург, 2002. – 608 с.
4. *Maxfield, C.* The Design Warrior's Guide to FPGAs / Maxfield C. – Amsterdam : Elsevier, 2004. – 541 pp.
5. *Соловьёв, В. В.* Логическое проектирование цифровых систем на основе программируемых логических интегральных схем / В. В. Соловьёв, А. С. Климович. – М. : Горячая линия-Телеком, 2008. – 376 с.
6. *DeMicheli, G.* Synthesis and Optimization of Digital circuits / DeMicheli G. – New York : McGraw Hill, 1994. – 541 pp.
7. FPGA, CPLD, and ASIC from Altera [Electronic resource]: база данных содержит сведения о микросхемах ПЛИС фирмы Altera. – Электрон. дан. – Режим доступа : http://www.altera.com. – Загл. с экрана.
8. FPGA and CPLID Solutions from Xilinx, Inc [Electronic resource]: база данных содержит сведения о микросхемах ПЛИС фирмы Xilinx. – Электрон. дан. – Режим доступа: www.xilinx.com. – Загл. с экрана.
9. Cypress Semiconductor [Electronic resource]: база данных содержит сведения о микросхемах ПЛИС фирмы Cypress. – Электрон. дан. – Режим доступа: www.cypress.com. – Загл. с экрана.
10. *Баркалов, А. А.* Принципы оптимизации логической схемы микропрограммного автомата Мура // Кибернетика и системный анализ. – 1998. – № 1. – С. 65–72.
11. *Баркалов, А. А.* Матричная реализация автомата Мура с расширением кодов состояний перехода/ А. А. Баркалов, Р. В. Мальчева, К. А. Солдатов // Научные труды Донецкого национального технического университета. Серия «Информатика, кибернетика и вычислительная техника» (ИКВТ-2010). Выпуск 11 (164). – Донецк : ГВУЗ «ДонНТУ». – 2010. – С. 79–83.
12. *Baranov, S.* Logic Synthesis for Control Automata / Baranov S. – New York : Kluwer Academic Publishers, 1994. – 312 pp.

Стаття надійшла до редакції 23.01.2012.

Баркалов О. О., Мальчева Р. В., Солдатов К. А.

ОПТИМІЗАЦІЯ СХЕМИ АВТОМАТА МУРА, ЩО РЕАЛІЗУЄТЬСЯ В БАЗИСІ ПЛІС

У статті пропонується метод, призначений для зменшення числа вхідних змінних і проміжних термів в реалізованих системах булевих функцій. Запропонований метод заснований на розширенні кодів станів переходу і заміні логічних умов. Застосування запропонованого методу дозволяє до 20 % зменшити загальну кількість макрокомірок в блоках БЛУ та БФП.

Ключові слова: автомат Мура, ПЛІС, ГСА, псевдоеквівалентні стани, заміна логічних умов.

Barkalov A. A., Malcheva R. V., Soldatov K. A.

OPTIMIZATION OF MOORE FINITE STATE MACHINE IMPLEMENTED ON THE PROGRAMMABLE LOGIC

This article is proposed a method which is designed to reduce the number of input variables and intermediate terms of Boolean functions. The method is based on the extended codes of states and replacement of logic conditions. Application of the proposed method allows up to 20% reduction in the total count of macrocells in blocks BLC and BFM.

Key words: Moore FSM, Programmable Logic, GSA, pseudoequivalent states, replacement of logic conditions.

УДК 004:519.2

Кириченко Л. О.[1], Демерчян К. А.[2], Кайали Э.[3], Хабачёва А. Ю.[4]

[1]Канд. техн. наук, доцент Харьковского национального университета радиоэлектроники
[2, 4]Студент Харьковского национального университета радиоэлектроники
[3]Аспирант Харьковского национального университета радиоэлектроники

МОДЕЛИРОВАНИЕ ТЕЛЕКОММУНИКАЦИОННОГО ТРАФИКА С ИСПОЛЬЗОВАНИЕМ СТОХАСТИЧЕСКИХ МУЛЬТИФРАКТАЛЬНЫХ КАСКАДНЫХ ПРОЦЕССОВ

В работе рассматривается моделирование реализаций телекоммуникационного трафика, обладающего мультифрактальными свойствами, на основе математической модели мультипликативного стохастического каскада, весовые коэффициенты которого имеют бета-распределение вероятностей.

Ключевые слова: стохастический каскадный процесс, модель телекоммуникационного трафика, самоподобный процесс, мультифрактальный процесс.

ВВЕДЕНИЕ И ПОСТАНОВКА ЗАДАЧИ

Экспериментальные и численные исследования, проведенные в последние десятилетия, свидетельствуют, что трафик во многих мультимедийных сетях обладает фрактальными свойствами. Такой трафик имеет особую структуру, сохраняющуюся на многих масштабах, – в реализации всегда присутствует некоторое количество очень больших выбросов при относительно небольшом среднем уровне трафика. Эти выбросы вызывают значительные задержки и потери пакетов, даже когда суммарная потребность всех потоков далека от максимально допустимых значений. Причина такого эффекта заключается в особенностях распределения файлов по серверам, их размерах, в типичном поведении пользователей, и в значительной степени связана с изменениями сетевых ресурсов и топологии сети [1–4].

Самоподобные свойства трафика привели к появлению ряда моделей трафика на основе самоподобных (монофрактальных) стохастических процессов [2, 3]. В последнее десятилетие интенсивно изучаются мультифрактальные свойства трафика. Мультифрактальный трафик определяется как расширение самоподобного трафика за счет учета скейлинговых свойств статистических характеристик второго и выше порядков. Использование мультифрактальных стохастических процессов для моделирования телекоммуникационного трафика достаточно ново, и список мультифрактальных моделей значительно короче [4–6].

Целью представленной работы является разработка модели реализаций телекоммуникационного трафика, обладающего мультифрактальными свойствами, на основе математической модели мультипликативного стохастического каскада.

ХАРАКТЕРИСТИКИ САМОПОДОБНЫХ И МУЛЬТИФРАКТАЛЬНЫХ СЛУЧАЙНЫХ ПРОЦЕССОВ [6–9]

Самоподобие случайных процессов заключается в сохранении статистических характеристик при изменении масштаба времени. Стохастический процесс $X(t)$ является самоподобным с параметром H, если процесс $a^{-H}X(at)$ описывается теми же конечномерными законами распределений (Law), что и :

$$\mathrm{Law}\{a^{-H}X(at)\} = \mathrm{Law}\{X(t)\}, \quad \forall a > 0,\ t > 0. \quad (1)$$

Параметр H, $0 < H < 1$, называемый показателем Херста, представляет собой степень самоподобия. Наряду с этим свойством, показатель H характеризует меру долгосрочной зависимости стохастического процесса. В случае $0,5 < H < 1$ процесс обладает длительной памятью: если в течение некоторого времени в прошлом наблюдались положительные приращения процесса, т. е. происходило увеличение, то и впредь в среднем будет происходить увеличение. В случае $0 < H < 0,5$ высокие значения процесса следуют за низкими, и наоборот. При $H = 0,5$ отклонения процесса от среднего являются действительно случайными и не зависят от предыдущих значений.

Можно показать, положив в (1) $a = 1/t$, что для самоподобного процесса выполняется следующее равенство:

$$\mathrm{Law}\{X(t)\} = \mathrm{Law}\left\{\left(\frac{1}{t}\right)^{H} X(1)\right\} = \mathrm{Law}\{t^{H} X(1)\}. \quad (2)$$

Учитывая (2), начальные моменты самоподобного случайного процесса можно выразить как:

$$\mathrm{M}\left[|X(t)|^{q}\right] = \mathrm{M}\left[\left|t^{H}X(1)\right|^{q}\right] = t^{qH}\,\mathrm{M}\left[|X(1)|^{q}\right] = C(q)\cdot t^{qH}, \quad (3)$$

где величина $C(q) = \mathrm{M}\left[|X(1)|^{q}\right]$.

Для мультифрактальных процессов рассматривается более общее соотношение:

$$\mathrm{Law}\{X(at)\} = \mathrm{Law}\{M(a)\cdot X(t)\},$$

где $M(a)$ – независимая от $X(t)$ случайная функция. В случае самоподобного процесса $M(a) = a^{H}$. Мульти-

фрактальные процессы проявляют более гибкие скейлинговые закономерности для моментных характеристик:

$$\mathrm{M}\left[\left|X(t)\right|^{q}\right] = c(q)\cdot t^{\tau(q)+1}, \qquad (4)$$

где $c(q)$ – некоторая детерминированная функция, $\tau(q)$ – скейлинговая экспонента, в общем случае нелинейная функция, для которой значение $\frac{\tau+1}{q}$ при $q = 2$ совпадает со значением степени самоподобия H. Для временных рядов, которые отвечают монофрактальному процессу, скейлинговая экспонента $\tau(q)$ линейна.

МЕТОД МАКСИМУМОВ МОДУЛЕЙ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ [7, 10–12]

Одним из самых популярных инструментов мультифрактального анализа является метод максимумов модулей непрерывного вейвлет-преобразования (ММВП). Он базируется на вейвлет-анализе, который называют «математическим микроскопом» из-за способности сохранять хорошее разрешение на разных масштабах. Поскольку вейвлет-функции являются локализованными по времени и частоте, метод ММВП является мощным инструментом статистического описания нестационарных процессов.

Непрерывное вейвлет-преобразование фукции $X(t)$ имеет вид $W(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} X(t)\psi_{ab}(t)dt$, где $\psi_{ab}(t)$ – вейвлет-функция с параметрами масштаба a и сдвига b. Функция $W(a,b)$ называется вейвлет-спектром и может быть представлена как поверхность вейвлет-коэффициентов в трехмерном пространстве. Наиболее важная информация содержится в линиях локальных экстремумов поверхности $W(a,x)$, поиск которых проводится на каждом масштабе a.

Метод ММВП позволяет численно получить статическую сумму:

$$Z(q,a) = \sum_{l\in L(a)}\left(\sup_{a'\le a}\left|W(a', x_l(a'))\right|\right)^{q},$$

где $L(a)$ – множество всех линий l максимумов модулей вейвлет-коэффициентов на масштабе a; $x_l(a)$ – расположение максимума на этом масштабе. Для вычисления $Z(q,a)$ выбирается максимальное значение модуля вейвлет-коэффициентов вдоль каждой линии на масштабах, меньших заданного значения масштаба a.

В этом случае выполняется зависимость:

$$Z(q,a) \approx a^{\tau(q)},$$

де $\tau(q)$ – скейлинговая экспонента из формулы (4), которую определяют для каждого значения q путем вычисления наклона $\ln Z(q,a)$ от $\ln a$.

СТОХАСТИЧЕСКИЕ МУЛЬТИФРАКТАЛЬНЫЕ КАСКАДНЫЕ ПРОЦЕССЫ

Простейшей моделью мультифрактального процесса с заданными свойствами является детерминированный биномиальный мультипликативный каскад [6, 8, 9]. При его построении первоначальный единичный отрезок делится на два равных интервала, которым приписываются весовые коэффициенты p_1 и $p_2 = 1 - p_1$ соответственно. Затем с каждым из интервалов проделывается аналогичная процедура. В результате на втором шаге имеется 4 интервала с весовыми коэффициентами p_1^2, p_1p_2, p_2p_1 и p_2^2. При числе шагов $n \to \infty$ и $p_1 \neq p_2$ мы приходим к предельной мере, являющейся неоднородным фрактальным множеством. На рис. 1, *а* показаны временные ряды значений биномиального каскада при значениях $p_1 = 0,6$ (вверху) и $p_1 = 0,8$ (внизу). Число итераций $n = 10$, т.е. длина реализации равна 2^{10} значений. Очевидно, что с увеличением первоначального весового коэффициента p_1 увеличивается неоднородность временного ряда.

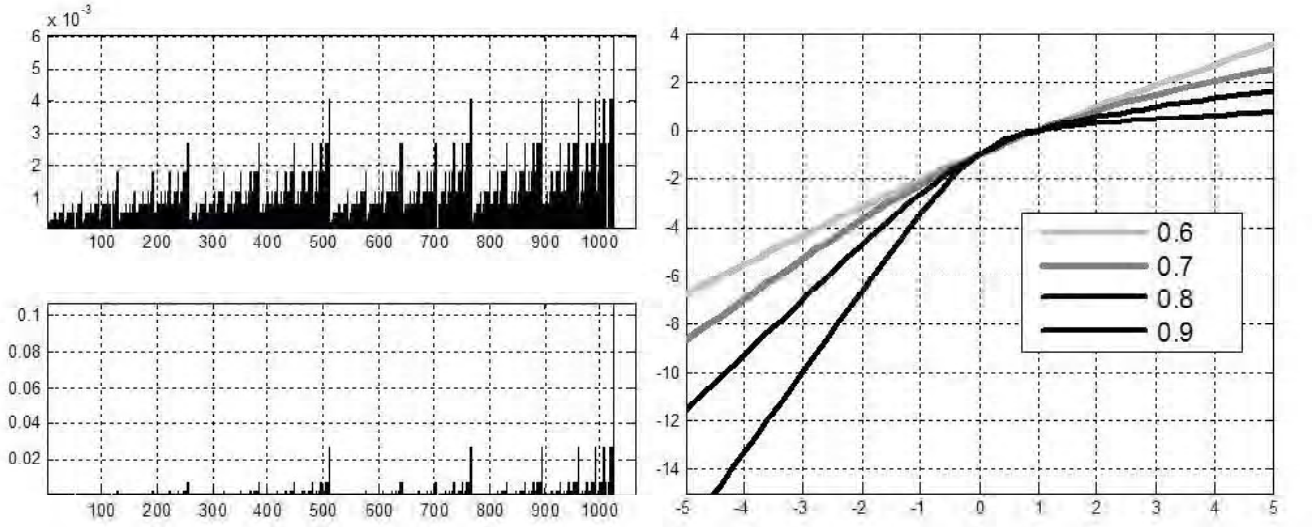


Рис. 1. Реализации каскада (*а*) и скейлинговые экспоненты $\tau(q)$ для разных p_1 (*б*)

В детерминированном случае мультифрактальная характеристика $\tau(q)$ для биномиального процесса зависит только от весового коэффициента p_1: $\tau(q) = \frac{-\ln(p_1^q + p_2^q)}{\ln 2}$. На рис. 1, *б* представлены теоретические скейлинговые экспоненты $\tau(q)$ для значений $p_1 = \{0,6; 0,7; 0,8; 0,9\}$.

Реализации детерминированного каскада полностью определяются величиной p_1, что неприемлемо для моделирования случайных процессов. При построении стохастических каскадов весовыми коэффициентами являются независимые значения некоторой заданной случайной величины W [4, 6, 9]. Случайная величина выбирается таким образом, чтобы математическое ожидание суммы весовых коэффициентов на каждой итерации равнялось единице. Если выбрать случайную величину, определенную на интервале [0,1], то сумма коэффициентов на каждой итерации будет равной единице.

В этом случае первым двум интервалам будут приписаны весовые коэффициенты w_1 и $1 - w_1$ соответственно. На втором шаге добавляются два новых независимых случайных значения w_2 и w_3. Получится 4 интервала с весовыми коэффициентами $w_1 w_2$, $w_1(1 - w_2)$, $(1 - w_1)w_3$ и $(1 - w_1)(1 - w_3)$. При $n \to \infty$ мы приходим к предельной мере, являющейся неоднородным фрактальным множеством.

В работе предложено в качестве случайной величины, порождающей весовые коэффициенты, использовать случайную величину, имеющую бета-распределение. Бета- распределением с параметрами $a > 0$, $b > 0$, называется распределение с плотностью вероятностей:

$$p(x) = \begin{cases} \frac{1}{B(a,b)}(1-x)^{b-1}, & x \in (0,1); \\ 0, & x \notin (0,1), \end{cases}$$

где $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$ – бета-функция. Для бета-распределения с одинаковыми значениями параметров $a = b$, у которого функция плотности распределения симметрична, можно аналитически определить скейлинговую экспоненту $\tau(q)$ [6, 9]:

$$\tau(q) = -\log_2 \frac{\mathrm{Beta}(\alpha+q,\alpha)}{\mathrm{Beta}(\alpha,\alpha)} - 1. \qquad (5)$$

На рис. 2, *а* приведены различные виды графиков плотности распределения вероятностей, для симметричного бета-распределения при значениях $a = \{0,5; 1; 1,5; 3\}$. При значениях параметров $a = b = 1$ мы получаем случайную величину, имеющую равномерное распределение на интервале [0,1]. На рис. 2, *б* представлены графики скейлинговых экспонент $\tau(q)$ для соответствующих значений параметра a симметричного бета-распределения.

Очевидно, что с увеличением значения параметра a происходит ослабление мультифрактальных свойств временного ряда. На рис. 3 показаны соответствующие реализации биномиальных каскадов.

В случае симметричного бета-распределения мультифрактальные свойства каскада полностью определяются параметром a. Показатель Херста H, учитывая формулу (5), в этом случае равен:

$$H = \frac{\tau(2)+1}{2} = -\log_2 \frac{\mathrm{Beta}(\alpha+q,\alpha)}{2\,\mathrm{Beta}(\alpha,\alpha)}.$$

В работе проведены исследования мультифрактальных свойств каскадов, порождаемых бета-распределениями с разными значениями параметров a и b. Получены численные зависимости, которые значениям параметра H ставят в соответствие различные функции скейлинго-

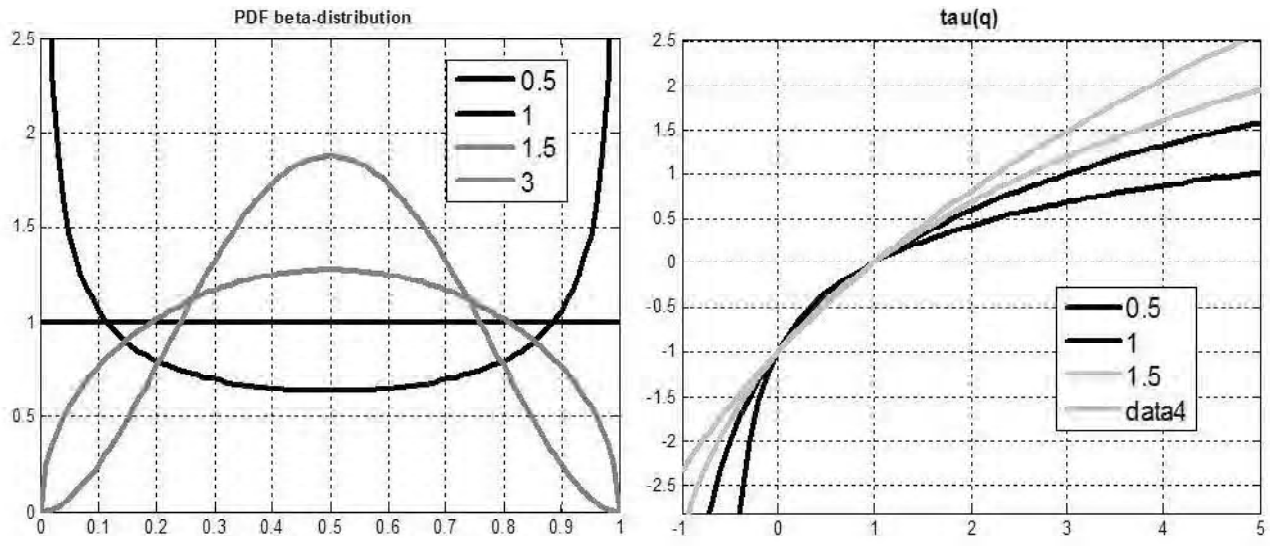
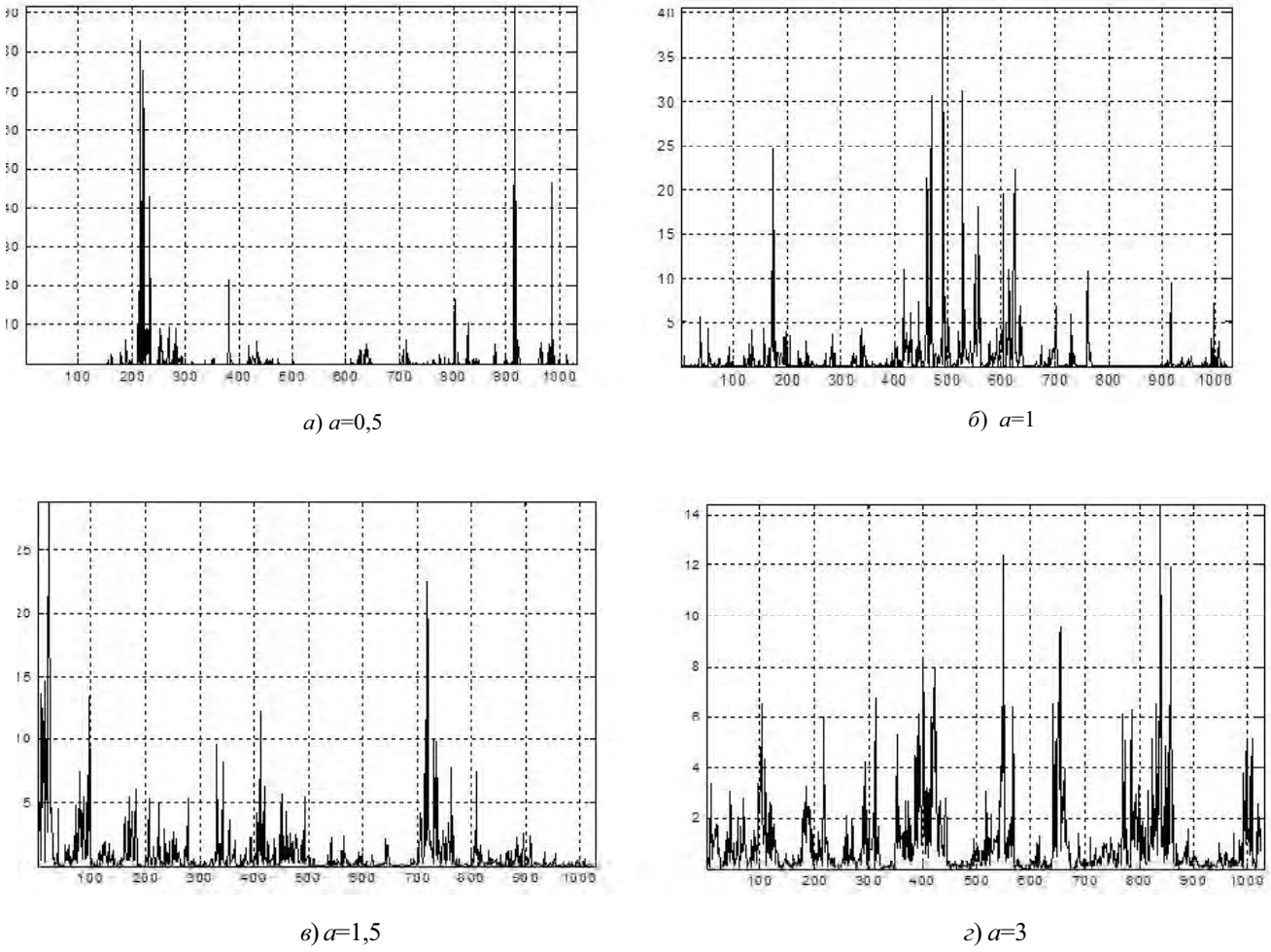


Рис. 2. Плотности распределения (а) и скейлинговые экспоненты $\tau(q)$ для разных значений параметра a симметричного бета-распределения (*б*)



а) a=0,5 *б*) a=1

в) a=1,5 *г*) a=3

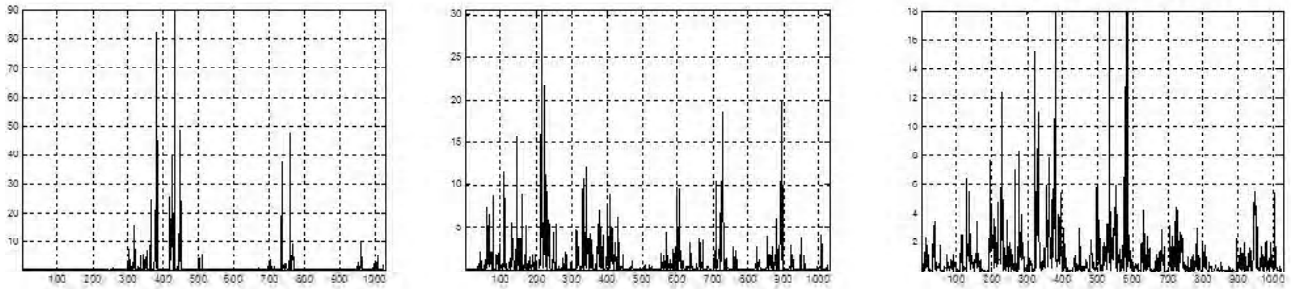
Рис. 3. Реализации биномиального каскада для разных значений a

вых экспонент $\tau(q)$. В этом случае можно выбрать каскад не только с определенной скейлинговой экспонентой, но и с заданным показателем Херста, который определяет степень долгосрочной зависимости временного ряда. На рис. 4 приведены реализации каскадных процессов с показателем $H = 0{,}8$ (вверху) и различными мультифрактальными свойствами $\tau(q)$ (посредине), которые определяются плотностью бета-распределения (внизу) разных значений a и b.

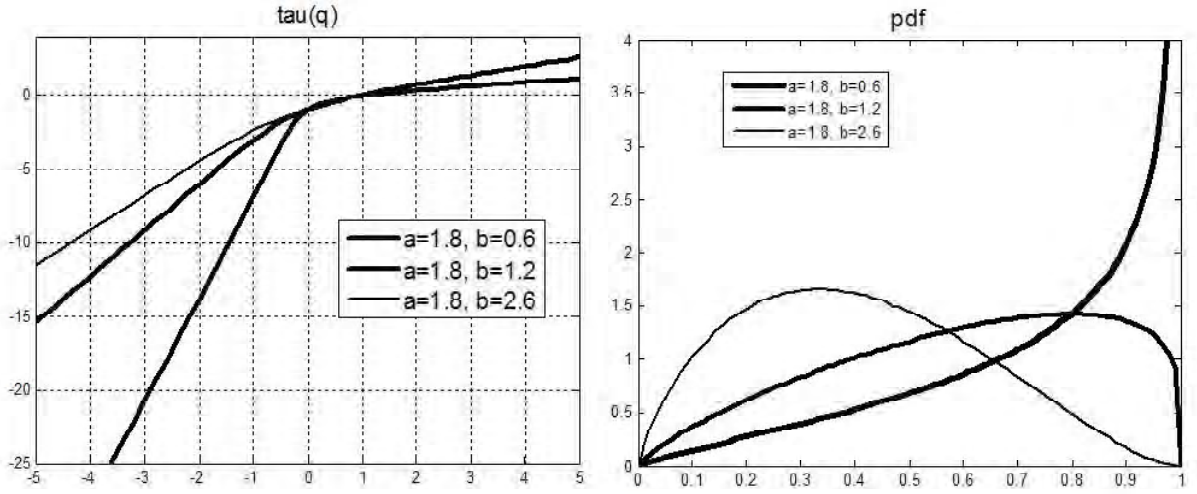
ПОСТРОЕНИЕ МОДЕЛЬНЫХ РЕАЛИЗАЦИЙ TCP-ТРАФИКА

Предложенная в работе модель мультифрактального трафика имеет три основных параметра $\left(I, H, \tau(q)\right)$, где I – интенсивность (среднее значение) трафика, H – показатель Херста, который определяет степень долгосрочной зависимости (степенное убывание корреляционной функции), $\tau(q)$ – скейлинговая экспонента, определяющая неоднородность (выбросы) реализации. Для построения модельной реализации необходимо оценить соответствующие параметры телекоммуникационного трафика и выбрать подходящий закон бета-распределения, генерирующий весовые коэффициенты мультифрактального каскада.

В работе были проведены исследования реализаций трафиков различных протоколов, которые показали их явные мультифрактальные свойства. На рис. 5, *а* приведен график выборочной реализации трафика TCP-протокола. Рассчитанная с помощью метода ММВП скейлинговая экспонента $\tau(q)$ представлена на рис. 5, *б*. Для данной реализации оценка показателя Херста $H = 0{,}83$. Каскады с такими мультифрактальными свойствами могут быть получены на основе бета-распределения с параметрами $a = 2{,}3$, $b = 2{,}5$, плотность которого показана на рис. 5, *г*. Одна из модельных реализаций каскада с данными мультифрактальными свойствами приведе на рис. 5, *в*.

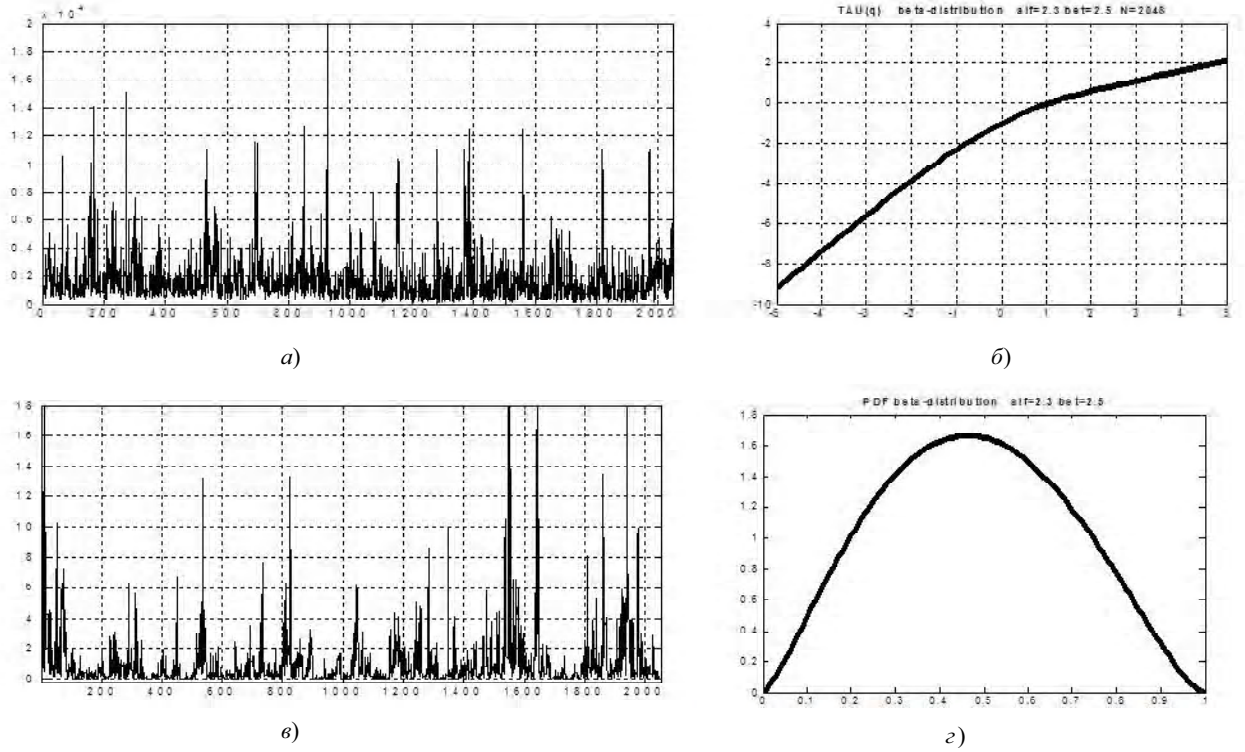


а) *a*=1,8, *b*=0,6 *б*) *a*=1,8, *b*=1,2 *в*) *a*=1,8, *b*=2,6



г) *д*)

Рис. 4. Реализации каскадов для разных значений *a* и *b* (*а*–*в*), соответствующие скейлинговые экспоненты (*г*) и плотности бета-распределений (*д*)



а) *б*)

в) *г*)

Рис. 5. Реализация трафика (*а*), выборочная скейлинговая экспонента τ(*q*) (*б*), плотность бета-распределения *a*=2,3, *b*=2,5 (*г*), модельная реализация (*в*)

ВЫВОДЫ

В работе были исследованы свойства стохастических мультипликативных каскадных процессов с функциями бета-распределения случайных весов. Предложена математическая модель трафика, параметрами которой являются средняя интенсивность, показатель Херста и скейлинговая экспонента. Показано, что модели трафика, полученные с помощью стохастических мультипликативных каскадов, позволяют гибко представлять мультифрактальные свойства реального телекоммуникационного трафика.

СПИСОК ЛИТЕРАТУРЫ

1. *Leland, W. E.* On the self-similarnature of ethernet traffic / [W. E. Leland, M. S. Taqqu, W. Willinger, D. V. Wilson] // IEEE/ ACM Transactions of Networking. – 1994. – № 2(1). – P. 1–15.
2. *Sheluhin, O. I.* Similar processes in telecommunications / O. I. Sheluhin, S. M. Smolskiy, A. V. Osin. – John Wiley & Sons Ltd, England, 2007. – 337 p.
3. *Столлингс, В.* Современные компьютерные сети 2-е изд. / В. Столлингс – С. Пб. : Питер, 2003. – 784 с.
4. *Шелухин, О. И.* Мультифракталы. Инфокоммуникационные приложения приложения / О. И. Шелухин . – М. : Горячая Линия -Телеком, 2011. – 578 с.
5. *Veitch, D.* Multifractality in TCP/IP traffic: the case against / D. Veitch, N. Hohn, P. Abry // Computer Networks-2005. – № 48(3). – P. 293–313.
6. *Riedi R. H.* Multifractal processes / Riedi R. H., Doukhan P., Oppenheim G., Taqqu M. S. (Eds.) // Long Range Dependence: Theory and Applications: Birkhuser. – 2002. – P. 625–715.
7. *Kantelhardt, J. W.* Fractal and Multifractal Time Series. – 2008 [Электронный ресурс]. – Режим доступа: http:// arxiv.org/abs/0804.0747. – Загл. с экрана.
8. *Федер, Е.* Фракталы / Е. Федер. – М. : Мир, 1991. – 254 с.
9. *Calvet, L.* Large Deviationsand the Distribution of Price Changes / L. Calvet, A. Fisher, B.B. Mandelbrot // Cowles Foundation Discussion Paper. – 1997. –N. 1165. –P. 1–30.
10. *Малла, С.* Вэйвлеты в обработке сигналов / С. Малла. – М. : Мир, 2005. – 671 с.
11. *Muzy, J. F.* Multifractal formalism for fractal signals: the structure-function approach versus the wavelet-transform modulus-maxima method / Muzy J. F., Bacry E., Arneodo A. // Phys. Rev. E. – 1993. –V. 47. – P. 875–884.
12. *Павлов, А. Н.* Мультифрактальный анализ сигналов / А. Н. Павлов, В. С. Анищенко // Известия Саратовского университета. Серия «Физика». – 2007. – Т. 7, Вып. 1. – С. 3–25.

Стаття надійшла до редакції 16.01.2012
Після доробки 14.02.2012.

Кіріченко Л. О., Демерчян К. А., Кайалі Е., Хабачова А. Ю.
МОДЕЛЮВАННЯ ТЕЛЕКОМУНІКАЦІЙНОГО ТРАФІКУ З ВИКОРИСТАННЯМ СТОХАСТИЧНИХ МУЛЬТИФРАКТАЛЬНИХ КАСКАДНИХ ПРОЦЕСІВ

В роботі розглядається моделювання реалізацій телекомунікаційного трафіку, що володіє мультифрактальними властивостями, на основі математичної моделі мультиплікативного стохастичного каскаду, вагові коефіцієнти якого мають бета-розподіл ймовірностей.

Ключові слова: стохастичний каскадний процес, модель телекомунікаційного трафіку, самоподібний процес, мультифрактальний процес.

Kirichenko L. O., Demerchan K. A., Kayali E., Habachyova A. Yu.
MODELING TELECOMMUNICATIONS TRAFFIC USING STOCHASTIC MULTIFRACTAL CASCADE PROCESS

In the work the simulation of telecommunications traffic has been examined, which has multifractal properties, based on a mathematical model of the stochastic multiplicative cascade, the weights of which are beta probability distribution.

Key words: a stochastic cascade process, the model of telecommunications traffic, self-similar process, multifractal process.

УДК 519.24:62-50

Кошевой Н. Д.[1], Сухобрус Е. А.[2]

[1]Д-р. техн. наук, профессор, заведующий кафедрой Национального аэрокосмического университета им. Н. Е. Жуковского «ХАИ»
[2] Аспирант Национального аэрокосмического университета им. Н. Е. Жуковского «ХАИ»

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ ОПТИМИЗАЦИИ МНОГОУРОВНЕВЫХ ПЛАНОВ МНОГОФАКТОРНОГО ЭКСПЕРИМЕНТА

Предложен метод поиска оптимального или близкого к оптимальному по стоимости реализации многоуровневого плана многофакторного эксперимента. Для автоматизации процесса поиска с использованием предложенного метода разработано программное обеспечение. Проведен сравнительный анализ разработанного программного обеспечения с программой поиска оптимальных многоуровневых комбинаторных планов многофакторного эксперимента, реализующей метод генерации перестановок с минимальным числом транспозиций соседних элементов.

Ключевые слова: программное обеспечение, симплекс-метод, быстродействие.

ПОСТАНОВКА ЗАДАЧИ

Изменение порядка проведения опытов существенно влияет на стоимость реализации эксперимента. При увеличении количества рассматриваемых вариантов усложняется поиск плана с наименьшей стоимостью. Трудность поиска вызвана быстрым ростом вариантов пе-

рестановок в зависимости от количества факторов и уровней. В связи с этим актуальной является проблема проведения анализа многоуровневых планов многофакторного эксперимента в оптимальные временные промежутки.

АНАЛИЗ ПОСЛЕДНИХ ИССЛЕДОВАНИЙ И ПУБЛИКАЦИЙ

Известна программа поиска оптимальных многоуровневых комбинаторных планов многофакторного эксперимента [1]. В основе работы программы лежит генерация комбинаторных планов многофакторного эксперимента, оценка их характеристик, отбор оптимального по стоимости варианта. Программа реализована на языке «Turbo Pascal». Предусмотрено два режима поиска оптимального решения: случайный поиск и последовательная генерация возможных вариантов перестановок. Количество рассматриваемых вариантов может быть задано, а процесс поиска может быть прерван.

Схема алгоритма программы поиска оптимальных многоуровневых комбинаторных планов многофакторного эксперимента представлена на рис. 1. Алгоритм работы программного обеспечения следующий.

1. Осуществляется ввод имени файла исходных данных, содержащего информацию о количестве анализируемых факторов, матрицу исходного плана эксперимента и стоимости изменений значений факторов. Производится чтение исходных данных.

2. Выбирается режим поиска: случайный поиск или анализ перестановок.

3. Производится выбор количества анализируемых вариантов путем введения заданного количества *N* или введения «0». В последнем случае процесс поиска может быть прерван при нажатии клавиши «ESC».

4. Рассчитывается стоимость исходной матрицы планирования эксперимента.

5. Осуществляется генерация перестановок строк исходной матрицы. В режиме случайного поиска генерация производится с использованием функции Randomize. В режиме анализа перестановок используется алгоритм генерации перестановок с минимальным числом транспозиций соседних элементов.

6. Производится расчет стоимости полученной матрицы и последующее сравнение её со стоимостью исходной матрицы. Если полученное значение стоимости меньше, чем стоимость исходной матрицы, то оно признается оптимальным. В противном случае полученное значение стоимости признается максимальным.

7. Производится сравнение количества проанализированных вариантов с заданным количеством N. Если заданное количество анализируемых вариантов достигнуто или произошло прерывание процесса поиска при нажатии клавиши «ESC», то осуществляется создание файла результата и переход к окончанию процесса поиска. В противном случае происходит переход к этапу 5 и повторяется аналогичная процедура.

Основным недостатком данной программы является необходимость анализа большого количества вариантов перестановок и, следовательно, снижение быстродействия процесса поиска при увеличении количества уровней факторов.

ЦЕЛЬ СТАТЬИ

Разработка метода поиска, оптимизированного по стоимости реализации, многоуровневого плана многофакторного эксперимента, позволяющего проводить поиск без полного перебора всех вариантов перестановок, и программы для реализации метода, обеспечивающего сокращение времени вычислений на ЭВМ.

ОСНОВНАЯ ЧАСТЬ

Предложен метод поиска на основе симплекс-метода, позволяющий получать оптимальные и близкие к оптимальным по стоимости реализации многоуровневые планы многофакторного эксперимента. Суть метода заключается в том, что многофакторный план представляется в виде выпуклого многогранника в многомерном пространстве, вершины которого соответствуют значениям уровней факторов плана. Поиск оптимального по стоимости реализации многоуровневого плана многофакторного эксперимента с использованием предложенного метода осуществляется в следующем порядке.

1. В качестве первой строки плана выбирается та, переход на которую максимален по стоимости.

2. Осуществляется поиск минимального по стоимости перемещения уровня фактора. В случае наличия нескольких равноценных по стоимости перемещения уровней факторов, выбирается первый из них по порядковому номеру.

3. Выполняется проверка: приводит ли перемещение уровня фактора, определенного на шаге 2, к появлению повторяющейся строки. Если полученная строка не встречалась ранее, то она записывается в выходную матрицу планирования. Если полученная строка уже встречалась, то перемещается следующий по возрастанию стоимости перемещения уровень фактора или, в случае наличия равноценных по стоимости перемещения уровней факторов, следующий по порядковому номеру.

4. Производится циклический повтор шагов 2 и 3 до тех пор, пока не будет получено требуемое количество строк матрицы планирования. В случае, если все строки, в которые возможно осуществить переход, уже повторялись, а требуемое количество строк матрицы планирования не достигнуто, осуществляется возврат к строке, содержащей равноценные по стоимости перемещения уровни факторов, и перемещается следующий по порядковому номеру уровень фактора.

Для автоматизации процесса поиска с использованием предложенного метода разработано программное обеспечение, позволяющее получать оптимальные или близкие к оптимальным по стоимости реализации многоуровневые планы многофакторного эксперимента без необходимости перебора всех вариантов перестановок, что позволяет значительно сократить время поиска. Схема алгоритма поиска оптимального плана эксперимента на основе симплекс-метода представлена на рис. 2.

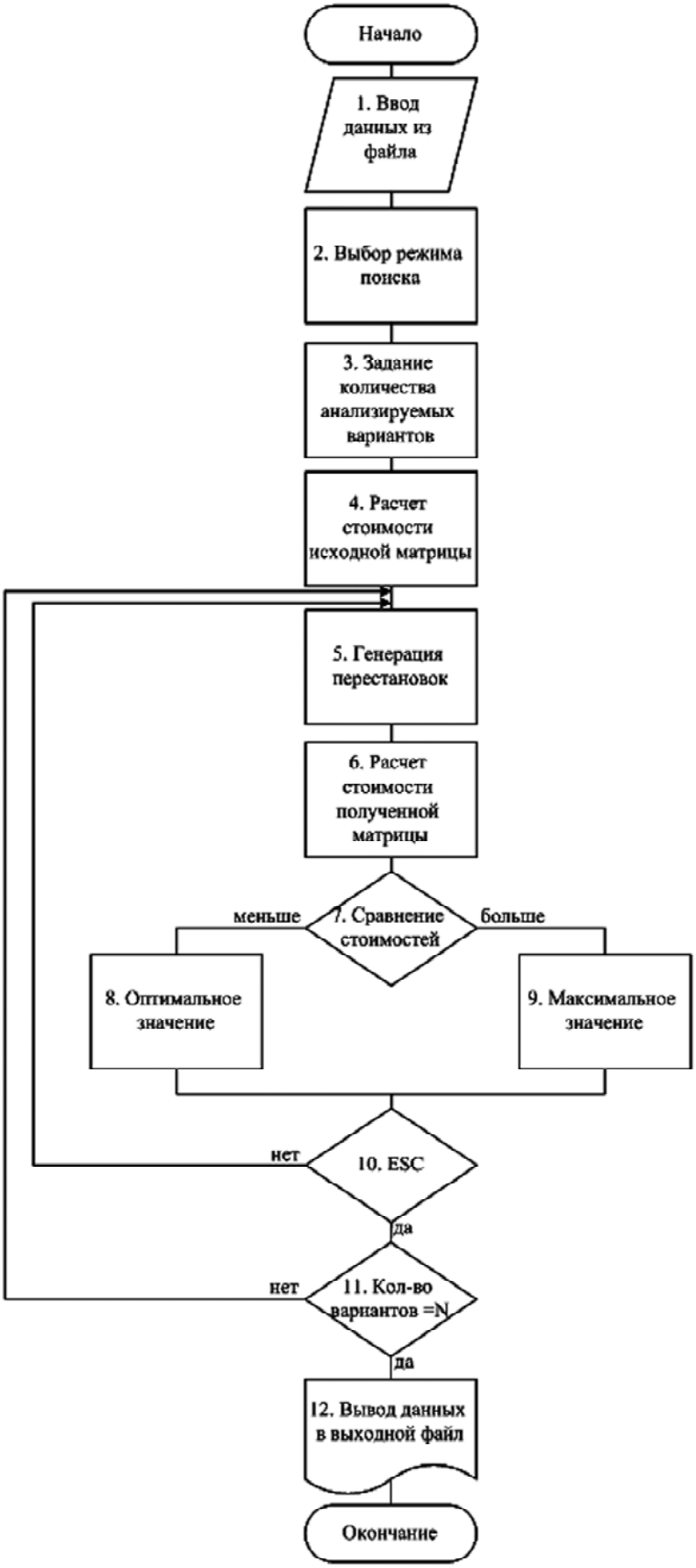


Рис. 1. Схема алгоритма работы известного программного обеспечения