НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
НАВЧАЛЬНО-НАУКОВИЙ КОМПЛЕКС
«ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ» НТУУ «КПІ»
МОН ТА НАН УКРАЇНИ

СИСТЕМНІ ДОСЛІДЖЕННЯ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

МІЖНАРОДНИЙ НАУКОВО–ТЕХНІЧНИЙ ЖУРНАЛ

№ 4 2011

ЗАСНОВАНИЙ У ЛИПНІ 2001 р.

РЕДАКЦІЙНА КОЛЕГІЯ:

Головний редактор

М.З. ЗГУРОВСЬКИЙ, акад. НАН України

Заступник головного редактора

Н.Д. ПАНКРАТОВА, проф., д. т. н.

Члени редколегії:

П.І. АНДОН, чл.-кор. НАН України
А.В. АНІСІМОВ, проф., д. ф.-м. н.
В.Г. БОНДАРЕНКО проф., д. ф.-м. н.
Ж. БОННІН, проф.., Франція
Г.-В. ВЕБЕР, проф.., Турція
Р. ВОРОНКА, проф.., США
О.Д. ГВІШІАНІ, проф., д. ф.-м. н., Росія
В.М. КУНЦЕВИЧ, акад. НАН України
О.С МАКАРЕНКО проф., д. ф.-м. н.
О.А. ПАВЛОВ, проф., д.т.н.
Т.Л. СААТІ, проф.., д. ф.-м. н.
Л. САКАЛАУСКАС, проф., Литва
І.В. СЕРГІЄНКО, акад. НАН України
Е.О. ФАЙНБЕРГ проф., США
Я.С. ЯЦКІВ акад. НАН України

АДРЕСА РЕДАКЦІЇ:

03056, м. Київ,
пр. Перемоги, 37, корп. 35,
Навчально-науковий комплекс
«Інститут прикладного системного аналізу»
НТУУ «КПІ» МОН та НАН України
Тел.: **406-81-44**; Факс: **406-81-44**
E-mail: journal@mmsa.ntu-kpi.kiev.ua
http://journal.iasa.com.ua

У номері:

• Автоматизовані системи управління

• Прогресивні інформаційні технології, високопродуктивні комп'ютерні системи

• Проблеми прийняття рішень і управління в економічних, технічних, екологічних і соціальних системах

• Математичні методи, моделі, проблеми і технології дослідження складних систем

© *Навчально-науковий комплекс «Інститут прикладного системного аналізу» НТУУ «КПІ» МОН та НАН України*

NATIONAL ACADEMY OF SCIENCES OF UKRAINE
EDUCATIONAL-SCIENTIFIC COMPLEX
«INSTITUTE FOR APPLIED SYSTEM ANALYSIS»
AT THE NATIONAL TECHNICAL UNIVERSITY OF UKRAINE «KPI»
OF MES AND NAS OF UKRAINE

SYSTEM RESEARCH AND INFORMATION TECHNOLOGIES

INTERNATIONAL SCIENTIFIC AND TECHNICAL JOURNAL

№ 4 2011

IT IS FOUNDED IN JULY, 2001

EDITORIAL BOARD:

The editor – in – chief

M.Z. ZGUROVSKY, Academician of NASU

Deputy editor – in – chief

N.D. PANKRATOVA, Prof., Ukraine

Associate editors:

F.I. ANDON, Correspondent member of NASU
A.V. ANISIMOV, Prof., Ukraine
V.G. BONDARENKO Prof., Ukraine
Z. BONNIN, Prof., France
E.A. FEINBERG Prof., USA
A.D. GVISHIANI Prof., Russia
V.M. KUNTHEVICH, Academician of NASU
A.S. MAKARENKO Prof., Ukraine
A.A. PAVLOV, Prof., Ukraine
T.L. SAATY, Prof., USA
L. SAKALAUSKAS Prof., Lithuania
I.V. SERGIENKO, Academician of NASU
R. VORONKA, Prof., USA
G.-W. WEBER Prof., Turkey
Ya.S. YATSKIV Academician of NASU

THE EDITION ADDRESS:

03056, Kyiv,
av. Peremogy, 37, building 35,
Educational-Scientific Complex
«Institute for Applied System Analysis» at the
National Technical University of Ukraine «KPI»
of MES and NAS of Ukraine
Phone: **406-81-44**; Fax: **406-81-44**
E-mail: journal@mmsa.ntu-kpi.kiev.ua
http://journal.iasa.com.ua

In the issue:

• Automated control systems

• Progressive information technologies, high-efficiency computer systems

• Decision making and control in economic, technical, ecological and social systems

• Mathematical methods, models, problems and technologies for complex systems research

Шановні читачі!

Навчально-науковий комплекс «Інститут прикладного системного аналізу» НТУУ «КПІ» МОН та НАН України видає міжнародний науково-технічний журнал

«СИСТЕМНІ ДОСЛІДЖЕННЯ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ».

Журнал публікує роботи теоретичного і прикладного характеру в широкому спектрі проблем, пов'язаних із системними дослідженнями та інформаційними технологіями.

Провідні тематичні розділи журналу:

Теоретичні та прикладні проблеми і методи системного аналізу; теоретичні та прикладні проблеми інформатики; автоматизовані системи управління; прогресивні інформаційні технології, високопродуктивні комп'ютерні системи; проблеми прийняття рішень і управління в економічних, технічних, екологічних і соціальних системах; теоретичні та прикладні проблеми інтелектуальних систем підтримки прийняття рішень; проблемно і функціонально орієнтовані комп'ютерні системи та мережі; методи оптимізації, оптимальне управління і теорія ігор; математичні методи, моделі, проблеми і технології дослідження складних систем; методи аналізу та управління системами в умовах ризику і невизначеності; евристичні методи та алгоритми в системному аналізі та управлінні; нові методи в системному аналізі, інформатиці та теорії прийняття рішень; науково-методичні проблеми в освіті.

Головний редактор журналу — ректор Національного технічного університету України «КПІ», директор Навчально-наукового комплексу «Інститут прикладного системного аналізу» НТУУ «КПІ» МОН та НАН України, академік НАН України Михайло Захарович ЗГУРОВСЬКИЙ.

Журнал «Системні дослідження та інформаційні технології» включено до переліку фахових видань ВАК України.

Статті публікуються українською, російською та англійською мовами.

Журнал можна передплатити. **Наш індекс 23918.** Якщо ви не встигли передплатити журнал, його можна придбати безпосередньо в редакції. Наша адреса: 03056, м. Київ, пр. Перемоги, 37, корп. 35.

Завідуюча редакцією **С.М. Шевченко**
Редактор **А.О. Млечко**
Молодший редактор **В.М. Литвин**
Комп'ютерна верстка та набір **А.А. Патіоха**

Підписано до друку 15.12.2011. Формат 70×108/16.
Папір офс. Офс. друк. Ум. друк. арк. 14. Ціна договірна.
Свідоцтво про реєстрацію КВ № 13804–2778 ПР від 17.03.2008 р.

Видавництво та поліграфія ВПК «Політехніка»
ВПІ НТУУ «КПІ»
03056, Київ–56, пр. Перемоги, 37, корп. 15

Dear Readers!

The Educational-scientific complex «Institute for applied system analysis» at the National technical university of Ukraine «KPI» of MES and NAS of Ukraine is published of the international scientific and technical journal

«SYSTEM RESEARCH AND INFORMATION TECHNOLOGIES».

The Journal is printing works of a theoretical and applied character on a wide spectrum of problems, connected with system researches and information technologies.

The main thematic sections of the Journal are the following:

Theoretical and applied problems and methods of system analysis; theoretical and applied problems of computer science; automated control systems; progressive information technologies, high-efficiency computer systems; decision making and control in economic, technical, ecological and social systems; theoretical and applied problems of intellectual systems for decision making support; problem- and function-oriented computer systems and networks; methods of optimization, optimum control and theory of games; mathematical methods, models, problems and technologies for complex systems research; methods of system analysis and control in conditions of risk and uncertainty; heuristic methods and algorithms in system analysis and control; new methods in system analysis, computer science and theory of decision making; scientific and methodical problems in education.

The editor-in-chief of the Journal is Rector of the National Technical University of Ukraine «KPI», Director of the Educational-scientific complex «Institute for applied system analysis» at the National technical university of Ukraine «KPI» of MES and NAS of Ukraine, academician of the NASU Michael Zaharovich Zgurovsky.

The articles to be published in the Journal in Ukrainian, Russian and English languages are accepted. Information printed in the Journal is included in the Catalogue of periodicals of Ukraine.

СИСТЕМНІ ДОСЛІДЖЕННЯ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

4 • 2011

ЗМІСТ

АВТОМАТИЗОВАНІ СИСТЕМИ УПРАВЛІННЯ

Романенко В.Д., Мілявський Ю.Л. Синтез багатовимірних координуючих систем керування з різнотемповою дискретизацією в детермінованому середовищі 7

ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ, ВИСОКОПРОДУКТИВНІ КОМП'ЮТЕРНІ СИСТЕМИ

Петренко А.І., Свістунов С.Я., Свірін П.В. Алгоритми балансування навантаження в Грід-системах 21

Алексейчук А.Н., Гришаков С.В. Неасимптотические оценки эффективности случайного кодирования в системе передачи информации по двоичному симметричному каналу связи с отводом 37

ПРОБЛЕМИ ПРИЙНЯТТЯ РІШЕНЬ І УПРАВЛІННЯ В ЕКОНОМІЧНИХ, ТЕХНІЧНИХ, ЕКОЛОГІЧНИХ І СОЦІАЛЬНИХ СИСТЕМАХ

Кривонос Ю.Г., Крак Ю.В., Бармак О.В. Комп'ютерне моделювання української жестової мови 48

Москаленко В.В., Кондращенко В.В. Математические модели управления процессами финансирования инвестиционных проектов 61

Сімашко В.Й. Методи і технології зниження пікових навантажень у корпоративних інформаційних системах із централізованим збереженням даних 74

Левченко Є.Г., Демчишин М.В., Рабчун А.О. Математичні моделі економічного менеджменту інформаційної безпеки 88

Джамалов А.Т., Рагимов Р.М. Прогнозирование эрозионных процессов почв на маршрутах прокладки магистральных трубопроводов на основе геоинформационных технологий и космических снимков высокого разрешения..... 97

МАТЕМАТИЧНІ МЕТОДИ, МОДЕЛІ, ПРОБЛЕМИ І ТЕХНОЛОГІЇ ДОСЛІДЖЕННЯ СКЛАДНИХ СИСТЕМ

Балабанов О.С. Від коваріацій до каузальності. Відкриття структур залежностей у даних 104

Кірік О.Є. Проблема існування розв'язку в задачах розподілу потоків 119

Спекторский И.Я. Применение сетей Петри для анализа КС-грамматик 129

Горбань Н.В. Довгострокові прогнози функцій стану квазілінійних гіперболічних систем в 134

Реферати 140

Відомості про авторів 145

Зміст журналу «Системні дослідження та інформаційні технології» за 2011 р. 147

Автори статей за 2011 р. 149

SYSTEM RESEARCH AND INFORMATION TECHNOLOGIES

4 • 2011

CONTENTS

AUTOMATED CONTROL SYSTEMS

Romanenko V.D., Milyavsky Y.L. Synthesis of multivariate coordinating multidimensional coordination control systems with multirate sampling in deterministic enviromemt 7

PROGRESSIVE INFORMATION TECHNOLOGIES, HIGH-EFFICIENCY COMPUTER SYSTEMS

Petrenko A.I., Svistunov S.Ya., Svirin V.I. Algorithm of balancing the load in the Grid-systems 21

Alekseychuk A.N., Gryshakov S.V. Unasymptotic estimates of random coding efficiency in the system of information transfer channel with take-off 37

DECISION MAKING AND CONTROL IN ECONOMIC, TECHNICAL, ECOLOGICAL AND SOCIAL SYSTEMS

Kryvonos Iu.G., Krak Iu.V., Barmak O.V. Computer-based modeling of Ukrainian sigh language 48

Moskalenko V.V., Kondrashchenko V.V. Mathematical models of investment projects funding processes management 61

Simashko V.Yo. Methods and technologies of reducing peak-loads in the corporate information systems with the centralized data preservation 74

Levchenko Y.H., Demchyshyn M.V., Rabchun A.O. The mathematical models of economic management of information security 88

Djamalov A.T., Ragimov R.M. Forecasting of soil erosion processes on the routes of the pipelines based on geoin formation technologies and high resolution satellite images 97

MATHEMATICAL METHODS, MODELS, PROBLEMS AND TECHNOLOGIES FOR COMPLEX SYSTEMS RESEARCH

Balabanov O.S. From covariation to causation. Discovery of structures of dependency in data 104

Kirik O.E. The problem of solution existence in the flow distribution problems 119

Spectorsky I.Ya. Application of Petri nets for the analysis of CF-grammars 129

Gorban N.V. Long-time forecasts for state functions of quasilinear hyperbolic systems in $\mathbb{R}^n$ 133

References 140

Information about the authors 145

Contents of journal «System Research & Information technologies» over 2011 147

Information about the authors over 2011 149

УДК 62-50

СИНТЕЗ БАГАТОВИМІРНИХ КООРДИНУЮЧИХ СИСТЕМ КЕРУВАННЯ З РІЗНОТЕМПОВОЮ ДИСКРЕТИЗАЦІЄЮ В ДЕТЕРМІНОВАНОМУ СЕРЕДОВИЩІ

В.Д. РОМАНЕНКО, Ю.Л. МІЛЯВСЬКИЙ

Поставлено та вирішено задачу керування співвідношеннями між вихідними координатами багатовимірного різнотемпового об'єкта в детермінованому середовищі. Запропоновано новий алгоритм керування, що дозволяє одночасно забезпечувати виконання задачі слідкування і задачі координації в перехідному та усталеному режимах. Наведено результати чисельного моделювання, що підтверджують практичну цінність цієї розробки.

ВСТУП

Задача керування співвідношеннями, відома як проблема координуючого керування [1], почала розв'язуватись на досить загальному теоретичному рівні відносно недавно, хоча технічні системи, для яких принциповим є дотримання заданих співвідношень, відомі давно. До таких систем належать системи синхронного керування електрогенераторами, двигунами, стержнями ядерних реакторів, системи автоматичного дозування речовин тощо. У наш час до області застосування методів координації відносять також дослідження природних (біологічних, екологічних) та соціально-економічних систем, що складаються з великої кількості елементів, що взаємодіють, і функціонування яких полягає в дотриманні певних співвідношень між вихідними величинами.

Загальноприйняті підходи теорії керування, як виявилось, не дозволяють забезпечувати достатню точність керування співвідношеннями між вихідними координатами об'єкта в динаміці. Більшість методів, розроблених на сьогодні в теорії координуючого керування, на жаль, підходять лише для досить вузьких класів об'єктів або типів співвідношень [2, 3], і тому не можуть розглядатись як універсальне вирішення проблеми. У [1] розглядається задача координації на узагальненому рівні, але лише для неперервних систем. У цій роботі задача координуючого керування розглядається для дискретних систем, причому із різнотемповою дискретизацією.

Нині теорія дискретних різнотемпових систем характеризується інтенсивним розвитком. Це зумовлено наявністю широких класів об'єктів і підсистем у хімічній технології, енергетиці, економіці, екології та інших сфе-

© *В.Д. Романенко, Ю.Л. Мілявський*, 2011

рах, яким властиві різні швидкості перебігу процесів. Наприклад, у фінансово-економічній сфері актуальними є побудова й аналіз різнотемпових моделей динаміки курсів валют, процентних ставок, рівня безробіття, ВВП, обсягу виробництва чи витрат тощо, оскільки різні показники, що входять до цих моделей, змінюються і вимірюються з різною частотою. У теорії керування вже відомо досить багато цифрових алгоритмів, що забезпечують управління об'єктами з різнотемповою дискретизацією (тобто з різною частотою дискретизації входів та виходів) [4]. Але задача керування співвідношеннями для систем з різнотемповою дискретизацією досі не розглядалась. Проте її актуальність є очевидною вже тому, що на практиці майже завжди різні величини, співвідношення між якими необхідно забезпечувати, вимірюються в різні моменти часу, із різною частотою.

Отже, у цій роботі вперше поставлено задачу координуючого керування для систем із різнотемповою дискретизацією та запропоновано новий метод синтезу регуляторів для таких систем.

ПОСТАНОВКА ЗАДАЧІ КООРДИНУЮЧОГО КЕРУВАННЯ ПРИ РІЗНОТЕМПОВІЙ ДИСКРЕТИЗАЦІЇ

Нехай є дискретний багатовимірний різнотемповий об'єкт керування з n входами та n виходами, причому вхідні сигнали $u_j(kT_0),\ j=1,...,n$ усі мають період дискретизації T_0, а вихідні сигнали $y_i(l_ih_i)$ мають у загальному випадку різні періоди дискретизації $h_i=m_iT_0,\ m_i\ge 1,\ l_i=\left[\dfrac{k}{m_i}\right],\ i=1,...,n$, де [ ] — позначення цілої частини числа. Динаміка цього об'єкта задається за допомогою матричної дискретної передаточної функції (МДПФ) у формі z-перетворення $W(z,z_1,...,z_n)=\{W_{ij}(z,z_i)\},\ i,j=1,...,n$, де $W_{ij}(z,z_i)=$ $=\dfrac{y_i(z_i)}{u_j(z)},\ z_i=z^{m_i}$. Оскільки z^{-1} — оператор зворотного зсуву на період T_0, то $z_i^{-1}=z^{-m_i}$ — оператор зворотного зсуву на період $m_iT_0=h_i$. Якщо підставити $z_i=z^{m_i}$ у вираз для $W_{ij}(z,z_i)$, отримаємо $W_{ij}(z)=\dfrac{y_i(z^{m_i})}{u_j(z)}$, і тоді можна розглядати $W(z,z_1,...,z_n)$ як $W(z)=\{W_{ij}(z)\},\ i,j=1,...,n$.

У роботі розглядається проблема синтезу регулятора для цього об'єкта, що має забезпечити виконання одразу двох задач, які надалі будемо називати задачею слідкування та задачею координації. Під задачею координації розумітимемо вимогу виконання в динаміці системи лінійних співвідношень між вихідними координатами

$$AY=b, \tag{1}$$

де $Y=(y_1,...,y_n)^T$, A — задана матриця розмірності $r\times n,\ r<n$, що має повний ранг, b — заданий вектор розмірності r. Оскільки координати $y_i(l_ih_i)$ вимірюються в різні моменти часу, то припускаємо, що

$y_i(l_i h_i + k'T_0) = y_i(l_i h_i)$, $0 \le k' < m_i$, $i = 1,...,n$, та будемо вимагати виконання рівності (1) при кожному k у сенсі $AY(kT_0) = b$, де $Y(kT_0) = (y_1(kT_0),..., y_n(kT_0))^T$.

Під задачею слідкування розумітимемо класичну задачу приведення вихідного сигналу Y до рівня сигналу задаючого діяння $G = (g_1,..., g_n)^T$ (якщо $G = \text{const}$, то задача слідкування стає задачею стабілізації). Очевидно, що для того, щоб задачі слідкування та координації були узгодженими, G слід подавати спеціальним чином. Запропонуємо дві схеми. Згідно з рис. 1, задаючі діяння одразу формуються так, щоб задовольняти (1), тобто $AG(kT_0) = b$. При цьому вони формуються без зворотного зв'язку (рис. 1). Перевагою схеми на рис. 1 є простота і «чистота» задаючого сигналу, не пов'язаного з якістю роботи системи. Недоліком є те, що задаюче діяння не враховує, чи виконується реально співвідношення між вихідними координатами, тобто вимагається виконання співвідношення лише між ідеальними, бажаними сигналами, а не між дійсними.

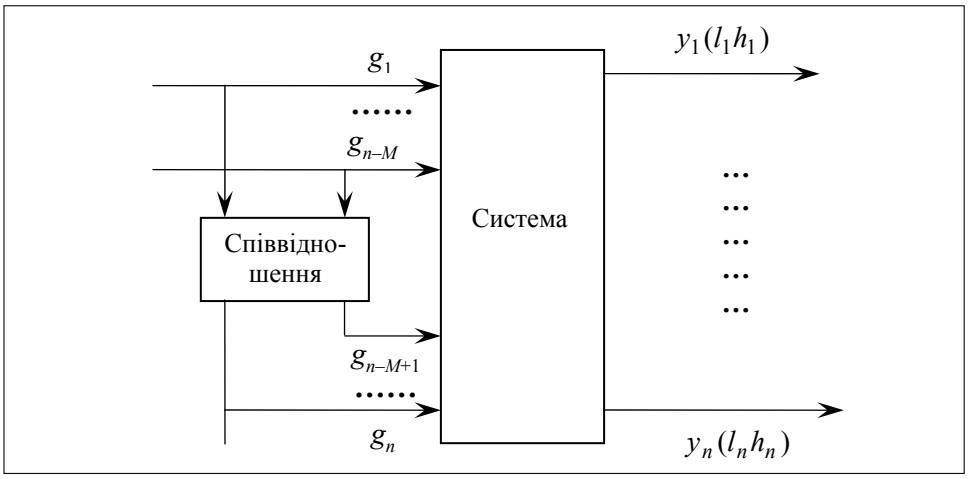

Рис. 1. Схема 1 координуючого керування

Щодо частоти дискретизації задаючих діянь, розглянемо два варіанти. У варіанті 1 всі задаючі діяння дискретизуються з періодом T_0, тобто $G(kT_0) = (g_1(kT_0),..., g_n(kT_0))^T$. У варіанті 2 частота дискретизації задаючого діяння дорівнює частоті відповідної вихідної координати, тобто $G = (g_1(l_1h_1),..., g_n(l_nh_n))^T$. У цьому разі для виконання при кожному k $AG(kT_0) = b$ припускається, що $g_i(l_i h_i + k'T_0) = g_i(l_i h_i)$, $0 \le k' < m_i$, $i = 1,..., n$.

Згідно зі схемою на рис. 2 із фізичних міркувань серед вихідних змінних обирається $n - M$ ведучих і M ведених координат. Оскільки за припущенням усі співвідношення лінійно незалежні, то ведені координати можна виразити однозначно як лінійні комбінації ведучих. По прямих каналах ведучих координат подаються певні задаючі діяння (без обмежень), а замість задаючих діянь для ведених координат подаються відповідні лінійні комбінації ведучих вихідних координат (рис. 2). Перевагою схеми 2 є той факт,

що ми в явному вигляді вимагаємо від системи, щоб ведені вихідні координати відповідали потрібним співвідношенням ведучих вихідних координат, тобто виконання співвідношення тепер є не додатковою умовою, поряд із відпрацюванням вхідних сигналів, а основною умовою. Недоліком схеми є відносна складність та внесення деякого запізнення в задаюче діяння по веденому каналу внаслідок інерційності системи. За умови застосування схеми на рис. 2 для обчислення задаючих діянь для ведених вихідних координат доведеться шукати лінійні комбінації ведучих вхідних координат із різною частотою дискретизації. Для простоти будемо, як завжди, припускати, що вони не змінюють своє значення в проміжку часу між моментами квантування, тоді в загальному випадку найзручнішим для схеми 2 буде варіант дискретизації 1 (з малим періодом для всіх $g_1,...,g_n$), хоча в конкретних ситуаціях можливі й інші варіанти дискретизації.

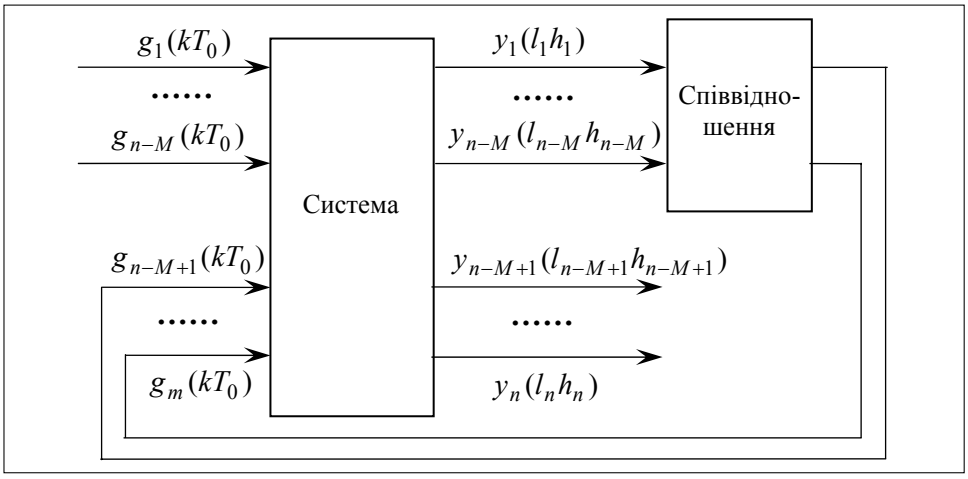

Рис. 2. Схема 2 координуючого керування

Далі ми не будемо уточнювати, згідно з якою схемою подається задаюче діяння, оскільки це не впливає на подальші викладки.

РОЗРОБКА ЕТАЛОННОЇ МОДЕЛІ ДИНАМІКИ ЗАМКНЕНОЇ СИСТЕМИ З РІЗНОТЕМПОВОЮ ДИСКРЕТИЗАЦІЄЮ. СИНТЕЗ СЛІДКУЮЧОГО РЕГУЛЯТОРА

Запропонуємо контур слідкуючого керування, зображений на рис. 3, де позначено $E = (e_1,...,e_n)^T = G - Y$ — вектор помилок керування.

Для проектування регулятора $W^p(z)$ використовуватимемо метод безпосереднього синтезу регулятора, в одновимірному випадку описаний у [6]. Для цього спочатку задамо бажану МДПФ замкненої системи. Її вигляд залежить від обраного варіанта дискретизації задаючих діянь. Очевидно, що для забезпечення автономності її потрібно обирати діагональною. При варіанті 1 матимемо $H(z,z_1,...,z_n) = \text{diag}\{H_i(z,z_i)\}$, $i = 1,...,n$, де $H_i(z,z_i) = \dfrac{y_i(z_i)}{g_i(z)}$. Ця різнотемпова передаточна функція має ті ж властивості, що

і розглянена вище передаточна функція об'єкта керування $W_{ij}(z, z_i)$, і так само приводиться до вигляду $H_{ij}(z)$. Під час використання варіанта 2 отримаємо $H(z_1,...,z_n) = \text{diag}\{H_i(z_i)\}$, $i = 1,...,n$, де $H_i(z_i) = \dfrac{y_i(z_i)}{g_i(z_i)}$, тобто передаточна функція по *i*-му каналу буде однотемповою з періодом дискретизації h_i. Зрозуміло, що і в цьому випадку за потреби можна підставити $z_i = z^{m_i}$ та отримати $H(z) = \text{diag}\{H_i(z)\}$, $i = 1,...,n$, $H_i(z) = \dfrac{y_i(z^{m_i})}{g_i(z^{m_i})}$.

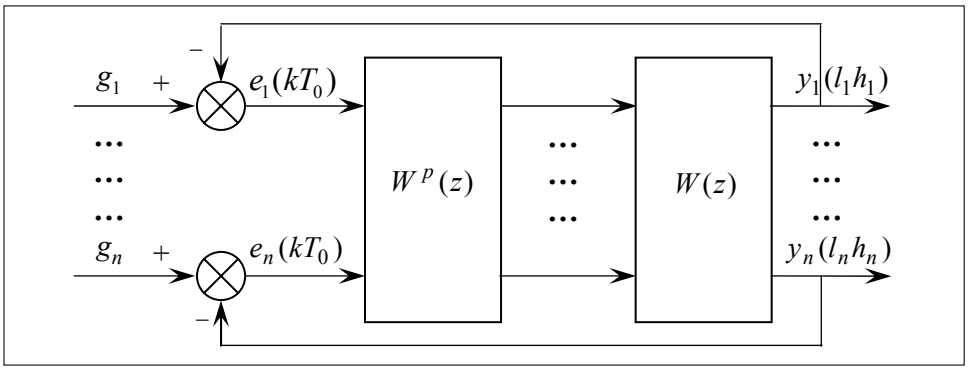

Рис. 3. Контур слідкуючого керування

Запропонуємо конкретний вигляд бажаної МДПФ замкненої системи для слідкуючого контуру. Найчастіше бажаним є перехідний процес, що моделюється в неперервному випадку аперіодичною ланкою першого порядку $\dfrac{1}{Ts+1}$, де T — постійна часу, що обирається із міркувань швидкодії. Якщо на вході системи знаходиться екстраполятор нульового порядку з частотою квантування T_0, тобто задаюче діяння змінюється в моменти часу kT_0, $k = 1, 2, \ldots$, то дискретна передаточна функція відповідної цифрової системи має вигляд $\dfrac{(1-\alpha)z^{-1}}{1-\alpha z^{-1}}$, де $\alpha = e^{-\lambda T_0}$, $\lambda = \dfrac{1}{T}$ [5]. У загальному випадку, коли задаюче діяння дискретизується з періодом $h = mT_0$, $m > 1$, а вихідна координата — з періодом T_0, передаточну функцію аперіодичної ланки першого порядку можна представити (шляхом домноження чисельника та знаменника на $1 + z^{-1} + z^{-2} + ... + z^{-m}$) як

$$\frac{(1-\alpha)(z^{-1} + z^{-2} + ... + z^{-m-1})}{1 + (1-\alpha)(z^{-1} + z^{-2} + ... + z^{-m}) - \alpha z^{-m-1}}. \qquad (2)$$

У випадку, коли, навпаки, вихідна координата дискретизується з періодом $h = mT_0$, $m > 1$, а вхідна — з періодом T_0, шляхом домноження чисельника та знаменника передаточної функції аперіодичної ланки на $1 + \alpha z^{-1} + \alpha^2 z^{-2} + ... + \alpha^{m-1} z^{-m}$, отримуємо її різнотемповий аналог:

$$\frac{(1-\alpha)(z^{-1}+\alpha z^{-2}+...+\alpha^{m-1}z^{-m})}{1-\alpha^m z^{-m}}=\frac{(1-\alpha)(z^{-1}+\alpha z^{-2}+...+\alpha^{m-1}z^{-m})}{1-\alpha^m z_1^{-1}}, \quad (3)$$

де $z_1 = z^m$.

Таким чином, підставляючи потрібне *m*, отримуємо бажані дискретні різнотемпові передаточні функції для кожного з каналів. Ці передаточні функції, об'єднані в єдину МДПФ, утворюють еталонну модель динаміки замкненого контуру, що виступає як критерій оптимальності для синтезу регулятора.

Припустимо, що МДПФ об'єкта невироджена, тобто $\det W(z, z_1,..., z_n) \neq 0$. Згідно з методом безпосереднього синтезу отримуємо МДПФ слідкуючого регулятора такого вигляду:

$$W^p(z) = W^{-1}(z)(I - H(z))^{-1}H(z). \quad (4)$$

Покажемо, що за умови застосування такого регулятора замкнена система дійсно матиме бажану МДПФ *H*(*z*). Для цього знайдемо МДПФ замкненого контуру:

$$[I + W(z)W^p(z)]^{-1}W(z)W^p(z) =$$

$$= [I + W(z)W^{-1}(z)(I - H(z))^{-1}H(z)]^{-1}W(z)\times$$

$$\times W^{-1}(z)(I - H(z))^{-1}H(z) =$$

$$= [I + (I - H(z))^{-1}H(z)]^{-1}(I - H(z))^{-1}H(z) =$$

$$= [(I - H(z))(I + (I - H(z))^{-1}H(z))]^{-1}H(z) =$$

$$= [I - H(z) + H(z)]^{-1}H(z) = H(z).$$

Отже, регулятор (4) забезпечує задану через $H(z)$ динаміку контуру слідкування.

Таким чином, побудовано регулятор, що має забезпечувати відслідковування вихідними координатами Y задаючих діянь G. Але задача координації вирішена тільки частково, в усталеному стані, оскільки і в схемі 1, і в схемі 2 задаюче діяння задовольняє співвідношенню (1), і тому, коли після завершення перехідного процесу виконається $Y = G$, то рівність (1) буде задовольнятись і для вихідних координат. Але цього, як правило, недостатньо. На практиці найчастіше вимагається, щоб співвідношення (1) виконувалось по можливості також і під час перехідного процесу, а не лише в усталеному стані. Іншими словами, задача координації має вищий пріоритет і тому має бути вирішена «швидше», ніж задача слідкування.

СИНТЕЗ КООРДИНУЮЧОГО ТА КОМБІНОВАНОГО КОНТУРІВ КЕРУВАННЯ

У цій роботі пропонується синтезувати додатковий контур координації. Оскільки співвідношення за припущенням мають задовольнятись у кожний

момент часу kT_0, цей контур буде однотемповим з періодом T_0. Введемо в розгляд уявний об'єкт керування з МДПФ $AW(z)$. Оскільки за визначенням $W(z)U(kT_0) = Y(kT_0)$, де $U(kT_0) = (u_1(kT_0),...,u_n(kT_0))^T$ — сигнал керування, то вихідним сигналом уявного об'єкта буде $AW(z)U(kT_0) = AY(kT_0)$. Розглянемо контур координації, в якому вектор *b* буде відігравати роль задаючого діяння, уявний об'єкт — роль об'єкта керування, а $E^{\kappa}(kT_0) = (e_1^{\kappa}(kT_0),...,e_r^{\kappa}(kT_0))^T = b - AY(kT_0)$ — помилки керування (рис. 4).

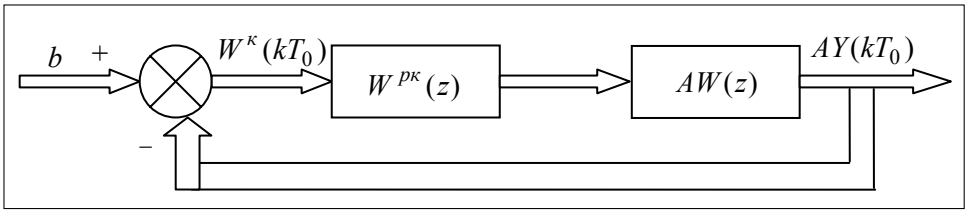

Рис. 4. Контур координації

Необхідно синтезувати координуючий регулятор $W^{p\kappa}(z)$, який буде підтримувати «вихідну координату» уявного об'єкта $AY(kT_0)$ на рівні задаючого діяння *b*, тобто забезпечуватиме виконання (1). Виберемо бажану МДПФ контуру координації $H^{\kappa}(z) = \text{diag}\{H_s^{\kappa}(z)\}$, $s = 1,...,r$, де $H_s^{\kappa}(z) = \frac{x_s(z)}{b_s}$, $AY(z) = (x_1(z),...,x_r(z))^T$, $b = (b_1,...,b_r)^T$.

Суттєвою особливістю цього контуру є те, що об'єкт і регулятор мають різні кількості входів і виходів, а саме об'єкт має *n* входів і r виходів, а регулятор — r входів та *n* виходів, $r < n$. Отже, не існує оберненої матриці до $AW(z)$, і метод безпосереднього синтезу застосувати неможливо.

Виберемо деяку опорну матрицю C розмірності $r \times n$ так, щоб $\det(AW(z)C) \neq 0$. За аналогією з (4) запропонуємо такий регулятор:

$$W^{p\kappa}(z) = C(AW(z)C)^{-1}(I - H^{\kappa}(z))^{-1}H^{\kappa}(z). \qquad (5)$$

Знайдемо МДПФ замкненого контуру координації:

$$[I + AW(z)W^{p\kappa}(z)]^{-1}AW(z)W^{p\kappa}(z) =$$

$$= [I + AW(z)C(AW(z)C)^{-1}(I - H^{\kappa}(z))^{-1}H^{\kappa}(z)]^{-1}AW(z)C \times$$

$$\times (AW(z)C)^{-1}(I - H^{\kappa}(z))^{-1}H^{\kappa}(z) =$$

$$= [I + (I - H^{\kappa}(z))^{-1}H^{\kappa}(z)]^{-1}(I - H^{\kappa}(z))^{-1}H^{\kappa}(z) =$$

$$= [(I - H^{\kappa}(z))(I + (I - H^{\kappa}(z))^{-1}H^{\kappa}(z))]^{-1}H^{\kappa}(z) =$$

$$= [I - H^{\kappa}(z) + H^{\kappa}(z)]^{-1}H^{\kappa}(z) = H^{\kappa}(z).$$

Отже, регулятор (5) забезпечує задану через $H^{к}(z)$ динаміку контуру координації. В якості еталонної моделі візьмемо

$$H^{к}(z) = \mathrm{diag}\,\{H_i^{к}(z)\},\ i = 1,...,r,\ H_i^{к}(z) = \frac{(1 - e^{-\lambda_i^{к} T_0}) z^{-1}}{1 - e^{-\lambda_i^{к} T_0} z^{-1}},$$

де $\lambda_i^{к} = \dfrac{1}{T_i^{к}}$, $T_i^{к}$ — стала часу *i*-го каналу замкненого контуру координації.

Тепер об'єднаємо контури слідкування і координації в єдиний контур керування. На вхід реального об'єкта будемо подавати суму керуючих діянь, що виробляються слідкуючим та координуючим регуляторами, тобто $U(kT_0) = W^{p}(z)E(kT_0) + W^{pк}(z)E^{к}(kT_0)$ (рис. 5).

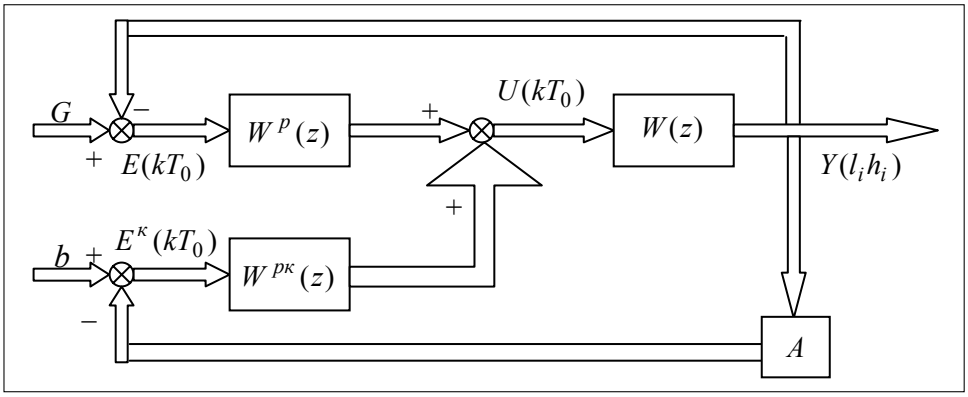

Рис. 5. Комбінований контур керування

Вимога щодо більшої швидкодії контуру координації в порівнянні з контуром слідкування може бути забезпечена відповідним вибором бажаних $H(z)$ і $H^{к}(z)$. Можна також явно ввести деяке $\alpha < 1$ і обчислювати $U(kT_0) = \alpha W^{p}(z)E(kT_0) + W^{pк}(z)E^{к}(kT_0)$, але для простоти вважатимемо, що $H^{к}(z)$ одразу вибрано більш швидкодіючим по всіх каналах, ніж $H(z)$, тобто відповідні сталі часу $H^{к}(z)$ менші, ніж $H(z)$.

Найголовніша умова, яку мають задовольняти всі параметри системи, — це стійкість комбінованого замкненого контуру. Виведемо цю умову математично. Очевидно, що контури слідкування і координації є стійкими за побудовою (тому що бажані МДПФ вибираються стійкими), але після їх об'єднання система може втратити стійкість. Для знаходження умови стійкості комбінованого контуру виведемо спочатку його рівняння динаміки, представивши все в однотемповій формі:

$$Y(z) = W(z)U(z) = W(z)(W^{p}(z)E(z) + W^{pк}(z)E^{к}(z)) =$$

$$= W(z)W^{p}(z)(G(z) - Y(z)) + W(z)W^{pк}(z)(b - AY(z)) =$$

$$= W(z)W^{p}(z)G(z) + W(z)W^{pк}(z)b - (W(z)W^{p}(z) + W(z)W^{pк}(z)A)Y(z);$$

$$Y(z)=[I+W(z)W^{p}(z)+W(z)W^{p\kappa}(z)A]^{-1}[W(z)W^{p}(z)G(z)+W(z)W^{p\kappa}(z)b].$$

Отже, характеристичне рівняння замкненого комбінованого контуру буде

$$\det[I+W(z)W^{p}(z)+W(z)W^{p\kappa}(z)A]=0 \qquad (6)$$

або, після підстановки (4), (5)

$$\det[I+(I-H(z))^{-1}H(z)+W(z)C(AW(z)C)^{-1}(I-H^{\kappa}(z))^{-1}H^{\kappa}(z)A]=0.$$

Тоді стійкість замкненого контуру забезпечується розміщенням коренів рівняння (6) всередині кругу одиничного радіусу на комплексній площині.

ПРИКЛАД: СИНТЕЗ ДВОВИМІРНОЇ КООРДИНУЮЧОЇ СИСТЕМИ З РІЗНОТЕМПОВОЮ ДИСКРЕТИЗАЦІЄЮ. РЕЗУЛЬТАТИ МОДЕЛЮВАННЯ

Розглянемо для прикладу двовимірний об’єкт керування з частим квантуванням на першому виході та вдвічі рідшим на другому. Нехай різнотемпова МДПФ об’єкта має вигляд:

$$W(z,z_1)=\begin{pmatrix} W_{11}(z) & W_{12}(z) \\ W_{21}(z,z_1) & W_{22}(z,z_1)\end{pmatrix},\ \text{де}\ z_1=z^2\ \text{та}$$

$$W_{11}(z)=\frac{y_1(z)}{u_1(z)}=\frac{0{,}013z^{-1}+0{,}01148z^{-2}}{1-1{,}64565z^{-1}+0{,}6703z^{-2}},$$

$$W_{12}(z)=\frac{y_1(z)}{u_2(z)}=\frac{0{,}039z^{-1}+0{,}016z^{-2}}{1-1{,}2875z^{-1}+0{,}37334z^{-2}},$$

$$W_{21}(z,z_1)=\frac{y_2(z_1)}{u_1(z)}=\frac{0{,}13z^{-1}+0{,}012873z^{-2}+0{,}0076z^{-3}+0{,}00769z^{-4}}{1-1{,}16756z_1^{-1}+0{,}2493z_1^{-2}},$$

або інакше

$$W_{21}(z)=\frac{y_2(z^2)}{u_1(z)}=\frac{0{,}13z^{-1}+0{,}012873z^{-2}+0{,}0076z^{-3}+0{,}00769z^{-4}}{1-1{,}16756z^{-2}+0{,}2493z^{-4}},$$

$$W_{22}(z,z_1)=\frac{y_2(z_1)}{u_2(z)}=\frac{0{,}013z^{-1}+0{,}032873z^{-2}+0{,}0276z^{-3}+0{,}00769z^{-4}}{1-1{,}36756z_1^{-1}+0{,}4493z_1^{-2}},$$

або інакше

$$W_{22}(z)=\frac{y_2(z^2)}{u_2(z)}=\frac{0{,}013z^{-1}+0{,}032873z^{-2}+0{,}0276z^{-3}+0{,}00769z^{-4}}{1-1{,}36756z^{-2}+0{,}4493z^{-4}}.$$

Нехай задаюче діяння для першої координати є незалежним і квантується з подвійним періодом, а задаюче діяння для другої координати обчислюється як лінійна функція від першого виходу (рис. 2) і, відповідно, квантується з малим періодом. Виберемо для двовимірної системи бажану МДПФ $H(z)=\begin{pmatrix} H_1(z) & 0 \\ 0 & H_2(z)\end{pmatrix}$.

У нашій системі H_1 та H_2 мають бути різнотемповими: H_1 має великий період дискретизації на вході та малий на виході, H_2 — навпаки. Підставимо в (2) $m=2$ та будемо мати $H_1(z)=\dfrac{(1-\alpha)(z^{-1}+z^{-2}+z^{-3})}{1+(1-\alpha)(z^{-1}+z^{-2})-\alpha z^{-3}}$. Потім підставимо в (3) $m=2$ і отримаємо $H_2(z,z_1)=\dfrac{(1-\alpha)(z^{-1}+\alpha z^{-2})}{1-\alpha^2 z_1^{-1}}$ або $H_2(z)=\dfrac{(1-\alpha)(z^{-1}+\alpha z^{-2})}{1-\alpha^2 z^{-2}}$.

Отже, при $\alpha=e^{-\lambda T_0}$, вибравши деякі λ_1, λ_2 (обернені до сталих часу еталонних моделей слідкуючого контуру по обох каналах), отримуємо:

$$H(z)=\begin{pmatrix} \dfrac{(1-e^{-\lambda_1 T_0})(z^{-1}+z^{-2}+z^{-3})}{1+(1-e^{-\lambda_1 T_0})(z^{-1}+z^{-2})-e^{-\lambda_1 T_0}z^{-3}} & 0 \\ 0 & \dfrac{(1-e^{-\lambda_2 T_0})(z^{-1}+e^{-\lambda_2 T_0}z^{-2})}{1-e^{-2\lambda_2 T_0}z^{-2}} \end{pmatrix}.$$

Таким чином, за формулою (4) можна синтезувати слідкуючий регулятор $W^{p}(z)$.

Нехай потрібно забезпечити виконання співвідношення $y_1(kT_0)-2y_2(kT_0)=1$, тобто $A=(1-2)$, $b=1$. Побудуємо (однотемповий) контур координації. Знову виберемо бажану передаточну функцію замкненого контуру координації (що в цьому прикладі виявився одновимірним) у формі $H^{к}(z)=\dfrac{(1-e^{-\lambda_к T_0})z^{-1}}{1-e^{-\lambda_к T_0}z^{-1}}$, де $\lambda_к=\dfrac{1}{T_к}$, $T_к$ — постійна часу контуру координації, і синтезуємо координуючий регулятор $W^{pк}(z)$ за формулою (5). Матрицю C, а також параметри $\lambda_1,\lambda_2,\lambda_к$ вибираємо так, щоб $\det(AW(z)C)\neq 0$ і виконувалась умова (6). У цьому прикладі вибрано $C=(2\ 1)^T$, $\lambda_1=0{,}3$, $\lambda_2=0{,}6$, $\lambda_к=1$, де виконується умова $\lambda_к>\lambda_1$, $\lambda_к>\lambda_2$, що забезпечує більшу швидкодію контуру координації в порівнянні зі слідкуючим контуром. Схему комбінованого контуру зображено на рис. 6.

Моделювання проводилось у середовищі Simulink 7.4 for Matlab 7.9. Нехай першим задаючим діянням є синусоїдний сигнал, а друге задаюче діяння формується згідно зі схемою на рис. 2, тобто, у цьому випадку, як $g_2(kT_0)=\dfrac{y_1(kT_0)-1}{2}$. На рис. 7 зображено ведучу координату та її задаюче діяння, на рис. 8 — ведену координату та її задаюче діяння, на рис. 9 — похибку співвідношення (тобто $e^{к}(kT_0)=1-y_1(kT_0)+2y_2(kT_0)$). Щоб переконатись у доцільності введення запропонованого в роботі додаткового контуру координації, наведемо на рис. 10–12 результати моделювання для тієї ж системи, але без цього додаткового контуру.

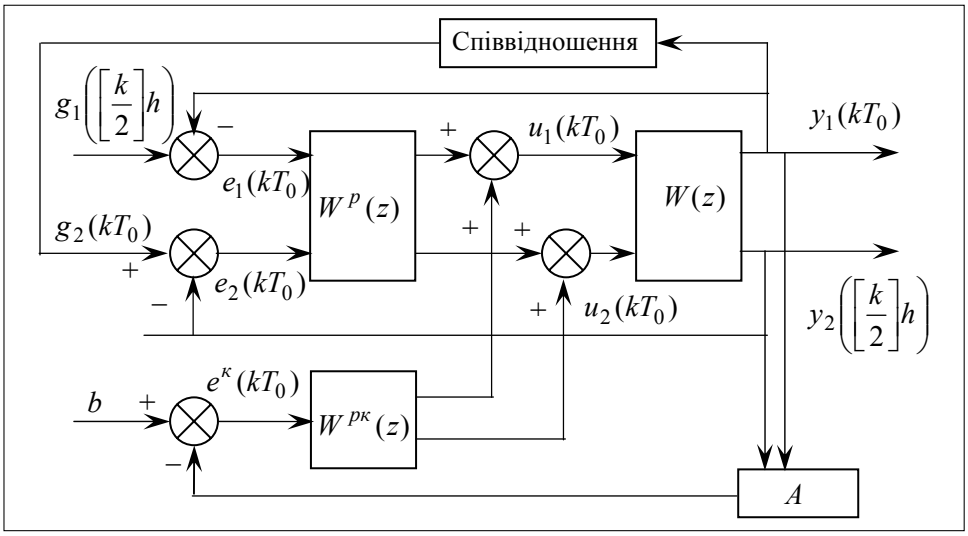

Рис. 6. Схема двовимірного різнотемпового контуру для моделювання

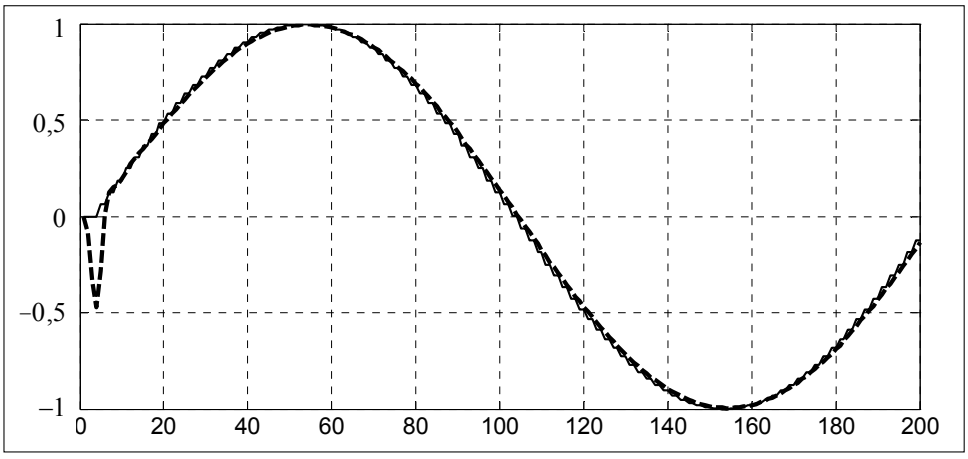

Рис. 7. Результати моделювання (неперервна лінія — задаюче діяння, пунктир — вихідна координата) для першого каналу

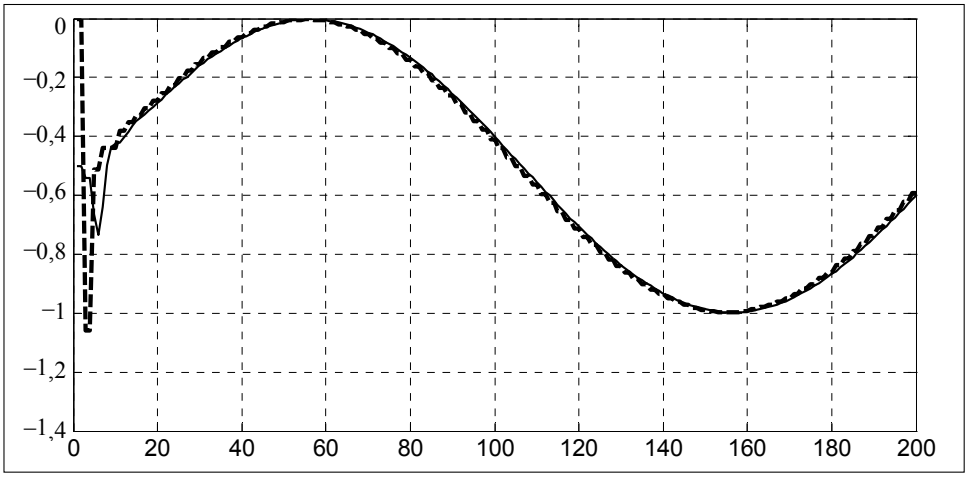

Рис. 8. Результати моделювання (неперервна лінія — задаюче діяння, пунктир — вихідна координата) для другого каналу

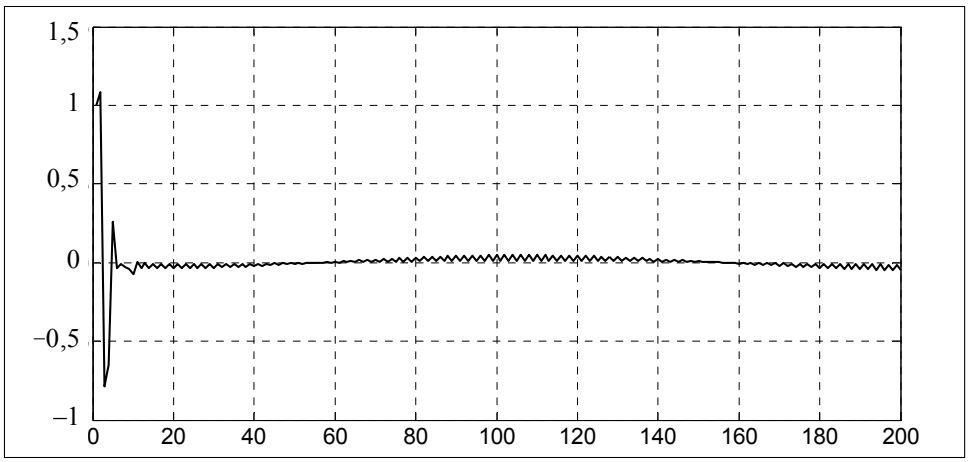

Рис. 9. Результати моделювання: графік похибки співвідношення

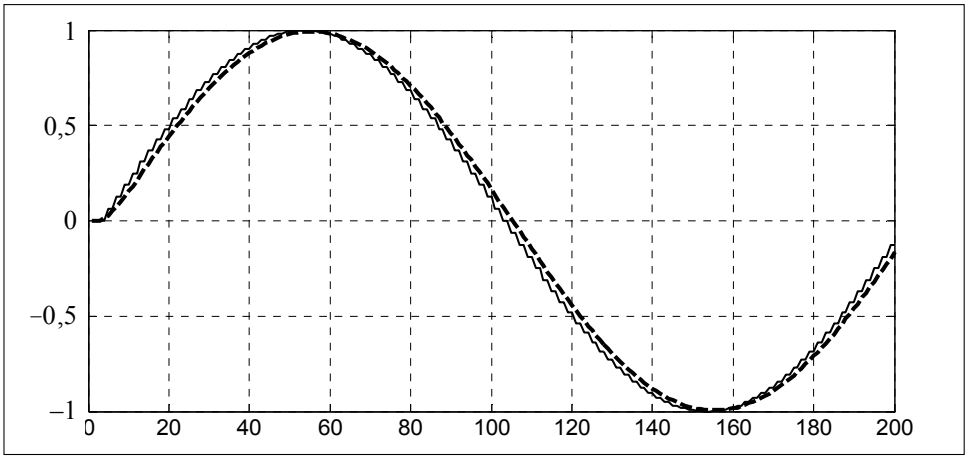

Рис. 10. Результати моделювання (неперервна лінія — задаюче діяння, пунктир — вихідна координата) для першого каналу за відсутності контуру координації

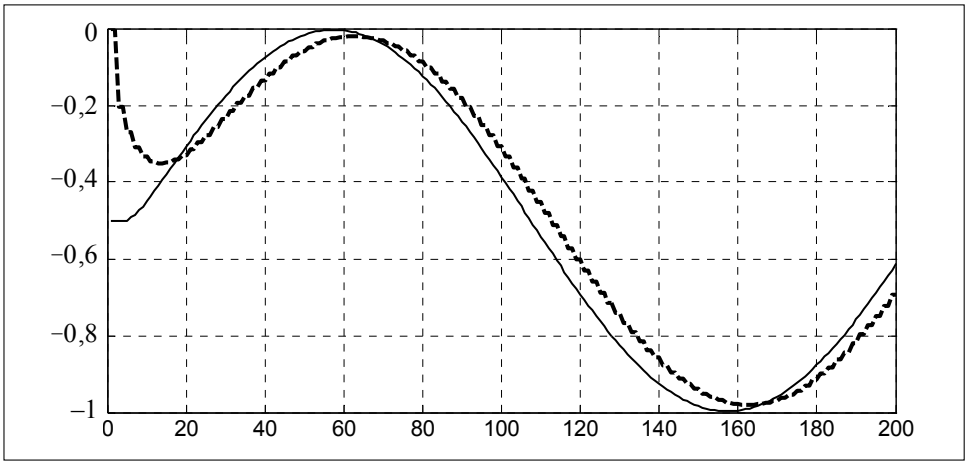

Рис. 11. Результати моделювання (неперервна лінія — задаюче діяння, пунктир — вихідна координата) для другого каналу за відсутності контуру координації

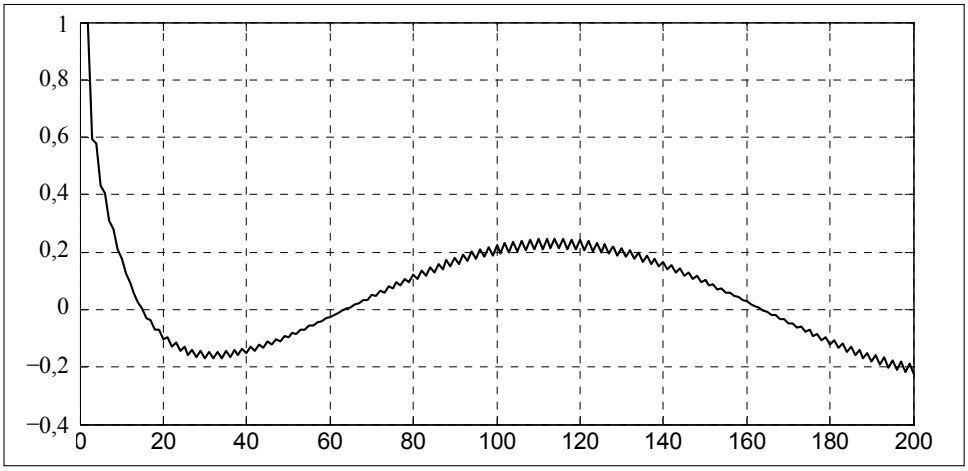

Рис. 12. Результати моделювання: графік похибки співвідношення за відсутності контуру координації

Із отриманих результатів випливає, що контур координації значно зменшує похибку співвідношення, що і було головною метою його введення. Крім того, виявляється, що якість слідкування також не тільки не погіршується, а навіть покращується, оскільки задаючі діяння задовольняють співвідношенню, і тому задачі слідкування і координації не суперечать, а доповнюють одна одну.

ВИСНОВКИ

У роботі досліджено проблему координуючого керування різнотемповим багатовимірним об’єктом у детермінованому середовищі.

1. Сформульовано математичну постановку задачі керування співвідношеннями (координуючого керування) між вихідними координатами об’єкта із різнотемповою дискретизацією та запропоновано декілька схем і варіантів формування задаючих діянь.

2. Розроблено різнотемпові критерії оптимальності, які подані у формі бажаних дискретних передаточних функцій замкненого контуру цифрового керування з різнотемповою дискретизацією.

3. Спроектовано слідкуючий регулятор на основі методу безпосереднього синтезу при різнотемповій матричній передаточній функції об’єкта керування та різнотемповій еталонній моделі.

4. Запропоновано спеціальний швидкодіючий контур координуючого керування, що забезпечує виконання заданих співвідношень під час перехідних процесів із достатньою точністю.

5. Розроблено метод синтезу цифрового регулятора для об’єкта типу «вхід–вихід» із різною кількістю входів і виходів та застосовано його для проектування координуючого регулятора.

6. Виведено умови стійкості комбінованого (слідкуючого та координуючого) контуру керування.

7. Проведено чисельне моделювання двовимірної координуючої системи з різнотемповою дискретизацією, що підтвердило ефективність та доцільність використання отриманих у роботі результатів.

ЛІТЕРАТУРА

1. *Бойчук Л.М.* Синтез координирующих систем автоматического управления. — М.: Энергоатомиздат, 1991. — 160 с.
2. *Мирошник И.В.* Согласованное управление многоканальными системами. — Л.: Энергоатомиздат, 1990. — 129 с.
3. *Кабальнов Ю.С., Лютов А.Г., Насибуллин Ф.Г.* Координированное управление группой автономных динамических объектов // Выч. техника и новые инф. технологии: межвуз. сб. науч. тр. — 1999. — С. 45–48.
4. *Згуровский М.З., Романенко В.Д.* Системы фильтрации и управления с разделяющимися разнотемповыми движениями. — Киев: Наук. думка, 1998. — 376 с.
5. *Романенко В.Д.* Методи автоматизації прогресивних технологій: підруч. — Київ: Вища шк., 1995. — 519 с.
6. *Chiu K.C., Corripio A.B., Smith C.L.* Digital control algorithms. Part 1. Dahlin algorithm // Instruments and control systems. — 1973. — № 10. — P. 57–59.

Надійшла 06.10.2010

УДК 004.724.2, 004.75

АЛГОРИТМИ БАЛАНСУВАННЯ НАВАНТАЖЕННЯ В ГРІД-СИСТЕМАХ

А.І. ПЕТРЕНКО, С.Я. СВІСТУНОВ, П.В. СВІРІН

Проведено аналіз стратегій балансування навантаження в Грід-системах та порівняльний аналіз наявних алгоритмів розподілу потоку завдань між обчислювальними ресурсами Грід-середовища.

ВСТУП

Необхідність застосування особливих механізмів управління ресурсами в *локально* і *глобально* розподілених середовищах є актуальною. До ресурсів відноситься все, що так чи інакше бере участь в обробці даних: обчислювальні кластери; сховища даних; файлові системи; програмне забезпечення (ПЗ); мережеве устаткування, яке забезпечує з'єднання ресурсів в єдину систему.

Векторні — паралельні комп'ютери, масивно-паралельні комп'ютери з розподіленою пам'яттю, паралельні комп'ютери із загальною пам'яттю та обчислювальні кластери з певною умовністю можна віднести до локальних розподілених обчислювальних середовищ «замкненої» конфігурації. Глобальне середовище метакомп'ютингу на сьогодні асоціюється з Грід-технологіями. Нинішня інтерпретація концепції Грід передбачає об'єднання в єдину обчислювальну систему як високопродуктивних суперкомп'ютерів, так і використання в якості обчислювальних ресурсів звичайних персональних комп'ютерів.

Для забезпечення вимог користувачів з продуктивності та ефективності виконання завдань Грід-система має реалізувати ефективний алгоритм розподілу завдань між доступними на даний час обчислювальними ресурсами. Основна мета такого балансування навантаження в Грід-системі — скоротити час виконання завдання користувача та забезпечити ефективність використання обчислювальних ресурсів для виключення ситуації у випадку, коли одні ресурси простоюють, а інші перевантажені виконанням завдань користувачів [1].

Незважаючи на те, що алгоритми балансування навантаження обчислювальних ресурсів у розподіленої системі досліджуються вже досить давно й існує багато готових алгоритмічних рішень, і програмних реалізацій, ін-

© *А.І. Петренко, С.Я. Свістунов, П.В. Свірін*, 2011

тенсивний розвиток Грід-технологій та вдосконалення проміжного ПЗ робить проблему балансування навантаження постійно актуальною, а інтерес до дослідження в цій галузі не зменшується, особливо для гетерогенних (неоднорідних) систем [2]. У цій роботі алгоритми балансування навантаження розглядаються з точки зору можливості застосування в українській Грід-інфраструктурі [3, 4, 5].

МОДЕЛЬ І СТРАТЕГІЇ РОЗПОДІЛУ НАВАНТАЖЕННЯ

Розглянемо загальну модель розподілу навантаження в Грід-системі G (рис. 1), яка складається з набору Грід-сайтів S_j, з'єднаних між собою звичайними каналами Інтернет [6]. Ресурси Грід-сайта включають у себе: обчислювальні ресурси, ПЗ, ресурси збереження даних і мережеву інфраструктуру. Кожен Грід-сайт S_j є обчислювальним кластером C_k з додатковими керуючими серверами із встановленим проміжним ПЗ. Це може бути проміжне ПЗ ARC проекту NorduGrid [7], gLite проекту EGI [8], UNICORE [9].

Обчислювальний кластер C_k є певною кількістю обчислювальних серверів CM_{jk}, з'єднаних комутаційним обладнанням, керуючого сервера з встановленою власної системою керування розподілом навантаження між обчислювальними серверами.

Цю модель, згідно з [11], назвемо *G/ S /C / M* моделлю, де G — Грід-система; S — кількість Грід-сайтів, що входять в Грід-систему; C — кількість кластерів, що входять у Грід-сайт; M — кількість обчислювальних серверів, що входять у кластер.

Ця загальна модель характеризується *неоднорідністю ресурсів* (як мережевих, так і обчислювальних) [12]; *автономністю ресурсів*, що належать окремим організаціям [13]; *динамічністю ресурсів* — коли обслуговуються як зовнішні, так і локальні користувачі, запити яких вишукуються в черги [14]; *спільним використанням ресурсів*, які можуть бути підключеними до багатьох Грід-систем одночасно та використовувати різне проміжне ПЗ; *різноманітністю додатків,* які можуть виконуватися паралельно чи послідовно. З урахуванням усіх аспектів завдань реалізація загальної мети балансування навантаження системи є вкрай складною [15].

У загальному вигляді стратегія балансування навантаження в Грід-системі має деревоподібну ієрархічну структуру (рис. 1), де введено такі рівні [11].

Рівень 0. Балансування навантаження між Грід-системами, які працюють під керуванням різного проміжного ПЗ. На даний час проблема функціональної сумісності та взаємодії Грід-систем не вирішена остаточно. Існуючі мета-планувальники забезпечують розподіл завдань користувачів у менш завантажену Грід-систему.

Рівень 1: Балансування навантаження всередині Грід-системи, коли планувальник збирає інформацію про поточне завантаження обчислювальних ресурсів у Грід-системі під час надходження завдань користувачів приймає рішення про розподіл завдань між Грід-сайтами.

Рівень 2: Балансування навантаження всередині Грід-сайту, коли розподілення навантаження виконується лише для завдань користувачів, які надійшли на цей Грід-сайт. Якщо в Грід-сайт входять декілька кластерів, то виконується розподіл завдань між цими обчислювальними кластерами.

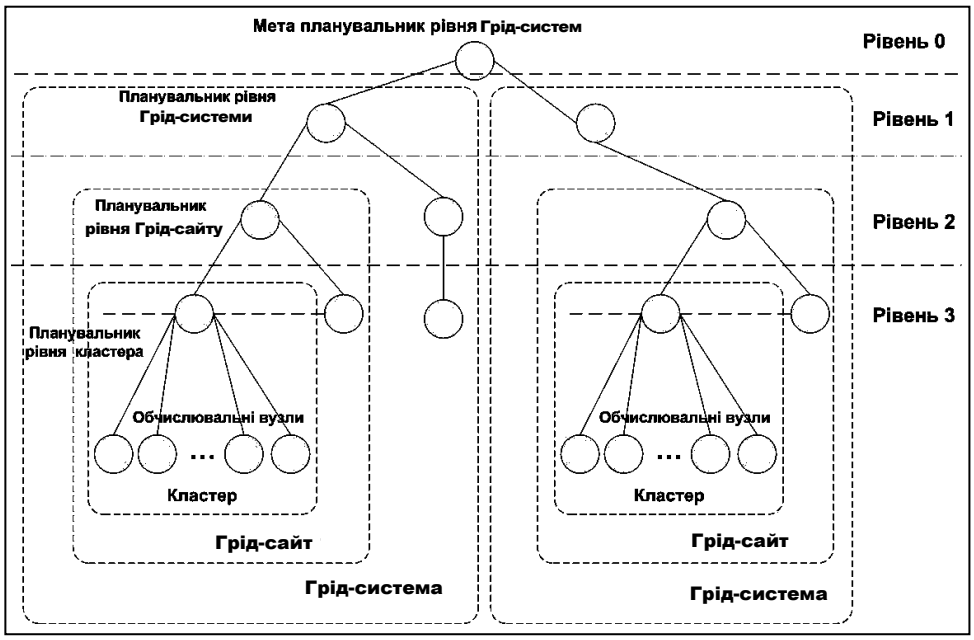

Рис. 1. Ієрархічна стратегія балансування навантаження в Грід-системі

Рівень 3: Балансування навантаження всередині кластера, коли завдання користувачів розподіляються між обчислювальними серверами, які входять у кластер, за допомогою локальної системи управління завданнями типу PBS, Condor тощо [15].

БАЛАНСУВАННЯ НАВАНТАЖЕННЯ МІЖ ГРІД-СИСТЕМАМИ

Для вибору ресурсів, моніторингу проходження завдань і доступу до даних з додатків застосовується *Resource Broker*, що відіграє роль зовнішнього *планувальника або мета-диспетчера робіт* і слугує інтерфейсом між Грід-службами і локальними планувальниками. Прикладами таких додатків є Nimrod/G [16], AppLeS Parameter Sweep Template (APST) [17], Gridbus Resource Broker (GRB) [18], GridWay [19].

При цьому для користувача забезпечується єдина точка входу до різних Грід-систем як до єдиного обчислювального ресурсу, функціональну сумісність яких засновано на використанні адаптерів і трансляторів. Для кожної Грід-структури в структурі мета-планувальника включено спеціальний адаптер, що змінює опис завдання користувача таким чином, щоб воно було зрозумілим іншій Грід-системі при наступній трансляції. У таблиці наведено характеристики досить простих алгоритмів, які використовуються в цих планувальниках ресурсів для забезпечення функціональної сумісності різних Грід-систем.

Таблиця. Характеристики алгоритмів планування ресурсів

Мета планувальника	**Інформаційна система**	**Застосований алгоритм планувальника**	**Особливості**
Nimrod/G	Інформаційний сервіс MDS Globus	В основу алгоритму покладено концепцію сервісу торгівлі ресурсами GRACE (Grid Architecture for Computational Economy — Грід-архітектура для «обчислювальної економіки»). Ключові параметри алгоритму: T — кінцевий термін виконання завдання, CR — умовна вартість ресурсу, CZ — умовний бюджет завдання. Алгоритми планування: вибір ресурсу найменшої вартості для виконання завдання в термін T (оптимізація вартості); вибір ресурсу для мінімізації терміну виконання завдання (оптимізація часу); оптимізація за критерієм час — вартість, яка подібна вартісній оптимізації, але за наявності множини ресурсів однакової вартості застосовується стратегія оптимізації часу; консервативна за часом стратегія, яка подібна оптимізації часу, але гарантує кожному завданню мінімум витрат на одне завдання	Для початку планування мають бути встановлено такі властивості завдання як «термін виконання» та «бюджет»
AppLeS Parameter Sweep Template (APST)	Network Weather Service (NWS), MDS, Ganglia	В якості метрик використовуються статичні характеристики (архітектура, обсяг пам'яті, дисковий обсяг), характеристики мережі (смуга пропускання, навантаження та затримка), пріоритет черги, надійність ресурсу, цінові характеристики ресурсу. Алгоритми планування: евристичні алгоритми Min-min, Max-min, Sufferage, и Xsufferage [20]; сторонні алгоритми планування виконання завдань [21]	Планування проводиться на короткий період часу з наступними перепланауваннями
Gridbus Resource Broker (GRB)	Grid Market Directory (GMD), Grid Index Information Service (GIIS)	Алгоритми планування проводять оптимізацію за критеріями мінімізації часу виконання завдання та ціни на основі параметрів наявних обчислювальних ресурсів та параметрів опису задачі (термін виконання, бюджет тощо)	Брокер не є Грід-сервісом, а встановлюється на кожній клієнтській машині

БАЛАНСУВАННЯ НАВАНТАЖЕННЯ ВСЕРЕДИНІ ГРІД-СИСТЕМИ

Час обробки завдання складається з часу: аналізу опису завдання, пошуку відповідного ресурсу, завантаження завдання на ресурс, очікування в черзі, виконання завдання на ресурсі та завантаження результатів користувачеві. Добре відомо, що задача оптимального розподілення робіт за ресурсами є NP-повною [36], тому існує лише наближені алгоритми балансування. Їх можна розділити на централізовані і децентралізовані, динамічні й статичні, періодичні й неперіодичні. Усі алгоритми балансування навантаження по-

будовано таким чином, щоб навантаження ресурсів було рівномірним, або використання ресурсів було максимальним при мінімізації загального часу виконання завдання. Прийнято розрізняти три типи таких алгоритмів, які здійснюють:

- *розподіл навантаження* у випадку розподілу навантажень на обчислювальний ресурс, який в даний час не використовується в обчисленнях, тобто вільний;
- *балансування навантаження*, коли завдання розподіляються між обчислювальними ресурсами таким чином, щоб забезпечити більш-менш рівномірне використання ресурсів або задовольнити певні критерії балансування (більше завдань на потужних кластерах);
- *вирівнювання навантаження*, коли не просто здійснюється рівномірний розподіл навантаження серед усіх ресурсів або вирівнювання навантаження, а планувальник прагне уникнути перенавантаження окремих ресурсів.

Вибір конкретного алгоритму балансування навантаження залежить від обраної моделі побудови брокера ресурсів, які можуть бути трьох таких типів.

Централізований брокер ресурсів — виконує вибір ресурсів за вимогами завдань користувачів для всієї Грід-системи, тобто повністю відповідає за розподіл усіх наявних ресурсів (рис. 2). Усю необхідну інформацію про динамічні характеристики ресурсів централізованого брокера він отримує від інформаційної системи. Прикладом реалізації такої моделі є проміжне забезпечення gLite [8], в якому WMS (Workfolw Management System — система управління потоком знань) відповідає за розподіл і управління завдань у межах наявних ресурсів [22–23].

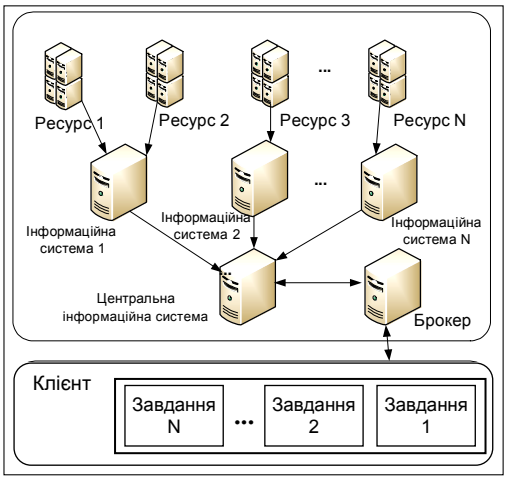

Рис. 2. Загальна архітектура Грід-системи з централізованим брокером

WMS також контролює процес виконання завдань та забезпечує отримання результатів виконання завдання користувачем. В gLite для балансування навантаження використовуються алгоритми «максимальної оцінки» та «нечіткої оцінки». Ці алгоритми проводять відбір ресурсів-кандидатів за допомогою оцінювання відповідності ресурсу вимогам завдання, яке надійшло в чергу.

Модель розподіленого брокера, в якому за результатами аналізу опису завдання на виконання кожен Грід-сайт формує свій перелік доступних обчислювальних ресурсів. Необхідну динамічну інформацію про стан Грід-системи такий брокер може отримувати від центральної інформаційної системи, а також шляхом запитів безпосередньо до доступних ресурсів (рис. 3), після чого починає пошук необхідного ресурсу для завдання користувача [24].

Прикладом реалізації такої моделі є проміжне забезпечення ARC [7, 38] проекту NorduGrid, у брокері якого реалізовано алгоритми вибору ресурсів, що забезпечують достатню масштабованість та надійність системи. У склад ARC включено алгоритми «Random» (проводить сортування кандидатів випадковим чином); «Benchmark» (проводить сортування згідно з оцінкою продуктивності ресурсу); «FastestQueueBroker» (сортує список кандидатів залежно від коефіцієнта наповненості черги ресурсу, наприклад, довжини черги, поділеної на загальну кількість процесорів); «Data» (сортує ресурси згідно з обсягом даних, що зберігаються в кеш обчислювального ресурсу).

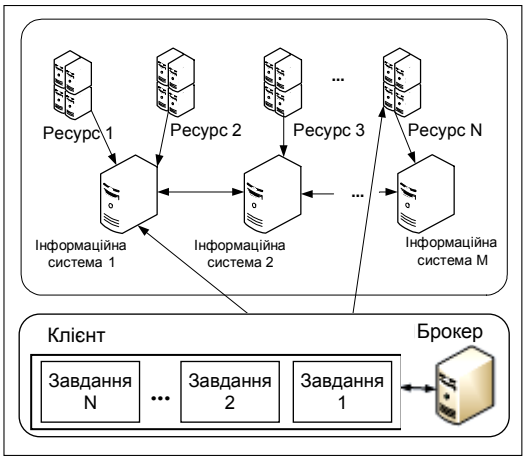

Рис. 3. Архітектура системи з розподіленим брокером ресурсів

Модель ієрархічного брокера поєднує в собі елементи централізованого та розподіленого підходів. На рис. 4 зображено узагальнену структуру ієрархічної моделі, що містить декілька екземплярів брокера ресурсів, які розподіляють свої функції за встановленими правилами. На відміну від одного великого ресурсного брокера існує декілька підходів для створення цієї ієрархічної структури, наприклад, віртуальна ієрархічна структура віртуальної організації, яку представлено в [25, 26], та різні ієрархічні структури використання Грід-систем на основі різного проміжного ПЗ [27].

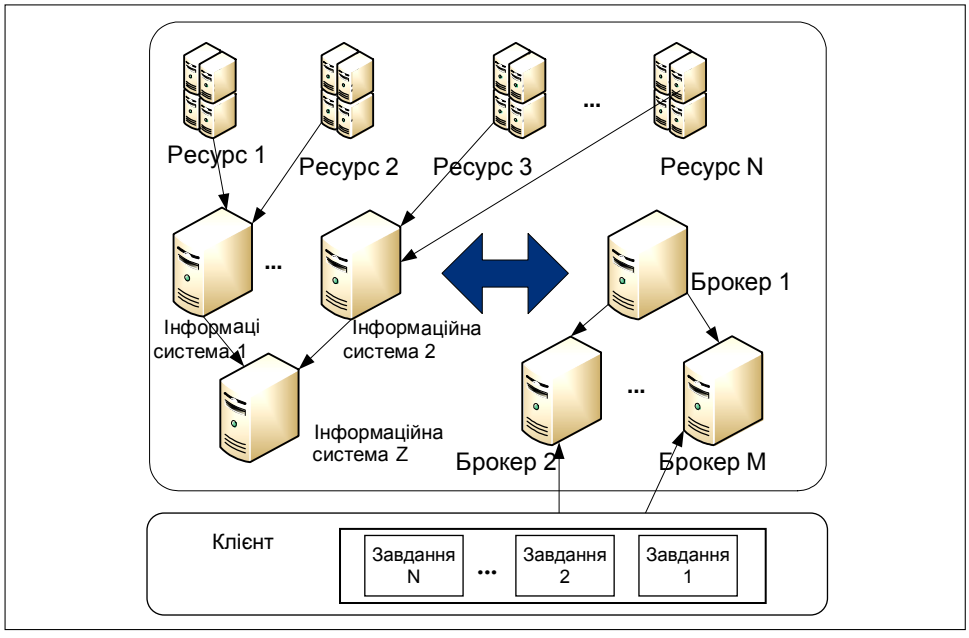

Рис. 4. Архітектура системи з ієрархічним брокером ресурсів

Одним із прикладів реалізації такої моделі є ієрархічна структура [27], для якої розроблено новий адаптер-планувальник, що дозволяє службі

управління Грід-ресурсами Globus GRAM відправляти завдання до GridWay [19], який потім використовує свій брокерський алгоритм для подання завдань до LRMS, або на Грід-системи з різним проміжним ПЗ, таким як AstorGrid та gLite.

Модель агентського брокера використовує спеціальні програмні компоненти (агенти), що реалізують алгоритм штучного інтелекту, автономні за своєю суттю, здатні до самовідновлення і прийняття рішень [28, 29, 30, 31]. Загальну структуру такої моделі показано на рис. 5. Динамічний характер агентів відрізняє цей підхід від традиційних, розглянених вище. Взаємодія між агентами у межах системи може бути децентралізованою, ієрархічною або представленою за схемою «головний–підлеглий».

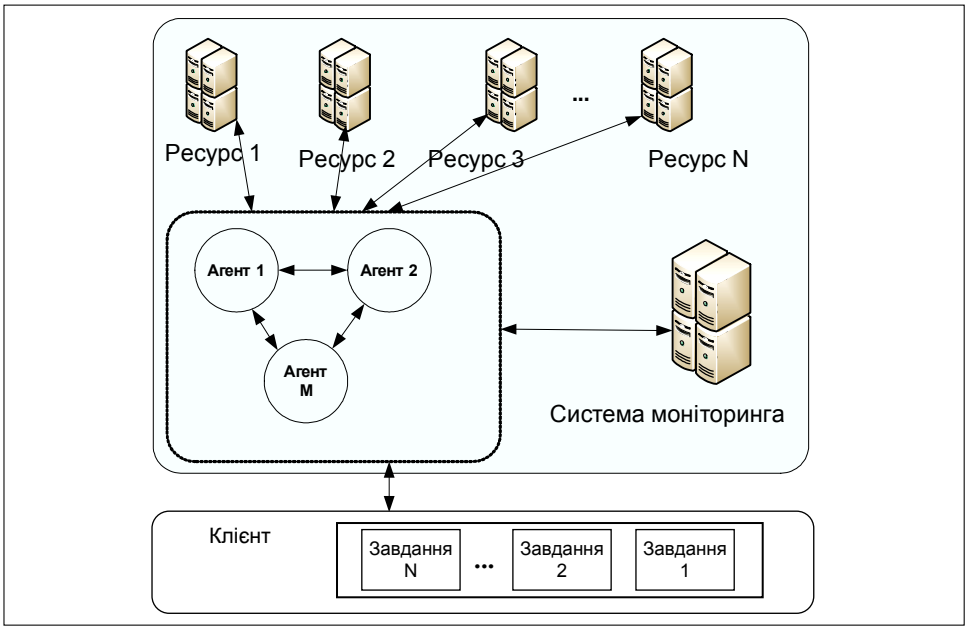

Рис. 5. Архітектура система з агентським брокером

Одним із прикладів реалізації агентської моделі є проміжне забезпечення AliEn [32], яке використовується для обробки даних в експерименті ALICE [33]. За виконання завдання на ресурсі відповідає Job Agent, який стартує на обчислювальному сервері кластера. Він звертається до центральної черги завдань; отримує JDL файл опису завдання, використовуючи сервіс PackMan, встановлює необхідні додатки; контролює виконання завдання та зберігання результатів виконання.

Іншим прикладом може бути Система Управління Ресурсами Агентів [26], в якої кожен агент відповідає за один або декілька ресурсів, що означає, що агенти в основному є постачальниками послуг. Інформація про ресурси, котрі надаються агентами, та зв'язок між ними забезпечується механізмом пошуку ресурсів.

БАЛАНСУВАННЯ НАВАНТАЖЕННЯ ВСЕРЕДИНІ КЛАСТЕРА

Завдання, призначені окремим обчислювальним ресурсам, виконуються на серверах кластера під наглядом локальної системи управління ресурсами, яку в різних джерелах називають по-різному: *JMS* (job management system —

система управління завданнями), *RMS* (resource management system або resource management and scheduling — система управління ресурсами або система управління та планування), *CMS* (cluster management software — система керування кластером) тощо. У ПЗ локальних систем можна виділити два компоненти: менеджер ресурсів і планувальник. Менеджер відповідає за розподіл процесорних вузлів (серверів), а планувальник визначає черговість виконання робіт (рис. 6). В основі багатьох локальних систем управління ресурсами лежить архітектура «клієнт-сервер». Найчастіше реалізується режим batch job (пакетної обробки).

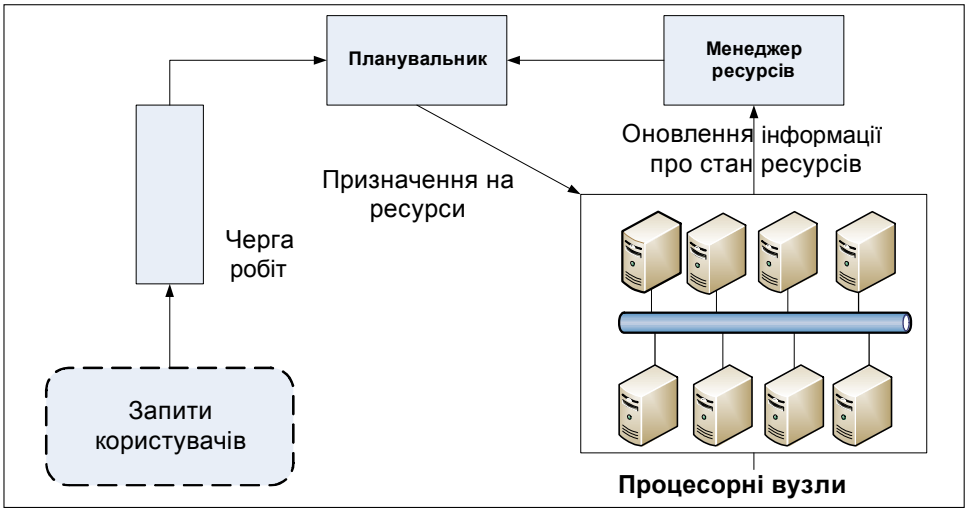

Рис. 6. Типова структура системи управління ресурсами

Під час управління ресурсами використовується *механізм черг* та *планування робіт*, або диспетчеризації. Два принципи організації черг завдань, які найчастіше використовуються, такі: FCFS (first come first serve — «першим прийшов — першим виконується») та LWF (least work first — «найменш трудомістка робота виконується першою»). У LWF трудомісткість може бути виражена через визначення необхідного числа процесорних вузлів та оцінку часу виконання роботи. Черги конфігуруються за допомогою спеціальних атрибутів.

Наприклад, у планувальнику Maui [34] кожна з робіт забезпечується атрибутами *geom* та *dux* [35]. Перший — описує необхідну конфігурацію ресурсів: запис $geom = 2,4$ означає, що для виконання роботи потрібно два вузли, в одному з яких задіюються два, а в іншому чотири процесори. Атрибут *dux* вказує на тривалість виконання завдання. Відповідно зі значеннями атрибутів алгоритм планування визначає черговість завдань і виділяє для них ресурси. Так, з черги може бути обрано завдання, для виконання якого достатньо наявних вільних процесорів.

Локальна система управління ресурсами, як правило, підтримує сукупність черг. Роботи можуть розподілятися за чергами в залежності від спеціальних вимог до ресурсів у запиті (необхідний доступ до високопродуктивних графічних станцій тощо). Ресурсний запит оформлюється у вигляді скрипту команди *qsub*. У цілому диспетчеризація в режимі пакетної обробки дозволяє ефективніше використовувати обчислювальні ресурси шляхом

перерозподілу початку виконання завдання та завантажувати трудомісткими, багатогодинними роботами ресурси, наприклад вночі.

АЛГОРИТМИ БАЛАНСУВАННЯ НАВАНТАЖЕННЯ

Основною метою алгоритмів балансування навантаження є прискорення виконання завдань користувачів на ресурсах, робоче навантаження яких може бути непередбачуваним під час виконання завдання. Отже, важливо встановити метрику, яка визначає робоче навантаження ресурсу. Кожен алгоритм динамічного балансування навантаження має своєчасно оцінювати інформацію про стан обчислювального ресурсу [6], яка зводиться до відповідей на такі питання:

- «Як виміряти робоче навантаження ресурсу?»;
- «Які критерії є основними для визначення цього робочого навантаження?»;
- «Як уникнути негативного впливу динамічності ресурсу на робоче навантаження?»;
- «Як враховувати неоднорідність ресурсів для того, щоб миттєво отримати середній показник робочого навантаження системи?».

Балансування навантаження в розподілених середовищах має враховувати два аспекти: різні обсяги обчислень для різних типів завдань і неоднорідність ресурсів, зокрема, неоднакову продуктивність процесорів. Критерій ефективності використання процесорів можна визначити таким чином. Нехай *i*-й процесор реалізує свій обсяг обчислень w_i за час T_i. Час виконання програми дорівнює

$$T_f = \max\{T_i\},\ i = 1, \ldots, P,$$

де P — число процесорів. Час T_i включає в себе чистий час T_c, що витрачається на обчислення, а також деякий час на обмін даними з базовим процесором. Час T_c визначається обсягом обчислень w_i і продуктивністю процесора $p_i(t_i)$, що залежить від стану (ступеня завантаженості) системи в момент часу t_i:

$$T_{ci} = \frac{w_i}{p_i(t_i)}.$$

Ефективність використання процесорів формально визначається так:

$$RE = \sum_{i=1}^{p} \frac{T_{ci}}{T_f} \times \frac{p_i(t_i)}{\sum_{j=1}^{p} p_j(t_j)}.$$

Тоді, завдання балансування навантаження полягатиме в максимізації значення RE на відрізку часу T_f при заданому числі P процесорів.

Успіх алгоритму розподілу навантаження залежить від кількості параметрів, які входять у формули обчислення, а саме: часу їх вимірювання, часу обчислення робочого навантаження ресурсів та загального часу прийняття рішення за алгоритмом розподілу навантаження. Варто зазначити, що час

реалізації (виконання) алгоритму розподілу навантаження є найважливішим показником для користувача. Важливим є також урахування витрат на передачу завдання між вузлами Грід-інфраструктури, визначених операцією балансування навантаження, однак отримання та верифікація цієї інформації пов'язані з великими труднощами [37].

Алгоритми балансування навантаження можна поділити на дві категорії: статичні і динамічні [40]:

Статичні алгоритми — які розподіляють завдання на обчислювальні ресурси, виходячи з завантаження обчислювальних ресурсів у момент надходження завдання, і в якості критерію балансування вибирається рівномірне навантаження ресурсів (рис. 7). Для статичного планування необхідні методи визначення обсягів робіт w_i для кожного з ресурсів, а також методи оцінки продуктивності $p_i(t_i)$.

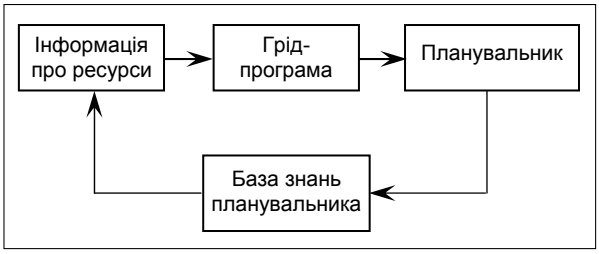

Рис. 7. Статичне балансування навантаження

Перевагою такого роду алгоритму є простота з точки зору програмної реалізації, оскільки немає необхідності постійно контролювати наявне завантаження обчислювальних ресурсів та накопичувати статистику завантаження [40]. До недоліків статичних алгоритмів слід віднести значну похибку в оцінках w_i та $p_i(t_i)$ при різних значеннях продуктивності процесорів обчислювальних ресурсів. Статичні алгоритми працюють добре тільки тоді, коли різниця в навантаженні обчислювальних ресурсів невелика. Корисні реалізації статичних методів балансування навантаження наведені в [16], серед яких слід визначити такі алгоритми:

- Round-Robin Algorithm (алгоритм кругового обслуговування) — завдання розподіляються за обчислювальними ресурсами в послідовному порядку з першого до останнього [39, 40];
- Randomized Algorithm (алгоритм випадкового вибору), коли обчислювальний ресурс для виконання завдання обирається за допомогою випадкового вибору [41];
- Simulated Annealing or Genetic Algorithms (генетичний алгоритм) забезпечує змішану процедуру розподілу завдань за обчислювальними ресурсами з використанням методів оптимізації [42].

До недоліків алгоритмів статичного балансування навантаження слід віднести труднощі з апріорною оцінкою часу виконання завдання.

Динамічні алгоритми вирішують задачу розподілення завдань на основі поточної інформації про завантаження всіх доступних обчислювальних ресурсів (рис. 8).

Як результат динамічні алгоритми виконують розподіл завдань коректніше, оскільки для призначення завдання на обчислювальний ресурс окрім статичної інформації про характеристики ресурсів, яка зберігається в інформаційній системі, використовується поточна інформація про реальне завант-

таження обчислювальних ресурсів, що надходить в інформаційну систему від постачальників у режимі реального часу (надходять усі зміни в значеннях параметрів оцінки стану ресурсів). У літературі [43] визначено три головних чинники, які зазвичай визначають стратегію та вибір конкретного алгоритму балансування завантаження. Ці чинники пов'язано з відповідями на такі основні питання:

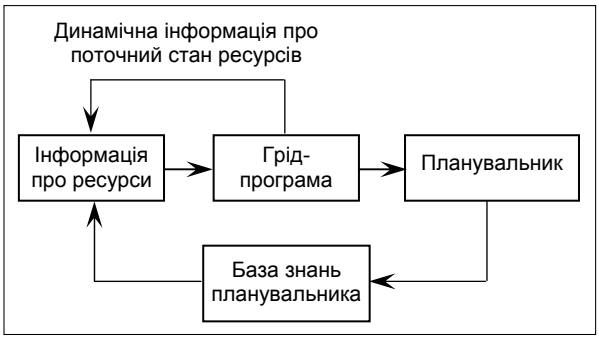

Рис.8. Динамічне балансування навантаження

- «Хто приймає рішення про розподіл завдань за ресурсами?»;
- «Яка інформація використовується для прийняття рішення про розподіл завдань за ресурсами?»;
- «Де приймається рішення в ієрархії структурних елементів проміжного ПЗ Грід-інфраструктури?».

Успішні реалізації динамічних методів балансування навантаження систематизовано в [3, 38], серед яких відзначимо такі:

- Sender-Initiated Strategy (балансування на вимогу клієнта) та Receiver-Initiated Strategy (балансування з серверного боку) — перевантажені ресурси намагаються перенести роботу до ненавантажених ресурсів (випадок балансування з серверного боку) рис. 9 чи, навпаки, ненавантажені ресурси шукають перенавантажені ресурси, з яких можна отримати роботу (випадок балансування на вимогу клієнта) (рис. 10) [45]. Методика балансування на вимогу клієнта ефективніша, ніж методика балансування з серверної сторони для ненавантажених Грід-систем, оскільки в цих умовах вірогідність знайти ненавантажений вузол вища, ніж знайти перенавантажений вузол [45, 46].

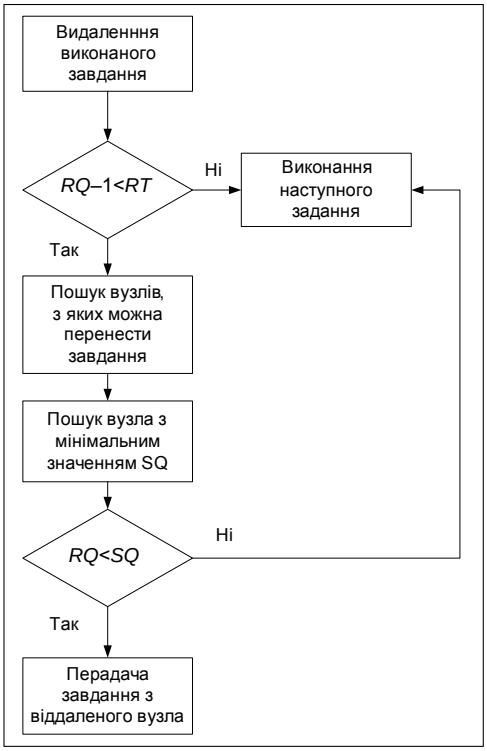

Рис. 9. Балансування з серверної сторони

- *Балансування на вимогу сервера* — обчислювальний ресурс проводить пошук іншого ресурсу, з якого можна перенести завдання у власну чергу і при цьому не перевершити максимальну розмірність черги T із урахуванням параметрів завдань. Якщо остання умова не виконується, то ресурс, що

ініціював пошук, не починає виконувати знайдене завдання і продовжує пошук, поки не будуть знайдено ресурси для завдання, що задовольняють наведеним умовам, або кількість спроб пошуку не досягне певного значення L, після якого пошук зупиняється. Цю стратегію із значенням $T=1$ було досліджено в [47].

- *Резервування* — з одного боку, подібно до стратегії балансування на вимогу клієнта переадресовуються лише нові завдання, а ресурси намагаються зарезервувати для кожного з наявних завдань деяку частину процесорного часу, з іншого — подібно до стратегії балансування з серверного боку ресурс намагається замінити завдання, що виконуються, якщо на ньому в стані виконання залишаються менше ніж T завдань. Обчислювальний ресурс, з якого можна перенести завдання, обирається випадковим чином і перевіряється, чи буде довжина черги на ньому після трансферу деякого завдання нижче певного значення T. Якщо ця вимога не виконується і немає інших резервацій на ініціюючому ресурсу, то для обраного завдання резервується процесорний час і воно передається на ініціюючий ресурс (якщо за цей час не надійшли на виконання інші завдання).

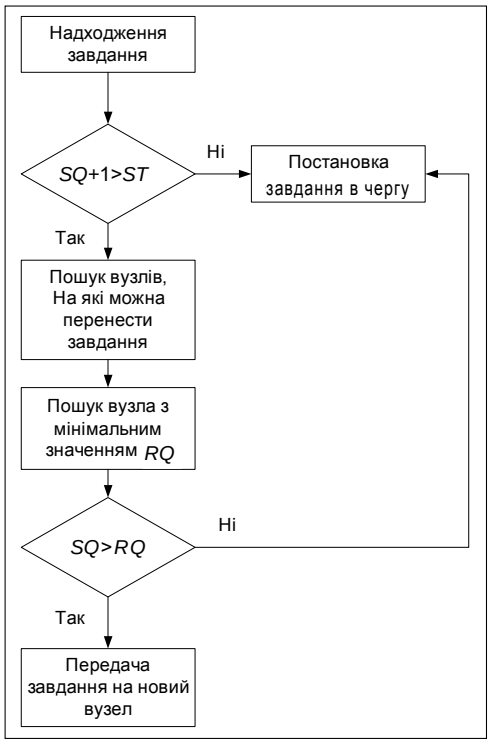

Рис. 10. Балансування на вимогу клієнта

Якщо умови розміру черги на віддаленому ресурсі не виконуються, то продовжується пошук іншого ресурсу, з якого можна перенести деяке завдання, поки не знайдеться відповідний ресурс або кількість спроб пошуку не досягне певного значення L. У цьому випадку ініціюючий ресурс має дочекатись, поки не буде завершено виконання іншого завдання, після чого процес розпочнеться з початку. Дослідження показують, що резервування поступається за продуктивністю двом попереднім підходам [40].

Якщо довжина черги є єдиним показником завантаження ресурсу, то може використовуватися трансферна політика пересилки завдань, яка базується на врахуванні співвідношення довжини вхідної черги (SQ) і її порогу ST після отримання нового завдання. Необхідна інформація про завантаження отримується опитуванням інших ресурсів клієнтом через broadcast-повідомлення про довжини їх черг. Отримавши необхідну інформацію, клієнт може обрати ресурс із найменшою довжиною черги (RQ), де значення RQ задовольняє нерівність $SQ > RQ$ (рис. 10).

- *Симетрично-ініційований алгоритм* — комбінація балансування з серверної та клієнтських сторін. Ресурс може використовувати алгоритми балансування з клієнтської сторони в тих випадках, коли довжина черги перевищує поріг ST, або алгоритми балансування з серверного боку, коли довжина черги знижується до значення, меншого за RT [48].

- *Евристичні алгоритми*, до яких відносяться такі алгоритми, як Min-min, Max-min, Sufferage, Xsufferage [20]. Алгоритм Min-min вибирає завдання з набору завдань, що має мінімальний час виконання на наявних ресурсах, і надсилає його на той ресурс, який за оцінкою забезпечує мінімальний час виконання. Алгоритм Max-min працює аналогічно, але першим опрацьовує завдання з максимальним оціночним часом, що може бути ефективно в наборі завдань, в якому домінують короткі завдання. В алгоритмах Sufferage та Xsufferage замість часу виконання використовується значення, яке розраховується, як різниця двох мінімальних оціночних [20].

ОСОБЛИВОСТІ РЕАЛІЗАЦІЇ АЛГОРИТМІВ БАЛАНСУВАННЯ НАВАНТАЖЕННЯ

Балансування є головною складовою існуючих підходів до динамічного планування, при чому самі підходи відрізняються за своєю точністю та кількістю інформації, необхідної для аналізу [45]. Алгоритм Джоу [46] проводить балансування навантаження, періодично вимагаючи від кожного ресурсу інформувати всі відомі йому ресурси про зміни в його навантаженні. Вілебек, Лем Маір та Рів [47] запропонували чотири політики планування, які динамічно балансують навантаження під час використання інформації про поточний стан інших ресурсів. Обидва підходи — централізований і децентралізований — вимагають досить частого обміну повідомленнями для збору інформації про поточний стан ресурсів і прийняття рішень. У повністю розподіленій системі з N обчислювальними ресурсами для прийняття рішення про планування ресурс має обмінятись $2(N-1)$ повідомленнями [50]. Гібридні алгоритми [45] об'єднують у собі переваги статичних і динамічних стратегій. У гібридних алгоритмах статичний алгоритм вважається «грубим» розподіленням завдань за ресурсами, а динамічний алгоритм — «тонким» регулюванням первинного розподілу. Під час використання тільки статичного алгоритму балансування навантаження може виникнути дисбаланс у Грід-системі. Коли це трапляється, динамічний алгоритм починає свою роботу і забезпечує збалансованість завдань у чергах по всій Грід-системі.

ВИСНОВКИ

Кожна з розглянутих вище моделей побудови архітектури брокера ресурсів має свої переваги та недоліки. Наприклад, широкі можливості функціонального використання централізованої моделі визначаються її масштабованістю та продуктивністю. Натомість, можливості розподіленого брокера звужені. Ієрархічний підхід може бути ефективним для розподілу завдань у Грід-середовищі, але потребує більше часу для визначення необхідних ресурсів. Використання автономних агентів у системі для отримання її поточного стану може ідеально дозволити керувати розподілом ресурсів у ефективний і надійний спосіб, але вимагає побудови чіткої системи зв'язків між ними.

Серед алгоритмів балансування навантаження ресурсів Грід-системи можна розрізнити статичні й динамічні залежно від параметрів, що враховуються під час прийняття рішення про призначення і виконання завдання. У статичному підході використовується середньостатистична інформація

про роботу системи, ігноруючи її поточний стан. Статичні алгоритми [49, 50] частіше застосовуються для балансування періодичних завдань із жорсткими вимогами щодо часу завершення.

Навпаки, динамічні алгоритми [51, 52] виконують балансування в реальному часі і динамічно визначають можливість поліпшення балансування навантаження з врахуванням вже призначених для оброблення розподілених задач [26]. Оскільки серед задач у Грід-середовищі періодичні є меншістю, то доцільніше використовувати динамічні алгоритми.

Під час обговорення згаданих вище брокерських структур і стратегій балансування було з'ясовано, що задача розподілу ресурсів пов'язана не лише з вибором механізму розподілу, але значною мірою визначається інформацією, доступною для брокера, і її актуальністю. Підтримка оновлення інформації про стан ресурсів може бути досить витратною і вартість її зростає разом із розміром системи. Чим ефективніше управління інформаційними потоками в Грід-системі, тим ефективніший процес розподілу ресурсів. Це положення використано авторами для побудови інноваційного брокера для нового покоління проміжного забезпечення NorduGrid, відомого як KnowARC [7], про який йтиметься в наступній статті.

ЛІТЕРАТУРА

1. *Yagoubi B., Slimani Y.* Task Load Balancing Strategy for Grid Computing // Journal of Computer Science . — 2007 — **3**, № 3. — P. 186–194.
2. *Heiss H.-U., Schmitz M.* Decentralized dynamic load balancing: The particles approach // Information Sciences. — 1995. — № 84. — P. 115–128.
3. *Martynov E., Zinovjev G., Svistunov S.* Academic segment of Ukrainian Grid infrastructure // System Research and Information Technologies. — 2009. — № 3. — P. 31–42.
4. *Загородний А., Зиновьев Г., Мартынов Е., Свистунов С.* Украинский академический Грид: Українсько-македонський наук. зб. — 4-ий вип. — Київ: Вид. Нац. б-ка України ім. В.І. Вернадського, 2009. — С. 140–150.
5. *Zagorodny A., Zgurovsky M., Zinovjev G., Petrenko A., Martynov E.* Integrating Ukraine into European Grid Infrastructure // Системні дослідження та інформаційні технології. — 2009. — № 2. — С. 35–49.
6. *Yagoubi B., Slimani Y.* Dynamic load balancing strategy for grid computing // Transactions on Engineering, Computing and Technology. — 2006. — № 13. — P. 260–265.
7. *Nordugrid* ARC. — http://www.nordugrid.org.
8. gLite. — http://glite.web.cern.ch/glite/.
9. UNICORE. — www.unicore.eu/.
10. *Nath R.* Efficient Load Balancing Algorithm in Grid Environment // Master Thesis, Thapar University, Patiala, May 2007. — 63 p.
11. *Петренко А.І.* Національна Grid — інфраструктура для забезпечення наукових досліджень і освіти. — http://netallted.cad.kiev.ua/downloads/Grid.pdf.
12. *Kenthapadi K., Manku G.* Decentralized algorithms using both local and random probes for P2P load balancing // Proceedings of SPAA, Las Vegas, Nevada, 2005. — P. 135–144.
13. *Buyya R., Murshed M*. GridSim: A Toolkit for the Modeling and Simulation of Distributed Resource Management and Scheduling for Grid Computing // Concurrency And Computation: Practice And Experience. — 2002. — **14**, № 13. — P. 1175–1220.

14. *Chapman C., Wilson P.* Condor services for the Global Grid: Interoperability between Condor and OGSA // Proceedings of the 2004 UK e-Science All Hands Meeting. — 2004. — № 2. — P. 870–877. — http://www.cs.wisc.edu/condor/doc/condor-ogsa-2004.pdf.
15. *Buyya R., Abramson D., Giddy J.* Nimrod/G: An architecture for a resource management and scheduling system in a global computational grid // Fourth International Conference on High Performance Computing in Asia — Pacific Region, Beijing, China. — 2000. — P. 283–289.
16. *Casanova H.* et al. The AppLeS Parameter Sweep Template: User Level Middleware for the Grid. — http://www.sc2000.org/techpapr/papers/pap.pap169.pdf.
17. *Venugopal S., Nadiminti K., Gibbins H., Buyya R.* Designing a resource broker for heterogeneous grids // Software: Practice and Experience. — 2008. — **38**, Issue 8. — P. 793–825. — http://www.buyya.com/papers/ gridbus-broker-design-spe.pdf.
18. GridWay. — http://www.gridway.org/.
19. *Ming Wu, Xian-He Sun.* The GHS Grid Scheduling System:Implementation and Performance Comparison. — http://www.cs.iit.edu/~scs/psfiles/GHS-ngs-Sun.pdf.
20. *Berman F., Wolski R., Casanova H.* et al. Adaptive Computing on the Grid Using AppLeS. — http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.58.7606&rep=rep1&type=ps.
21. *Laure E., Hemmer F.* et al. Middleware for the next generation grid infrastructure. — http://cdsweb.cern.ch/record/865715/files/p826.pdf?version=1.
22. *Cecchi M., Capannini F.* et al. The glite workload management system // GPC '09: Proceedings of the 4-th International Conference on Advances in Grid and Pervasive Computing. — Berlin: Heidelberg, Springer-Verlag, 2009. — P. 256–268.
23. *Пономаренко В.С., Листровой С.В., Минухин С.В., Знахар С.В.* Методы и модели планирования ресурсов в GRID-системах. — Х.: ИНЖЭК, 2008. — 408 с.
24. *Foster I., Kesselman C., Tuecke S.* The anatomy of the grid: Enabling scalable virtual organizations // International Journal of High Performance Computing Applications. — 2001. — **15**, № 3. — P. 200–222.
25. *Ying Li, Feng Hong* et al. A framework for price-based resource allocation on the grid // Lecture Notes in Computer Science. — 2005. — **3320/2005**(1). — P. 341–344.
26. *Huedo E., Montero R., Llorente I.* A recursive architecture for hierarchical grid resource management // Future Generation Computer Systems. — 2009. — **25**(4). — P. 401–405.
27. *Cao J., Jarvis S.* et al. Arms: An agent-based resource management system for grid computing // Scientific Programming. — 2002. — **10**(2). — P. 135–148.
28. *Cao J., Spooner D.* et al. Grid load balancing using intelligent agents // Future Generation Computer Systems. — 2005. — **21**, № 1. — P. 135–149.
29. *Dugenie P., Cerri S., Duvert F., Jonquet C.* Agent-grid integration ontology // Lecture Notes in Computer Science. — 2006. — **4277/2006**(1). — P. 136–146.
30. *Ganzha M., Gawinecki M.* et al. Agents as resource brokers in grids forming agent teams // Future Generation Computer Systems. — 2008. — **4818/2008**(1). — P. 489–491.
31. *Buncic P.* et al. The AliEn system, status and perspectives // Computing in High Energy and Nuclear Physics, 24–28 March, La Jolla, California, Proceedings of CHEP. — 2003. — http://www.slac.stanford.edu/econf/C0303241/proc/papers/MOAT004.PDF.
32. *Betev L.* et al. The ALICE physics data challenge and the ALICE distributed analysis prototype. — http://indico.cern.ch/getFile.py/access?contribId=506&sessionId=9&resId=2&materialId=paper&confId=0.
33. *Maui* cluster. — http://www.clusterresources.com/products/maui-cluster-scheduler. php.
34. *Аветисян А.И., Гайсарян С.С., Грушин Д.А., Кузюрин Н.Н., Шокуров А.В.* Эвристики распределения задач для брокера ресурсов Grid // Тр. Ин-та системного программирования РАН. — 2004. — Т. 5. — С. 269–280.

35. *Веб-сайт* СпбГУ, Факультет Информационных технологий и программирования, кафедра Компьютерных технологий. — http://rain.ifmo.ru/cat/view.php/theory/algorithm-analysis/np-completeness-2004.
36. *Leinberger W., Karypis G., Kumar V., Biswas R.* Load balancing across near-homogeneous multi-resource servers // 9-th Heterogeneous Computing Workshop. — 2000. — P. 60–71.
37. *Javier Bustos-Jimenez*. Robin hood: An active objects load balancing mechanism for intranet. — http://www.dcc.uchile.cl/~jbustos/Pub/rh.pdf.
38. *Abdur Razzaque1 Md., Hong C.S.* Dynamic Load Balancing in Distributed System: An Efficient Approach. — http://networking.khu.ac.kr/publications/data/ Dynamic%20Load%20 Balancing%20in%20 Distributed%20System%20An% 20 Efficient%20Approach.pdf.
39. *Ramos J., Sang J.* Simulation of large scale networks III: an improved computational algorithm for round-robin service // Winter Simulation Conference. — New Orleans, Louisiana, USA. — 2003. — P. 721–728.
40. *Топорков В.В.* Модели распределенных вычислений. — М.: ФИЗМАТЛИТ, 2004. — 320 с.
41. *Leland R., Hendrickson B.* An empirical study of static load balancing algorithms // Proceedings of the Scalable High-Performance Computing Conference. — 1994. — P. 682–685.
42. *Subrata R., Zomaya A.Y., Landfeldt B.* Artificial life techniques for load balancing in computational Grids // Journal of Computer and System Sciences. — 2007. — **73**, № 8. — P. 1176–1190.
43. *Zaki M., Li W., Parthasarathy S.* Customized Dynamic Load Balancing for a Network of Workstations // Proc. of 5-th IEEE International Symposium on High Performance Distributed Computing, 6–9 August, Syracuse, NY, USA, 1996. — P. 282–291.
44. *Taibi T., Abid A., Azahan E.F.E.* A Comparison of Dynamic Load Balancing Algorithms // Jordan Journal of Applied Sciences. — 2007. — **9**, № 2. — P. 125–133.
45. *Derek L. Eager.* A Comparison of Receiver-Initiated and Sender-Initiated Adaptive Load Sharing. — http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.77.4986&rep=rep1 &type=pdf.
46. *Livny M., Melman M.* Load Balancing in Homogeneous Broadcast Distributed Systems // ACM Computer Network Performance Symposium, April. — 1982. — P. 47–55.
47. *Malik S.* Dynamic Load Balancing in a Network of Workstations. // Research Report. — 2000. — www.cs.toronto.edu/~smalik/downloads/paper_515.pdf.
48. *Caminero A., Rana O., Caminero B., Carrión C.* Autonomic Network-Aware Metascheduling for Grids: A Comprehensive Evaluation // Advances in Grid Computing, Febrary. — 2011. — P. 49–72. — http://www.intechopen.com/source/pdfs/13943/InTech-Autonomic_network_aware_metascheduling_for_grids_a_comprehensive_ evaluation.pdf.
49. *Lo V.M.* Heuristic Algorithms for Task Assignments in Distributed Systems // IEEE Transactions on Computers. — 1988. — **37**, № 11. — P. 1384–1397.
50. *Bokhari S.* Partitioning Problems in Parallel, Pipelined, and Distributed Computing // IEEE Transactions Computers. — 1988. — **37**, № 1. — P. 48–57.
51. *Cavanaugh C., Radhika A.* Dynamic Resource Management Algorithm for a Distributed Real-time System // Proceedings of the 19-th IEEE International Parallel and Distributed Processing Symposium, 4–8 April. — Denver, Cjljrado, USA. — 2005. — P. 131.
52. *Zeng Z., Veeravalli B.* Divisible load scheduling on arbitrary distributed networks via virtual routing approach // Proceedings of 10-th International Conference on Parallel and Distributed Systems, 7–9 July. — Newport Beach, CA, USA. — 2004. — P. 161.

Надійшла 05.03.2011

УДК 621.391:519.2

НЕАСИМПТОТИЧЕСКИЕ ОЦЕНКИ ЭФФЕКТИВНОСТИ СЛУЧАЙНОГО КОДИРОВАНИЯ В СИСТЕМЕ ПЕРЕДАЧИ ИНФОРМАЦИИ ПО ДВОИЧНОМУ СИММЕТРИЧНОМУ КАНАЛУ СВЯЗИ С ОТВОДОМ

А.Н. АЛЕКСЕЙЧУК, С.В. ГРИШАКОВ

Исследована эффективность случайного кодирования при многократной передаче безизбыточных сообщений по двоичному симметричному каналу связи с отводом. Получены неасимптотические оценки надежности восстановления сообщений законным получателем информации и стойкости их защиты в отводном канале.

ВВЕДЕНИЕ

Математическая модель системы передачи информации по каналу связи с отводом предложена в [1] и в дальнейшем изучалась в [2–11] и ряде других работ. В настоящее время эта модель широко используется при разработке алгоритмических методов защиты информации от утечки по побочным каналам связи, квантово-криптографических протоколов распределения ключей, схем разделения секрета, а также при решении других криптографических задач [10, 11].

Традиционная система передачи информации по каналу связи с отводом состоит из двух статистически независимых каналов связи с общим входом. Выход одного (основного) канала наблюдает законный получатель, а выход другого (отводного) канала — противник. Для повышения стойкости защиты информации в отводном канале применяют случайное кодирование, при котором входные сообщения преобразуются в дискретные сигналы, выбираемые случайно и равновероятно из подходящих конечных множеств. В большинстве доступных публикаций, посвященных исследованию эффективности случайного кодирования, рассматривается случай, в котором основной канал связи является идеальным (не имеет помех), а отводной — двоичным симметричным каналом или каналом со стиранием [3, 5–11]. В частности, в [8, 9] были исследованы асимптотические свойства характеристик эффективности случайного кодирования при многократной передаче безизбыточного случайного сообщения по идеальному основному и двоичному симметричному отводному каналам связи, а также получены асимптотические оценки минимального количества передач, необходимого для восстановления противником переданного сообщения с заданной надежностью. Вместе с тем, асимптотический вид оценок, полученных в [8, 9], не позволяет применять их непосредственно к системам со случайным кодированием кодами фиксированной длины.

© *А.Н. Алексейчук, С.В. Гришаков*, 2011

Цель работы — исследовать эффективность случайного кодирования при многократной передаче сообщений по неидеальному основному каналу связи. В работе получены неасимптотические оценки вероятности правильного восстановления сообщений законным получателем и, соответственно, противником. Показано, что при определенных соотношениях между параметрами линейных кодов, применяемых для случайного кодирования, и вероятностями искажений в основном и отводном каналах можно обеспечить надежный прием сообщений законным получателем при высокой практической стойкости их защиты в отводе, используя сравнительно небольшое число передач.

ПОСТАНОВКА ЗАДАЧИ И ОСНОВНЫЕ РЕЗУЛЬТАТЫ

Рассмотрим систему передачи информации по каналу связи с отводом, состоящую из источника и двух статистически независимых каналов связи с общим входом: основного канала — от источника к законному получателю, и отводного канала — от источника к противнику [1, 2]. Предположим, что основной канал является двоичным симметричным каналом с вероятностью ошибки p_1, а отводной — двоичным симметричным каналом с вероятностью ошибки p_2, где $0 < p_1 < p_2 < 0{,}5$. Источник вырабатывает случайное сообщение S с равномерным распределением на множестве $V_k = \{0,1\}^k$, для передачи которого отправитель выбирает сюръективное линейное отображение $\sigma : V_n \to V_k$ и генерирует t независимых случайных векторов $X_1, ..., X_t$ с равномерным распределением на множестве $C_s = \sigma^{-1}(s)$, где $s \in V_k$ — значение случайного вектора S. Затем последовательность $X_1, ..., X_t$ передается по основному и отводному каналам связи. При этом она искажается и в основном канале преобразовывается в случайную последовательность $Y_1, ..., Y_t$, а в отводном канале — в случайную последовательность $Z_1, ..., Z_t$. Необходимо получить оценки вероятности правильного восстановления сообщения S по каждой из указанных последовательностей в предположении, что декодеры основного и отводного каналов строятся на основе тех или иных практически реализуемых алгоритмов.

Предположим, что законный получатель информации применяет следующий алгоритм декодирования по методу максимума правдоподобия.

Алгоритм 1. Пусть $y = y_1, ..., y_t$ — наблюдаемая реализация случайной последовательности $Y_1, ..., Y_t$. Тогда получатель находит последовательность $\sigma(y) = \sigma(y_1), ..., \sigma(y_t)$ и выбирает в качестве оценки искомого значения случайного вектора S произвольный элемент $s^* = s^*(y)$ с наибольшей частотой встречаемости в последовательности $\sigma(y)$.

Если противнику известен некоторый критерий отбраковки ложных значений сообщения S (например, S является ключом симметричной криптосистемы, и существует возможность его опробования), то алгоритм 1 можно модифицировать следующим образом.

Алгоритм 2. Пусть $z = z_1, ..., z_t$ — наблюдаемая реализация случайной последовательности $Z_1, ..., Z_t$. Тогда противник находит последовательность $\sigma(z) = \sigma(z_1),..., \sigma(z_t)$, составляет список ее членов, расположенных в порядке убывания их частот, и опробует в указанном порядке (до первого успеха) векторы из списка.

Обозначим $\lambda_1 = \lambda_1(t, p_1)$ и $\lambda_2 = \lambda_2(t, p_2)$ вероятности правильного восстановления сообщения S с использованием алгоритмов 1 и 2 соответственно. Для того, чтобы привести оценки параметров λ_1, λ_2, введем ряд дополнительных обозначений.

Обозначим C линейный $(n, n-k)$-код, равный ядру отображения σ. Для любого $p \in [0, 1]$ положим

$$\pi(C_s; p) = \sum_{u \in C_s} p^{\|u\|} (1-p)^{n-\|u\|}, \qquad (1)$$

где $C_s = \sigma^{-1}(s)$ — смежный класс (СК) кода C, соответствующий вектору $s \in V_k$;

$$\pi_0(C; p) = \pi(C_0; p),\ \pi_1(C; p) = \max\{\pi(C_s; p) : s \in V_k \setminus \{0\}\}. \qquad (2)$$

Отметим, что на основании тождества Мак-Вильямс [12]

$$\pi(C_s; p) = 2^{-k} \sum_{u \in C^{\perp}} (-1)^{au} \Delta^{\|u\|},\ s \in V_k, \qquad (3)$$

где $C^{\perp}$ — код, дуальный к C; a — произвольный вектор, принадлежащий СК C_s; au — булево скалярное произведение векторов $a, u \in V_n$, $\Delta = 1 - 2p$. Отсюда непосредственно следует, что $\pi_1(C; p) < \pi_0(C; p)$ для любого $0 < p < 0{,}5$.

Утверждение 1. Справедливы неравенства

$$\lambda_1(t, p_1) \geq 1 - 2^k \exp\left\{-\frac{t}{2}(\pi_0(C; p_1) - \pi_1(C; p_1))^2\right\}, \qquad (4)$$

$$\lambda_2(t, p_2) \leq 1 - (1 - \pi_0(C; p_2))^t. \qquad (5)$$

Доказательство. Как следует из описания алгоритма 2, вероятность его успешного завершения не превосходит вероятности события, состоящего в том, что случайный вектор S присутствует в последовательности $\sigma(Z)$. Последнее равносильно тому, что хотя бы один из векторов искажений, переводящих слово X_i в слово Z_i в отводном канале, принадлежит коду C, $i = \overline{1, t}$. Отсюда вытекает справедливость неравенства (5).

Докажем неравенство (4). Обозначим P_s условное распределение вероятностей на множестве значений случайной последовательности $Y = Y_1, ..., Y_t$ при условии $S = s$; $n_s(Y)$ — частоту появления вектора s в последовательности $\sigma(Y)$. Справедливо равенство

$$\lambda_1(t, p_1) = 2^{-k} \sum_{s \in V_k} P_s\{s^*(Y) = s\}\,. \qquad (6)$$

Заметим теперь, что для любого $s \in V_k$ событие $\{s^*(Y) \neq s\}$ влечет событие $\bigcup_{s' \neq s} \{n_{s'}(Y) > n_s(Y)\}$. Следовательно, для любого $A > 0$ выполняются следующие неравенства:

$$P_s\{s^*(Y) \neq s\} \le P_s\left(\bigcup_{s' \neq s} \{n_{s'}(Y) > n_s(Y)\}\right) \le$$

$$\le P_s\{n_s(Y) \le A\} + \sum_{s' \neq s} P_s\{n_{s'}(Y) > A\}\,. \qquad (7)$$

Далее, при выполнении условия $S = s$ случайная величина $n_s(Y)$ имеет биномиальное распределение с параметрами $(t, \pi_0(C; p_1))$, а случайная величина $n_{s'}(Y)$ — биномиальное распределение с параметрами $(t, \pi(C_{s \oplus s'}; p_1))$. Следовательно, в силу неравенств для вероятностей больших уклонений [13] при выполнении условия

$$\pi_1(C; p_1) < At^{-1} < \pi_0(C; p_1) \qquad (8)$$

для любых $s, s' \in V_k$, $s' \neq s$, справедливы следующие неравенства:

$$P_s\{n_s(Y) \le A\} \le \exp\{-2t(At^{-1} - \pi_0(C; p_1))^2\}\,, \qquad (9)$$

$$P_s\{n_{s'}(Y) > A\} \le \exp\{-2t(At^{-1} - \pi(C_{s \oplus s'}; p_1))^2\}\,. \qquad (10)$$

Подставляя оценки (9), (10) в формулу (7) и принимая во внимание второе соотношение (2), получим, что при выполнении условия (8)

$$P_s\{s^*(Y) \neq s\} \le \exp\{-2t(At^{-1} - \pi_0(C; p_1))^2\} +$$

$$+ (2^k - 1)\exp\{-2t(At^{-1} - \pi_1(C; p_1))^2\}\,. \qquad (11)$$

Наконец, полагая в формуле (11) $At^{-1} = \frac{1}{2}(\pi_0(C; p_1) + \pi_1(C; p_1))$, на основании равенства (6) получим соотношение

$$\lambda_1(t, p_1) \ge 1 - 2^k \exp\left\{-2t\left(\frac{\pi_0(C; p_1) - \pi_1(C; p_1)}{2}\right)^2\right\},$$

которое равносильно неравенству (4).

Утверждение доказано.

Отметим, что соотношения (4), (5) справедливы для любого линейного $(n, n-k)$-кода C. Чтобы оценивать с их помощью эффективность систем со случайным кодированием, построенных на основе конкретных кодов, необходимо иметь пригодные для практических вычислений границы парамет-

ров (2). Ряд таких границ, вытекающих, в основном, из результатов работ [7, 14–16], содержит следующее утверждение.

Утверждение 2. Пусть C — произвольный линейный $(n, n-k)$-код с минимальным расстоянием d и дуальным расстоянием d'. Тогда для любого $0 < p < 0{,}5$ справедливы следующие неравенства:

$$\pi_0(C; p) \le 2^{-k}(1 + (2^k - 1)\Delta^{d'}), \tag{12}$$

$$\pi_0(C; p) \le 2^{-\frac{kp}{1-p}}, \tag{13}$$

$$\pi_0(C; p) \le 2^{-k} + (1-\widetilde{p})^{-n}\begin{pmatrix}\left\lceil \frac{n}{2} \right\rceil + l \\ l\end{pmatrix}^{-1} \sum_{i=2l+1}^{n} \binom{n}{i} \widetilde{p}^{\,i}(1-\widetilde{p})^{n-i}, \tag{14}$$

где

$$\Delta = 1 - 2p,\ l = \left\lfloor \frac{d'-1}{2} \right\rfloor,\ \widetilde{p} = \frac{1-2p}{2(1-p)}. \tag{15}$$

Кроме того, справедливо неравенство

$$\pi_0(C; p) - \pi_1(C; p) \ge 2(1-p)^n \frac{\Delta^{n-k+1}}{1 + \Delta^{n-k+1}}, \tag{16}$$

а при условии $d \ge 3$ — неравенство

$$\pi_0(C; p) - \pi_1(C; p) \ge \Delta^{\frac{n+1}{2}}. \tag{17}$$

Обе границы (12), (17) достигаются, если C является кодом Хэмминга.

Доказательство. Справедливость оценки (12) и ее достижимость для кодов Хэмминга доказаны в [7]. Неравенство (13) следует из формулы

$$\pi_0(C; p) = (1-p)^n + P_{\text{но}}(C), \tag{18}$$

где

$$P_{\text{но}}(C; p) = \sum_{x \in C \setminus \{0\}} p^{\|x\|}(1-p)^{n-\|x\|} \tag{19}$$

есть вероятность необнаружения ошибки кодом C, и соотношения

$$P_{\text{но}}(C; p) \le \left(2^{-n} \mid C \mid\right)^{\frac{p}{1-p}} - (1-p)^n,$$

полученного в работе [16].

Для доказательства формулы (14) воспользуемся неравенством [15]

$$P_{\text{но}}(G; \widetilde{p}) \le \begin{pmatrix}\left\lceil \frac{n}{2} \right\rceil + l \\ l\end{pmatrix}^{-1} \sum_{i=2l+1}^{n} \binom{n}{i} \widetilde{p}^{\,i}(1-\widetilde{p})^{n-i},\ 0 < \widetilde{p} < 0{,}5 \tag{20}$$

справедливым для любого двоичного линейного кода G длины n, исправляющего l ошибок. Полагая в формуле (20) $G = C^{\perp}$, при выполнении (15) получим следующее неравенство:

$$P_{\text{но}}(C^{\perp};\widetilde{p}) \le \left(\begin{matrix}\left\lceil \frac{n}{2} \right\rceil + l \\ l\end{matrix}\right)^{-1} \sum_{i=2l+1}^{n} \binom{n}{i} \widetilde{p}^{\,i} (1-\widetilde{p})^{n-i}. \qquad (21)$$

Далее, согласно формуле (3) при $s=0$, выполняется равенство $\pi_0(C;p) = 2^{-k} \sum_{x\in C^{\perp}} \Delta^{\|x\|}$. Следовательно,

$$\pi_0(C;p) = 2^{-k} + \sum_{x\in C^{\perp}\setminus\{0\}} \Delta^{\|x\|} = 2^{-k} + \sum_{x\in C^{\perp}\setminus\{0\}} \left(\frac{\widetilde{p}}{1-\widetilde{p}}\right)^{\|x\|} =$$

$$= 2^{-k} + (1-\widetilde{p})^{-n} \sum_{x\in C^{\perp}\setminus\{0\}} \widetilde{p}^{\,\|x\|} (1-\widetilde{p})^{n-\|x\|} = 2^{-k} + (1-\widetilde{p})^{-n} P_{\text{но}}(C^{\perp};\widetilde{p}). \qquad (22)$$

Подставляя оценку (21) в формулу (22), получим неравенство (14).

Для доказательства неравенства (16) воспользуемся следующей оценкой [14]:

$$\frac{\pi_0(C;p)}{\pi_1(C;p)} \ge \frac{1+\Delta^{n-k+1}}{1-\Delta^{n-k+1}}. \qquad (23)$$

На основании (23) справедливы соотношения

$$\pi_0(C;p) - \pi_1(C;p) \ge \pi_0(C;p)\left(1 - \frac{1-\Delta^{n-k+1}}{1+\Delta^{n-k+1}}\right) = 2\pi_0(C;p)\frac{\Delta^{n-k+1}}{1+\Delta^{n-k+1}},$$

из которых в силу неравенства $\pi_0(C;p) \ge (1-p)^n$, вытекающего из формулы (18), следует оценка (16).

Убедимся, наконец, в справедливости неравенства (17). Зафиксируем элемент $s \in V_k$ такой, что $\pi_1(C;p) = \pi(C_s;p)$, и вектор $a \in C_s$. Обозначим

$$G = C^{\perp},\ G_0 = \{u \in G : ua = 0\},\ G_1 = \{u \in G : ua = 1\}.$$

На основании формулы (3) и неравенства между средним арифметическим и средним геометрическим справедливы соотношения

$$\pi_0(C;p) - \pi_1(C;p) = 2^{-k} \sum_{u\in G} \Delta^{\|u\|} - 2^{-k} \sum_{u\in G} (-1)^{au} \Delta^{\|u\|} =$$

$$= 2^{-(k-1)} \sum_{u\in G_1} \Delta^{\|u\|} \ge \Delta^{\,2^{-(k-1)} \sum_{u\in G_1} \|u\|}.$$

Следовательно, для обоснования оценки (17) достаточно показать, что при $d \ge 3$ выполняется следующее неравенство:

$$2^{-(k-1)} \sum_{u\in G_1} \| u \| \le \frac{n+1}{2}. \qquad (24)$$

Для любого $H \in \{G, G_0, G_1\}$ обозначим $s_i(H) = |\{u = (u_1,...,u_n) \in$ $\in H : u_i = 1\}|$, $i = \overline{1,n}$. Заметим, что

$$2^{-(k-1)} \sum_{x \in G_1} \| u \| = 2^{-(k-1)} \sum_{i=1}^{n} s_i(G_1). \qquad (25)$$

При этом, поскольку G и G_0 являются линейными кодами размерности k и $k-1$ соответственно, а G_1 — смежным классом кода G_0, то

$$s_i(G) \in \{0, 2^{k-1}\},\ s_i(G_0) \in \{0, 2^{k-2}\},\ s_i(G_1) \in \{0, 2^{k-2}, 2^{k-1}\},\ i = \overline{1,n}. \quad (26)$$

Предположим, что существует два различных значения $i, j \in \{1, 2, \dots, n\}$ таких, что $s_i(G_1) = s_j(G_1) = 2^{k-1}$. Тогда $s_i(G_0) =$ $= s_j(G_0) = 0$, согласно первым двум соотношениям (26). Следовательно, для любого слова $u = (u_1,...,u_n) \in G$ выполняется равенство $u_i = u_j$. Но в таком случае дуальное расстояние кода G (равное минимальному расстоянию d кода $C = G^{\perp}$) не превосходит 2, что противоречит условию утверждения.

Итак, существует не более одного значения $i \in \{1, 2, ..., n\}$, для которого $s_i(G_1) = 2^{k-1}$. Для остальных же значений $j \neq i$ в силу третьего соотношения (26) $s_j(G_1) \le 2^{k-2}$. Отсюда следует, что $2^{-(k-1)} \sum_{i=1}^{n} s_i(G) \le$ $\le 2^{-(k-1)}(2^{k-1} + 2^{k-2}(n-1)) = \frac{n+1}{2}$, из чего следует, что на основании формулы (25) справедливо неравенство (24), что и требовалось доказать.

Достижимость оценки (17) для $(n, n-k)$-кода Хэмминга C, $n = 2^k - 1$, вытекает из следующих равенств [3]:

$$\pi_0(C; p) = 2^{-k}(1 + (2^k - 1)\Delta^{2^{k-1}}),\ \pi_1(C; p) = 2^{-k}(1 - \Delta^{2^{k-1}}).$$

Утверждение доказано.

ЧИСЛЕННЫЕ ОЦЕНКИ ХАРАКТЕРИСТИК ЭФФЕКТИВНОСТИ СИСТЕМ ПЕРЕДАЧИ ИНФОРМАЦИИ ПО КАНАЛУ СВЯЗИ С ОТВОДОМ

Для сравнения полученных оценок были проведены численные расчеты значений в правых частях неравенств (12)–(14), (16), (17). Расчеты проводились для различных линейных кодов C, имеющих длину от нескольких десятков до нескольких сотен битов.

В качестве типового примера, иллюстрирующего полученные результаты, приведем оценки параметров $-\log_2(\pi_0(C; p_1) - \pi_1(C; p_1))$ и $-\log_2 \pi_0(C; p_2)$, рассчитанные для двух линейных кодов $C = G^{\perp}$ (табл. 1).

В табл. 1 после символа G указаны значения параметров n, k и d', равных соответственно длине, размерности и минимальному расстоянию кода G; символ «–» показывает, что соответствующая верхняя граница параметра $\pi_0(C; p_2)$ является тривиальной, то есть превосходит число 1.

Таблица 1. Оценки вероятности правильного восстановления сообщений в основном и отводном каналах

Код $G(63,45,7)$				
p_1	10^{-2}	10^{-4}		10^{-8}
Неравенство (16)	1,217	0,012		$1{,}2 \cdot 10^{-6}$
Неравенство (17)	0,933	0,009		$0{,}9 \cdot 10^{-6}$
p_2	0,1	0,2	0,3	0,4
Неравенство (12)	2,253	5,159	9,253	16,253
Неравенство (13)	5,000	11,250	19,286	30,000
Неравенство (14)	–	–	–	–
Код $G(256,27,106)$				
p_1	10^{-2}	10^{-4}		10^{-8}
Неравенство (16)	9,429	0,071		$0{,}7 \cdot 10^{-5}$
Неравенство (17)	3,745	0,037		$0{,}4 \cdot 10^{-5}$
p_2	0,1	0,2	0,3	0,4
Неравенство (12)	26,990	27,000	27,000	27,000
Неравенство (13)	3,000	6,750	11,571	18,000
Неравенство (14)	–	–	27,000	27,000

Как видно из табл. 1, неравенство (17) в обоих случаях доставляет более точное приближение параметра $\pi_0(C; p_1) - \pi_1(C; p_1)$ по сравнению с формулой (16). Для первого из указанных кодов (имеющего относительно большую скорость передачи $k/n = 45/63$) более точную аппроксимацию параметра $\pi_0(C; p_2)$ обеспечивает оценка (13), а для второго (имеющего относительно малую скорость $k/n = 27/256$) — оценка (12). При этом во втором случае значения $\pi_0(C; p_2)$ быстро приближаются к экстремальному значению 2^{-k} с ростом p_2 (например, при $p_2 = 0{,}1$ значение параметра $\pi_0(C; p_2)$ заключено в пределах от 2^{-27} до $2^{-26{,}990}$). Аналогичное поведение верхних границ (12)–(14) наблюдается для ряда других линейных кодов C, имеющих достаточно большие длину и дуальное расстояние.

В табл. 2 приведены численные значения параметров, характеризующих эффективность систем со случайным кодированием кодами $C_i = G_i^{\perp}$, $i = \overline{1,7}$. Как и в табл. 1, тут после символа G_i указаны значения параметров n, k и d', равных соответственно длине, размерности и минимальному расстоянию кода G_i, $i = \overline{1,7}$.

Таблица 2. Оценки эффективности систем передачи информации по каналу связи с отводом

Код $G(63,45,7)$												
p_1	10^{-2}				10^{-4}				10^{-8}			
t	245				68				67			
p_2	0,1	0,2	0,3	0,4	0,1	0,2	0,3	0,4	0,1	0,2	0,3	0,4
$\hat{\lambda}_2(t,p_2)$	0,0006	3,39	11,35	22,06	0,18	5,18	13,20	23,91	0,18	5,20	13,22	23,93
Код $G_2(63,51,5)$												
p_1	10^{-2}				10^{-4}				10^{-8}			
t	275				77				76			
p_2	0,1	0,2	0,3	0,4	0,1	0,2	0,3	0,4	0,1	0,2	0,3	0,4
$\hat{\lambda}_2(t,p_2)$	0,006	4,67	13,75	25,90	0,35	6,49	15,59	27,73	0,36	6,51	15,61	27,75
Код $G_3(256,203,84)$												
p_1	10^{-2}				10^{-4}				10^{-8}			
t	51440				302				287			
p_2	0,1	0,2	0,3	0,4	0,1	0,2	0,3	0,4	0,1	0,2	0,3	0,4
$\hat{\lambda}_2(t,p_2)$	11,39	46,25	95,39	179,39	18,80	53,67	102,80	186,80	18,88	53,74	102,88	186,87
Код $G_4(256,229,106)$												
p_1	10^{-2}				10^{-4}				10^{-8}			
t	57923				340				323			
p_2	0,1	0,2	0,3	0,4	0,1	0,2	0,3	0,4	0,1	0,2	0,3	0,4
$\hat{\lambda}_2(t,p_2)$	18,30	62,30	124,30	213,18	25,71	69,71	131,71	220,59	25,79	69,78	131,79	220,66
Код $G_5(255,247,3)$												
p_1	10^{-2}				10^{-4}				10^{-8}			
t	61162				366				348			
p_2	0,1	0,2	0,3	0,4	0,1	0,2	0,3	0,4	0,1	0,2	0,3	0,4
$\hat{\lambda}_2(t,p_2)$	11,54	45,85	89,96	148,77	18,93	53,23	97,34	156,15	19,00	53,31	97,41	156,22
Код $G_6(256,53,84)$												
p_1	10^{-2}				10^{-4}				10^{-8}			
t	14043				83				79			
p_2	0,1	0,2	0,3	0,4	0,1	0,2	0,3	0,4	0,1	0,2	0,3	0,4
$\hat{\lambda}_2(t,p_2)$	13,26	39,22	39,22	39,22	20,67	46,62	46,62	46,62	20,73	46,70	46,70	46,70
Код $G_7(256,27,106)$												
p_1	10^{-2}				10^{-4}				10^{-8}			
t	7560				45				43			
p_2	0,1	0,2	0,3	0,4	0,1	0,2	0,3	0,4	0,1	0,2	0,3	0,4
$\hat{\lambda}_2(t,p_2)$	14,11	14,12	14,12	14,12	21,50	21,51	21,51	21,51	21,56	21,57	21,57	21,57

При этом G_1 и G_2 являются кодами Боуза-Чоудхури-Хоквингема, а G_5 — кодом Хэмминга [12]. Параметры остальных линейных кодов взяты из таблиц, приведенных в [17] (отметим, что G_1 и G_7 совпадают с кодами, указанными в табл. 1). Символы p_1 и p_2 в табл. 2 обозначают вероятности ошибок в основном и в отводном каналах связи соответственно, а параметры $t = t(p_1)$ и $\hat{\lambda}_2(t, p_2)$ определяются следующим образом.

Обозначим $\mu_1(C; p_1)$ наибольшее из двух значений в правых частях неравенств (16) и (17) соответственно при $\Delta = 1 - 2p_1$; $\mu_2(C; p_2)$ — наименьшее из трех значений в правых частях неравенств (12), (13), (14) соответственно при $p = p_2$. Тогда t есть наименьшее натуральное число, удовлетворяющее условию

$$1 - 2^k \exp\left\{-\frac{t}{2}\mu_1(C; p_1)^2\right\} \geq 0{,}9, \tag{27}$$

а значение $\hat{\lambda}_2(t, p_2)$ определяется по формуле

$$\hat{\lambda}_2(t, p_2) = -\log_2(1 - (1 - \mu_2(C; p_2))^t), \tag{28}$$

где $C = G_i^{\perp}$, $i = \overline{1,7}$. Отметим, что на основании неравенств (4) и (27) вероятность правильного восстановления сообщения S законным получателем рассматриваемой системы передачи информации по каналу связи с отводом не меньше $0{,}9$. При этом, согласно формулам (5) и (28), вероятность правильного восстановления этого сообщения противником не превосходит $2^{-\hat{\lambda}_2(t, p_2)}$.

ВЫВОДЫ

В целом, полученные результаты позволяют сделать вывод о том, что при заметном отличии между вероятностями ошибок в основном и отводном каналах можно обеспечить достаточно надежный прием сообщений законным получателем при высокой практической стойкости их защиты в отводном канале, используя сравнительно умеренное число передач t. Так, в системе со случайным кодированием кодом $C_4 = G_4^{\perp}$ и вероятностями ошибок $p_1 = 0{,}01$, $p_2 = 0{,}3$ для передачи сообщения S длины 229 бит достаточно сформировать случайную последовательность $X_1, ..., X_t$, состоящую из $t = 57923$ слов, каждое из которых имеет длину 256 битов. Законный получатель, используя алгоритм 1, сможет восстановить S с вероятностью не менее $0{,}9$, в то время как противник, применяющий более надежный алгоритм 2, сумеет восстановить S с вероятностью не более чем $2^{-124{,}3}$ (при этом вероятность угадывания сообщения S равна 2^{-229}).

Уменьшить значения параметра t удается лишь за счет снижения скорости передачи информации. Например, в системе со случайным кодированием кодом $C_6 = G_6^{\perp}$ при тех же значениях p_1 и p_2 для передачи сообщения S длины 53 бита достаточно положить $t = 14043$. При этом законный

получатель восстановит переданное сообщение с вероятностью не менее $0{,}9$, в то время как противник — с вероятностью не более $2^{-39{,}22}$ (табл. 2). Для повышения скорости передачи информации (при заданных границах надежности ее приема в основном канале связи и стойкости защиты — в отводном) следует, по-видимому, использовать отличные от рассмотренного выше способы случайного кодирования.

ЛИТЕРАТУРА

1. *Wyner A.D.* The wire-tap channel // Bell System Technical Journal. — 1975. — **54**. — № 8. — P. 1355–1388.
2. *Csiszar I.*, *Korner J.* Broadcast channels with confidential messages // IEEE Transactions on Information Theory. — 1978. — **24**, № 3. — P. 339–348.
3. *Коржик В.И.*, *Яковлев В.А.* Неасимптотические оценки кодового зашумления одного канала // Проблемы передачи информации. — 1981. — Т. 17. — Вып. 4. — С. 11–18.
4. *Коржик В.И.*, *Яковлев В.А.* Пропускная способность канала связи с внутренним случайным кодированием // Проблемы передачи информации. — 1992. — Т. 28. — Вып. 4. — С. 24–34.
5. *Яковлев В.А.* Границы для оценки неопределенности в системе передачи со случайным кодированием // Радиотехника. — 1996. — № 12. — С. 58–63.
6. *Иванов В.А.* О методе случайного кодирования // Дискретная математика. — 1999. — Т. 11. — Вып. 3. — С. 99–108.
7. *Алексейчук А.Н.* Оценки эффективности кодовой защиты дискретных сообщений с использованием линейных кодов с большим дуальным расстоянием // Реєстрація, зберігання і обробка даних. — 2001. — Т. 3. — № 2. — С. 99–106.
8. *Иванов В.А.* Асимптотические характеристики критериев проверки гипотез по случайно преобразованной выборке // Тр. по дискретной математике: в 11 т. Т. 5. — М.: Физматлит, 2002. — С. 61–72.
9. *Иванов В.А.* Статистические методы оценки эффективности кодового зашумления // Тр. по дискретной математике: в 11 т. Т. 6. — М.: Физматлит, 2002. — С. 48–63.
10. *Алексейчук А.Н.*, *Гришаков С.В.* Нелинейное случайное кодирование в системах передачи информации по каналу связи с отводом // Правове, нормативне та метрологічне забезпечення системи захисту інформації в Україні. — 2004. — Вип. 8. — С. 133–140.
11. *Thangaraj A.*, *Dihidar S.*, *Calderbank A.R.*, *McLaughlin S.*, *Merolla J.-M.* Capacity achiving codes for the wire-tap channel with applications to quantum key distribution. — http://eprint.arXiv:cs.IT/0411003v1.
12. *Мак-Вильямс Ф.Дж.*, *Слоэн Н.Дж.А.* Теория кодов, исправляющих ошибки: пер. с англ. — М.: Связь, 1979. — 743 с.
13. *Ширяев А.Н.* Вероятность. — М.: Наука, 1989. — 638 с.
14. *Sullivan D.* A fundamental inequality between the probabilities of binary subgroups and cosets // IEEE Transactions on Information Theory. — 1967. — **13**, № 1. — P. 91–94.
15. *Kasami T.*, *Klove T.*, *Lin S.* Linear block codes for error detection // IEEE Transactions on Information Theory. — 1983. — **29**, № 1. — P. 131–136.
16. *Ashikhmin A., Gohen G., Krivelevich M., Litsyn S.* Bounds on distance distributions in codes of known size // IEEE Transactions on Information Theory. — 2005. — **51**, № 1. — P. 250–258.
17. *Влэдуц С.Г.*, *Ногин Д.Ю.*, *Цфасман М.А.* Алгеброгеометрические коды. Основные понятия. — М.: МЦНМО, 2003. — 504 с.

Поступила 09.10.2009

ПРОБЛЕМИ ПРИЙНЯТТЯ РІШЕНЬ І УПРАВЛІННЯ В ЕКОНОМІЧНИХ, ТЕХНІЧНИХ, ЕКОЛОГІЧНИХ І СОЦІАЛЬНИХ СИСТЕМАХ

УДК 004.946

КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ УКРАЇНСЬКОЇ ЖЕСТОВОЇ МОВИ

Ю.Г. КРИВОНОС, Ю.В. КРАК, О.В. БАРМАК

Запропоновано математичну модель та її комп'ютерну реалізацію для отримання та відтворення жестової мови за допомогою трьохвимірних моделей людей. Запропоновано технологію, яка досить реалістично відтворює рухи на трьохвимірній моделі, що отримані з відеозображення конкретної людини — носія жестової мови.

ВСТУП

Дослідження закономірностей і анімації рухів є актуальною задачею. Розв'язок цієї проблеми у формі інноваційних технологій може суттєво вплинути на такі сфери людської діяльності як медицина, робототехніка, комп'ютерна графіка та системи віртуальної реальності.

В Україні більше півмільйона дітей з вадами слуху, а кількість глухих людей, для яких потрібно розробляти сучасні засоби навчання та спілкування — становить багато сотень тисяч. Розвиток сучасної науки, комп'ютеризація суспільства, використання мультимедійних та Інтернет-технологій створили достатні умови для розробки комп'ютерних систем комунікації цих людей у формах і образах близьких і зрозумілих для них і для оточуючого середовища.

Люди з вадами слуху для спілкування між собою використовують жестову мову. Основними візуальними засобами, які відтворюють цю мову є фотографічні та відеозображення жестів. Створювати за допомогою цих засобів сучасні навчальні та комунікаційні комп'ютерні системи досить проблематично. Фотографічне зображення не відтворює потрібної динаміки жестів, а відеозображення досить громіздке і в ньому відсутня потрібна інтерактивність (не можливо подивитися на людину, яка відтворює жест з інших ракурсів, крім того, в якому проведений відеозапис).

Ці суттєві обмеження існуючих засобів відтворення жестової мови спонукають до розробки більш гнучких технологій, за допомогою яких можна було б створювати нові комп'ютерні системи навчання та комунікації для людей із вадами слуху. Для розвитку цього запропоновано концепцію [1, 2] інформаційної технології невербального спілкування людей з вадами слуху. Комплексна інформаційна технологія включає в себе і функціональність по

© *Ю.Г. Кривонос, Ю.В. Крак, О.В. Бармак*, 2011

синтезу рухів жестової мови глухих та дактильної абетки на тривимірній моделі людини.

ПОСТАНОВКА ЗАДАЧІ

Реалізація можливості генерації анімації процесу мовлення за допомогою жестової мови з використанням віртуальних моделей людей потребує розробки відповідних інформаційних та математичних моделей. Виходячи з цього, сформульована така постановка задачі:

- розробити інформаційну та математичну модель для фіксації морфем (мінімально значимих одиниць) жестової мови;
- у межах цієї моделі потрібно розробити технологію та відповідне програмне забезпечення для отримання, збереження та відтворення жестів;
- запропонувати алгоритмічне рішення для розрахунку людиноподібних траєкторій руху при переходах від жесту до жесту.

МОДЕЛЬ ДЛЯ ФІКСАЦІЇ МОРФЕМ МОВИ ЖЕСТІВ

Процес відтворення жесту трьохвимірною моделлю людини можна вважати анімацією з відповідною частотою різних станів скелетної моделі людини.

Скелетна модель людини спрощено відтворює скелет живої людини. Її можна формалізувати як ієрархічну структуру, яка складається з поєднаних кінематичних пар, які відтворюють основні кістки скелета людини. Сучасні пакети трьохвимірного моделювання (Poser, 3D Studio Max) можуть генерувати анімацію за допомогою віртуальної статичної моделі та інформації про зміну відповідних кутів скелета. Отже, для формального опису процесу фіксації жесту можна використовувати множину, що відображає спрощений скелет людини та зміни значення кутів Ейлера і порядок їх застосування для відповідних кісток цього скелета з плином часу (дискретно, з відповідною частотою — 1/30 с тощо):

$$H = \{H_i : H_i = \{k, d_i, M_i \in M\}\}, \tag{1}$$

де H_i — *i*-та кістка скелета ($i = 0, \cdots, N-1$, N — кількість кісток у скелеті); k — індекс кістки-предка; $d_i = [x_i, y_i, z_i]^T$ — координати точки — кінця кістки в системі координат, яка пов'язана з початком цієї кістки;

$$M = \{M_i : M_i = \{order_i, \theta_i^j\}\}, \tag{2}$$

де M_i — для *i*-ї кістки значення кутів Ейлера та порядок застосувань обертань для кістки з плином часу; $order_i \in \{1, \cdots, 6\}$ — порядок застосування обертання навколо відповідних координатних вісей для *i*-ї кістки (1–*XYZ*, 2–*XZY*, 3–*YXZ*, 4–*YZX*, 5–*ZXY*, 6–*ZYX*); $\theta = (\theta^j)$, $\theta^j = \{\theta_i^j : \theta_i^j = \{\varphi_i^j{}_X, \varphi_i^j{}_Y, \varphi_i^j{}_Z\}\}$ — множина зміни кутів Ейлера для *i*-ї кістки з плином часу ($j = 0, \ldots, K-1$, K — кількість кадрів відтворення руху із заданою частотою).

ТЕХНОЛОГІЯ ДЛЯ ОТРИМАННЯ ТА ЗБЕРЕЖЕННЯ ЖЕСТІВ

Історія трьохвимірної анімації нараховує вже не одне десятиліття, і прогрес у цій області очевидний. У цій роботі, для отримання та збереження жестових одиниць, використовується Motion Capture — технологія захвату рухів. Типова система Motion Capture є набором пристроїв-покажчиків, які прикріплені до тіла людини (або інтегровані в костюм). Інформація з пристроїв надходить до комп'ютера та опрацьовується відповідним програмним забезпеченням, завдяки чому відтворюється математичний опис руху актора, який, у свою чергу, використовується для керування віртуальною моделлю-актором. Основним недоліком існуючих Motion Capture-систем є досить велика вартість як самого обладнання, так і відповідних послуг з відцифровки рухів. Авторами запропоновано простішу реалізацію цієї технології.

Для отримання множини кутів, які характеризують зміни положення кісток від початкового стану скелета запропоновано таку технологічну схему.

1. За допомогою трьох камер проводимо зйомку людини, яка відтворює певний жест. Камери встановлено на однаковій фокусній відстані (L) від об'єкта зйомки та фіксують проекції — фронтальну, ліву та праву. Стани скелета характеризують N кадрів зйомки (із заданою частотою кадрів на секунду).

2. Для фіксації i-го стану ($i = 1 \ldots N$) скелета відтворимо відповідні кадри зйомки у вигляді відображення на трьох дзеркальних поверхнях типу «трюмо», які розташуємо на задньому плані трьохвимірної сцени (рис. 1). Скелет розташуємо перед цими дзеркалами і спроектуємо на них ті кістки,

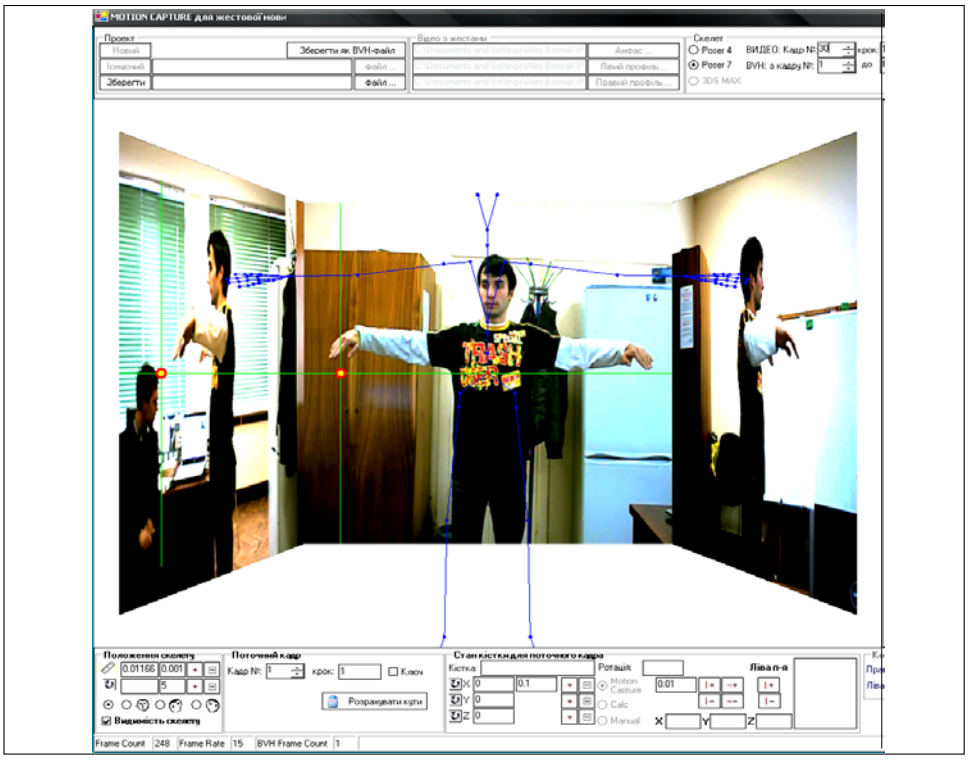

Рис. 1. Сцена для фіксації змін кутів у скелета

кути зміни для яких потрібно визначити (рис. 1). Співставимо точки проекцій кісток скелета з відповідним їм зображенням реальної людини (для фронтального, лівого та правого кадрів), тобто отримаємо нові координати.

3. Для *j*-го ($j = 1 \ldots N$) кадру на парі зображень (фронтальне і ліве, або фронтальне і праве) вказуємо точки зчленувань кісток, які змінили положення $((x_{i,\text{new}}^{\text{front}}, y_{i,\text{new}}^{\text{front}})$ та $(x_{i,\text{new}}^{\text{left or right}}, y_{i,\text{new}}^{\text{left or right}}))$ та визначаємо їх трьохвимірні координати (у системі координат, пов'язаній з кореневою точкою скелета).

Далі, маючи трьохвимірні координати нового положення точок зчленування кісток визначимо значення кутів Ейлера, повороти на які переводять кожну кістку з попереднього стану до поточного. Маємо систему координат XYZ і в ній два одиничних вектори: $r_1 = (x_1, y_1, z_1)$ та $r_2 = (x_2, y_2, z_2)$. Потрібно знайти кути Ейлера $\varphi_x, \varphi_y, \varphi_z$ обертання навколо відповідних вісей X, Y, Z такі, що переводять вектор r_1 у вектор r_2. Для цього:

1. Знайдемо два кути такі, щоб вектор $r_1 = (x_1, y_1, z_1)$ сумістився з віссю *oX*. Вектор r_1 є одиничним вектором, отже, $x_1^2 + y_1^2 + z_1^2 = 1$. Проведемо відрізок від початку координат у точку (x_1, y_1, z_1). Цей відрізок має одиничну довжину та орієнтований у напрямку вектора r_1. Опустимо перпендикуляри із точки (x_1, y_1, z_1) на кожну координатну вісь. Три направляючих кути — ϕ_x, ϕ_y, ϕ_z — це кути між вектором r_1 та координатними вісями. Між направляючими косинусами і компонентами r_1 існує співвідношення: $\cos(\phi_x) = x_1$, $\cos(\phi_y) = y_1$, $\cos(\phi_z) = z_1$. У силу виконання рівності $\cos^2(\phi_x) + \cos^2(\phi_y) + \cos^2(\phi_z) = 1$, незалежні тільки два кути. Знаючи значення направляючих косинусів, можна визначити і кути:

$$R_z(\varphi_1^{\text{п}}) = \begin{pmatrix} \cos(\varphi_1^{\text{п}}) & -\sin(\varphi_1^{\text{п}}) & 0 \\ \sin(\varphi_1^{\text{п}}) & \cos(\varphi_1^{\text{п}}) & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} x_1/d & -y_1/d & 0 \\ y_1/d & x_1/d & 0 \\ 0 & 0 & 1 \end{pmatrix} \Rightarrow$$

$$\Rightarrow \left.\begin{array}{l} \cos(\varphi_1^{\text{п}}) = \dfrac{x_1}{d} \\ \sin(\varphi_1^{\text{п}}) = \dfrac{y_1}{d} \end{array}\right\} \Rightarrow \varphi_1^{\text{п}} = \operatorname{arctg}\left(\frac{y_1}{x_1}\right) + n\pi;$$

$$R_y(\varphi_2^{\text{п}}) = \begin{pmatrix} \cos(\varphi_2^{\text{п}}) & 0 & \sin(\varphi_2^{\text{п}}) \\ 0 & 1 & 0 \\ -\sin(\varphi_2^{\text{п}}) & 0 & \cos(\varphi_2^{\text{п}}) \end{pmatrix} \begin{pmatrix} d & 0 & z_1 \\ 0 & 1 & 0 \\ -z_1 & 0 & d \end{pmatrix} \Rightarrow$$

$$\Rightarrow \left.\begin{array}{l} \cos(\varphi_2^{\text{п}}) = d \\ -\sin(\varphi_2^{\text{п}}) = -z_1 \end{array}\right\} \Rightarrow \varphi_2^{\text{п}} = \operatorname{arctg}\left(\frac{z_1}{\sqrt{x_1^2 + y_1^2}}\right).$$

Знаходимо базисні вектори для системи координат $X^{\text{п}}Y^{\text{п}}Z^{\text{п}}$:

$$oX^{r_1}: j_1^{r_1} = (j_1^{x_1}, j_1^{y_1}, j_1^{z_1}) = (x_1, y_1, z_1),$$

$$oY^{r_1}: j_2^{r_1} = (j_2^{x_1}, j_2^{y_1}, j_2^{z_1}) = R_y(\varphi_2^{r_1}) R_z(\varphi_1^{r_1})(0,1,0)^T = \left(-y_1, \frac{x_1}{\sqrt{x_1^2 + y_1^2}}, \frac{y_1 z_1}{\sqrt{x_1^2 + y_1^2}} \right),$$

$$oZ^{r_1}: j_3^{r_1} = (j_3^{x_1}, j_3^{y_1}, j_3^{z_1}) = R_y(\varphi_2^{r_1}) R_z(\varphi_1^{r_1})(0,0,1)^T = \left(z_1, 0, \sqrt{x_1^2 + y_1^2} \right).$$

2. Аналогічно знаходимо кут $\varphi_1^{r_2}$ обертання навколо вісі Z і кут $\varphi_2^{r_2}$ обертання навколо вісі Y такі, що переводять систему координат XYZ у систему координат $X^{r_2} Y^{r_2} Z^{r_2}$, в якій вісь oX^{r_2} співпадає з вектором r_2 та базисні вектори для цієї системи:

$$oX^{r_2}: j_1^{r_2} = (j_1^{x_2}, j_1^{y_2}, j_1^{z_2}) = (x_2, y_2, z_2),$$

$$oY^{r_2}: j_2^{r_2} = \left(j_2^{x_2}, j_2^{y_2}, j_2^{z_2} \right) = \left(-y_2, \frac{x_2}{\sqrt{x_2^2 + y_2^2}}, \frac{y_2 z_2}{\sqrt{x_2^2 + y_2^2}} \right),$$

$$oZ^{r_2}: j_3^{r_2} = \left(j_3^{x_2}, j_3^{y_2}, j_3^{z_2} \right) = \left(z_2, 0, \sqrt{x_2^2 + y_2^2} \right)^T.$$

3. Будуємо матрицю повороту T, що переводить вектори r_1 у r_2:

$$r_1 = T r_2, \text{ де } T = \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{pmatrix} = \begin{pmatrix} \cos(\alpha_{X^{r_1} X^{r_2}}) & \cos(\alpha_{X^{r_1} Y^{r_2}}) & \cos(\alpha_{X^{r_1} Z^{r_2}}) \\ \cos(\alpha_{Y^{r_1} X^{r_2}}) & \cos(\alpha_{Y^{r_1} Y^{r_2}}) & \cos(\alpha_{Y^{r_1} Z^{r_2}}) \\ \cos(\alpha_{Z^{r_1} X^{r_2}}) & \cos(\alpha_{Z^{r_1} Y^{r_2}}) & \cos(\alpha_{Z^{r_1} Z^{r_2}}) \end{pmatrix},$$

де, беручи до уваги, що вектори одиничні:

$$t_{11} = j_1^{x_1} j_1^{x_2} + j_1^{y_1} j_1^{y_2} + j_1^{z_1} j_1^{z_2} = x_1 x_2 + y_1 y_2 + z_1 z_2,$$

$$t_{21} = j_2^{x_1} j_1^{x_2} + j_2^{y_1} j_1^{y_2} + j_2^{z_1} j_1^{z_2} = -y_1 x_2 + \frac{x_1 y_2}{\sqrt{x_1^2 + y_1^2}} + \frac{y_1 z_1 z_2}{\sqrt{x_1^2 + y_1^2}},$$

$$t_{31} = j_3^{x_1} j_1^{x_2} + j_3^{y_1} j_1^{y_2} + j_3^{z_1} j_1^{z_2} = z_1 x_2 + z_2 \sqrt{x_1^2 + y_1^2},$$

$$t_{12} = j_1^{x_1} j_2^{x_2} + j_1^{y_1} j_2^{y_2} + j_1^{z_1} j_2^{z_2} = -x_1 y_2 + \frac{y_1 x_2}{\sqrt{x_2^2 + y_2^2}} + \frac{z_1 y_2 z_2}{\sqrt{x_2^2 + y_2^2}},$$

$$t_{22} = j_2^{x_1} j_2^{x_2} + j_2^{y_1} j_2^{y_2} + j_2^{z_1} j_2^{z_2} =$$

$$= y_1 y_2 + \frac{x_1 x_2}{\sqrt{x_1^2 + y_1^2}\sqrt{x_2^2 + y_2^2}} + \frac{y_1 z_1 y_2 z_2}{\sqrt{x_1^2 + y_1^2}\sqrt{x_2^2 + y_2^2}},$$

$$t_{32} = j_3^{x_1} j_2^{x_2} + j_3^{y_1} j_2^{y_2} + j_3^{z_1} j_2^{z_2} = -z_1 y_2 + \frac{y_2 z_2 \sqrt{x_1^2 + y_1^2}}{\sqrt{x_2^2 + y_2^2}},$$

$$t_{13} = j_1^{x_1} j_3^{x_2} + j_1^{y_1} j_3^{y_2} + j_1^{z_1} j_3^{z_2} = x_1 z_2 + z_1 \sqrt{x_2^2 + y_2^2},$$

$$t_{23} = j_2^{x_1} j_3^{x_2} + j_2^{y_1} j_3^{y_2} + j_2^{z_1} j_3^{z_2} = -y_1 z_2 + \frac{y_1 z_1 \sqrt{x_2^2 + y_2^2}}{\sqrt{x_1^2 + y_1^2}},$$

$$t_{33} = j_3^{x_1} j_3^{x_2} + j_3^{y_1} j_3^{y_2} + j_3^{z_1} j_3^{z_2} = z_1 z_2 + \sqrt{x_1^2 + y_1^2}\sqrt{x_2^2 + y_2^2}.$$

4. Далі визначимо кути Ейлера $\varphi_x, \varphi_y, \varphi_z$. Слід звернути увагу на те, що множення матриць не комутативне, тому потрібно розглянути шість випадків, які характеризують порядок обертання навколо вісей. Тобто, потрібно визначити кути для наступних порядків обертання навколо вісей: *ZYX*, *YXZ*, *YZX*, *XYZ*, *XZY*, *ZXY*.

4.1. Порядок обертання: *ZYX* ($c() = \cos(), \ s() = \sin()$):

$$\begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{pmatrix} =$$

$$= \begin{pmatrix} c(\varphi_y)c(\varphi_x) & c(\varphi_z)s(\varphi_x)s(\varphi_y) - c(\varphi_x)s(\varphi_z) & c(\varphi_x)c(\varphi_z)s(\varphi_y) + s(\varphi_x)s(\varphi_z) \\ c(\varphi_y)s(\varphi_z) & c(\varphi_x)c(\varphi_z) + s(\varphi_x)s(\varphi_y)s(\varphi_z) & -c(\varphi_z)s(\varphi_x) + c(\varphi_x)s(\varphi_y)s(\varphi_z) \\ -s(\varphi_z) & c(\varphi_y)s(\varphi_x) & c(\varphi_x)c(\varphi_y) \end{pmatrix}.$$

Звідси визначимо кути Ейлера шляхом вирішення таких систем:

$$\left.\begin{aligned} t_{11} &= \cos(\varphi_y)\cos(\varphi_z) \\ t_{21} &= \cos(\varphi_y)\sin(\varphi_z) \end{aligned}\right\} \Rightarrow \frac{t_{21}}{t_{11}} = \frac{\sin(\varphi_z)}{\cos(\varphi_z)} \Rightarrow \varphi_z = \operatorname{arctg}\left(\frac{t_{21}}{t_{11}}\right) + n\pi,$$

$$\left.\begin{aligned} t_{11} &= \cos(\varphi_y)\cos(\varphi_z) \\ t_{21} &= \cos(\varphi_y)\sin(\varphi_z) \\ t_{31} &= -\sin(\varphi_y) \end{aligned}\right\} \Rightarrow \varphi_y = \operatorname{arctg}\left(\frac{-t_{31}}{\sqrt{t_{11}^2 + t_{21}^2}}\right) + n\pi,$$

$$\left.\begin{aligned} t_{32} &= \cos(\varphi_y)\sin(\varphi_x) \\ t_{33} &= \cos(\varphi_x)\cos(\varphi_y) \end{aligned}\right\} \Rightarrow \varphi_x = \operatorname{arctg}\left(\frac{t_{32}}{t_{33}}\right) + n\pi.$$

4.2. Для інших порядків обертання кути визначаються аналогічно.

СИНТЕЗ ПРОЦЕСУ АНІМАЦІЇ ЖЕСТОВОЇ МОВИ

Для синтезу процесу анімації жестів та міміки трьохвимірною моделлю людини запропоновано такий формальний опис, який використовує відповідні множини параметрів та алгоритми роботи з ними. Трьохвимірна модель людини, на якій реалізуватиметься процес анімації жестів та міміки, має такі атрибути:

- $V = \{v_i : v_i = \{x, y, z\}\}$ — множина вершин трикутників для тріангуляції поверхні трьохвимірної моделі людини;
- $N = \{n_i : n_i = \{x, y, z\}\}$ — множина нормалей до вершин;
- $T = \{t_i : t_i = \{t, s\}\}$ — множина текстурних координат до вершин;

- $V^{\text{ind}} = \left\{ V_i^{\text{ind}} : V_i^{\text{ind}} = \{k_1, k_2, k_3\} \right\}$ — множина індексів, які вказують порядок побудови трикутників з множини вершин;
- $I = \left\{ I_i : I_i = \{img\} \right\}$ — множина фотографічних зображень елементів моделі — текстури.

Для моделювання скелетної анімації потрібно вміти розраховувати нові значення вершин трикутників (V). Для цього пропонується використовувати механізм скінінга. Скінінг можна означити як алгоритм прив'язки множини вершин трикутників, які означують поверхню моделі до значень кутів скелета. Тоді модель скелетної анімації можна формалізувати таким чином: $MH = \left\{ MH_i : HM_i = \left\{ k, \{l_1, \cdots l_m\}, d_i, Glb_i, Order_i \right\} \right\}$ — опис спрощеного скелета людини (ієрархія кісток) для реалізації скелетної анімації, де MH_i — i-та кістка скелета ($i = 0, \cdots, N-1$, N — кількість кісток у скелеті); k — індекс кістки-предка; $\{l_1 \cdots l_m\}$ — множина індексів нащадків кістки; $d_i = [x_i, y_i, z_i]^T$ — координати точки — кінця кістки в системі координат, яка пов'язана з початком цієї кістки; Glb — вектор для визначення координат кістки в глобальній системі координат; $Order_i$ — порядок застосування обертання. $Skin = \left\{ Skin_i : Skin_i = \{(IndexVertex_1, Weight_1), \cdots\} \right\}$ — множини вершин, які впливають на поточну вершину при зміні кутів.

Скінінг розраховується для кожної вершини V таким чином:

$$v'_j = \sum_{i=0}^{N} \{(v_j * IBM_{H_i} * JM_{H_i}) * JW_{H_i}\},$$

де n — кількість кісток, пов'язаних із вершиною v; IBM_{H_i} — інверсна bind-pose матриця для кістки H_i; JM_{H_i} — матриця переміщення для кіски H_i; JW_{H_i} — ваговий коефіцієнт для застосування впливу точок кістки H_i на вершину v.

Для моделювання анімації промовляння та емоційного забарвлення запропоновано використовувати механізм морфінгу. Морфінг можна означити як алгоритм плавного переходу від одного стану об'єкта до іншого. Під час використання морфінгу використовуються лише опорні стани, за допомогою яких розраховуються проміжні стани і моделюється процес анімації. Модель морфінгу міміки промовляння можна формалізувати таким чином.

Відображення або побудова міміки на обличчі трьохвимірної моделі досягається шляхом застосування відносного (сегментного) морфінгу до моделі голови та мімічного процесу. Формула відносного морфінгу для M морфів у формалізмі моделі матиме такий вигляд:

$$V' = V + \sum_{m=1}^{M} w_m TM_m,$$

де w_m — вагові коефіцієнти, V — меш базової моделі без морфінгу, TM_m — вхідний меш (морф) для блендінгу (blending). Результатом операції є лінійна комбінація мешів моделі та мімік.

Метод сегментного морфінгу дає змогу: формувати декілька виразів обличчя на основі невеликої кількості морфів та змінювати стан обличчя при анімації промовляння. Додатковою перевагою сегментного морфінгу є те, що можна анімувати щелепу незалежно від губ та очей (кліпання), незалежно від емоційного прояву на обличчі.

СИНТЕЗ ПРОМІЖНИХ ТРАЄКТОРІЙ-ПЕРЕХОДІВ МІЖ ЖЕСТАМИ

Лексичною одиницею у звичайній мові є слово, стале словосполучення або інша конструкція, що спроможна означати предмети, явища, їх ознаки тощо. У жестовій мові такою лексичною одиницею є жест. Він має вигляд знака або завершеного руху. У запропонованій моделі жесту-знака буде відповідати один стан спрощеного скелета людини H із відносно сталим значенням кутів M , а жесту-руху — послідовна зміна значення кутів та порядка застосувань обертань для кожної кістки з плином часу.

Суттєвою відмінністю жестового мовлення є його аморфність. Мовленєвий жест містить поняття, але не виражає форму числа, рід, відмінок, час тощо. Тобто, із досить обмеженої кількості жестів утворюються їх прості поєднання шляхом аглютинації (склеювання) у відомому порядку:

- дійова особа, предмет – дія ([Я] – [ПРАЦЮВАТИ]),
- дія – заперечення ([ХОТІТИ] – [НІ]),
- предмет – якість,
- стан ([ДИТИНА] – [ХВОРА] [ТЯЖКО]) тощо.

Беручи до уваги, що однією з важливих ознак жестової одиниці є локалізація (місце виконання жесту) виникає проблема розрахунку траєкторій-переходів від одного жесту до іншого, що подібні до рухів людей. У моделі жести отримуються шляхом оцифровування жестів реальних людей — носіїв жестової мови. Як екстенсивний, можна запропонувати підхід для отримання траєкторій-переходів шляхом фіксації та оцифрування всіх можливих траєкторій між областями-локалізаціями жестів. При такому підході залишиться задача зглажування стиків між жестом і початком траєкторії переходу. Це пов'язано з тим, що локалізація жеста досить відносне поняття — для групи жестів з однією локалізацією не можливо точно зафіксувати стан спрощеного скелета.

Виходячи з цього, сформульована така *постановка задачі*: потрібно, в межах запропонованої моделі, розв'язати задачу розрахунку траєкторій-переходів між двома довільними жестами.

У загальному випадку, задача планування рухів маніпуляційної системи в області з обмеженнями є досить складною проблемою. Розв'язок оберненої задачі (з відомого значеня трьохвимірних координат точки-цілі визначити значення узагальнених координат маніпуляційної системи) є задачею мінімізації функціонала в просторі з обмеженнями (оскільки таких значень узагальнених координат — множина та потрібно знайти оптимальне значення) і вона не має загального аналітичного розв'язку [3]. Існуючі чисельні підходи [4] пропонують оптимальний пошук траєкторії із множини можливих. У межах моделі, яка використовується, потрібно знайти функціональний перехід від кінцевого стану спрощеного скелета (вектор узагальнених координат — значення кутів обертання навколо відповідних вісей при кож-

ній ланці (кістці)) одного жесту до початкового стану (інший вектор узагальнених координат) іншого жесту. При цьому цей функціональний перехід має породжувати таку траєкторію, яка візуально близька до траєкторії людини при аналогічному переході.

Скористаємося тим, що можна організувати планування траєкторій маніпуляційної системи як аналога процедур навчання й прийняття рішень виконання маніпуляцій і локомоцій вищих організмів [5]. Можна сказати, що вищі організми для переміщення використовують таке керування, для здійснення якого потрібно зробити найменшу роботу, тобто найменшу кількість рухів. У нашому випадку найменшою кількістю рухів буде лінійна зміна кутів для переводу одного стану скелета в інший. Для формалізації жесту, а також з метою подальшого використання механізму аглютинації (для створення жестових конструкцій типу речень) використаємо таку множину:

$$G = \left\{ G_i : G_i = \{ f_i^{\text{beg}},\ f_i^{\text{end}},\ S_i^{\text{beg}},\ S_i^{\text{end}},\ \theta_i \} \right\}, \tag{3}$$

де G_i — конкретний жест ($i = 0, \cdots, \infty$), який відповідає послідовності зміни станів спрощеного скелета з дискретним (з відповідною частотою) плином часу (1)–(2); f_i^{beg} — номер стану, що відповідає початку жесту; f_i^{end} — номер стану, що відповідає кінцю жесту; $S_i^{\text{beg}} = \{(x, y, z)\}$ — координати кінців кісток спрощеного скелета для стану f_i^{beg} (у системі координат, пов'язаній з корневою кісткою спрощеного скелета); S_i^{end} — аналогічно для стану f_i^{end}; θ_i — множина змін кутів Ейлера (із (2)).

Для розрахунку траєкторії (θ) між жестами G_l та G_m пропонується розрахувати значення узагальнених координат таким чином:

$$\theta_0 = \theta_{\text{end}}(G_l),$$

$$\theta_i = \theta_{i-1} + \left(\theta_{\text{beg}}(G_m) - \theta_{\text{end}}(G_l) \right) \Big/ k, \quad i = 1, \cdots, k-1, \tag{4}$$

де $\theta_{\text{end}}(G_l)$ — значення узагальнених координат для останнього стану (f_i^{end}) жесту G_l; $\theta_{\text{beg}}(G_m)$ — значення узагальнених координат для першого стану (f_i^{beg}) жесту G_m; k — кількість станів між цими двома жестами.

Кількість станів між двома жестами є фактично часом, за який потрібно перейти з кінцевого положення одного стану до початкового положення іншого. Для автоматичного визначення кількості станів пропонується обраховувати коефіцієнт, який є відношенням кількості станів переходу між двома жестами до відстані, на яку зроблено при цьому переміщення деякої точки скелета, тобто:

$$koef = \frac{K}{L}, \tag{5}$$

де K — кількість станів між двома положеннями (отриманими з людиниоснія жестової мови при фіксованій швидкості показування жесту); L — відстань, на яку перемістилась деяка точка скелета.

Виходячи з того, що оцифрування жестів G проводилось в одному (середньому) темпі, розрахувавши коефіцієнт (5) для різних жестів, можливо отримати його середнє значення та використати його для розрахунку кількості перехідних станів (K) між двома жестами.

Маючи коефіцієнт $koef$, та знаючи для кожного жесту координати початку кісток спрощеного скелета S_i^{beg} для стану f_i^{beg} та S_i^{end} для стану f_i^{end} можливо визначити кількість станів переходу від одного жесту до іншого (K):

$$K = koef\, L_{\max}, \tag{6}$$

де $L_{\max} = \max\left\{\sqrt{(S_i^{\text{end}}(G_l) - S_i^{\text{beg}}(G_m))^2}\right\}$ — максимальна відстань між двома відповідними точками-кінцями кожної кістки скелета ($i = 0, \cdots, N-1$, N — кількість кісток у скелеті) між станами f_i^{end} (для жесту G_l) та f_m^{beg} (для жесту G_m).

Беручи до уваги те, що існують деякі доволі складні жести (що відтворюють двома руками складні асоціації з реального життя) можливі ситуації, коли отримана означеним чином траєкторія (4) приведе до колізії при відтворенні просторовою моделлю (неприродні перетини рук, занурення однієї руки в іншу тощо). Для подолання таких ситуацій пропонується обраховувати траєкторії з використанням проміжних станів. Тобто, для таких складних рухів вводитимуться додаткові проміжні стани, розрахунок траєкторії через які не приводитиме до колізій.

ТЕХНОЛОГІЯ ДЛЯ ВІДТВОРЕННЯ ПРОЦЕСУ АНІМАЦІЇ

Для відтворення процесу анімації жестів та міміки трьохвимірною моделлю людини створено комплекс програм, який реалізує скелетну (для відтворення жесту) та морфемну (для відтворення міміки промовляння та емоцій) анімацію. Реалізована відповідна програмна функціональність, яка, використовуючи трьохвимірне API OpenGL, відтворює по означеним атрибутам модель людини (рис. 2), та з використанням алгоритмів скінінгу та морфінгу відтворює анімацію (рис. 3).

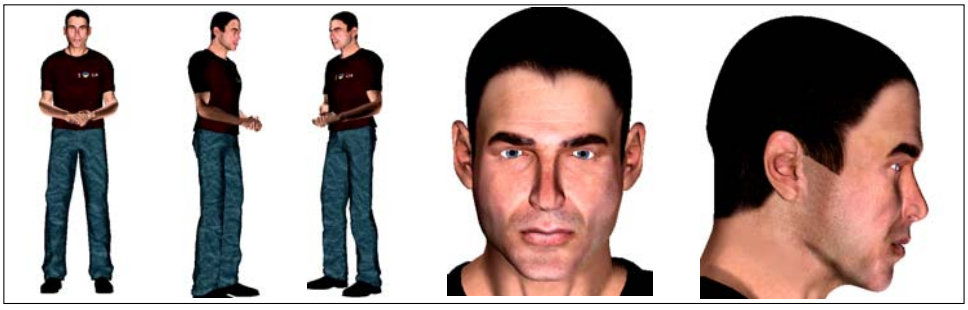

Рис. 2. Відтворення трьохвимірної моделі людини

Для реалізації української жестової мови було створено програму (рис. 4), яка відтворює методику викладання жестової мови в спеціальних

загальноосвітніх школах для нечуючих дітей. За основу було взято рекомендовану Міністерством освіти програму [6] для початкових класів.

Рис. 3. Кадри анімації жесту «Ти»

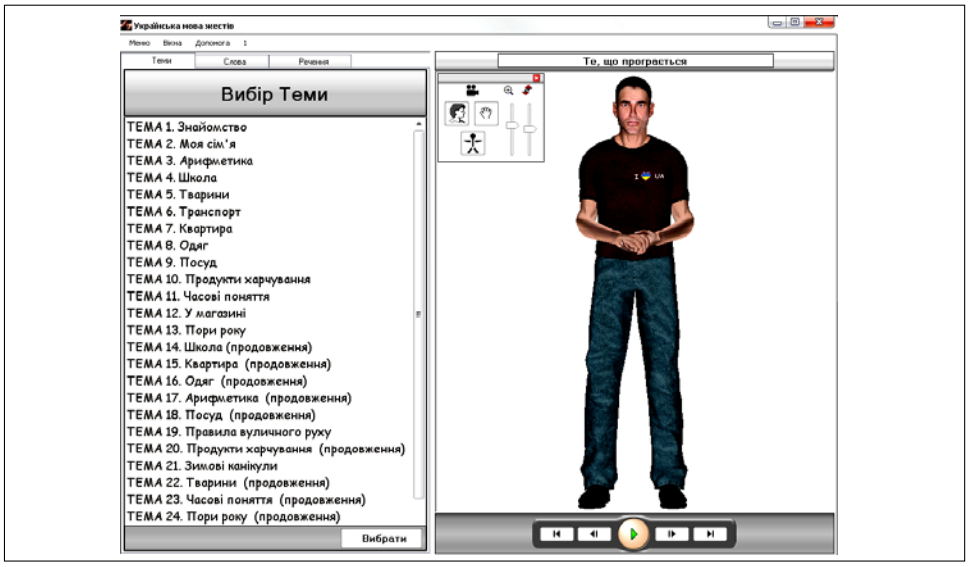

Рис. 4. Програма «Українська жестова мова»

Функціонально програма складається з трьох інформаційних блоків «Теми», «Слова» та «Речення» і блока відтворення жесту віртуальною моделлю. Основним блоком є блок «Теми». У ньому сконцентровано основні методичні відомості, які розглядаються на уроці: що потрібно у цій темі сформувати в учнів; які навички потрібно засвоїти, які особливості морфології жестової мови та, які види синтаксичних конструкцій містить матеріал. Наведено список нових жестів з теми, що розглядається, та речення з цими словами, які містять граматичні та синтаксичні конструкції (рис. 5).

Блоки «Слова» та «Речення» (рис. 6, 7) мають допоміжну функцію. У них сконцентровано всі жести та приклади речень (які використовують розглянуті жести).

Блок відтворення жесту віртуальною моделлю у цій реалізації виконує особливу функцію. Саме за допомогою нього можливо демонструвати в навчальному процесі динаміку жесту. Беручи до уваги, що жести оцифровуються носіями жестової мови, вони, по суті, стають стандартами відтворення жесту. Реалізована у блоці можливість покадрового показу жесту слугує засобом, за допомогою якого стане можливим вивчення жесту без особливостей показу конкретним викладачем. Це стане основою того, що вивчені дітьми жести будуть однаковими для різних шкіл, регіонів. Фактично, жестова мова (в основній своїй множині) стане стандартом на території України.

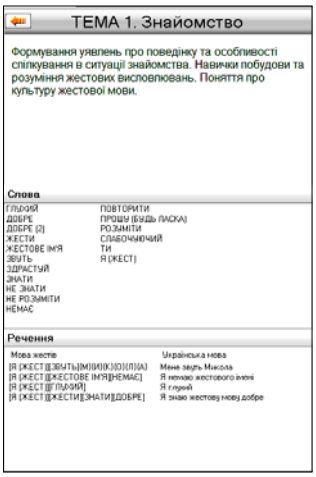

Рис. 5. Блок «Теми», тема «Знайомство»

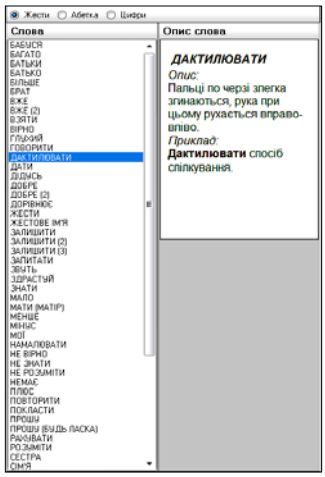

Рис. 6. Блок «Слова»

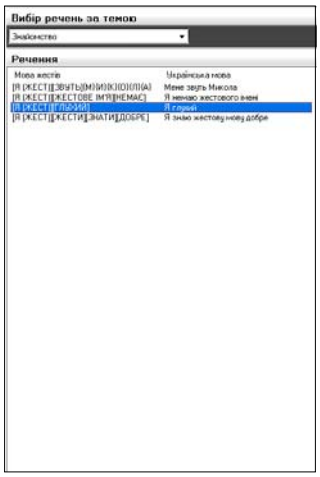

Рис. 7. Блок «Речення»

ВИСНОВКИ

Використовуючи модель для фіксації рухів, які відтворюють українську жестову мову було оцифровано множину з 50 жестів. Відтворення жестів із цієї множини (за допомогою моделі для відтворення анімації) показало спроможність запропонованої технології досить реалістично відтворювати на трьохвимірній моделі рухи, що отримані з відеозображення конкретної людини — носія жестової мови.

Було реалізовано програмне забезпечення, яке відтворює декілька уроків української жестової мови. Було взято уроки із програми спеціальних загальноосвітніх шкіл для нечуючих дітей [6].

Розроблене програмне забезпечення є основою для розробки та реалізації певних стандартів жестової мови. Це дає можливість вирішити проблему відмінностей для одних і тих самих жестів, яка виникає через їх інтерпретацію конкретним викладачем.

Подальші дослідження мають бути спрямовані на вдосконалення запропонованої технології:

- наповнення бази даних жестів основною множиною жестів української жестової мови — створення стандарту жестової мови;
- створення засобу для семантичного зв'язування речень українською мовою з реченнями на жестовій мові.

Відзначимо, що реалізація всього словника української жестової мови, використовуючи розроблений підхід і створену комп'ютерну технологію, займе невеликий проміжок часу (рік–два) і дозволить зробити впровадження такої системи в межах усієї держави.

ЛІТЕРАТУРА

1. *Кривонос Ю.Г., Крак Ю.В., Бармак О.В.* та ін. Інформаційна технологія невербального спілкування людей з вадами слуху // Штучний інтелект. — 2008. — № 3. — С. 325–331.

2. *Кривонос Ю.Г., Крак Ю.В., Бармак О.В.* та ін. Інформаційна технологія для моделювання української мови жестів // Штучний інтелект. — 2009. — № 3. — С. 186–198.
3. *Кириченко М.Ф., Крак Ю.В., Сорока Р.О.* Оптимізація маніпуляційних роботів. — Київ: Либідь, 1990. — 144 с.
4. *Бармак А.В.* Информационная компьютерная технология для моделирования и управления манипуляционными и виртуальными системами // Проблемы управления и информатики. — 2003. — № 5. — С. 143–154.
5. *Бернштейн Н.А.* Очерки по физиологии движений и физиологии активности. — М.: Медицина, 1966. — 350 с.
6. *Грищенко Є.С., Стьопкін В.В.* Українська жестова мова 1–4 класи. — Київ: Богдана, 2004. — 100 с.

Надійшла 07.06.2010

УДК 519.816

МАТЕМАТИЧЕСКИЕ МОДЕЛИ УПРАВЛЕНИЯ ПРОЦЕССАМИ ФИНАНСИРОВАНИЯ ИНВЕСТИЦИОННЫХ ПРОЕКТОВ

В.В. МОСКАЛЕНКО, В.В. КОНДРАЩЕНКО

Рассмотрена проблема построения оптимальной схемы финансирования инвестиционного проекта с привлечением внешних источников финансирования. Предложен комплекс математических моделей, соответствующих конкретным формам финансирования, каждая из которых позволяет рассчитать оптимальную схему финансирования и получить оценку ее экономической эффективности с учетом специфики каждой конкретной формы финансирования.

ВВЕДЕНИЕ

В практике управления инвестиционными проектами особое место занимает проблема сокращения затрат на реализацию проекта. Специфика инвестиционных проектов зачастую приводит к тому, что стоимостной аспект является решающим при принятии решения о целесообразности реализации проекта. Наряду с традиционными путями минимизации затрат, которые основаны на корректировке сроков, объемов работ, оптимизации объемов необходимых ресурсов, в данной работе предлагается уделить внимание проблеме выбора формы и источника финансирования проекта.

Цель работы — предложить комплекс математических моделей, которые позволяют субъекту, реализующему инвестиционный проект, оценить экономическую эффективность привлечения внешнего финансирования из того или иного источника, а также рассчитать наиболее эффективную схему привлечения и погашения займов.

Прежде всего, данная проблема актуальна для проектов, в которых значительную долю финансирования необходимо привлечь извне. В зависимости от специфики конкретного проекта, объема оборотных средств предприятия, занимающегося разработкой проекта, привлекательность той или иной формы финансирования будет меняться. Правильно выбранная форма и источник финансирования позволят значительно сократить затраты на использование внешнего капитала, что, в свою очередь, повышает рентабельность и привлекательность проекта.

Существенным отличием данной работы от целого комплекса работ является то, что проблема оптимального выбора рассматривается с точки зрения предприятия, которое нуждается в финансировании, а не с точки зрения инвестора.

ПОСТАНОВКА ЗАДАЧИ

Решение задачи выбора наиболее подходящей формы и источника финансирования предлагается проводить на основе решения комплекса оптимизаци-

© *В.В. Москаленко, В.В. Кондращенко*, 2011

онных задач математического программирования (ЗМП). Предлагается смоделировать процесс финансирования (как совокупность финансовых потоков) для различных форм привлечения средств. Ввиду существенных различий в характере процесса финансирования для различных его форм, необходимо рассматривать комплекс моделей. На данном этапе работы мы не будем рассматривать возможность одновременного привлечения средств из различных по форме источников (например, выпуск акций и привлечение внешнего целевого инвестирования), поскольку корреляция финансовых потоков будет существенно влиять на параметры источника финансирования. Например, выпуск акций подразумевает учреждение акционерного общества и контроль над деятельностью предприятия со стороны держателей акций, что, в свою очередь, влияет на ожидаемые риски и рентабельность инвестиций целевого назначения. Подобные ситуации сложно формализовать, и потому для решения практической задачи совместного финансирования будем использовать поэтапную человеко-машинную процедуру, при которой каждый источник финансирования рассматривается отдельно и в зависимости от решения, полученного на первом этапе, формируется задача для второго и т.д.

Целью моделирования процесса финансирования является расчет оценки прибыльности инвестиционного проекта, построение схемы привлечение заемных средств, схемы вложения собственных оборотных средств предприятия, схемы распределения ожидаемых будущих поступлений от реализации проекта, а также расчет финансового риска невыполнения проекта.

ОБЩАЯ СХЕМА РЕШЕНИЯ ЗАДАЧИ

Основная идея, лежащая в основе построения моделей, состоит в поэтапном рассмотрении процесса финансирования значения для каждого из этапов рассчитываются в динамике, т.е. финансовые решения, принятые на более ранних этапах, влияют на дальнейшее развитие проекта. Оценку финансового риска удобно выполнять, применяя нечетко-множественный подход, при котором вычисляется коэффициент устойчивости NPV (net present value — чистая приведенная стоимость) на основе интервально-симметричных оценок исходных данных проекта.

На данном этапе работы предлагается рассмотреть три формы финансирования: банковский кредит, внешние прямые инвестиции и выпуск акций. Каждая форма характеризуется некоторым набором параметров, которые зачастую имеют трудно формализуемый характер. Ниже приводится краткий обзор выбранных форм финансирования и их параметров [1, 2].

Банковский кредит — кредит, представляемый банками в денежной форме. Он имеет строго целевой и срочный характер. Рассматриваемые параметры: процентная ставка (%/год), минимальная сумма кредита (д.е. — денежные единицы), максимальная сумма кредита (д.е.), схема погашения кредита, вид и стоимость обеспечения.

Внешнее прямое инвестирование — целевое выделение денежных средств организацией-инвестором на реализацию инвестиционного проекта. Рассматриваемые параметры: максимальный размер инвестиций (д.е.), минимальная рентабельность инвестиций, риск невыполнения проекта и т.д.

Выпуск акций — выпуск долевых ценных бумаг, закрепляющих права на получение прибыли в виде дивиденда, на участие в управлении предприятием и на получение части имущества, оставшегося после его ликвидации. Рассматриваемые параметры: текущая рыночная цена, номинальная стоимость акции, будущая стоимость акции. Выпуск акций может осуществлять предприятие, являющееся только акционерным обществом.

Общая схема решения задачи выбора оптимальной формы и источника финансирования состоит в следующем (рисунок):

- Осуществляется сбор данных о возможных источниках финансирования, которые группируются по формам финансирования.
- Для каждого источника $\{A_1,\ldots,A_n,B_1,\ldots,B_m\}$ выполняется анализ и подготовка данных, необходимых для применения соответствующего инструмента расчета, где $A_1,\ldots,A_n$ — источники банковского кредитования, т.е набор возможных банковских кредитов, $B_1,\ldots,B_m$ — источники прямого инвестирования.
- Если предприятие является акционерным обществом, то выполняется анализ фондового рынка и подготовка данных, необходимых для реализации акций, при этом рассматриваются два варианта: C_1 — продажа на фондовом рынке собственных акций, которые находятся в распоряжении у предприятия; C_2 — дополнительная эмиссия акций и продажа их на рынке.

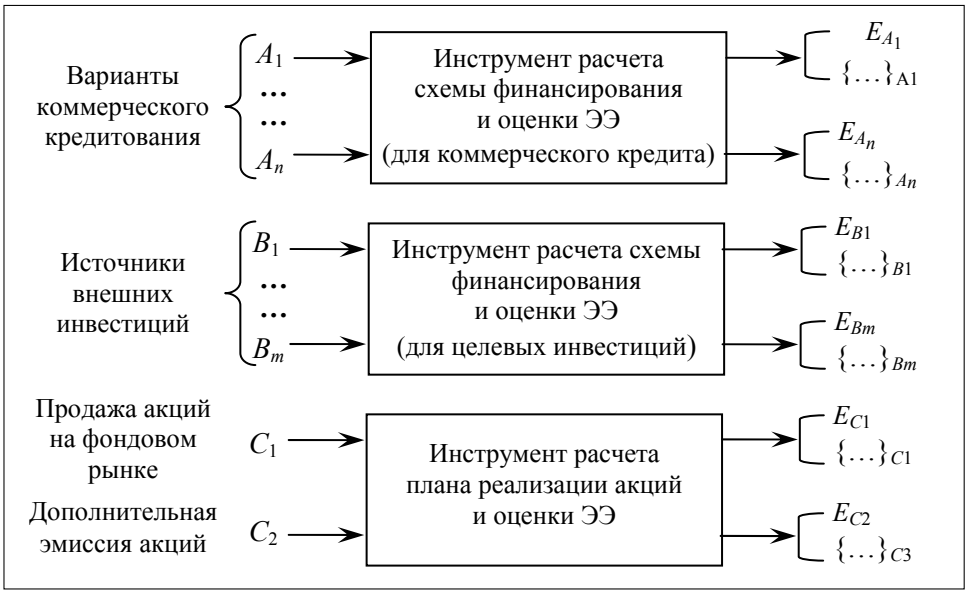

Рисунок. Общая схема выбора оптимального источника финансирования

- Для каждого источника $\{A_1,\ldots,A_n,B_1,\ldots,B_m\}$ применяется инструмент расчета схемы финансирования (в соответствии с формой источника финансирования) и производится оценка $\{E_{A_1},\ldots,E_{A_n},E_{B_1},\ldots,E_{B_m}\}$ от реализации проекта на конец рассматриваемого периода. При возможности использования инструментов фондового рынка, аналогичная процедура используется и для способов C_1 и C_2, т.е. определяются E_{C_1} и E_{C_2}.

- Производится сравнение полученных значений $\{E_{A_1},\ldots,E_{A_n}, E_{B_1},\ldots,E_{B_m},E_{C_1},E_{C_2}\}$, в результате которого выбирается такой способ $k\in\{A_1..A_n,B_1..B_m,C_1,C_2\}$, при котором было достигнуто максимальное E_k.
- Выбранному способу k будет соответствовать схема финансирования данного инвестиционного проекта.

Использование приведенной схемы определения наилучшего способа внешнего финансирования накладывает определенные ограничения на ту информацию, которую необходимо знать о проекте и о рассматриваемых источниках финансирования.

Построение моделей основывается на необходимости непрерывного выполнения проекта в соответствии с графиком работ. Для этого предлагается весь период выполнения проекта разбить на N равных этапов. Для каждого этапа известны объем средств d_k $(k=\overline{1,N})$, необходимый для вложения (затраты), и ожидаемый чистый доход от реализации проекта p_k $(k=\overline{1,N})$ на данном этапе. Предполагается, что все ресурсы, необходимые для выполнения проекта, представлены в стоимостном выражении. Таким образом, в математической модели требование непрерывности выполнения выражается в том, что в начале каждого этапа денежных средств должно быть достаточно для покрытия расходом этого этапа.

ФИНАНСИРОВАНИЕ ИНВЕСТИЦИОННОГО ПРОЕКТА С ПРИВЛЕЧЕНИЕМ БАНКОВСКОГО КРЕДИТА

Из всех возможных форм финансирования проектов банковское кредитование является наиболее универсальным и гибким. Это связано с тем, что банковский кредит предоставляется на строго определенный срок и может быть привлечен в любое удобное заемщику время. С другой стороны, кредит удобен с точки зрения его погашения, поскольку при известной схеме погашения кредита можно четко контролировать предстоящие расходы.

Принятие решения на каждом из этапов о взятии кредита, его размере и длительности использования, определяется запасом оборотных средств на каждом из этапов. Обозначим уровень запасов в конце *k*-го этапа s_k. Тогда, $s_0=Q$, где Q — объем целевых средств.

Уровень запасов на *k*-м этапе имеет следующий вид [3]:

$$s_k = s_{k-1} - d_k + p_k + x_k - \sum_{i=1}^{k-1} Y_i ,$$

где s_{k-1} — уровень запасов в конце $(k-1)$-го этапа; d_k — затраты в *k*-м этапе; p_k — поступления в *k*-м этапе; x_k — размер кредита, взятого в начале *k*-го этапа; $\sum_{i=1}^{k-1} Y_i$ — суммарные затраты на погашение кредитов, взятых в предыдущих этапах.

Затраты в *k*-м этапе на погашение кредита, взятого в *i*-м этапе, определяются как:

$$Y_i = \begin{cases} Y_i(x_i, n_i), \text{ если } (k-i) \le n_i \text{ и } k > i, \\ 0, \text{ если } (k-i) > n_i \text{ или } k < i, \end{cases}$$

где вид функции Y_i определяется типом схемы погашения кредита t_i. Предлагается рассматривать схему погашения долга равными срочными уплатами, равными выплатами основного долга и выплатами, изменяющимися в арифметической или геометрической прогрессиях. Каждой схеме погашения поставим в соответствие номер от 1 до 4, тогда $t_i \in \{1, 2, 3, 4\}$.

Для поддержания уровня оборотных средств, достаточного для покрытия затрат на каждом этапе, необходимо, чтобы выполнялось условие $s_k \ge 0$.

Критерием оптимальности набора кредитов является минимум суммарных начисленных процентов по всем кредитам. Величину начисленных процентов по кредиту, взятому на *k*-м этапе, можно рассчитать по формуле: $(x_k(1 + r(n_k))^{n_k} - x_k$.

Тогда суммарные начисленные проценты будут равны сумме начисленных процентов по кредитам, взятым на каждом из этапов:

$$W = \sum_{k=1}^{N} [(x_k(1 + r(n_k))^{n_k} - x_k],$$

где $x_k = 0$ для этапов, в которых кредит брать не нужно, и $x_k \ne 0$ в противном случае.

Учет ограниченности размера кредита произведем путем введения констант $x^{\min}$ и $x^{\max}$. Аналогично учет ограниченности длительности использования кредита произведем путем введения констант $n^{\min}$ и $n^{\max}$.

Таким образом, можно построить следующую математическую модель:

$$W = \sum_{k=1}^{N} [(x_k(1 + r(n_k))^{n_k} - x_k] \to \min, \tag{1}$$

$$s_k = s_{k-1} - d_k + p_k + x_k - \sum_{i=1}^{k-1} Y_i^{t_i}, \quad k = \overline{1, N}, \tag{2}$$

$$x^{\min} \le x_k \le x^{\max}, \quad k = \overline{1, N}, \tag{3}$$

$$n^{\min} \le n_k \le n^{\max}, \quad k = \overline{1, N}, \tag{4}$$

$$t_k \in \{1,\ 2,\ 3,\ 4\}, \quad k = \overline{1, N}, \tag{5}$$

$$s_0 = Q, \tag{6}$$

$$s_k \ge 0, \quad k = \overline{1, N}, \tag{7}$$

где Y_i соответствует схеме погашения кредита.

Имеем задачу нелинейного программирования с дискретными переменными. Она построена на основе классической задачи динамического программирования управления запасами, однако имеет существенное отличие, а именно: в задаче управления запасами рассматривается фиксированный расход на каждом этапе, а в нашем случае расход является переменным, поскольку зависит от объемов средств, необходимых для погашения взятых кредитов.

На основании математической модели ставится следующая оптимизационная задача: определить x_k, n_k и t_k ($k = \overline{1, N}$), которые удовлетворяют условиям (2)–(7) и доставляют минимум функции (1).

Очевидно, что задачу (1)–(7), как задачу нелинейного программирования с дискретными переменными решать затруднительно, поэтому предлагается представить базовую модель в виде задачи динамического программирования и для ее решения применить одну из схем динамического программирования, например, метод последовательного анализа вариантов.

ФИНАНСИРОВАНИЕ ИНВЕСТИЦИОННОГО ПРОЕКТА С ПРИВЛЕЧЕНИЕМ ПРЯМЫХ ИНВЕСТИЦИЙ

В финансировании проекта с привлечением прямых внешних инвестиций непосредственно участвуют две заинтересованных в прибыли стороны: предприятие, занимающееся реализацией проекта, и внешний инвестор (далее инвестор). Таким образом, будем рассматривать задачу формирования схемы финансирования проекта на основе критериев максимизации суммарной приведенной прибыли обеих сторон. Другими словами, необходимо произвести расчет долей финансирования между инвестором и предприятием по этапам реализации инвестиционного проекта (в работе не рассматриваются единовременные инвестиции, а инвестирование осуществляется по этапам реализации проекта).

Выделим следующие критерии оптимальности схемы финансирования проекта:

- Максимизация суммарной (по всем этапам реализации проекта) прибыли инвестора. Суммарная прибыль инвестора будет равна его суммарному доходу от реализации проекта минус его суммарные инвестиции.
- Максимизация суммарной прибыли предприятия от реализации проекта. Затраты предприятия складываются из двух составляющих: целевые вложения предприятия на выполнение проекта, и отчисления части дохода на финансирование следующего этапа.
- Минимизация стоимости капитала, затраченного на выполнение проекта. Так как финансирование проекта осуществляется из двух источников (средства предприятия и средства инвестора), то выбор приоритетного источника может быть осуществлен на основании оценки стоимости капитала источника финансирования. Поэтому, стоимость капитала предприятия может быть оценена коэффициентом рентабельности альтернативного вложения средств, а стоимость капитала инвестора — минимальной величиной рентабельности данного проекта.

Целью моделирования является определение для каждого этапа выполнения проекта оптимальных значений следующих величин:

- объем внешних инвестиций. Если на некотором этапе инвестиции не требуются, то их объем считается равным нулю;
- часть дохода от внедрения проекта, которая отчисляется инвестору;
- часть дохода от внедрения проекта, которая отчисляется предприятию;
- часть дохода предприятия, которую необходимо вложить на следующем этапе выполнения проекта.

Основное требование к реализации проекта состоит в необходимости его непрерывного выполнения в соответствии с запланированным графиком работ. В математической модели это требование выражается в том, что в начале каждого этапа денежных средств должно быть достаточно для покрытия расходов на этом этапе.

Для этого могут быть задействованы:

- q_k $(k=\overline{1,N})$ — целевые средства, выделенные предприятием на реализацию проекта (объемы целевых средств планируются предприятием для каждого этапа и являются фиксированными).
- x_k $(k=\overline{1,N})$ — внешние инвестиции, объем которых на каждом этапе является переменной величиной и определяется в ходе решения задачи (величина x_k может быть ограничена некоторым фиксированным максимальным значением $x^{\max}$).
- p_k^C $(k=\overline{1,N})$ — объем внутренних инвестиций проекта, выделенных предприятием из чистого дохода *k*-го этапа в $(k+1)$-й этап (данная величина является переменной).

Для обеспечения непрерывного финансирования проекта необходимо выполнение условия:

$$x_k + p_{k-1}^C = d_k - q_k \quad (k=\overline{1,N}). \tag{8}$$

Предприятие и инвестор получают доход за счет поступлений от внедрения проекта p_k $(k=\overline{1,N})$. Поступления на *k*-м этапе представим в виде суммы двух составляющих: дохода инвестора p_k^{I} и дохода предприятия p_k^{Π}.

$$p_k^{\text{I}} + p_k^{\Pi} = p_k \quad (k=\overline{1,N}). \tag{9}$$

Предприятие может вложить часть своего дохода *k*-го этапа в следующий $(k+1)$-й этап проекта, сократив при этом внешние инвестиции. Размер отчислений обозначим p_k^C, причем

$$p_k^C \le p_k^{\Pi} \quad (k=\overline{1,N}). \tag{10}$$

Целесообразность самоинвестирования на том или ином этапе будем определять на основании критерия минимизация «стоимости» капитала:

$$V = \sum_{k=1}^{N} \frac{c^{\mathrm{I}} x_k + c^{\Pi}(p_{k-1}^{C} + q_k)}{(1+r)^k} \to \min, \tag{11}$$

где c^{I} — коэффициент, характеризующий стоимость внешних инвестиций для предприятия; c^{Π} — коэффициент, характеризующий стоимость собственного капитала предприятия; r — рыночная стоимость капитала.

Как предприятие, так и инвестор, стремятся максимизировать свою прибыль от реализации проекта. Прибыль будем рассчитывать на основе *NPV*.

Стремление инвестора получить максимальную прибыль можно представить в виде следующего критерия:

$$NPV^{\mathrm{I}} = \sum_{k=1}^{N} \frac{p_k^{\mathrm{I}} - x_k}{(1+r)^k} \to \max, \tag{12}$$

где x_k — объем инвестиций в *k*-м этапе (равен нулю, если инвестиции не вкладывались).

Интересы предприятия можно представить как:

$$NPV^{\Pi} = \sum_{k=1}^{N} \frac{p_k^{\Pi} - p_{k-1}^{C} - q_k}{(1+r)^k} \to \max. \tag{13}$$

В результате имеем модель инвестирования, состоящую из критериев (4)–(6) при ограничениях (1)–(3), а также

$$x_k \le x^{\max} \quad (k = \overline{1, N}); \tag{14}$$

$$p_k^{\mathrm{I}} \ge 0, \quad p_k^{\Pi} \ge 0, \quad p_k^{C} \ge 0, \quad x_k \ge 0 \quad (k = \overline{1, N}), \tag{15}$$

причем $p_0^{C} = 0$, $p_0^{\Pi} = 0$.

Тогда задача состоит в определении таких векторов $x = \{x_k\}$, $p^{\mathrm{I}} = \{p_k^{\mathrm{I}}\}$, $p^{\Pi} = \{p_k^{\Pi}\}$, $p^{C} = \{p_k^{C}\}$, которые удовлетворяют ограничениям (8)–(10), (14)–(15) и доставляют экстремумы функциям (11)–(13).

Данная задача относится к классу многокритериальных задач линейного программирования с непрерывными переменными. Алгоритм ее решения подробно описан в [4].

ФИНАНСИРОВАНИЕ ИНВЕСТИЦИОННОГО ПРОЕКТА С ИСПОЛЬЗОВАНИЕМ ИНСТРУМЕНТОВ ФОНДОВОГО РЫНКА

Инструменты фондового рынка (эмиссия и размещение пакетов акций на фондовых рынках) могут применяться предприятиями, которые по форме собственности являются акционерными обществами. В данной работе фондовый рынок рассматривается как один из источников привлечения внешнего капитала для реализации конкретного инвестиционного проекта. При этом подразумевается, что предприятие имеет юридические основания для проведения подобных операций, располагает некоторым объемом акций,

которые оно предполагает реализовать на фондовой бирже без утраты контрольного пакета акций.

Расчет плана финансирования с использованием инструментов фондового рынка предлагается выполнить на основе оптимизационной математической модели. В качестве исходных данных предлагается использовать следующую специфическую информацию: q_i — объем целевых внутренних инвестиций, выделяемых предприятием в i-м этапе на реализацию проекта; c_i — среднее прогнозное значение тренда стоимости акций предприятия в i-м этапе ($i \in \overline{1,N}$). Для упрощения дальнейшей корректировки тренда в зависимости от «решений», принимаемых в ходе моделирования процесса финансирования, c_i будет представлено в виде суммы двух компонент c_i^{I} и c_i^{II}, где c_i^{I} — составляющая тренда, обусловленная объективной конъюнктурой рынка акций в данной отрасли, а c_i^{II} — составляющая тренда, существенно зависящая от данного конкретного предприятия. Такое разделение позволяет производить корректировку только составляющей c_i^{II}, притом, что c_i^{I} не будет существенно меняться в зависимости от действий предприятия на фондовом рынке. G — затраты на размещения пакета акций на данной конкретной фондовой бирже. Будем считать, что взнос G предприятие обязано уплачивать каждый раз при выставлении на продажу пакета акций. $f(e)$ — функция стоимости дополнительной эмиссии пакета акций объемом e. V_i — количество акций, которыми предприятие располагает для продажи в начале i-го этапа. Соответственно, V_0 — количество акций для продажи, которыми предприятие располагает в начале планового периода. От этапа к этапу значение V_i изменяется в зависимости от количества акций, проданных и дополнительно эмитированных в предыдущих этапах. v_i — количество акций, проданных на фондовом рынке в i-м этапе ($i \in \overline{1,N}$). $k(c_j^{\mathrm{II}}, j, i, v_i, e_i)$ — функция корректировки составляющей c_i^{II} тренда стоимости акций предприятия в j-м этапе в зависимости от объема акций v_i проданных, и объема акций e_i эмитированных в i-м этапе.

Тогда

$$\forall_{j>i}\, c_j = c_j^{\mathrm{I}} + k(c_j^{\mathrm{II}}, j, i, v_i, e_i).$$

На основании этих данных можно составить следующее рекуррентное соотношение, характеризующее прибыль от инвестиционной и финансовой деятельности предприятия в i-м этапе:

$$S_i = S_{i-1} + q_i + p_{i-1} - d_i + c_i v_i - g_i - f(e_i),$$

где $S_0 = 0$, $S_i \geq 0$, $g_i = \begin{cases} 0, & v_i = 0, \\ G, & v_i > 0. \end{cases}$

Объем акций, который может быть продан на i-м этапе, ограничивается необходимостью сохранения контрольного пакета акций за предприятием:

$$v_i \le V_{i-1} + e_i, \quad V_i = V_{i-1} + e_i - v_{i-1},$$

$$V_i \ge \frac{V_{\text{ext}} + V_0 + \sum_{j=1}^{i} e_i}{2} + 1,$$

где V_0 и V_{ext} — количество акций, которые, соответственно, принадлежали и не принадлежали предприятию на начало планового периода. Очевидно, что $V_0 + V_{\text{ext}}$ составляют общее количество акций эмитированных на начало планового периода.

Отсюда, можно выразить следующее ограничение:

$$v_i \le \frac{V_{\text{ext}} + V_0 + \sum_{j=1}^{i} e_i}{2} - \sum_{j=1}^{i-1} v_j - 1.$$

Таким образом, проект на каждом этапе может быть профинансирован из двух источников: нераспределенной прибыли от финансовой деятельности предприятия S_{i-1} и дохода от продажи акций предприятия на фондовой бирже $c_i v_i$. Вклад каждого источника в процесс финансирования на каждом этапе определяется из ограничения на рентабельность привлечения денежных средств от продажи акций, которая должна быть не меньше рентабельности нераспределенной прибыли от финансовой деятельности R:

$$\frac{c_i v_i - (g_i + f(e_i))}{g_i + f(e_i)} \ge R.$$

Критерием оптимальности схемы финансирования проекта с привлечением средств от реализации акций на фондовом рынке является максимизация суммарной рентабельности инвестиционной и финансовой деятельности

$$\sum_{i=1}^{N} \frac{S_i}{d_i + g_i + f(e_i)} \to \text{MAX}.$$

Таким образом, имеем оптимизационную математическую модель с одним критерием оптимальности следующего вида:

$$\sum_{i=1}^{N} \frac{S_i}{d_i + g_i + f(e_i)} \to \text{MAX} \tag{16}$$

при ограничениях

$$S_i = S_{i-1} + q_i + p_{i-1} - d_i + c_i v_i - g_i - f(e_i), \tag{17}$$

$$S_0 = 0, \quad S_i \ge 0, \tag{18}$$

$$g_i = \begin{cases} 0, & v_i = 0, \\ G, & v_i > 0, \end{cases} \tag{19}$$

$$v_i \le \frac{V_{\text{ext}} + V_0 + \sum_{j=1}^{i} e_i}{2} - \sum_{j=1}^{i-1} v_j - 1, \tag{20}$$

$$\frac{c_i v_i - (g_i + f(e_i))}{g_i + f(e_i)} \ge R, \tag{21}$$

$$\underset{j>i}{\forall} c_j = c_j^I + k(c_j^{II}, j, i, v_i, e_i). \tag{22}$$

Тогда задача состоит в определении таких значений v_i и e_i, $i \in \overline{1,N}$, которые удовлетворяют ограничениям (17)–(22) и доставляют максимум (16). Решением данной задачи является множество троек значений $\{v_i, c_i, e_i\}$, $i \in \overline{1,N}$, которые определяют поэтапный план продажи и дополнительной эмиссии акций. При заданных исходных данный и допущениях это решение позволят своевременно реализовать инвестиционный проект при максимальной суммарной рентабельности инвестиционной деятельности с привлечением инструментов фондового рынка. Необходимо отметить, что дисконтирование финансовых потоков в модели не применяется.

АНАЛИЗ МОДЕЛЕЙ

Рассмотренные модели были апробированы для построения схем финансирования нескольких инвестиционных проектов. На основе приведенной схемы выбора оптимального источника финансирования для построенных схем с использованием трех рассмотренных моделей был проведен анализ стоимости привлечения и эффективности проектов. Таким образом, можно сделать следующие выводы о применимости предложенных моделей в отдельных случаях. Использование кредитования как вариант финансирования проектов имеет свои преимущества и недостатки. Привлечение банковского кредита не требует учета интереса банка в объекте инвестирования и дополнительных выгод с инвестируемым объектом, как в случае с прямыми инвестициями. Это связано с тем, что финансовый интерес банка выражается в параметрах займа, таких как процентная ставка, схема погашения и длительность погашения. Ввиду того, что задача построения схемы финансирования рассматривается с точки зрения предприятия, занимающего реализацией проекта, основной акцент делается на максимизации экономической выгоды именно предприятия. Поэтому набор параметров кредитования может изменяться в заданных пределах, однако в работе не рассматривается задача наиболее эффективных параметров с точки зрения банка. Так как кредитование сопровождается обязательным обеспечением кредита и требованиями к кредитоспособности предприятия как объекта кредитования, то это может существенно сузить набор рассматриваемых кредитов для финансирования и/или увеличить затраты на оформление кредита, что, в свою очередь, увеличит затраты по проекту и уменьшит эффективность проекта. Поэтому, кредитование как вид финансирования инвестиционных проектов является выгодным только для финансово устойчивых предприятий с бла-

гоприятной кредитной историей, имеющих ликвидные активы в своем распоряжении. В других случаях необходимо применять прямое инвестирование, что обязано сопровождаться заинтересованностью инвестора в объекте инвестирования, в его участии в процессе внедрения проекта. Это возможно в случае производственной кооперации предприятий, вертикальной интеграции, когда инвестируемый объект завязан в цепочку «поставка-производство-сбыт», а также в других особых случаях, когда присутствует не только коммерческий интерес, но и другая экономическая целесообразность инвестирования.

Однако, как показывает зарубежный опыт инвестирования, поиск прямого инвестора связан не только со временем его поиска, но и с другими трудностями, такими как невозможностью страхования рисков (либо большими затратами на страхование). Поэтому в большинстве случаев в экономически развитых странах предприятия прибегают к непрямому инвестированию, т.е. используется финансовое инвестирование с привлечением инструментов фондового рынка. Однако применение такого способа также ограничено. Предприятие должно быть акционерным обществом, а также иметь возможности дополнительной эмиссии или должно иметь в своем распоряжении пакет собственных акций, который может быть продан для привлечения денежных средств без ущерба потери контроля над предприятием. Возможности в дополнительной эмиссии выражаются в следующем: предприятие обязано получить разрешение на дополнительную эмиссию от государственных структур, регулирующих выпуск ценных бумаг; рыночная стоимость акций, обращающихся на рынке, должна быть приемлемой с точки зрения доходности и ликвидности; предприятие должно пройти проверку фондовой биржи на предмет ликвидности и платежеспособности и т.д. Таким образом, при рассмотрении этого варианта финансирования должны быть учтены вопросы риска потери контроля над предприятием со стороны мажоритарных акционеров в случае дополнительной эмиссии. Поэтому можно сделать вывод, что для небольших проектов до одного года, которые внедряются на предприятиях, имеющих акционерную форму собственности, приемлемым является привлечение инвестиций за счет использования фондовых инструментов. Однако финансирование долгосрочных проектов также может осуществляться данным способом, но при условии, что часть проектов будут финансироваться за счет привлечения прямых инвесторов. Так как дополнительная эмиссия увеличивает собственный капитал предприятия, то может рассматриваться схема с прямыми инвестициями. Например, для долгосрочных проектов может рассматриваться модель прямого финансирования (8)–(15), в которой собственные денежные средства, выделяемые предприятием для финансирования q_k $(k = \overline{1, N})$, формируются с помощью модели с использованием инструментов фондового рынка (16)–(22).

Таким образом, при организации финансирования инвестиционных проектов должны быть учтен ряд факторов: организационно-правовая форма предприятия, внедряющего данный проект; состояние фондового рынка и рыночная стоимость акций предприятия, ее прогноз на рассматриваемом

периоде; состояние кредитного рынка и стоимость кредитов, их виды и возможности получения кредита предприятием; рентабельность собственного капитала и возможности самофинансирования проекта и целесообразность отвлечения денежных средств из оборота предприятия на финансирование проекта и другие факторы. Поскольку все сказанное выше предполагает сбор и обработку множества разнородной информации, в работах [4, 5] предложена система поддержки принятия финансовых решений, в которой предложенные модели входят в подсистему «Модели и алгоритмы финансирования». Было также разработано программное обеспечение, реализующее расчет планов финансирования, и предоставляющее инструменты для удобной работы с исходными данными и результатами расчета [6, 7].

ЗАКЛЮЧЕНИЕ

Предложенный комплекс математических моделей охватывает распространенные формы финансирования инвестиционных проектов. На их основе предлагается построить схемы финансирования проекта и выбрать оптимальную схему с точки зрения экономической эффективности. Все это даст количественную оценку управленческим решениям о выборе формы и источника финансирования проекта.

ЛИТЕРАТУРА

1. *Ван Хорн К.* Основы управления финансами. — М.: «Финансы и статистика», 1997. — 800 с.
2. *Бертонеш М., Найт Р.* Управление денежными потоками. — СПб.: Питер, 2004. — 240 с.
3. *Кондращенко В.В*. Математическая модель расчета оптимальной схемы финансирования инвестиционного проекта с привлечением коммерческого кредита // Тр. V междунар. науч.-практ. конф. — М.: МИИТ, 2008. — С. 117–119.
4. *Годлевский М.Д., Москаленко В.В., Кондращенко В.В.* Система поддержки принятия решений процесса финансирования инвестиционного проекта // Вестн. НТУ «ХПИ». Тем. сб. «Системный анализ, управление и информационные технологии». — 2007. — № 5. — С. 75–88.
5. *Godlevskiy M.D., Moskalenko V.V., Kondrashchenko V.V.* Analytical data modeling of investment project financing process // Information systems technology and its applications: 6-th international conference ISTA–2007, May 23–25. — Kharkiv. — 2007. — P. 78–90.
6. *Кондращенко В.В., Москаленко В.В., Захарова Т.В.* Архитектура СППР для построения схемы финансирования инвестиционного проекта // Вісн. НТУ «ХПІ». Тем. випуск «Системний аналіз, управління та інформаційні технології». — 2010. — № 9. — С. 149–156.
7. *Кондращенко В.В.* Программная реализация СППР для построения схемы финансирования инвестиционного проекта // Вісн. НТУ «ХПІ». Тем. випуск «Системний аналіз, управління та інформаційні технології». — 2010. — № 9. — С. 157–168.

Поступила 27.05.2009

УДК 004.9 (075)

МЕТОДИ І ТЕХНОЛОГІЇ ЗНИЖЕННЯ ПІКОВИХ НАВАНТАЖЕНЬ У КОРПОРАТИВНИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ ІЗ ЦЕНТРАЛІЗОВАНИМ ЗБЕРЕЖЕННЯМ ДАНИХ

В.Й. СІМАШКО

Описано пікові навантаження, які виникають у великих корпоративних інформаційних системах (КІС) із централізованим збереженням даних. Запропоновано методи і технології, які дозволяють знизити або ліквідувати пікові навантаження на сервери та на мережу системи.

ВСТУП

Сучасні тенденції розвитку корпоративних інформаційних систем (КІС) демонструють бурхливе зростання обсягів інформації, яка зберігається, обробляється і передається по мережі КІС. Із цим також пов'язане стрімке зростання кількості персональних комп'ютерів, які входять у корпоративну мережу передачі даних (КМПД) підприємства [1]. У великих КІС одним із сучасних напрямів розвитку є тенденція до централізації зберігання та обробки даних [1]. Насамперед це відбувається в білінгових системах, у системах економічного призначення (банківських, планово-фінансових і системах контролінгу) та деяких інших КІС великих підприємств. У процесі експлуатації таких КІС виникає низка проблем, зокрема:

- великі обсяги бази даних (БД) корпорації;
- значні обсяги обчислень і пов'язана з цим недостатня потужність як робочих місць користувачів, так і серверів;
- великий мережевий трафік між центральним серверним сегментом КМПД та робочими місцями користувачів;
- великий обсяг робіт з підтримки функціональності робочих місць (контроль та поновлення версій драйверів, додатків для користування, шаблонів документів, антивірусних баз тощо).

ПОСТАНОВКА ЗАДАЧІ

Дослідження та публікації з питань побудови КІС великих підприємств [1, 2, 3] недостатньо висвітлюють шляхи вирішення цих проблем. В якості універсального засобу їх вирішення майже завжди пропонується [1, 2] перехід до трьохрівневої структури обробки даних у корпоративній інформаційній системі: сервер БД, сервер додатків і клієнтське робоче місце. Проте практика показує, що така схема повністю не усуває жодної з описаних вище проблем, а лише частково знижує обсяг робіт з підтримки функціональності клієнтських робочих місць.

© *В.Й. Сімашко*, 2011

Описані вище проблеми періодично проявляються у різкому зниженні продуктивності роботи КІС у періоди так званих пікових навантажень: під час підготовки місячних, квартальних і річних звітів; при закритті розрахункових періодів; у випадку розповсюдження на комп'ютери корпорації об'ємних поновлень (нових версій додатків, документів, антивірусних баз тощо); під час великих обсягах пакетного (файлового) завантаження інформації тощо. Надмірне пікове навантаження може призвести до часткової чи повної втрати дієспроможності інформаційних систем підприємства, наслідки цього можуть бути досить відчутними для корпорації. Вирішення цього питання вимагає комплексного підходу.

Мета роботи — описати деякі методи і технологій, які дозволяють суттєво знизити пікові навантаження на КІС із централізованим збереженням та обробкою даних, і таким чином підвищити стійкість і надійність функціонування такої КІС.

ПРОБЛЕМИ, ЩО ВИНИКАЮТЬ ПІД ЧАС ЦЕНТРАЛІЗАЦІЇ ОБРОБКИ ДАНИХ, МЕТОДИ І ТЕХНОЛОГІЇ ЇХ ВИРІШИННЯ

У великих КІС централізація обробки даних полягає не лише в перенесенні всіх існуючих БД в єдиний серверний сегмент корпорації або в єдиний центр обробки, а й у зберіганні інформації. Централізація зазвичай передбачає:

- Створення єдиної БД для всіх підсистем КІС. Зазвичай єдина БД створюється на основі потужної промислової реляційної системи керування БД (РСКБД), такої як Oracle, MS SQL, DB2 тощо. Недаремно одним із важливих якісних показників РСКБД є продуктивність роботи БД при дуже великих обсягах збереженої інформації.
- Програмування безпосередньо в БД якомога більшої кількості бізнес-правил корпорації, а також контроль за дотриманням бізнес-логіки безпосередньо на рівні процедур та функцій БД. Ця тенденція в сучасних КІС спонукала розробників промислових РСКБД за останнє десятиріччя суттєво розширити функціональні можливості процедурних мов програмування БД (таких як — Oracle PL/SQL, MS Tsansact-SQL тощо).
- Перенесення на сервер додатків функцій з виконання обчислень, звітів, поновленню версій додатків, збереженню шаблонів документів, виконаних звітів і т.ін.

Не менш важливими факторами, які спонукають до такої централізації — економічні, а саме: велика вартість утримання висококваліфікованих адміністраторів, супроводу та підтримки промислових БД; зведення до мінімуму ймовірності втрати даних. Але централізація збереження та обробки інформації не проходить без наслідків для корпорації. Негативні наслідки цього описано нижче.

Ще донедавна використання реляційних БД визначало застосування технології «клієнт-сервер» у її класичному розумінні. «Класична» технологія «клієнт-сервер» полягає в наступному. Клієнтський додаток, написаний зазвичай на мові програмування високого рівня, формує SQL-запити. Сервер БД виконує ці запити та повертає додатку результати їх виконання.

Отримані з сервера дані обробляються клієнтським додатком та зберігаються у вигляді файлів, звітів, документів тощо. Отже, під час застосування класичної технології «клієнт-сервер» основний обсяг роботи (з обробки даних, з реалізації бізнес-правил та бізнес-процесів) здійснювався клієнтським додатком. У сучасних умовах різкого зростання обсягів інформації, яка обробляється в КІС, до клієнтського робочого місця з БД передається все більший обсяг інформації, а додаток виконує все більший обсяг обчислень. Унаслідок цього корпорація має постійно вкладати кошти в розвиток пропускної здатності мережі та в підвищення потужності клієнтських комп'ютерів. Такою є плата за централізацію зберігання інформації.

До двох щойно описаних недоліків «класичної» технології «клієнт-сервер» (зростаючий обсяг обчислень, що здійснюють клієнтські комп'ютери; зростаючий мережевий трафік між сервером і клієнтськими комп'ютерами), слід додати третій суттєвий недолік, пов'язаний із все більшою популярністю WEB-інтерфейсів у КІС, які йдуть на заміну традиційним додаткам робочого столу — проблематична реалізація складних бізнес-правил та бізнес-процесів за допомогою WEB-додатків через обмежені можливості мов WEB-програмування.

Для усунення описаних вище недоліків виробники сучасних реляційних БД інтенсивно розвивають і вдосконалюють такі дві технології:

- інтерфейси БД із WEB-додатками та WEB-орієнтованими мовами програмування;
- використання збережених у БД процедур та функцій (stored procedures and functions).

З метою сприяння ефективному написанню клієнтських WEB-додатків для роботи з реляційними БД виробники розширяють підтримку JAVA та NET. Та все ж вирішальну роль у сучасних КІС відіграє розвиток і вдосконалення процедурних мов, на яких програмуються реляційні БД та створюються збережені в БД процедури, функції, тригери та інші об'єкти.

Розглянемо детальніше використання збережених у БД процедур. Згідно зі своєю назвою, процедура зберігається в БД, і незалежно від способу її ініціалізації (виклику) виконується безпосередньо сервером БД. Можливості сучасних процедурних мов БД такі, що результатом виконання процедури може бути не лише традиційний набір вихідних параметрів, але й: масив даних будь-яких типів; модифікація інформації в БД; створення, зміна та знищення будь-яких об'єктів БД; XML-документ тощо. Якщо використати можливість процедури зі збереження результатів її виконання в БД, то мережевий трафік між сервером та клієнтським додатком, і навантаження на клієнтське робоче місце зводяться до мінімуму. Клієнтському додатку необхідно сформувати та відправити на сервер SQL-запит виконання збереженої процедури, дочекатися результату виконання та відобразити цей результат на дисплеї. Якщо при цьому результат виконання великого обсягу (великий документ, звіт у вигляді значного масиву даних тощо), доцільно зберегти результат у БД, щоб дати можливість користувачеві отримати результат пізніше, у зручний для нього час. Така технологія особливо актуальна для клієнтських робочих місць КІС, віддалених від сервера, які працюють по виділених лініях, або в режимі дозвону (dial-up) по телефонних лініях — так звані «тонкі клієнти». Працівникам віддалених офісів не потрібно очікувати

звільнення каналу зв'язку та робити кілька спроб, щоб вдало виконати складний розрахунок чи звіт, збережена процедура гарантовано буде виконана з першої спроби, а от із переглядом результату, можливо, доведеться зачекати.

Принципово важливо, що під час застосування цієї технології вся бізнес-логіка (алгоритми розрахунків, бізнес-правила тощо) реалізовані в БД, а не в клієнтському додатку. Тип додатку (WEB-додаток або додаток робочого столу) та мова програмування, на якому він написаний, відходять на другорядний план. За клієнтським додатком залишається мінімум функцій, а саме:

- введення даних користувачем;
- формування та відправка на сервер SQL-запитів;
- відображення для користувача даних, отриманих із сервера.

Із такими завданнями впорається навіть мова програмування із дуже обмеженими можливостями.

Також суттєво те, що під час застосування цієї технології будь-яка зміна бізнес-логіки, а також виправлення помилок у програмному забезпеченні (ПЗ) — реалізуються в БД, а не в клієнтському додатку. Тобто, у більшості випадків для заміни версії ПЗ достатньо замінити збережену процедуру (частіше — пакет збережених процедур) в єдиному місці — безпосередньо на сервері в БД. Для великих КІС це в сотні разів швидше, дешевше і надійніше, ніж заміна модулів, що виконуються та бібліотек на кожному комп'ютері корпоративної мережі.

Отже, під час застосування збережених процедур суттєво спрощуються (а отже, здешевлюються) клієнтські додатки; знижується навантаження на клієнтські комп'ютери і на корпоративну мережу (саме тому, знижуються капіталовкладення в обчислювальну техніку і в мережеве обладняння); спрощується контроль і заміна версій ПЗ КІС та виникає низка інших переваг. Звичайно, такі суттєві переваги не даються дарма. Натомість виникають дві інші значніші проблеми:

- різке зростання складності програмування БД;
- різке зростання навантаження на сервер БД.

Іншими словами, можна сказати, що «дешевші комп'ютери та мережа — дорожчі за сервер і БД».

Звичайно, розглядати можливі варіанти побудови структурних схем КІС та приймати рішення про використання тієї чи іншої технології обробки інформації необхідно окремо в кожному конкретному випадку. Проте все більше та більше КІС будуються саме за технологією з основним акцентом на серверне ПЗ [1, 2]. Очевидно, у багатьох випадках це економічно доцільно. Про такі світові тенденції свідчить також велика увага, яку розробники сучасних реляційних систем керування БД приділяють розвитку та вдосконаленню процедурних мов програмування БД. Безумовний лідер з потужності й можливостей — мова Oracle PL/SQL. Засобами PL/SQL можна реалізувати практично будь-який алгоритм, бізнес-правило чи обмеження цілісності. При цьому застосовуються такі об'єкти БД, як процедури, функції, пакети процедур і функцій, тригери тощо. Також БД Oracle мають потужну підтримку мови JAVA. Відразу за Oracle процедурні мови і підтримку

JAVA розвивають виробники практично всіх РСКБД. Інший потужний виробник — фірма MicroSoft, яка для свого MS SQL-сервера розробляє процедурну мову Transact-SQL. Для розширення можливостей збережені процедури цієї РСКБД можуть навіть включати фрагменти коду на мові С.

Підсумовуючи, можна зробити висновок, що сучасні процедурні мови програмування БД мають усі необхідні можливості для написання збережених процедур будь-якої складності.

Друга проблема — різке зростання навантаження на сервер БД. Вона частково вирішується нарощуванням потужності апаратної частини сервера, а частково — застосуванням кластерної структури серверного сегмента. Проте потужність сервера все одно не буде безмежною. Технологія, орієнтована на збережені процедури, передбачає виконання левової частки обчислень саме сервером БД. Сервер додатків не виконує обчислень, за ним залишається лише функція форматування результатів обчислень у вигляді документів, графіків тощо. Тому в моменти пікових навантажень на БД (під час підготовки періодичних звітів, у зв'язку із закриттям розрахункових періодів, під час виконання масових розрахунків або пакетних завантажень тощо) сервер буде працювати на межі можливостей. Навантаження на процесори, дискову систему, пам'ять сервера БД так чи інакше досягатимемо 100 %. І саме це є насправді основною проблемою, на яку рано чи пізно натрапляють адміністратори корпоративної БД та КІС із централізованим збереженням даних. Ця проблема інколи може виникати одразу після введення КІС у промислову експлуатацію. Але зазвичай ця проблема виникає через деякий час, при досягненні «критичних» позначок: критичного обсягу БД, критичної кількості користувачів, критичного набору функціональних модулів.

Дефіцит потужності сервера БД виявляється непостійно, а в моменти пікових навантажень, коли занадто багато користувачів починають одночасно виконувати складні або довготривалі розрахунки, звіти тощо. Усі розрахунки в такі періоди уповільнюються в десятки разів. Починають виникати ситуації, коли сервер БД не може забезпечити цілісність даних на досить тривалому проміжку часу — через модифікацію інформації іншими користувачами під час виконання розрахунку або звіту. У цьому випадку розрахунок (звіт) завершується помилкою, тому його необхідно повторювати знову.

Нестача потужності сервера БД відображається на роботі всіх користувачів системи, а не лише тих, хто «замовляє» складні звіти. Зокрема, оператори КІС, які виконують короткочасні інтерактивні транзакції (обслуговують клієнтів корпорації, працюють із одиночними документами чи проводками тощо) помічають, що транзакції, які зазвичай виконувалися долі секунд, у моменти пікових навантажень виконуються хвилини, або завершуються помилкою, або ж взагалі ніколи не закінчуються («зависають»). Це призводить до значних втрат робочого часу персоналу. Також це може призвести до явних збитків — коли йдеться про прийом готівкових платежів із використанням касових апаратів, або про реєстрацію в БД документів тих осіб, які бажають стати клієнтами корпорації і т.ін.

Перш за все, ця проблема має вирішуватись оптимізацією серверної частини КІС: використанням сучасних можливостей мови SQL та процедур-

ної мови, індексацією даних, налаштуванням параметрів БД, усуненням причин взаємного блокування сесій користувачів тощо. Проте далеко не завжди вдається уникнути дефіциту потужності сервера БД навіть при якісно спроектованій і налаштованій БД, а також грамотному написанні SQL-запитів і збережених процедур. Яким би потужним не був сервер, рано чи пізно виникне ситуація, коли він занадто повільно оброблятиме зростаючі потреби користувачів.

Враховуючи зазначене вище, зрозуміло, що для суттєвого покращення роботи сервера БД та КІС у цілому необхідно за можливістю рівномірно розподілити в часі виконання довготривалих розрахунків та складних звітів. Для цього існує кілька методів. Найпростіший і найдешевший метод (оскільки він не потребує доробки ПЗ системи) — організаційний. Він полягає в інвентаризації прав доступу до системи. Практика показує, що чим більше підприємство, тим більше диференційовані його працівники за рівнем кваліфікації. Особливо це критично для інженерного персоналу, задіяного у важливих технологічних процесах КІС. Досвід показує, навіть приблизно за однакового рівня освіти та досвіду роботи технологічна дисципліна різних виконавців дуже відрізняється. На жаль, не всі відповідальні особи до кінця розуміють, які процеси відбуваються в інформаційній системі, яка послідовність дій необхідна для ефективної роботи, результати яких розрахунків впливають на які звіти тощо. Тому згодом керівники корпорації починають розуміти, що до всіх важливих процесів (розрахунків, звітів, пакетних завантажень тощо) доступ необхідно жорстко обмежувати. Перевірка на практиці показала, що на великих підприємствах, значно розділених у просторі, але із централізованими БД, значно ефективніше централізувати також і групу фахівців з експлуатації КІС, залишивши операторам на місцях право лише на інтерактивні операції по роботі з клієнтами корпорації та із одиночними проводками документів. У такому випадку значно легше контролювати технологічні процеси та розподілити графік їх виконання в часі, знизивши тим самим пікові навантаження на систему.

Організаційні заходи можуть принести бажане зменшення пікових навантажень на сервер БД, але розмір виграшу буде різним у різних конкретних КІС. Зменшення навантаження може бути тимчасовим (враховуючи зростання розмірів БД, кількості користувачів та функціональних модулів КІС), або взагалі невідчутним (якщо специфіка підприємства така, що більшість звітів мають виконувати працівники віддалених підрозділів корпорації).

Тому кардинально вирішити проблему можна лише завдяки певним доробкам ПЗ КІС. Ідея полягає в поєднані технологій черги, планувальника процесів та розкладу виконання завдань.

Пропонується така технологія роботи інформаційної системи. Процедура БД (розрахунок, звіт тощо) не виконується одразу під час запиту на ініціалізацію її виконання, а ставиться у чергу завдань сервера БД. Для цього назва процедури, код користувача, який замовив виконання та її вхідні параметри та всі вихідні дані виконання процедури мають бути збережені у БД. Якщо результатом виконання є складний документ, то він зберігається в БД у вигляді XML, якщо текстовий звіт або розрахунок — то збереження результатів відбувається у звичайних реляційних таблицях.

Із таких запитів автоматично формується одна, або декілька черг завдань БД — у залежності від параметра, заданого адміністратором КІС. Оптимальна кількість окремих черг завдань залежить від кількох факторів: апаратних параметрів сервера БД, кількості вузлів у кластері, можливістю підтримки сервером РСКБД багатопроцесорності, масштабами та обсягами КІС тощо. Цей параметр підбирається в кожному конкретному випадку під час експлуатації системи.

Кожному конкретному завданню має бути присвоєний його унікальний номер — ідентифікатор. Працівнику корпорації, який ініціював той чи інший звіт або розрахунок, у клієнтський додаток передається посилання (ідентифікатор завдання), по якому він може відкривати результат у зручний для нього час. Якщо декілька користувачів замовили один і той же звіт або розрахунок (з однаковими вхідними параметрами), завдання буде виконане лише один раз, а всім «замовникам» буде передане одне і те ж посилання. Кожне завдання має свій статус: «очікування», «виконання», «виконано успішно», «виконано із помилками». Статус визначає, які дії користувач або адміністратор системи може виконувати із завданням.

Виконання завдань по черзі можна реалізувати кількома способами. Важливу роль у цьому відіграє вибір системи керування БД. Якщо підприємство може придбати та здійснити підтримку сучасної промислової РСКБД, такої як Oracle, то воно отримує у своє розпорядження досить зручний інструмент завдань БД (JOB-и). У цьому випадку виконання завдань по черзі реалізується саме на завданнях БД. Якщо масштаби та вартість КІС такі, що визначають застосовування простіших серверів РСКБД, для реалізації черги необхідне написання нескладного додатку. Цей додаток (так званий «монітор черги завдань КІС») виконується на сервері БД або на сервері додатків, періодично перечитує таблицю черги, і за відсутності завдань зі статусом «виконання» та за наявності в ній завдань зі статусом «очікування» ініціює виконання такого завдання.

Алгоритм формування черги завдань також може бути реалізований кількома способами. Найпростіший — «по мірі надходження». Постановка завдання в чергу відбувається завжди «у кінець» тієї черги, яка містить найменше завдань. Такий алгоритм прийнятний для малих і середніх КІС, де пікові навантаження виникають досить рідко.

Для великих систем необхідно застосовувати складніші алгоритми:

По-перше, для кожного конкретного типу завдання в БД має фіксуватися рейтинг (або пріоритет), наприклад: «звичайне», «важливе» і «негайне». У залежності від цього визначається, у яке місце в черзі буде поставлене те чи інше завдання. Доцільно застосувати таке правило: завдання ставиться в ту чергу, в якій є менше завдань із рівним або вищим пріоритетом. Місце завдання в обраній черзі визначається так: наступне після останнього завдання із рівним або вищим пріоритетом.

По-друге, для кожного конкретного типу завдання в БД можна зафіксувати часові межі (інтервал годин доби від … до …), в які або не можна це завдання виконувати, або завдання можна виконувати виключно в заданий часовий період. Наприклад, можливі складні розрахункові процеси, для яких є бажаним незмінний стан довідників системи протягом усього виконання розрахунку, або виконання яких у робочий час унеможливлює нормальну

роботу інтерактивних користувачів КІС. Такі завдання не можна виконувати в робочий час, коли проводиться робота з клієнтами корпорації. Інший приклад: заміна версій серверного та клієнтського ПЗ КІС може проводитися тільки у тому випадку, коли в системі не працюють жодні користувачі та не проводяться жодні розрахунки. Такі завдання необхідно виконувати в суворо визначений, зазвичай у нічний час. У залежності від «заборонених» або «дозволених» часових інтервалів завдання може очікувати своєї черги (точніше, «свого часу»), навіть якщо воно розташоване першим у черзі.

По-третє, алгоритм постановки у чергу може бути адаптивним. Для його реалізації по кожному типу завдання у БД накопичується й усереднюється статистична інформація (наприклад, середня тривалість успішного виконання). У цьому випадку завдання з однаковим пріоритетом, які виконуються миттєво, доцільно ставити у чергу перед завданнями, які виконуються кілька годин (особливо, якщо прогноз показує, що результат цього довготривалого завдання користувач зможе переглянути лише наступного дня).

Також слід зазначити способи, якими можливо формувати чергу. Програмний модуль, який реалізує описані в попередньому абзаці алгоритми, можна вбудувати в клієнтський додаток, яким користувачі системи ініціюють виконання тих чи інших розрахунків, пакетних завантажень чи звітів. Проте більш природно реалізувати цю логіку як процедуру або пакет процедур, збережений у БД, а клієнтський додаток має працювати із цим пакетом. Другий варіант має перевагу з точки зору можливості оперативної зміни алгоритму формування черги під час виникнення критичних ситуацій.

Чим складніша та об’ємніша інформаційна система, і чим відчутніша в ній проблема нестачі ресурсів сервера, тим важливішу роль відіграє можливість застосування технології черги завдань. У будь-якому випадку універсальних рекомендацій з алгоритму формування черги немає. Для кожної конкретної КІС замовником і розробником мають узгоджуватися свої специфічні алгоритми. Також слід зазначити, що для гнучкої експлуатації системи її адміністратори завжди повинні мати можливість ручного впливу на чергу завдань. Тобто, адміністратор повинен мати змогу видалити завдання, перемістити його в черзі, змінити пріоритет тощо.

Під час використання технології черги завдань сервером БД одночасно може виконуватися лише така кількість завдань (розрахунків, звітів, об’ємних пакетних завантажень), яка відповідає кількості окремих черг у цій системі. Практика показує, що під час оптимального підбору кількості черг та завдяки вмілому написанню алгоритму формування черги, пікові навантаження на сервер БД можна звести до мінімуму або взагалі до нуля. Це означає, що навіть у найнапруженіші для КІС періоди експлуатації, ресурси процесорного часу, оперативної пам’яті та дискової системи не будуть наближатися до 100 % зайнятості. Практично це призведе до значно швидшої, а значить, ефективнішої роботи інформаційної системи зокрема, та підприємства в цілому.

Описані вище підходи стосувалися технологій і методів, які знімають пікові навантаження як із серверів КІС, так із робочих станцій. Далі зупинимося детальніше на методах зменшення навантажень на КМПД.

Як зазначено вище, зростання обсягів даних, які обробляються і зберігаються в КІС, а також тенденція до централізації збереження інформації призводять до різкого зростання трафіку між серверним сегментом корпорації та кожним комп'ютером КМПД. Для прикладу розглянемо типові завдання, що виконуються звичайним клієнтським місцем КІС на початку кожного дня:

- після увімкнення комп'ютера операційній системі (ОС) необхідно зареєструватися в домені, отримати з сервера домену політики безпеки, пакети поновлень ОС, та, можливо, підключитися через мережу до загальнодоступних ресурсів;
- з антивірусного сервера корпорації необхідно завантажити зміни в антивірусних базах, які надійшли від розробника за минулу добу. Досить часто корпоративні версії антивірусів вимагають щоденного завантаження на комп'ютер не останніх змін, а повних антивірусних баз;
- із серверів КІС необхідно завантажити для клієнтських програм нові версії файлів: модулі, що виконуються, динамічні бібліотеки, шаблони документів тощо.

І це далеко не повний перелік обов'язкових процедур, кожна з яких зводиться до завантаження по мережі певного обсягу даних. Помноживши сумарний обсяг на кількість комп'ютерів, наприклад, в одному підрозділі підприємства, можна оцінити щоденний технологічний трафік між підрозділом та серверним сегментом підприємства. Такий пік навантаження спостерігається на початку кожного робочого дня — ще до того, як працівники розпочали виконувати роботу. Також пікові навантаження можуть виникати і в інший час, коли на сервери підприємства надходять об'ємні поновлення для ОС або антивірусних баз, нові версії клієнтських додатків тощо. Якщо цей підрозділ пов'язаний із серверами корпорації слабким каналом зв'язку, то зрозуміло, що на час пікових навантажень мережі корисна пропускна здатність каналу різко знижується. Крім втрат робочого часу, це призводить і до фінансових втрат, якщо канали зв'язку не є власністю підприємства і за них потрібно платити. Або, якщо канали зв'язку належать підприємству, але вони використовуються також і для надання послуг стороннім користувачам на комерційній основі.

Оскільки постійно нарощувати пропускну здатність КМПД неможливо, необхідно вжити заходів до зниження трафіку. Один з ефективних методів досягнення цієї мети — побудова ієрархічної структури КМПД.

Для ефективної побудови ієрархічної структури необхідно:

- здійснити поділ КІС підприємства на відносно автономні сегменти за їх територіальним розташуванням;
- здійснити поділ інформації на таку, яка важлива для підприємства в цілому, і тому має зберігатися централізовано, і таку, яка важлива для кожного робочого місця, і має бути щоденно та гарантовано доставлена на кожен комп'ютер.

Після такого поділу кожен відносно автономний сегмент КМПД має отримати свій окремий локальний сервер структурного підрозділу. Назвемо його «сервером поновлень структурного підрозділу», або ж «сервером поновлень».

Немає жодного сенсу завантажувати напряму з центрального сервера корпорації окремо на кожне робоче місце, скажімо, антивірусні бази розміром у десятки мегабайтів, або нові версії файлів для клієнтських програм. Це породжує невиправдано високі пікові обсяги мережевого трафіку. Оскільки такого роду інформація однакова для абсолютно всіх робочих місць, то достатньо завантажити інформацію на сервер поновлень структурного підрозділу, а всі робочі місця отримають її вже по локальній мережі. Це знизить технологічний трафік поновлення версій між серверним сегментом корпорації та конкретним структурним підрозділом рівно в стільки разів, скільки комп'ютерів є у локальній мережі цього структурного підрозділу.

Практика показує, що у якості сервера поновлень можна використати звичайний персональний комп'ютер. Єдина вимога — безперебійне живлення та цілодобовий режим роботи. Зазвичай на сервер поновлень, у першу чергу, інсталюють корпоративний антивірусний сервер, налаштовують його таким чином, щоб поновлення антивірусних баз надходили на нього із центрального антивірусного сервера. Антивірусні клієнтські програми всіх робочих місць підрозділу налаштовують таким чином, щоб поновлення антивірусних баз надходили на них із сервера поновлень цього підрозділу. Також на сервері поновлень інсталюються FTP та HTTP-сервери. Вони використовуються всіма інформаційними системами як проміжні «перевалочні» вузли для надходження в локальну мережу підрозділу нових версій клієнтських програм: модулів, що виконуються, динамічних бібліотек, шаблонів документів тощо. Також FTP та HTTP-сервери використовуються для внутрішнього обміну файлами між комп'ютерами локальної мережі структурного підрозділу.

Велика корпорація, яка охоплює значну територію, повинна мати кілька рівнів ієрархії КМПД. Наприклад підприємство, яке охоплює масштаби цілого регіону чи країни, матиме такі рівні ієрархії:

- центральний серверний сегмент корпорації;
- сервери поновлень обласних центрів, кожен із яких підпорядковується центральному серверному сегменту;
- сервери поновлень районних центрів, кожен із яких підпорядковується своєму серверу поновлень обласного центру;
- звичайні робочі місця, кожне із яких підпорядковується своєму серверу поновлень районного або обласного центру.

Нові версії всіх файлів (антивірусних баз, клієнтських програм, шаблонів документів тощо) автоматично розповсюджуються «зверху вниз» — від центрального серверного сегмента корпорації до сервера поновлень найнижчого рівня. Завантаження нових версій файлів із сервера поновлень на кожен комп'ютер КМПД ініціюється відповідними клієнтськими додатками під час їх запуску.

Така ієрархічна структура мережі передачі даних забезпечує мінімізацію технологічного трафіку по розповсюдженню нових версій файлів, а отже, зменшення прямих затрат на КМПД. Не менш важливо, що внаслідок зникнення пікових навантажень мережа, а значить і КІС у цілому, починає працювати значно стабільніше та надійніше. Платою за ці переваги є необхідність підтримки ієрархічної структури серверів поновлень. Проте

за сучасних можливостях віддаленого адміністрування ці додаткові затрати можуть бути значно менші, ніж отриманий виграш.

ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ ЗАПРОПОНОВАНИХ ТЕХНОЛОГІЙ

Для оцінювання ефективності запропонованих методів було здійснено їх експериментальну перевірку на двох промислових інформаційних системах. Аналітичним способом оцінити очікуваний виграш від впровадження описаних вище технологій складно. Розрахунок навантаження на серверний сегмент підприємства — це складна багатопараметрична задача, яка майже не піддається формалізації, а спроби вивести деякі аналітичні залежності носять емпіричний характер [2]. Було проведено експеримент на двох промислових системах:

- Дворівнева («класична» — «клієнт-сервер») інформаційна система. Сервер БД: ОС Windows 2003 Standart Edition Service Pack 2, РСКБД Oracle 9.2.0.7, апаратна платформа Intel E7505, 2 процесори Intel Xeon 2.4 ГГц, ОЗУ 2 ГБ, дисковий масив на базі SATA RAID контролера, обсяг БД 80 ГБ, кількість зареєстрованих користувачів приблизно 250.
- Трирівнева інформаційна система. Сервер БД: ОС Windows 2003 SE SP2, Oracle 10.2.0.4, апаратна платформа Intel E7525, 4 процесори Intel Xeon 3.4 ГГц, ОЗУ 4 ГБ, дисковий масив на SCSI RAID контролері, обсяг БД 700 ГБ, кількість зареєстрованих користувачів приблизно 450.

В обох системах було реалізовано найпростіший варіант організації черги: без пріоритету завдань, без часових інтервалів виконання завдань та без адаптивних алгоритмів (рис. 1). Згідно із цією блок-схемою, кожне завдання отримує свій незмінний ідентифікатор, а також має поточний статус (від «очікування виконання» до «виконано успішно»). Користувач, який в інтерактивному режимі замовляє той чи інший звіт чи ініціює процес, одразу отримує значення ідентифікатора. По цьому ідентифікатору в будь-який час можна переглянути не лише всі параметри завдання, а й його поточний статус, і якщо статус рівний «виконано успішно», отримати результат виконання. Спосіб отримання результатів завдання (документи, файли, інтерактивний перегляд) залежить від назви (тобто від типу) завдання. Крім того, користувач має можливість вибрати з БД ідентифікатори і назви всіх завдань, які цей користувач ініціював будь-коли.

Виконання завдань по черзі реалізує проста програма на сервері додатків, яка виконується постійно, шукає в кожній черзі завдання зі статусом «очікування», і стартує їх одне за одним.

Експериментальні дані збиралися таким чином. Засобами Oracle в БД постійно фіксувалися всі збої, що виникали (неможливість розпочати сесію, невиконання запиту тощо) і причини цих збоїв (коди помилок Oracle); а тричі на хвилину реєструвалася кількість активних сесій користувачів. Засобами ОС також тричі на хвилину у файлах на сервері БД реєструвалися: середній для всіх процесорів відсоток зайнятості, обсяг вільної оперативної пам'яті, середній для всіх жорстких дисків відсоток активності.

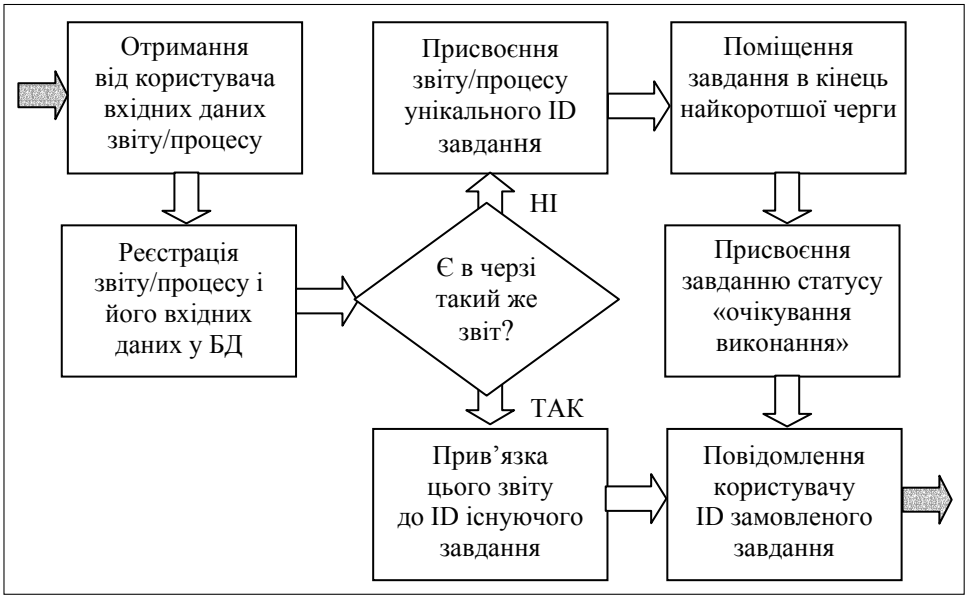

Рис. 1. Блок-схема

Накопичення даних було проведене таким чином. Протягом двох місяців реєструвалися та зберігалися всі експериментальні дані під час промислової експлуатації систем без організації черги завдань. Потім протягом місяця в обох системах було проведено впровадження описаної вище технології черги. На цьому етапі дослідним шляхом проведено вибір оптимальних налаштувань для обох систем, але експериментальні дані на цьому етапі не реєструвалися і не зберігалися. Після організації черги завдань знову реєструвалися і зберігалися всі експериментальні дані, також протягом двох місяців.

По завершенню збору інформації всі дані вимірів були завантажені в БД, синхронізовані в часі, нормовані (кількість збоїв — до кількості сесій, вільна пам'ять — до її загального обсягу тощо). Після нормування всі дані було усереднено по часу і по двох інформаційних системах. Для усереднення було використано не всі календарні дні, а лише ті, на яких спостерігалося суттєве навантаження на БД: крім вихідних та періоду «затишшя». У цих конкретних системах «затишшя» припадає на період із 10 по 25 число місяця, у ці дні складних звітів та масових операцій майже немає. Після усереднення будувалися графіки і діаграми, розраховувалися інтегральні показники, і на основі їх аналізу робилися висновки.

Перед початком збору даних для системи № 1 була підібрана оптимальна кількість черг, рівна 3, а для системи № 2 — рівна 7 окремих черг. За більшої кількості черг зниження пікових навантажень не відбувається. Таким чином виявлено залежність оптимальної кількості черг від кількості процесорів (тобто, ядер процесорів): у системі № 1 — 4 ядра, у № 2 — 8 ядер. Очевидно, що одне ядро обслуговує потреби ОС, а всі інші — вільні для обслуговування БД.

Графіки залежності вільної пам'яті ОС виявилися неінформативними, і тому в цій роботі не наводяться. Це пов'язано із особливостями ОС Windows 2003 Standart Edition, яка не виділяє під кожен окремий процес (та-

кож і під Oracle) оперативну пам'ять, більшу певної фіксованої величини. Нестача пам'яті виявляється не на рівні системи, а на рівні самого сервера Oracle, і стандартними засобами цю нестачу виміряти не вдалося.

Показником нестачі пам'яті в межах квоти, виділеної ОС серверу Oracle, є графік «Кількість збоїв» (рис. 2). Йдеться про збої БД, які (як показав аналіз кодів помилок сервера Oracle) у 90 % випадків виникали через неможливість сервером виділити область пам'яті під поточну сесію або для виконання запиту. На графіку «Кількість збоїв» нормована до поточної кількості активних сесій користувачів, і в деякі моменти цей показник досягав 30 %, тобто кожна третя сесія завершувалася аварійно через нестачу пам'яті. Можна побачити, що до впровадження черги середньостатистична частота збоїв сервера БД у моменти найбільшого навантаження рівна приблизно 10…12 %, а після впровадження черги збої припинилися взагалі. За допомогою черги завдань проблему нестачі пам'яті для сервера Oracle було повністю вирішено.

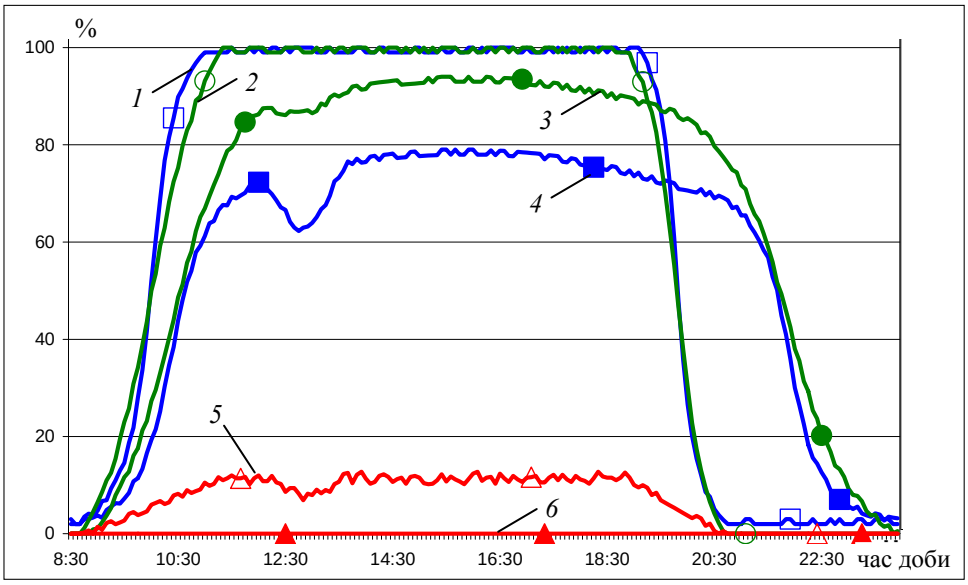

Рис. 2. Усереднені показники пікових навантажень на сервер БД, %: *1* — зайнятість процесів без черги, *2* — активність дисків без черги, *3* — активність дисків із чергою, *4* — зайнятість процесів із чергою, *5* — кількість збоїв без черги, *6* — кількість збоїв із чергою

Графіки зайнятості процесорів та дисків виявилися дуже подібними. Після впровадження черги ці показники знизилися, їх усереднені значення жодного разу не досягли 100 %. Проте в часі графіки розтягнулися приблизно на 2,5 години завдяки тому, що виконання звітів рівномірніше розподілилося в часі. Розраховано інтегральний показник: відношення площі під усередненим графіком до і після впровадження черги. Для зайнятості процесорів ця частка рівна 1,16; для активності дисків — 1,06. Отже, у періоди найбільших навантажень на БД середньостатистична завантаженість сервера в даних конкретних інформаційних системах знизилася в середньому на 11 %. Це відбулося завдяки тому, що кожен процес став успішно завершуватися з першої спроби, а ідентичні звіти виконувалися лише по одному разу.

Окрім навантаження на сервер, експериментальним шляхом проведено оцінку мережевого трафіку у двох напрямках: «від» серверного сегмента та «до» серверного сегмента корпоративної мережі. Засобами моніторингу апаратного мережевого маршрутизатора щодоби реєструвався сумарний (вхідний і вихідний) мережевий трафік у точці входу до серверного сегмента підприємства. Реєстрація проводилася протягом 30 діб до та 30 діб після повного переходу КМПД на ієрархічну схему розповсюдження всіх поновлень системного та прикладного ПЗ. Було введено лише один рівень ієрархії — сервери поновлень структурних підрозділів. Ці сервери завантажують усі поновлення безпосередньо із центральних серверів підприємства та «віддають» файли поновлень усім робочим місцям «свого» підрозділу. Завдяки зниженню цього виду технологічного трафіку середньомісячний обсяг інформації, переданої через точку входу в серверний сегмент мережі, знизився майже на 7 %.

ВИСНОВКИ

У роботі розглянуто сучасні тенденції побудови та розвитку КІС із централізованим збереженням та обробкою даних. Описано технічні та технологічні проблеми, що виникають у більшості таких КІС. Для вирішення типових проблем запропоновано поєднання організаційних, технологічних і технічних методів. Експериментально підтверджено, що запропоновані методи і технології дозволяють уникнути вузьких місць у КІС, знизити або взагалі ліквідувати пікові навантаження на сервер БД і на мережу КІС, і в результаті знизити капіталовкладення в КІС.

ЛІТЕРАТУРА

1. *Остер Д.* Архитектура систем планирования и управления ресурсами корпорации. — М.: Вершина, 2006. — 208 с.
2. *Спириус Ф.* Корпоративные базы данных: Проектирование и оптимизация. — М.: Вершина, 2007. — 316 с.
3. *Пфаулер Дж.* Корпоративные программные приложения. — СПб.: BHV, 2008. — 512 с.

Надійшла 12.05.2009

УДК 004.681

МАТЕМАТИЧНІ МОДЕЛІ ЕКОНОМІЧНОГО МЕНЕДЖМЕНТУ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ

Є.Г. ЛЕВЧЕНКО, М.В. ДЕМЧИШИН, А.О. РАБЧУН

Проведено порівняльний аналіз кількісних моделей економічного менеджменту інформаційної безпеки. Встановлено умови, за яких цільові функції розглянених моделей співпадають або досить близькі. Виконані розрахунки ілюструють методику визначення кількості втраченої інформації в залежності від співвідношення ресурсів нападу і захисту.

ВСТУП

Основним завданням економічного менеджменту інформаційної безпеки є пошук оптимальних рішень. При цьому ми стикаємось з низкою проблем, які мають як об'єктивний, так і суб'єктивний характер. Головна з них — побудова математичної моделі, яка включає:

- вибір критеріїв оптимальності, які враховують вибрані пріоритети, тобто важливість для підприємства таких показників, як ризик втрати інформації, сума витрат на її захист, рентабельність витрат тощо;
- визначення параметрів розрахунку та функціональних залежностей, які входять в математичну модель.

Основні задачі, що стоять перед менеджментом:

- визначення оптимальної кількості ресурсів захисту, які мінімізують сумарні витрати, що включають в себе потенційні втрати від витоку інформації і витрати на її захист із врахуванням відповідних вагових коефіцієнтів;
- оптимізація розподілу ресурсів між об'єктами, які містять різні обсяги інформації, характеризуються різним рівнем вразливості та певним ступенем корельованості, а також між окремими ступенями захисту;
- визначення оптимального розподілу ресурсів в умовах комплексного протистояння в конкурентній боротьбі, коли кожна сторона захищає свою інформацію й одночасно спрямовує свої зусилля на здобуття інформації конкурента, причому частина ресурсів може бути спрямована на розвідку;
- визначення зміни станів інформаційної безпеки з часом із врахуванням можливих дій суперників;
- розробка методики управління ресурсами в динамічному режимі, в якій враховано наведені ситуації та показники.

Перший крок у дослідженні процесів протистояння — це розробка аналітичної математичної моделі, яка має задовольняти двом суперечливим вимогам: максимально відображати найважливіші аспекти протистояння двох сторін в інформаційній сфері й одночасно уникати зайвого ускладнення, яке могло б ускладнити отримання практичних результатів.

Побудову моделі можна поділити на декілька етапів.

© *Є.Г. Левченко, М.В. Демчишин, А.О. Рабчун*, 2011

- Визначення параметрів та показників системи захисту інформації: кількість l об'єктів захисту; обсяг g_k інформації на кожному об'єкті (k — номер об'єкта); початкова вразливість υ_k об'єкта; виділений ресурс Y захисту; відношення до ризику.
- Оцінка дій суперника: характер атак (націлені, ненацілені); виділений ресурс X нападу; імовірність p_k нападу на об'єкт; імовірності виділення нападом ресурсів x_k на кожний об'єкт.
- Формування цільової функції, яке включає вибір цільового показника і незалежної змінної та встановлення між ними функціональної залежності: цільовий показник (кількість втраченої інформації; сумарні витрати ресурсів, які включають втрати від витоку інформації та витрати на її захист; ефективність інвестування, яку визначаємо як частку двох величин — зменшення обсягу втраченої інформації та витрат на захист); незалежні змінні (ресурси нападу і захисту — x_k та y_k; динамічна вразливість $\upsilon_k(t)$); вид функціональної залежності цільового показника від незалежної змінної (степенева, показникова).

Мета роботи — порівняльний аналіз моделей економічного менеджменту інформаційної безпеки.

РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ

Під час побудови математичної моделі ключовим питанням є формування цільової функції. Перша відома нам кількісна модель наведена в [1], де подана проблема була піддана ґрунтовному аналізу. У моделі Гордона-Лоеба (ГЛ) цільова функція визначає зменшення втрат в інформаційній системі (завдяки внесенню інвестицій) з відрахуванням витрат y на її захист (в [1] ці витрати позначені через z). У функцію як параметр входить вразливість υ об'єкта, тобто імовірність того, що напад буде успішним при $y = 0$. Авторами [1] запропоновано два широких класи функцій $S(y;\upsilon)$, які визначають імовірність втрати інформації:

$$S^I(y;\upsilon) = \frac{\upsilon}{(\alpha y + 1)^{\beta}}; \tag{1}$$

$$S^{II}(y;\upsilon) = \upsilon^{\alpha y + 1}, \tag{2}$$

де параметри $\alpha > 0$, $\beta \geq 1$ характеризують продуктивність інформаційної безпеки. Цільова функція має вигляд:

$$E(y) = [\upsilon - S(y,\upsilon)]L - y, \tag{3}$$

де L — потенційні втрати інформації при здійсненні нападу.

У (3) перша складова (υL) визначає кошти, втрачені в результаті нападу за відсутності системи захисту інформації (СЗІ), друга $(S(y;\upsilon)L)$ — при введенні СЗІ, третя (y) — інвестиції в СЗІ. Загалом (3) визначає кошти, збережені завдяки введенню СЗІ (в економічній термінології — прибуток від інвестицій).

Метою аналізу в [1] є визначення оптимальних витрат y^0 при різних значеннях вразливості. Показником оптимальності є максимум прибутку від

інвестицій, що виражається умовою $E'_y(y) = 0$. Показано, що вид залежності $y^0(\upsilon)$ відрізняється для двох класів функцій $S(y;\upsilon)$, і для розробки рекомендацій із визначення раціональної кількості інвестицій, крім вибору виду функції $S(y;\upsilon)$, необхідно встановити рівень вразливості об'єкта.

Останнім часом з'явилась низка робіт [2–5], спрямованих на розвиток моделі ГЛ. Зокрема, в [2] зосереджена увага на тому, що інвестиції в інформаційну безпеку можуть не тільки зменшувати можливі втрати, але й відлякувати потенційного порушника і в результаті зменшувати імовірність загрози. Розглядаючи ці явища (зменшення втрат і зменшення загрози), можна виокремити три варіанти їх взаємодії: відсутність впливу, позитивний вплив і негативний вплив. Для детальнішого дослідження цього питання в [2] введені поняття продуктивності зменшення вразливості й продуктивності зменшення загрози, а також простору продуктивності, який об'єднує ці показники. В залежності від їх значень випливають висновки щодо вибору стратегій, які забезпечують оптимальні витрати на захист інформації.

У [3] використані дещо змінені залежності $S(y;\upsilon)$:

$$S^I(y;\upsilon) = \frac{p\upsilon}{\alpha y + 1}, \tag{4}$$

$$S^{II}(y;\upsilon) = p \cdot \upsilon^{\alpha y + 1}, \tag{5}$$

де p — імовірність здійснення нападу. Предметом дослідження в [3] є визначення оптимального розподілу ресурсів нападу між об'єктами, на які здійснюють численні напади, в умовах обмеження бюджету й у випадку, коли такі обмеження відсутні. Показано, що існують граничні значення вразливості, за межами яких інвестиції недоцільні, значно простіше відшкодовувати втрати. Численні припущення під час побудови моделі [3], за думкою авторів, обмежують застосування цієї моделі.

В [4] проведено деяке узагальнення сімейства функцій $S(y;\upsilon)$ і запропоновано їх нові, складніші формулювання. Проте ступінь їх корисності ще варто встановити. В [5] зроблено спробу здійснення динамічного аналізу інвестування в інформаційну безпеку, причому порушення інформації розглядається як випадковий процес, який характеризується показниками дрейфу та волатильності. Наведено результати розрахунків на основі моделі ГЛ, в яких використано гіпотетичні значення параметрів.

У [6] запропоновано інший підхід до цієї проблеми. Математична модель [6] передбачає використання цільової функції $i(x;y)$, де i — віднесена до загальної кількості інформації вартість втраченої інформації, x та y — ресурси нападу i, відповідно, захисту. Ця функція в загальних ознаках має вигляд:

$$i(x,y) = \sum_{k=1}^{l} i_k(x,y) = \sum_{k=1}^{l} g_k p_k q_k(x,y) f_k(x,y), \tag{6}$$

де $k = \overline{1,l}$ — номер об'єкта; g_k — обсяг інформації на об'єкті; p_k — імовірність нападу на об'єкт; $q_k(x;y)$ — щільність імовірності виділення напа-

дом ресурсів x на k-й об'єкт; $f_k(x;y)$ — залежність частки втраченої інформації від співвідношення x та y, яку можна розглядати як імовірність втрати інформації при заданих значеннях x та y.

В якості залежностей $f(x,y)$ запропоновано два класи функцій:

$$\text{степеневі} \quad f(x,y)=\frac{a(x/y)^n}{b(x/y)^n+c} \tag{7}$$

$$\text{та показникові} \quad f(x,y)=d(1-e^{-m(x/y)^n}), \tag{8}$$

де параметри a,b,c,d,n,m приймають додатні значення і визначають положення та нахил кривих.

В [6] запропоновано два можливих види залежностей $q(x)$ у вигляді $q(x)=Nx^n e^{-h^2x^2}$: розподіл Максвела $q_M(x)=Nx^2e^{-h^2x^2}$ і розподіл Релея $q_P(x)=Nxe^{-h^2x^2}$, де N — нормовочний коефіцієнт, а константи n,h визначають положення максимуму залежності і ступінь її асиметрії. У співставленні цих розподілів суттєва для нас їх відмінність полягає в тому, що для $q_M(x)$ в початковій області $x \gtrsim 0$ опуклість направлена вниз, а для $q_P(x)$ — вгору.

Порівнюючи зазначені моделі, спробуємо окреслити їх відмінності та спільні ознаки.

- У моделі ГЛ за мету ставиться оптимізація витрат y^0 на захист інформації, у моделі [6] — оптимізація розподілу ресурсів y_k між окремими об'єктами.
- Цільова функція ГЛ визначає зменшення втрат від можливого витоку інформації з відрахуванням витрат на її захист, у моделі [6] — вартість втраченої інформації.
- У моделі ГЛ основним показником системи захисту інформації, який входить у цільову функцію та впливає на оптимальне значення y^0, є вразливість υ, у моделі [6] — співвідношення ресурсів нападу і захисту (x та y).
- Функції, які визначають імовірність втрати інформації, у моделі ГЛ задаються виразами (1), (2), у моделі [6] — виразами (7), (8).
- Імовірність нападу й імовірність втрати інформації в моделі ГЛ входять у цільову функцію у вигляді параметрів α та β, які є мірою продуктивності витрат, у моделі [6] імовірність нападу задається в явному вигляді, а імовірність втрати інформації при здійсненому нападі — залежностями (7), (8).
- У моделі ГЛ величини, які входять у розрахунок, даються в абсолютному вимірі, у моделі [6] — у відносному.
- Спільним в обох моделях є те, що параметри розрахунку (υ, α, β та, відповідно, p, q) і функції, які визначають залежність частки вилученої інформації від вкладених ресурсів ($S(y;\upsilon)$ та $f(x;y)$), не можуть бути встановлені точно, а знаходяться в результаті аналізу статистичних даних, а в разі їх відсутності — на основі експертних оцінок.

Слід зазначити, що оптимізація розподілу ресурсів між об'єктами певною мірою зумовлена можливими стратегіями нападу. Напади суперника можуть бути ненацілениими (шкідливе програмне забезпечення, віруси, фішинг, спам тощо) і націленими (хакерські атаки на банківську базу даних з метою вилучення коштів або вибір об'єкта нападу в системі, яка містить певну кількість об'єктів). Характер нападів може бути зумовлений, в одних випадках, — цільовою спрямованістю зловмисника, в інших, — наявною кількістю ресурсів. Останній варіант спостерігається, зокрема, коли об'єкти однотипні, і конкурент розглядає доцільність розподілу обмежених ресурсів між окремими об'єктами. Націлені атаки трапляються рідше, ніж ненацілені, проте їх наслідки можуть бути серйознішими для підприємства, що, звичайно, слід враховувати під час розподілу ресурсів захисту. В [3] зроблено припущення, що клас функцій (4) краще описує націлені атаки, а клас (5) — ненацілені.

Порівняємо тепер залежності $S(y;\upsilon)$ та $i(x;y)$, які виражають практично одну і ту ж величину. За цього порівняння величину x у виразі $i(x;y)$ вважатимемо сталою, і для спрощення всі величини, крім $f(x;y)$, покладемо рівними одиниці. Таким чином, ми зведемо задачу до порівняння залежностей $S(y;\upsilon)$ (де υ — параметр) та $f(x;y)$ (x — параметр). Поклавши $x=1$, функція $f(x;y)$ (7) матиме вигляд:

$$f(y)=\frac{a}{b+cy^n}. \qquad (9)$$

Порівнюючи її із залежністю $S^I(y;\upsilon)=\dfrac{\upsilon}{(\alpha y+1)^\beta}$, можемо зробити такі висновки. При $\alpha=\beta=1$ в (1) та $a=\upsilon$, $b=c=n=1$ у (9) ці залежності співпадають повністю: $f(y,\upsilon)=S^I(y;\upsilon)=\dfrac{\upsilon}{y+1}$. При $a=\upsilon$, $b=1$, $c=\alpha$, $n=\beta$ маємо замість (9): $f(y,\upsilon)=\dfrac{\upsilon}{\alpha y^\beta+1}$. Значення цієї величини перевищують значення (1) при всіх α, β та y, а при $\beta=1$ співпадають. Таким чином, модель [6] із цільовою функцією (7) тим ближча до моделі ГЛ із цільовою функцією (1), чим ближчі α та β у (7) до одиниці.

Зазначимо, що чим більше значення n у (7), тим більший інтервал Δy, в якому $f(y)$ залишається максимальним («поличка» в області $y\gtrsim 0$ на рис. 1, *а*), де $f(y)=0{,}9/(1+cy^n)$. Таким чином, величину n можна сприймати як один із показників вразливості системи. Криві $f(y)$ з різними значеннями n перетинаються в точці $y=1$. Положення цієї точки по осі $f(y)$ визначається величиною c. Величина n впливає на кривизну залежностей $f(y)$, причому їх опуклість в області $y=0..1$ при $n>1$ направлена вгору, при $n\le 1$ — вниз. Така різноманітність форми надає певну свободу дій після накопичення достатньої кількості статистичних даних, на основі яких з'явиться можливість встановити форми залежності $f(y)$ для реальних ситуацій.

У залежностях $S^I(y)$ «поличка» з $S = \text{const}$ відсутня, як і точка перетину кривих, але криві за формою досить схожі (рис. 1, *б*), де $S(y) = 0{,}9/(\alpha y + 1)^{\beta}$. Вплив параметрів α та β на їх крутизну видно з рис. 1

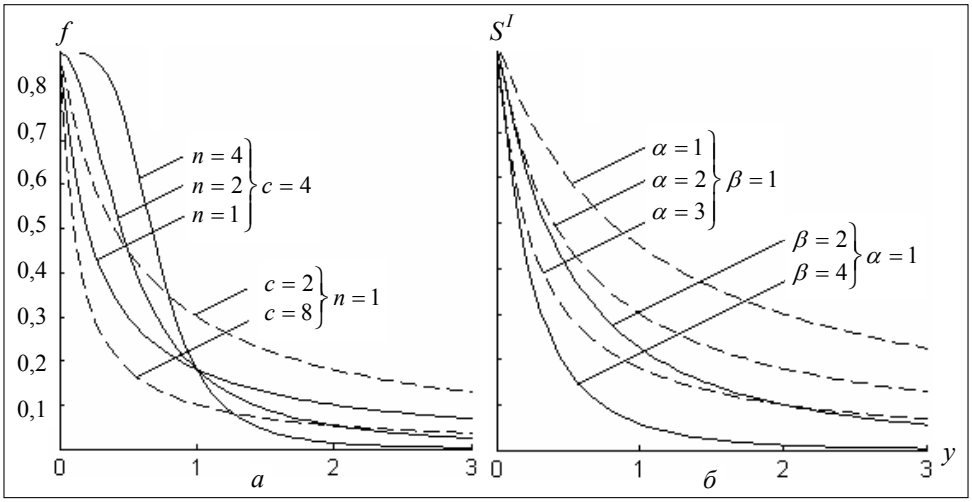

Рис. 1. Порівняння степеневих функцій $f(x;y)$ та $S^I(y;\upsilon)$

Порівняння залежностей $S^{II}(y;\upsilon)$ та $f(x;y)$ викликає ще більший інтерес, оскільки в [2], спираючись на статистичні дані [7], зазначається, що функції $S^{II}(y;\upsilon)$ знайшли своє емпіричне підтвердження. Хід кривих $f(x;y)$ під час використання показникових функцій (8) у вигляді $f(y) = 0{,}9(1 - e^{-m/y^n})$, де $x = 1$, а $\upsilon = 0{,}9$ (при $\upsilon = 1$ (2) втрачає сенс) видно з рис. 2, *а*. Для цих кривих також характерна «поличка» і точка перетину кривих із різним n при $y = 1$, в якій значення $f(y)$ визначається величиною m. Параметр n впливає на крутизну $f(y)$. Вплив параметра α на хід кривих $S^{II}(y;\upsilon)$ під час використання залежності (2) показано на рис. 2, *б*, де

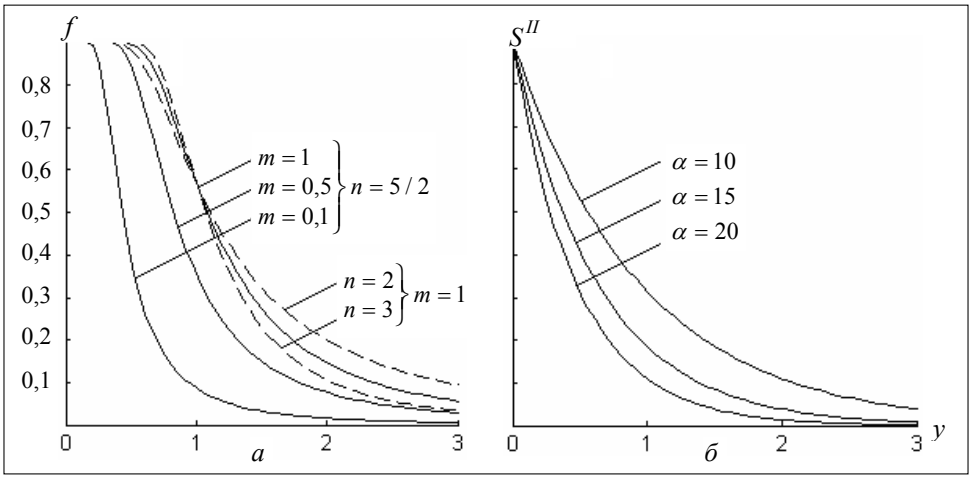

Рис. 2. Порівняння показникових функцій $f(x;y)$ та $S^{II}(y;\upsilon)$

$S(y,\upsilon)=\upsilon^{\alpha y+1}$, $\upsilon=0{,}9$. Для більш наочного співставлення результатів, розрахованих по залежностям (1) і (7) та, відповідно, (2) і (8) на рис. 3 наведене попарне порівняння цих залежностей, де суцільними і штриховими лініями зображено функції $f(y)$ при різних значеннях параметрів: на рис. 3, *а* $f(y)=0{,}9/(1+cy^n)$, на рис. 3, *б* $f(y)=0{,}9/(1-e^{-m/y^n})$, а штрих-пунктирними — функції $S(y)=0{,}9(\alpha y+1)^{\beta}$ на рис. 3, *а* та $S(y)=0{,}9^{\alpha y+1}$ на рис. 3, *б*.

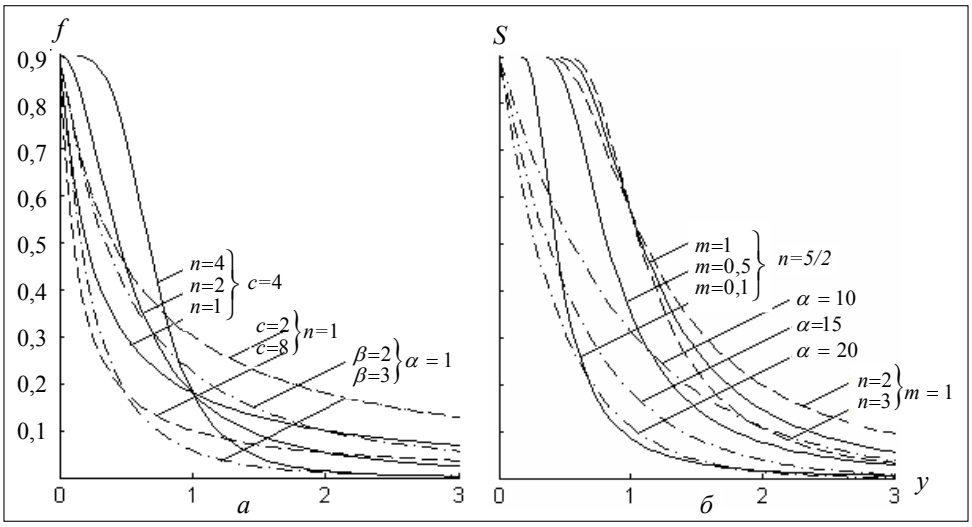

Рис. 3. Попарне порівняння степеневих та показникових функцій $f(x;y)$ та $S(y;\upsilon)$

Спробуємо підібрати параметри в (8), а також у (7) так, щоб ці залежності були в максимальній степені схожі на (2). Покладемо $\upsilon=0{,}9$, а параметр α в (2) виберемо з таких міркувань. Вважатимемо, що ресурси захисту y (відношення коштів, виділених на захист, до вартості інформації) знаходиться в межах $0..0{,}2$, а кількість вилученої інформації (тобто значення функцій $S(y;\upsilon)$ та $i(x;y)$), яка є небезпечною для підприємства, становить 15–20 % і катастрофічною — 20–30 %. Для вилучення такої кількості інформації для суперника буде доцільним виділити ресурс $(x/y)<3$. При цьому вважаємо, що при потрійному перевищенні ресурсів нападу $((x/y)\approx 3)$ суперник здатен вилучити значну кількість інформації. Значення $S^{II}(y;\upsilon)$, які відповідають цим міркуванням, наведені в табл. 1 і досягаються при $\alpha=30$:

$$S^{II}(y;v)=0{,}9^{30y+1}. \qquad (10)$$

Таблиця 1. Значення величин $S^{II}(y;\upsilon)$, $f_1(x;y)$, $f_2(x;y)$ при різних y та $x=1$

y	0	0,3	0,6	0,9	1,2	1,5	1,8	2,1	2,4	2,7	3
$S^{II}(y;\upsilon)$	0,9	0,348	0,135	0,052	0,020	0,007	0,003	0,001	0,0005	0,0002	0,00007
$f_1(x;y)$	0,9	0,500	0,120	0,045	0,022	0,012	0,008	0,005	0,004	0,003	0,0023
$f_2(x;y)$	0,9	0,453	0,136	0,054	0,027	0,016	0,010	0,007	0,005	0,003	0,002

Близькі до наведених значень $S^{II}(y;\upsilon)$ величини (які практично співпадають у найбільш важливому інтервалі $y = 0{,}5..1$, що відповідає $(x/y) = 1..2$) досягаються для показникової функції (8) у вигляді

$$f_1(y) = 0{,}9(1 - e^{-0{,}04(1/y)^{5/2}}) \tag{11}$$

і степеневої функції (7)

$$f_2(y) = 0{,}9\frac{(1/y)^{5/2}}{(1/y)^{5/2} + 20}, \tag{12}$$

значення цих функцій подано в табл. 1.

У вибраному інтервалі значень y степенева функція (12) ближча до функції $S^{II}(y;\upsilon)$ (10), ніж показникова функція (11). Це дозволяє припустити, що за певних значень параметрів степеневі функції (7) теж будуть відповідати емпіричним залежностям.

Врахуємо тепер імовірності $q(x,y)$, які оберемо у вигляді розподілу Максвела і використаємо в трьох формах:

$$q_1(x,y) = 2{,}26\left(\frac{x}{y}\right)^2 e^{-\left(\frac{x}{y}\right)^2}, \quad q_2(x,y) = 0{,}68\left(\frac{x}{y}\right)^2 e^{-0{,}44\left(\frac{x}{y}\right)^2},$$

$$q_3(x,y) = 0{,}36\left(\frac{x}{y}\right)^2 e^{-0{,}25\left(\frac{x}{y}\right)^2},$$

де максимум функції досягається при значенні $x_{1m} = 1$, $x_{2m} = 1{,}5$ та $x_{3m} = 2$ відповідно, а нормовочний коефіцієнт визначається з умови $\int_0^3 q_j(x)dx = 1$, $j = \overline{1,3}$. Під час розрахунку кількості вилученої інформації використаємо (7) у формі залежності від x (тобто у виразі (7) покладемо y рівним одиниці). Значення $i_j(x) = q_j(x)f_2(x)$, де $f_2(x) = 0{,}9\frac{x^{5/2}}{x^{5/2} + 20}$, наведені в табл. 2.

Таблиця 2. Кількість вилученої інформації з врахуванням імовірності $q(x)$ виділення нападом ресурсів x

$i_j(x)$ \ x	$x = 1$	$x = 1{,}5$	$x = 2$
$i_1(x)$	0,0356	0,0584	0,0328
$i_2(x)$	0,0187	0,0619	0,0928
$i_3(x)$	0,0120	0,0503	0,1051
$\bar{i}(x)$	0,0221	0,0569	0,0769

Середні значення $\bar{i}(x)$ показують кількість вилученої інформації при кожному значенні x за умови, що приведені залежності $q_j(x)$ рівноімовірні. Якщо в результаті експертної оцінки виявиться, що одна із залежностей

$q_j(x)$ має більш ваговий коефіцієнт c_j, ніж інші, то розподіл $i(x) = \sum_j c_j i_j(x)$ зміститься, очевидно, у бік цієї залежності.

ВИСНОВКИ

Наведені значення дозволяють менеджменту підприємства зробити висновок щодо достатності виділених коштів чи доцільності їх збільшення. Це залежить, звичайно, від допустимих величин $i(x, y)$, які, у свою чергу, визначаються з суб'єктивної оцінки топ-менеджера та його схильності до ризику. Остання ознака змушує дослідників розробляти різні моделі для випадків суворого обмеження ризиків і певного рівня їх допустимості [3]. При цьому модель ГЛ вважається побудованою на базі нейтрального відношення до ризику, за якого розглядається лише кінцевий результат (рівень необхідних інвестицій) без врахування рівня ризику.

Зазначимо, що подібна задача — мінімізація сумарних втрат (від витоку інформації та витрат на її захист, що рівноцінно підходу [3]) — розглядалась у [8], де проаналізовано загальні властивості цільової функції, і на основі аналізу зроблено висновки щодо максимально доцільного відсотку витрат на захист інформації. Порівняння методики в [8] із наведеними моделями потребує окремого дослідження.

ЛІТЕРАТУРА

1. *Gordon L.A., Loeb M.P.* The Economics of Information Security Investment, ACM // Transactions on Information and System Security. — 2002. — **5**, № 4. — P. 438–457.
2. *Matsuura K.* Productivity Space of Information Security in an Extension of the Gordon-Loeb's Investment Model // The Seventh Workshop on the Economics of Information Security (WEIS), 25–28 June, Hanover, USA. — 2008. — http://weis2008.econinfosec.org/papers/Matsuura.pdf .
3. *Huang C.D., Hu Q., Behara R.S.* Economics of Information Security Investment in the Case of Simultaneous Attacks // Proceeding of the Fifth Workshop on the Economics of Information Security, 26–28 June, Cambridge, England. — 2006. — http://weis2006.econinfosec.org/docs/15.pdf .
4. *Willemson J.* Extending the Gordon and Loeb Model for Information Security Investment, The Fifth International Conference on Availability, Reliability and Security ARES 2010 // Instityte of Electrical and Electrinics Engineers. — 2010. — http://research.cyber.ee/~jan/publ/aresGL.pdf .
5. *Tatsumi K., Goto M.* Optimal Timing of Information Security Investment: A Real Options Approach // The Eighth Workshop on the Economics of Information Security (WEIS 2009), UK, 24–25 June, University College London. — 2009. — P. 211–228. — http://weis09.infosecon.net/files/112/paper112.pdf .
6. *Левченко Є.Г., Рабчун А.О.* Оптимізаційні задачі менеджменту інформаційної безпеки // НТЖ «Сучасний захист інформації». — 2010. — № 1. — С. 16–23.
7. *Liu W., Tanaka H., Matsuura K.* Empirical-Analysis Methodology for Information-Security Investment and Its Application to a Reliable Survey of Japanese Firms // Information Processing of Japan Digital Courier. — 2007. — **3**. — P. 585–599.
8. *Задірака В.К., Олексюк О.С., Смоленюк Р.П., Штабалюк П.І.* Фінансування витрат на захист інформації в економічній діяльності // Університетські наук. зап. — 2006. — № 3–4 (19–20). — С. 479–490.

Надійшла 27.04.2010

УДК 528.852:531.012

ПРОГНОЗИРОВАНИЕ ЭРОЗИОННЫХ ПРОЦЕССОВ ПОЧВ НА МАРШРУТАХ ПРОКЛАДКИ МАГИСТРАЛЬНЫХ ТРУБОПРОВОДОВ НА ОСНОВЕ ГЕОИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И КОСМИЧЕСКИХ СНИМКОВ ВЫСОКОГО РАЗРЕШЕНИЯ

А.Т. ДЖАМАЛОВ, Р.М. РАГИМОВ

Исследуется возможность прогнозирования эрозионных процессов в грунте на маршрутах прокладки магистральных трубопроводов (МТ), что позволяет выбрать наиболее оптимальный маршрут прокладки МТ с минимальным риском. Составлено Универсальное уравнение потерь почвы (УУПП), определены все переменные, входящие в УУПП. Получена растровая карта модели УУПП для выбранного тестового участка и таблица оценок интенсивности потерь почвы.

ВСТУП

В результате дешифрования результатов аэрокосмической съемки решаются разные задачи, важнейшими из которых, на наш взгляд, являются: выявление аварийно-опасных участков и неблагоприятных природных явлений вблизи магистральных трубопроводов (МТ), развитие эрозионных и склоновых процессов, воздействие подземных и поверхностных вод, а также различные смещения, определяемые действиями гравитационных и других сил.

В данной работе исследуется возможность оценки риска эрозионных процессов на участках прокладки МТ на основе геоинформационных технологий и космических снимков высокого разрешения.

Карты эрозионно-опасных земель являются основой при решении задач по защите инфраструктур, в частности, технических коридоров прокладки магистральных нефтегазопроводов и должны давать объективное представление о распространении и степени эрозионной опасности земель, а также отвечать ряду требований, зависящих от стадии проектно-изыскательских работ. Решение этой проблемы достигается количественной оценкой основных факторов эрозии и использованием математических зависимостей между процессами и факторами их определяющими.

Для оценки эрозионной опасности могут использоваться различные модели эрозии, для которых должна быть создана соответствующая справочно-информационная база, охватывающая исследуемую зону региона.

Картирование почвенной эрозии, при использовании ГИС, дает возможность идентифицировать зоны потенциального риска интенсивной почвенной эрозии и позволяет дать количественную оценку потери почв в различных участках технического коридора для прокладки МТ.

Геоинформационные технологии основываются на использовании пространственной информации, как основы для моделирования различных процессов и объектов в геотехнической системе, включая эрозионные процес-

© *А.Т. Джамалов, Р.М. Рагимов*, 2011

сы. В последнее время были разработаны (в комбинации с ГИС программами) несколько моделей почвенной эрозии [1], одной из которых является Уточненное универсальное уравнение потерь почвы (УУПП).

В УУПП эродируемость определяется двумя группами факторов. К первой относятся основные характеристики почвы, такие как гранулометрический состав, содержание органического вещества, инфильтрационная способность и другие. Ко второй группе факторов относятся обработка почвы до посадки растений и в период ухода за ними, рельеф местности, характер проводимых противоэрозионных мероприятий и т.д. (рис. 1).

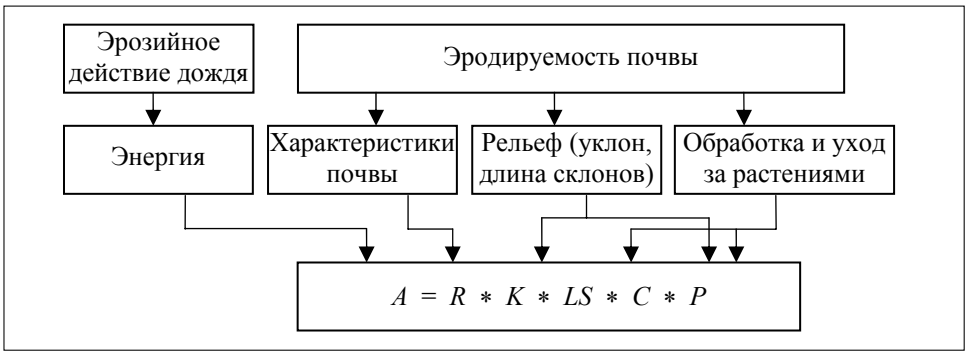

Рис. 1. Факторы, определяющие водную эрозию почв

Переменные, входящие в УУПП модель, описанную ниже, характеризуют почвенные данные для генерирования оценки риска эрозии.

На рис. 2–4 показаны примеры карт изучаемого региона (рис. 2 — векторизированная почвенная карта; рис. 3 — векторизированная климатическая карта; рис. 4 — цифровая модель рельефа).

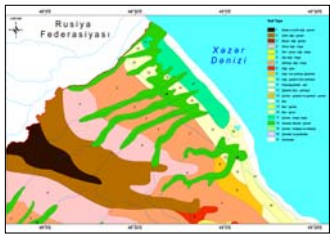

Рис. 2

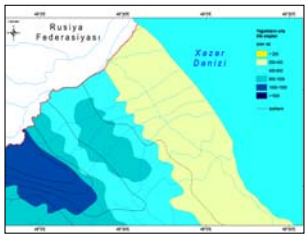

Рис. 3

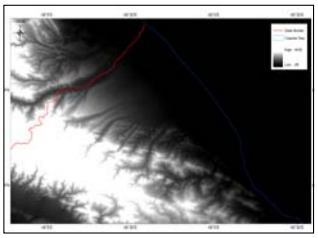

Рис. 4

ОБЗОР МЕТОДОВ РАСЧЕТА ЭРОЗИИ ПОЧВ. ЭМПИРИЧЕСКАЯ МОДЕЛЬ

В основе методов расчета склоновой водной эрозии лежит следующая зависимость:

$$A = f(R, K, L, S, C, P), \qquad (1)$$

где A — средние годовые потери почвы на единицу площади в год; R, K, L и S — индексы, учитывающие влияние энергии на величину смыва почв; интенсивности ливней (R), типа и состояния почвы (K), длины (L) и величины крутизны склона (S); C — индекс, отражающий почвенно-растительный покров; P — индекс влияния противоэрозионных почвозащитных мероприятий.

Модель УУПП позволяет прогнозировать эродируемость почв и эффективна при определении пространственных изображений потерь почвы в исследуемом регионе. При этом ГИС позволяет выделить и определить роль каждой переменной выражения (1) при определении среднегодовой эрозии почв, которая определяется как

$$A = R * K * L * S * C * P, \qquad (2)$$

где $*$ — функция умножения.

Переменные уравнения (2) могут быть разделены на две части [2]:

- R, L, S и K факторы — переменные окружающей среды (переменные относительно стабильные по времени).
- C и P факторы — переменные управления и их изменения носят сезонный характер (могут в течение года достаточно часто меняться).

УУПП модель имеет ряд ограничений, которые условно можно разделить на три категории:

- расположение исследуемого участка, в котором применяется УУПП;
- ограничения, связанные с математическими вычислениями;
- ограничения в масштабе.

Для проверки функционирования эмпирической модели и оценки процессов эрозии был выбран тестовый участок, расположенный в бассейне реки Куручай (Куба-Кусарский район) Азербайджана.

Бассейн реки Куручай имеет площадь приблизительно 19947 га (рис. 5) и, в свою очередь, состоит из нескольких водоразделов, которые имеют площадь от 282 до 12641 га. Выбранный в качестве тестового участок имеет площадь порядка 8975 га.

Обоснованием выбора данного тестового участка (ТУ) являются следующие причины: доступность пространственной и текстовой информации по почвам; различные виды почвенно-растительного покрова; топографические изменения по высоте и по наклону.

Данные модели почвы включают: почвенные данные; данные по рельефу; состояние почвенно-растительного покрова; данные о водных ресурсах и количестве атмосферных осадков; данные по почвенно-защитным мероприятиям; данные о границах водных бассейнов.

Для исследования тестового участка мы использовали космический снимок высокого разрешения со спутника Quick Bird (рис. 6). В результате дешифрирования космического снимка, дополненного полевыми исследованиями была составлена легенда тестового участка, включающая дороги, кустарники, огороды, водохранилища, русла рек, сельскохозяйственные угодья, леса, жилые массивы (рис. 7).

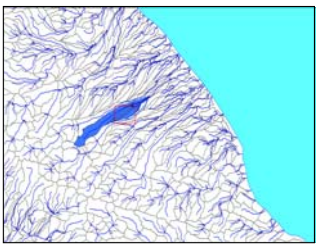

Рис. 5. Пример карты бассейна реки Куручай с тестовым участком

Рис. 6. Пример QuickBird снимка тестового участка

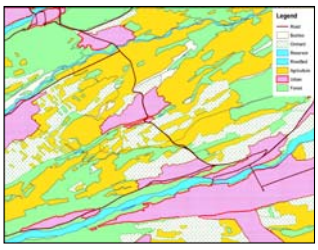

Рис. 7. Пример почвенно-растительного покрова тестового участка

Для исследования процесса эрозии почв на выбранном тестовом участке на маршруте прокладки магистрального газопровода необходимо составить эмпирическую модель расчета УУПП (уравнение 2) и определить значения входящих в нее переменных параметров.

Для генерации *K*-фактора (подверженность эрозии почвы) использовалась пространственная и справочная информация о составе почв данного ТУ.

Составленная цифровая модель рельефа (ЦМР) с разрешением в плане 75 м, дает информацию о высоте, что позволяет определить L и S факторы (длина и крутизна склона).

Цифровая информация о почвенно-растительном покрове, являющимся решающим фактором влияющим на эрозию почв, в соответствии с MLCCS (Minnesota Land Cover Classification System — система классификации растительного покрова штата Миннесота), использовалась для генерации C фактора. Данные о количестве атмосферных осадков использовались для генерации R фактора (дождевые и другие виды осадков). Информация об P факторе (почвенно-защитные мероприятия) для данного ТУ отсуствовала.

Для генерации модели УУПП была спроектирована ГИС УУПП с пространственным разрешением 75 м на основе программного обеспечения ESRI (Environmental Systems Research Institute — пакет программ, разработанный Институтом по исследованиям систем окружающей среды).

Для реализации проекта использовалась географическая проекция UTM, Zone 39N и все исходные данные были переведены в эту проекцию для дальнейшей обработки.

Следующим этапом была осуществлена процедура преобразования картографических слоев данных из векторного формата в растровый с помощью программы Spatial Analyst. Выходные растровые файлы были генерированы с пространственным разрешением 75 м (рис. 8 и 9).

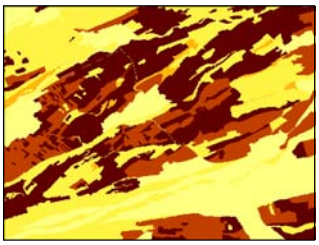

Рис. 8. Пример растра MLCCS, соответствующий значению *C*-фактора (почвенное покрытие)

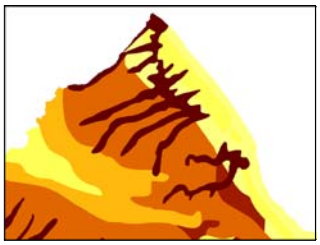

Рис. 9. Пример растра STATSGO, соответствующий значению *K*-фактора (эродируемость почвы)

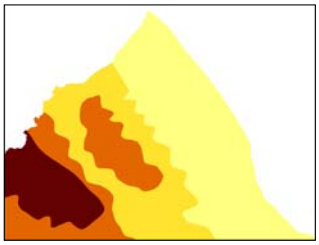

Рис. 10. Пример растра *R*-фактора, соответствующий значению атмосферных осадков

На рис. 10 показан растр *R*-фактора, соответствующий значению количества атмосферных осадков, на рис. 11 — растр *S*-фактора, соответствующий значению крутизны склона.

Были использованы специальные функции программы (ESRI) для создания выходного растра необходимого для определения направления водных течений исследуемого участка и для определения участков скопления воды. Растр сгенерирован размером пикселя в 75 м (рис. 12).

После выполнения описанных шагов, выходные данные о состоянии почвенно-растительного покрова, типе почвы, R-фактор, P-фактор, L и S факторы были включены в УУПП.

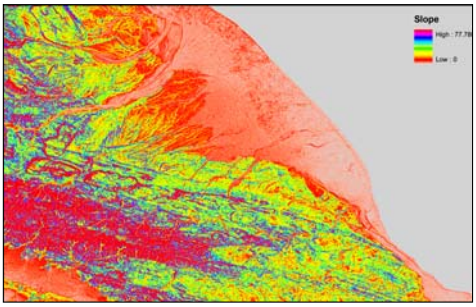

Рис. 11. Пример растра крутизны склона

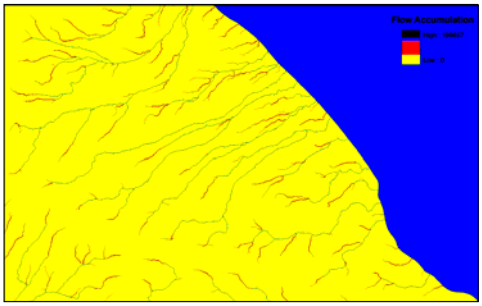

Рис. 12. Пример растра направления течений

Для преобразования векторных данных о почвенно-растительном покрове, типе почв, R и P факторах в каждый картографический слой добавлялись атрибутивные данные, которые содержат соответствующие значения фактора для каждого ТУ.

Значения C-фактора (эродируемость почв определенная в зависимости от типа покрытия) были выбраны из табл. 1 классификации MLCCS в соответствие с картой почвенно-растительного покрова выбранного тестового участка.

C-фактор имеет цифровое значение от нуля до единицы, и при значениях, близких к нулю, почва имеет низкий показатель эродируемости.

K-фактор, соответствующий типу и состоянию почвы, был выбран из табл. 2 согласно классификации STATGO (State Soil Geographic — Государственная географическая база данных по почвам).

Таблица 1. MLCCS классификация почвенно-растительного покрова и соответствующие им значения C-фактора

MLCCS	
Land Cover Type	*C* Factor
Cattail marsh permanently flooded	0,001
Coniferous forest	0,001
Cultivated herbaceous vegetation	0,320
Deciduous forest	0,002
Deciduous shrubland	0,025
Fruit trees on upland soils	0,110
Floodplain forest	0,010
Floodplain forest swamp while oak subtype	0,015
Grassland with sparse deciduous trees	0,010
Hydric soil row cropland	0,470
Lowland hardwood forest	0,001
Maple – basswood forest	0,002
Medium – tall grassland	0,012

Продолжение табл. 1

Medium – tall grassland altered/non-native dominated	0,015
Mixed coniferous-deciduous forest	0,001
Mixed coniferous-deciduous woodland	0,002
Oak forest with 4 % to 10 % impervious cover	0,004
Oak forest	0,002
Oak forest dry subtype	0,002
Oak woodland-brushland	0,005
Paulstrine open water	0,000
Planed or maintained grasses	0,140
Planed or maintained grasses with sparse tree cover	0,100
Planed, maintained or cultivated deciduous trees	0,070
Sand gravel pits with 0% to 10% impervious cover	0,000
Unconsolidated material (soil, sand and ash)	1,000
Upland coniferous forest	0,001
Upland coniferous woodland	0,001
Upland deciduous forest	0,002
Upland deciduous shrubland	0,005
Upland deciduous woodland	0,002

Таблица 2. Типы почв согласно USDA NRCS STATGO и соответствующие им значения *K*-фактора

STATGO Soil Dataset	Soil Type	MN240	MN216	MN221	MN225	MN212	MN216	MN228	MN214	MN240
	K Factor	0,3440	0,3240	0,3730	0,3340	0,3030	0,3240	0,1840	0,2950	0,3440

МОДЕЛИРОВАНИЕ УУПП

После того, как значения *C*, *K*, *R* и *P* факторов были добавлены в таблицы атрибутов и картографические слои MLCCS, STATSGO, факторы длины и крутизны склона и водных скоплений преобразованы из векторной формы в растровый формат, производится вычисление модели УУПП.

L и *S* факторы (длина и крутизна склона) были вычислены, используя ранее генерированные растры склона и водных скоплений. Очевидно, что чем длиннее и круче склон, тем большее значение имеет фактор аккумуляции поверхностного стока, что приводит к повышенной эрозии почв.

Для вычисления *LS*-фактора было использовано эмпирическое уравнение, опубликованное в USDA Agriculture Handbook № 537 [3] с соответствующими добавлениями [4].

Эмпирическое уравнение, использованное для вычисления *LS*-фактора в нашем проекте имеет вид:

$$LS = (\text{Flow Accumulation grid} * \text{cell size} / 22{,}13)^{0{,}4} * 0{,}01745) / 0{,}0896)^{1{,}4} * 1{,}4 .$$

На рис. 13 показана растровая карта длины и крутизны склона.

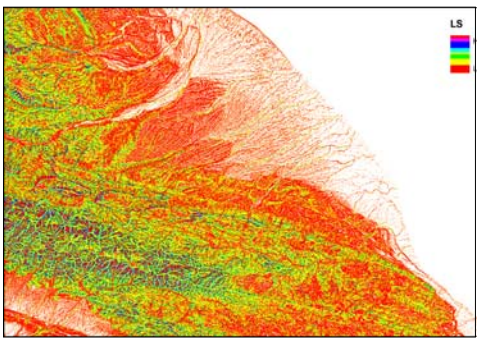

Рис. 13. Пример растра длины и крутизны склона

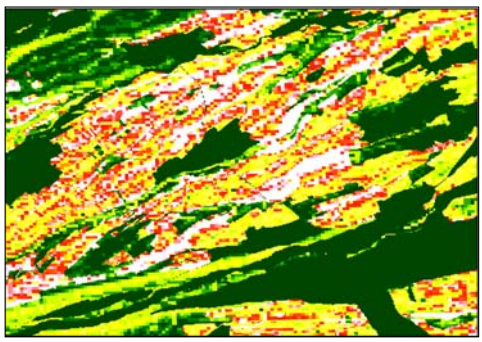

Рис. 14. Пример растра карты риска эрозии почвы вызванной водными накоплениями

ВЫВОДЫ

1. Конечным результатом исследований является полученный растр модели УУПП для определения риска эрозии почв в водосборе реки Куручай. Результат был использован при проведении проектных работ по прокладке северного маршрута магистрального газопровода (рис. 14).

2. Получены оценки интенсивности потерь количества почвы на тестовом участке (табл. 3).

Таблица 3. Примерная оценка интенсивности потерь почвы на тестовом участке

Диапазон интенсивности смыва, тонн/га·год	Площадь, га	Среднее количество смываемой почвы, т
< 0,5	3547,69	886,92
0,5 – 1,0	445,63	334,22
1,0 – 5,0	1074,31	2685,78
5,0 – 10,0	716,31	5372,34
10,0 – 25,0	1269,75	22220,63
25,0 – 50,0	894,5	33543,75
50,0 – 75,0	365,19	22824,22
75,0 – 100,0	218,88	19151,56
> 100,0	442,88	44287,50

Для более детального анализа распространения эрозии требуется ЦММ с более высоким пространственным разрешением из-за соизмеримости длины склонов с пространственным разрешением исходной цифровой модели местности, которое составляет 75 м.

ЛИТЕРАТУРА

1. *Wilson J.P.* GIS-based Land Surface/Subsurface Modeling: New Potential for New Models. — 2003. — 349 p.
2. *Hickey R., Burns E.* Development of a Statewide Erosion Vulnerability Screening Tool for Oregon: NY, 2005. — 450 p.
3. *Wischmeier W.H., Smith D.D.* Predicitng Rainfall Erosion Losses: A Guide to Conservation Planning // Agriculture Handbook U.S. Department of Agriculture. — Washington, DC. — 1978. — № 537. — 543 p.
4. *Moore I., Burch G.* Physical Basis of the Length-Slope Factor in the Universal Soil Loss Equation // Soil Science Society of America Journal. — 1986. — № 50:1294–1298. — 250 p.

Поступила 22.06.2010

Данная статья публикуется в редакции автора.

УДК 681.3.00:007

ВІД КОВАРІАЦІЙ ДО КАУЗАЛЬНОСТІ. ВІДКРИТТЯ СТРУКТУР ЗАЛЕЖНОСТЕЙ У ДАНИХ

О.С. БАЛАБАНОВ

Проаналізовано сучасну методологію виводу каузальних моделей та структур систем імовірнісних залежностей із статистичних даних пасивних спостережень. Висвітлено можливості, проблеми, застереження та обмеження методів індуктивної ідентифікації каузальних відношень в апараті марковських властивостей та баєсових мереж. Виділено кілька ступенів каузальних моделей згідно з рівнем їх обґрунтованості та адекватності джерелу даних. Сформульовано статистичний паттерн, який зводить обґрунтування висновку про каузальний характер зв'язку двох змінних до тестування набору статистичних фактів (не)залежності.

ВСТУП

Серед задач, що розглядаються в контексті методології виділення знань з даних, мабуть, найцікавішою та когнітивно значущою видається ідентифікація каузальних відношень на основі апарату багатовимірних імовірнісних моделей залежностей та марківських властивостей [1–8]. Можливість ідентифікації каузальних відношень у такий спосіб викликала гострі заперечення. Деякі підсумки дискусії підведені в [8, 9]. Традиційно в емпіричних науках каузальні відношення встановлювали через контрольовані рандомізовані експерименти, тобто на основі даних активних експериментів. Натомість методи виділення знань з даних, як правило, використовують дані пасивних спостережень. Задача виведення каузальних моделей саме з даних пасивних спостережень є дуже привабливою й актуальною для багатьох сфер діяльності та досліджень, де активні експерименти неможливі практично або невиправдані економічно, або з етичних міркувань. Така ситуація постає, приміром, у соціально-економічних та медико-біологічних дослідженнях, в астрономії, метеорології, геології та іноді навіть у техніці. Крім того, може бути невідома темпоральна (часова) послідовність змінних. Новітня емпірико-індуктивна методологія [1, 3, 5, 8] відкриває реальні можливості розв'язання задач за таких умов.

Каузальні моделі посідають, так би мовити, найвищий рівень в ієрархії емпіричних моделей. В англомовній літературі каузальними мережами (діаграмами) називають імовірнісні моделі на основі ациклічних орієнтованих графів (АОГ), або їх розширення. Синтаксис АОГ-моделей дозволяє виразити каузальну семантику. Однак питання полягає в іншому: чи можна (без

© *О.С. Балабанов*, 2011

апріорних знань) вивести з даних АОГ-модель, яка адекватно відображає каузальні відношення в предметній галузі, звідки взято статистичні дані.

Мета роботи — характеристика можливостей і обмежень методів виведення каузальних моделей з даних пасивних спостережень.

Провадиться не зовсім строга, але корисна класифікація рівня обґрунтованості висновків про причинно-наслідковий характер виведених зв'язків у моделі.

Оскільки зазначена тематика розвивалася кількома школами і до неї залучилися різні наукові спільноти — філософи, статистики, інформатики, дослідники з природничих та соціальних дисциплін, то в літературі існують розбіжності в трактуванні визначень і термінології. Намір автора — стисло і цілісно подати принципи та ключові положення предмета, а також вказати на менш відомі результати і запропонувати деякі уточнення.

Адекватність і точність каузальної моделі, виведеної з даних, детермінується двома складовими (компонентами) методу виведення. Перша складова — це базові процедури аналізу даних і тестування умовних незалежностей. Друга складова, яка і є предметом розгляду — «логічний» аналіз та синтез статистичних відношень, ідентифікованих першою компонентою.

Елементами синтаксису моделі є: вершина (змінна), дуга $X \to Y$ і ребро $X — Y$. Ребро — це дуга, орієнтація (спрямування) якої невідома або несуттєва в цьому контексті. Введемо якісні градації рівня адекватності каузальної моделі. Тобто, будемо виділяти кілька ступенів (ґатунків) моделі та її фрагментів згідно з адекватністю відображення каузальних відношень об'єкта моделювання. Пропонуємо такі градації моделей: *псевдокаузальні*, *протокаузальні*, *субкаузальні* та *каузальні*. Ґатунок моделі (рівень адекватності) визначається такими властивостями:

- надійність ідентифікації ребер моделі (топології статистичних зв'язків);
- обґрунтованість і надійність ідентифікації орієнтацій ребер моделі (спрямування впливів);
- обґрунтованість інтерпретації ребра моделі як каузального зв'язку визначеного напрямку.

Якщо всі ребра моделі адекватно відображають топологію безпосередніх зв'язків (залежностей) між змінними (відношення суміжності), будемо називати модель протокаузальною. Якщо при цьому ще і вістря орієнтованих дуг адекватно відображають автентичні напрямки залежностей (впливів) між змінними, то називатимемо модель субкаузальною. При цьому зв'язок змінних $X \to Y$ інтерпретується так: або змінна X, або якась асоційована з нею H має каузальний вплив на Y, а змінна Y певно не має каузального впливу на X. І нарешті, якщо пред'явлено переконливі аргументи, що змінна X каузально впливає на Y, то називатимемо модель каузальною (у цій частині моделі).

Припустимо, що домен («фрагмент світу»), який моделюється, описується системою ймовірнісних залежностей, структурованих ациклічним орієнтованим графом, тобто у формі АОГ-моделі. Ор-ациклічність каузальних структур виправдовується фізичною інтуїцією. Клас АОГ-моделей призначений для репрезентації систем усталених залежностей (на макрорівні).

АОГ-модель залежностей визначена як (G,θ), де G — АОГ, причому кожній змінній відповідає своя вершина графу, а θ — сукупність локально заданих параметрів. Параметри задано у формі $p(X \mid F(X))$, де $F(X)$ — множина всіх «батьків» для X, тобто тих вершин Z, від яких прямує дуга $Z \to X$. Якщо є орієнтований шлях $R \to \cdots \to W$, то змінна R називається «предком» для W, а W — «нащадком» для R. Прийнято виокремлювати три випадки (різновиди) АОГ-моделей. Коли модель визначена на дискретних (номінальних) змінних, її називають баєсовою мережею. Гаусові мережі визначено на дійсних змінних і лінійних залежностях (за нормально-розподілених відхилень). Подекуди їх називають також рекурсивними системами лінійних структуральних рівнянь. Гібридні мережі включають змінні різних типів. Для відображення каузальних відношень реальних систем потрібно враховувати можливість латентних змінних. При існуванні латентних змінних модель формально виходить за межі АОГ-моделей і стає певним розширенням, зокрема, класом моделей на основі анцестральних графів [10]. Латентна змінна відображається у моделі біорієнтованим ребром $X \leftrightarrow Y$.

Наявні дані є багатомірною статистичною відбіркою, яку можна розглядати як типову відбірку, генеровану з адекватної АОГ-моделі. Ідентифікувати модель — означає відтворити з наявних даних генеративну модель (з точністю до класу еквівалентності). Оскільки нас цікавить відкриття якісних відношень («знань»), доцільно спиратися на сепараційний (constraint-based) підхід до ідентифікації структур моделі. Цей підхід базується на тестах умовної незалежності і дає обґрунтування кожного фрагмента моделі. До найвідоміших методів і алгоритмів цього підходу відносяться алгоритми PC, IC, SGS та FCI [1–8]. Методи сепараційного підходу виводять модель у три етапи. На першому етапі ідентифікують усі ребра моделі, на другому — визначають орієнтації (спрямування) ребер, на третьому — обчислюють параметри. Стислий опис деяких методів можна знайти, зокрема, у [11–13].

ІДЕНТИФІКАЦІЯ ПРОТОКАУЗАЛЬНИХ ВІДНОШЕНЬ

Протокаузальна модель має адекватно відображати топологію безпосередніх залежностей (інформаційних зв'язків) між змінними в системі ймовірнісних залежностей, тобто сукупність ребер моделі. Ідентифікація ребер ґрунтується на виявленні відношень умовної незалежності. Умовну незалежність X від Y за фіксації (кондиціонування) набору змінних $\mathbf{Z}$ позначатимемо формулою $\mathrm{Ind}(X;\mathbf{Z};Y)$, де $X,Y \notin \mathbf{Z}$. Така незалежність означає, що $p(Y \mid X,\mathbf{Z}) = p(Y \mid \mathbf{Z})$, або еквівалентно $p(XY \mid \mathbf{Z}) = p(X \mid \mathbf{Z})p(Y \mid \mathbf{Z})$. Для систем лінійних залежностей умовну незалежність зазвичай виражають через частковий коефіцієнт кореляції $\rho_{XY.\mathbf{Z}} = 0$.

Нехай чинне твердження умовної незалежності $\mathrm{Ind}(X;\mathbf{S};Y)$. Тоді для взаємної інформації між X та Y за кондиціонування $\mathbf{S}$ легко отримати рівняння $\mathrm{Inf}(X;Y \mid \mathbf{S}) = 0$, а також $\mathrm{Inf}(Y;X\mathbf{S}) = \mathrm{Inf}(Y;\mathbf{S})$, $\mathrm{Inf}(X;Y\mathbf{S}) = \mathrm{Inf}(X;\mathbf{S})$. Це інтерпретується як відсутність безпосереднього зв'язку між X та Y. Отже, факти умовної незалежності відображають структуру інформаційних потоків (статичних зв'язків) між змінними.

Відношення умовної незалежності в моделі визначаються критерієм *d*-сепарації і ґрунтуються на каузальній марківській умові [1, 3, 5]. Резонно вимагати, аби модель відображала всі інформаційні та каузальні взаємозв'язки в явному вигляді своєю графічною структурою. Тобто, кожна умовна незалежність, чинна в розподіленні ймовірностей, має виникати тільки у випадку відсутності відкритого (активного) шляху між відповідними вершинами графу моделі. Цей принцип називають припущенням необманливості (faithfulness) сумісного розподілу ймовірностей [1], або вимогою мінімальності та стабільності моделі [3]. Коли всі відношення умовної незалежності репрезентовано в моделі графічно, модель задовольняє принципу структурно-поведінкої адекватності. Утім, для коректного виведення всіх ребер моделі з даних можна навіть послабити припущення необманливості: безпосередній зв'язок (ребро) між X та Y має бути наявний у моделі тоді і тільки тоді, коли змінні X та Y взаємозалежні в усіх контекстах (умовах). Але навіть таке послаблене припущення іноді порушується. Це трапляється за певних «маскувальних» властивостей співвідношень параметрів. Побіч того, інші обставини також можуть перешкоджати виведенню адекватних протокаузальних моделей. Причинами помилок можуть бути: недостатній обсяг даних; гамір та похибки у даних (зокрема, неточні вимірювання, округлення змінних); латентні змінні; невдалий вибір функціональної форми залежності.

Можна виділити такі форми неадекватності протокаузальної моделі: зайві (псевдокаузальні) ребра та шляхи; втрачені ребра та шляхи; не відображення часткового суміщення шляхів; неадекватне відображення емпіричної умовної незалежності (комбінація названих вище випадків).

Помилки суміжності через вади даних (синдром зайвих та «загублених» ребер). За малих відбірок даних (через значне ухилення ймовірностей) можуть порушуватись співвідношення, які виражають умовну незалежність. Тоді гіпотеза умовної незалежності може бути помилково відкинута під час тестування (ризик помилок збільшується в тестах високого рангу [1, 11]). У результаті структура моделі перенасичується зайвими (псевдокаузальними) ребрами. Крім того, фальсифіковане ребро може виникнути внаслідок округлення змінних чи загрублення шкали вимірювання. Припустимо, в автентичній моделі має чинність $\mathrm{Ind}(R;Z;W)$, але під час вимірювання чи запису дискретної змінної Z декілька значень було об'єднано (агреговано). Тоді у відбірці даних незалежність $\mathrm{Ind}(R;Z;W)$ може не виконуватись, і виникне зайве ребро $R — W$, аналогічне можливе для змінної Z дійсного типу. Зауважимо, що внаслідок округлення змінних можлива і протилежна помилка, коли ребро втрачається.

Обманні (превалюючі) асоціації та замасковані ребра. Багато методів статистичного аналізу даних розпочинають із обчислення парних коваріацій (асоціацій). У методах ідентифікації структур моделей без циклів суміжності (зокрема, в алгоритмі Chow&Liu та інших [14–16]) ребра встановлюють згідно з величинами парних асоціацій (кореляцій). Дійсно, якщо в структурі без циклів є пара ребер $(X — Y)$ та $(Y — Z)$, асоціація змінних X та Z (у типовій відбірці) буде слабшою за асоціації пар змінних (X, Y) та (Y, Z). Однак у структурах із циклами суміжності це правило не

діє, і опосередкована асоціація може перевершити будь-яку «реберну» асоціацію з-поміж тих, на яких вона ґрунтується [16, 17]. Наприклад, нехай маємо гаусову мережу, де змінні X та Y мають кілька спільних «батьків» $Z_1, Z_2, \ldots, Z_k$ (рис. 1). Легко підібрати параметри такої моделі, щоб коваріація змінних X та Y була в кілька разів більша, ніж для кожної пари змінних, що поєднані ребром. Зокрема, нехай ця модель описується рівняннями:

$$X = Z_1 + Z_2 + \ldots + Z_k + \varepsilon_X,$$

$$Y = Z_1 + Z_2 + \ldots + Z_k + \varepsilon_Y,$$

$$\varepsilon_X \sim N(0,\sigma);\ \ \varepsilon_Y \sim N(0,\sigma).$$

Тоді коваріація X та Y у k разів більша реберних. Для баєсової мережі з такою самою структурою можна підібрати такі параметри, що величина $\mathrm{Inf}(X;Y)$ наблизиться до максимально можливого значення (коли дві змінні статистично тотожні), у той час як усі $\mathrm{Inf}(X;Z_i)$ та $\mathrm{Inf}(Y;Z_i)$ можуть як завгодно зменшуватися і навіть досягати нуля. Тобто, ребра такої моделі не проявляються парними залежностями (і не ідентифікуються наївними методами). Це — синдром «зникаючих» безпосередніх зв'язків (замаскованих ребер). Утім моделі з таким синдромом параметрично нестабільні. Достатньо трохи змінити розподіли змінних Z_i, і ребра перестануть бути замаскованими (хоча можуть бути слабкими). Зауважимо, що розглянена структура (рис. 1) належить до підкласу монопотокових моделей і може бути відтворена за допомогою методу «Proliferator-D», який використовує тести умовної незалежності виключно першого та нульового рангу [16].

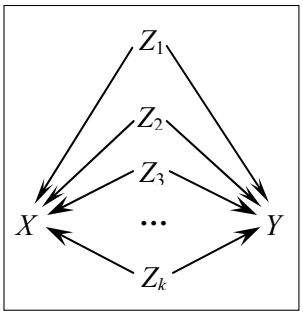

Рис. 1. Модель із «двійниковою» асоціацією

Ідемпотентні залежності, або ланцюг із «пересувними» ребрами. Нехай маємо дискретну модель ланцюгової структури $X \to Y \to Z$; тоді виконується $\mathrm{Ind}(X;Y;Z)$. У [18] показано, що можна так параметризувати цю структуру, що буде також виконувитися $\mathrm{Ind}(Y;X;Z)$. Остання незалежність відповідає структурі моделі у вигляді ланцюга Y — X — Z, де змінні розташовано в іншому порядку порівняно з тим, за яким насправді генероване сумісне розподілення ймовірностей. Більше того, сумісне розподілення ймовірностей (X, Y, Z) однаково добре узгоджується з ланцюговою структурою з будь-яким порядком розташування трьох змінних, зокрема, зі структурою Y — Z — X. Іншими словами, із трьох потенційно можливих ребер можна довільно обрати два. Можна уявити, що два ребра не мають певного місця. Звідси — образна назва «гумка-перевертень». Усього отримуємо 9 еквівалентних варіантів структури моделі. Утім, слід визнати, що така модель є аномальною і дещо штучною, бо передбачає спеціальні патерни значень параметрів. Аналогічний феномен (еквівалентність усіх послідовностей змінних у ланцюзі) також спостерігається у функціональних моделях (тобто для детермінованих залежностей). Однак там необхідна од-

накова значність усіх змінних, а також і взаємна однозначність функцій. А вказана модель цього не вимагає і має вільні параметри. Зауважимо, що таку модель можна замінити на структуру у вигляді «зірки» з латентною змінною в центрі [19].

ПРОЯВИ ЛАТЕНТНИХ ЗМІННИХ

У багатьох реальних дослідженнях задача виведення моделі з даних потерпає від нестачі релевантних змінних (факторів), тобто через відсутність деяких змінних у даних. Змінну, яка не репрезентована в даних, не у всіх випадках називають латентною. Неприсутню в даних змінну можна назвати: прихованою, оминутою або неспостережуваною. Апарат АОГ-моделей розраховано на ситуацію, коли кожна «оминута» змінна H_i впливає тільки на одну наявну змінну X_i і є незалежною від неприсутніх змінних H_j, які діють на інші модельні змінні. Такі неприсутні змінні не вважаються латентними. Латентна змінна — це неприсутня змінна, яка накладає суттєвий, якісний відбиток на розподіл імовірностей наявних змінних. Латентна змінна мусить впливати принаймні на деякі дві спостережувані змінні (тобто надходити до моделі двома дугами).

Можна виділити спеціальні випадки оминутої змінної згідно зі схемою її «контакту» з моделлю: «проміжна», коренева, комбінована змінна. Проміжна прихована змінна H у наведених прикладах (рис. 2, *а*, 2, *в*) не тягне ніяких змін марківських властивостей моделі. Але в багатьох випадках дуже бажано, аби змінна у проміжній позиції була наявна в даних. Це корисно для застосування моделі, зокрема, для квантифікацій величини (сили) каузального ефекту [3].

Власне, латентною змінною зазвичай вважають таку «невидиму» змінну, яка є кореневою та впливає на дві (чи більше) явних змінних моделі (рис. 2, *б,* 2, *г*). Але й така латентна змінна може не відбиватися на марківських властивостях (рис. 2, *б*). Латентна змінна в різних контекстах накладає різні паттерни залежностей. Коли латентна змінна формує фрагмент *W*-образної конфігурації (рис. 2, *г*), факт її існування проявляється статистично [7]. А саме, для випадку рис. 2, *г* латентна змінна ідентифікується на підставі нетранзитивності емпіричної залежності через ребро $X \text{—} Y$ в обох напрямках, тобто на підставі такої комбінації статистичних фактів: залежність пар змінних (X, R), (X, Y) та (Y, Z), і незалежність пар змінних (X, Z), (R, Y) та (R, Z).

Латентна змінна з комбінованою роллю, навіть у простому варіанті (рис. 3, *а*), породжує складності відображення відношень у моделі. Двореберний сурогат структури (рис. 3, *б*) для такої ситуації є неадекватним, тому що тягне умовну незалежність $\mathrm{Ind}(Y; X; Z)$, що не узгоджується з даними і чого немає в генеративній моделі. Триреберний сурогат (рис. 3, *в*) є неадекватним у тому, що залучає «нестандартне» (біорієнтоване) ребро і не відображає часткового суміщення шляхів. Крім того, є певні складності для параметризації такої моделі. У спеціальних випадках така латентна структура може бути ідентифікована [19].

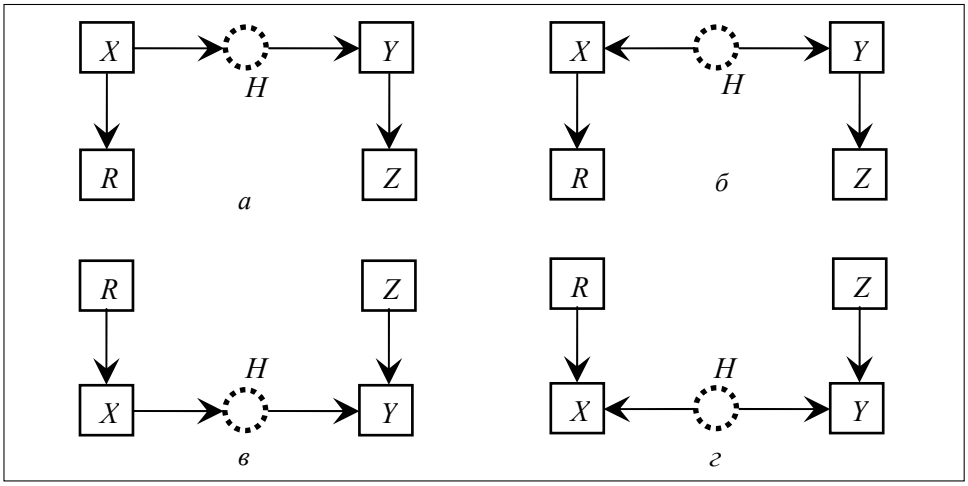

Рис. 2. Випадки прихованих змінних

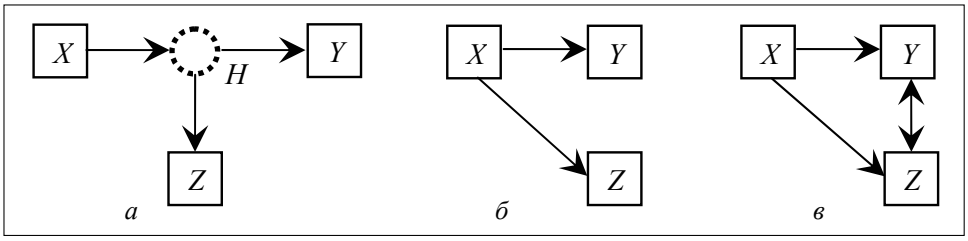

Рис. 3. «Нестандартна» латентна змінна

Латентні змінні допомагають спрощувати модель. Наприклад, фрагмент моделі у вигляді дворядного графу з перехресними ребрами (рис. 4, *а*) можна замінити на модель з одною «вузловою» латентною змінною (рис. 4, *б*). Унаслідок введення латентної змінної кількість зв'язків та параметрів кардинально зменшується. У роботі [20] наявність фрагмента моделі такого вигляду (рис. 4, *а*) пропонується розглядати як свідоцтво існування латентної змінної. Утім, таке свідоцтво — лише евристичне (слабке) і не є конфірмативним.

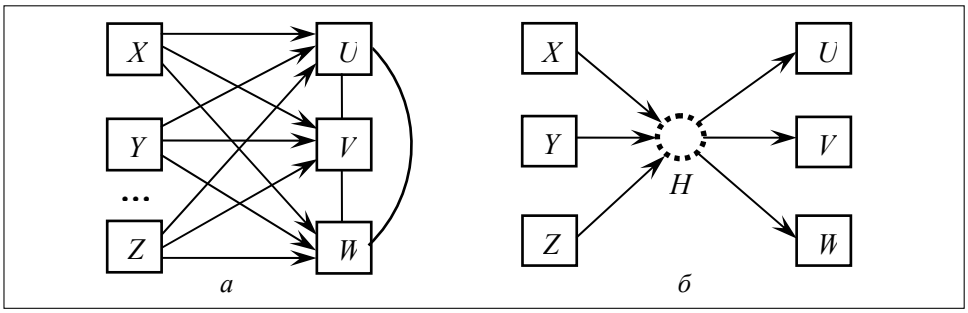

Рис. 4. Модель *а* спрощується під час введення латентної змінної *б*

КАУЗАЛЬНІ ТА СУБКАУЗАЛЬНІ МОДЕЛІ ТА ВІДНОШЕННЯ

Для трактовки моделі як каузальної чи субкаузальної необхідно, щоб модель адекватно відтворювала не тільки множину ребер, а також і адекватні (автен-

тичні) орієнтації цих ребер. Потенційно така можливість ґрунтується на експресивних можливостях АОГ-моделей і на властивостях відбірки даних та їхнього джерела. Фундаментом для ідентифікації каузальних моделей є постулат, відомий як принцип Рейхенбаха. Статистична асоційованість (взаємозалежність) змінних X та Y може пояснюватися трьома схемами механізму виникнення:

- X є причиною для Y; в апараті АОГ-моделей це означає, що від X до Y веде строго орієнтований шлях $X \to \cdots \to Y$;
- Y є причиною для X, тобто в моделі має бути шлях $X \leftarrow \cdots \leftarrow Y$;
- існує деяка спільна причина U для змінних X та Y, тобто в моделі має бути шлях $X \leftarrow \cdots \leftarrow U \to \cdots \to Y$ (змінна U може бути латентною).

Третя схема може співіснувати з першою або другою.

Можливості індуктивної ідентифікації орієнтацій ребер. Задача ідентифікації спрямованості впливу повною мірою постає, коли відсутня темпоральна інформація про змінні моделі. Розпізнання справжніх орієнтацій ребер базується на властивості колізорів та на теоремі еквівалентності АОГ-моделей. Колізор (колайдер) — це фрагмент орграфу вигляду $X \to Y \leftarrow Z$. Цей колізор називається шунтованим (екранованим), якщо у графі є ребро $X — Z$, інакше — нешунтованим. Змінну Y, яка входить до складу нешунтованого колізора $\to Y \leftarrow$, назвемо колізорною.

Для індуктивного виводу структур АОГ-моделей фундаментальним результатом є наступна теорема еквівалентності [7]: дві АОГ-моделі є еквівалентними, якщо і тільки якщо вони мають одні і ті ж ребра, а також одні і ті ж нешунтовані колізори.

Звідси відразу випливає, що в деревах та у повно зв'язаних АОГ-моделях орієнтація жодного ребра не ідентифікується. У методах сепараційного підходу процедура орієнтації ребер виконується на другому етапі алгоритму, коли всі ребра встановлено. Базовий принцип орієнтації ребер полягає у розпізнанні нешунтованих колізорів. Принцип формулюється як «колізорне правило»: якщо маємо $X — Y — Z$ та чинне $\mathrm{Ind}(X;\mathbf{S};Y)$, де $Y,\ X, Z \notin \mathbf{S}$, то орієнтуємо ребра $X \to Y$ та $Y \leftarrow Z$. Фізичний сенс цього принципу: тільки колізорна орієнтація такого двореберного шляху може пояснити нерозповсюдження інформації між X та Z через проміжну Y. Дуги, орієнтовані за ознакою колізора, вважаємо субкаузальними.

Неколізорні ребра теж можуть бути однозначно орієнтовані, а саме тоді, коли їх орієнтація логічно випливає з вимог орациклічності та уникнення несанкціонованих колізорів. Дійсно, якщо колізорне правило вже не вдається застосовувати і маємо дугу $X \to Y$, а також ребро $Y — Q$, причому Q не суміжна до X, то ребро $Y — Q$ має бути орієнтоване як $Y \to Q$. Отже, ребро, яке дотичне до стрілки, але не задовольняє ознакам нешунтованого колізора, має бути орієнтоване так, щоб запобігти виникненню необґрунтованого колізора. Це правило можна назвати пост-колізорним.

Зазначеними правилами не орієнтуються ті ребра, які не утворюють нешунтованих колізорів, не дотичні нащадкам колізорних вершин, а також ребра між вершинами, які мають ідентичні набори «батьків». Як приклад останнього випадку, неорієнтованими залишаються ребра, що з'єднують

«братів» (ребро 2–3 на рис. 5, *а*). Не вдається орієнтувати ребра між «предками» колізорних вершин (у кореневій частині орграфу). Фрагмент моделі, який не вдається орієнтувати зазначеними правилами, назвемо фрагментом з варіабельними орієнтаціями (еквівалентні моделі мають альтернативні орієнтації ребер цього фрагмента). Кожне ребро у складі фрагмента з варіабельними орієнтаціями допускає довільну орієнтацію (коли розглядається окремо), однак сукупність суміжних ребер не може бути орієнтована довільно. Тобто, усі припустимі варіанти орієнтації ребер зв'язного фрагмента підлягають певним «системним» («топологічним») обмеженням.

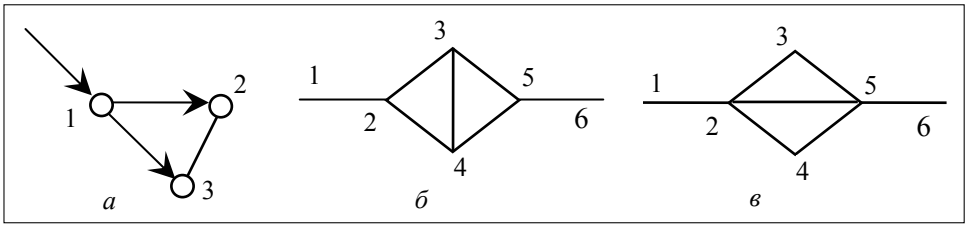

Рис. 5. Приклади неідентифікованості спрямування зв'язків

Нехай маємо кореневий фрагмент із варіабельними орієнтаціями. Відсутність нешунтованих колізорів означає, що в цьому фрагменті всі неорієнтовані цикли довжиною більше трьох ребер є трикутизованими (мають хорду). Вершину X цього фрагмента назвемо коренем, якщо в обраній моделі з класу еквівалентності всі ребра фрагмента дотичні до X, є спрямованими від вершини X, тобто мають вигляд $\leftarrow X \rightarrow$. «Системні» обмеження на орієнтації імплікуються такими синтаксичними вимогами: у кожній зв'язаній компоненті фрагмента має бути не менше одного кореня; усі корені — несуміжні; у кожному циклі має бути принаймні один колізор (шунтований) і стільки ж коренів.

Наприклад, розглянемо рис. 5, *б*. Припустимо, що ребра такої моделі не орієнтуються. Тоді, хоча граф такої моделі не є деревом, у ньому мусить бути тільки одна коренева вершина (цим коренем може виявитися довільна вершина). Те саме стосується графу моделі, зображеного на рис. 5, *в*. Граф на рис. 5, *б* репрезентує (охоплює) 14 еквівалентних моделей, а граф на рис. 5, *в* — 16 моделей.

У випадку, коли фрагмент із варіабельними орієнтаціями — некореневий, вказані обмеження узагальнюється так: у складі фрагмента існує вершина, яка є «предком» усіх інших вершин фрагмента.

Якщо є потреба ідентифікувати автентичні орієнтації ребер, які не орієнтуються колізорним та пост-колізорним правилами, аналітику доведеться вдатися до «слабких» методів, які не гарантують статистично спроможних (коректних) результатів. У певних ситуаціях можуть бути корисними евристичні критерії. Наприклад, коли є один домінуючий фактор впливу на задану змінну, можна розпізнати напрямок за допомогою міри дельта-детермінації [15]. Коли задача зводиться до вибору кореня дерева, то для підвищення шансів розпізнання кореня доцільно застосувати дельта-детермінацію в «колективному» режимі.

Концепція індуктивної каузальності. Деякі скептики заперечують каузальну інтерпретацію моделей, виведених виключно статистичними засобами з даних пасивних спостережень. Вони радять покладатися на рандомі-

зовані експерименти або на апріорну інформацію. Питання з'ясовується через аналіз та уточнення процедур.

Обґрунтовану концепцію індуктивної каузальності першими, мабуть, запропонували Дж. Перл і Т. Верма [3, 6]. Як передумову для каузальної інтерпретації моделей на основі АОГ, вони сформулювали два загальних принципи (або вимоги до моделі): мінімальність та стабільність моделі. Перша вимога відповідає відомому у філософії принципу «Леза Оккама». За умови адекватності простіша модель має перевагу, тому що легше піддається поясненню та верифікації. Друга вимога забезпечує стійкість виведеної структури до варіювання параметрів. Відтак дістаємо визначення: змінна X має справжній каузальний вплив на Y, тільки якщо в кожній мінімальній структурі моделі, узгодженій з даними, існує строго орієнтований шлях від X до Y. При цьому застережено спеціальні умови ідентифікації дуг вказаного шляху. Згідно з цими вимогами розроблено алгоритм IC*, який на основі розподілення ймовірностей відтворює структуру моделі й ідентифікує дуги графу, які відображають «справжні» каузальні впливи (алгоритм маркує їх зірочкою). Алгоритм припускає існування латентних змінних, кожна з яких є кореневою та впливає лише на дві спостережувані змінні.

У [3, 6] визначено умови, за яких можна розпізнати окремий «справжній» каузальний вплив між двома змінними в термінах впливу та незалежності. Виходячи з цих умов, для випадку відсутності темпоральної інформації пропонуємо строго формалізоване визначення паттерну каузального впливу (воно уточнює подане раніше в [18]). Для позначення того, що X та Y є залежні за фіксації умови $\mathbf{S}$ (розуміється, що $X, Y \notin \mathbf{S}$) введемо нотацію $\mathrm{Dep}(X;\mathbf{S};Y) \equiv \neg\,\mathrm{Ind}(X;\mathbf{S};Y)$. Відтак пропонується таке статистичне (індуктивне) визначення.

Визначення. Змінна X має справжній каузальний вплив на змінну Y, якщо

$$\exists Z\,\exists R\,\exists \mathbf{S}(X \notin \mathbf{S}):\{\mathrm{Ind}(Z;\mathbf{S};R)\ \&\ \mathrm{Dep}(X;\mathbf{S};Z)\ \&\ \mathrm{Dep}(X;\mathbf{S};R)\ \&$$

$$\&\ \mathrm{Dep}(X;\mathbf{S};Y)\ \&\ \mathrm{Dep}(Z;\mathbf{S};Y)\ \&\ \mathrm{Ind}(Z;X\mathbf{S};Y)\}. \qquad (1)$$

Каузальний вплив, виявлений через статистичний паттерн (1), може бути безпосереднім або опосередкованим. Можливий випадок опосередкованого каузального впливу відображено на рис. 6 пунктиром. Каузальний зв'язок є безпосереднім (у контексті заданої множини змінних), якщо додатково виконується умова $\forall \mathbf{Q}(X, Y \notin \mathbf{Q}):\mathrm{Dep}(X;\mathbf{Q};Y)$. Для безпосереднього зв'язку можна сформулювати точніший критерій каузальності. Крім того, у три останні терми формули (1) не обов'язково має входити та сама множина $\mathbf{S}$, що і в перші три терми. Зауважимо, що з огляду на припущення необманливості, визначення (1) містить надлишковий терм.

Формулювання статистичного паттерну (1) переводить питання верифікації каузального відношення в процедуру тестування композитної статистичної гіпотези. Таким чином, на основі симетричного поняття (не)залежності вдається ідентифікувати несиметричне відношення каузального впливу. Наведене визначення каузальності кореспондується з інтуїтивним розумінням. Зміст його можна передати так: «X є причиною для Y,

якщо існує інструментальна (керуюча) змінна (можливо, «невидима», а її індикатором є Z), і ця інструментальна змінна впливає на Y через посередництво X». У цьому формулюванні виділяються три ознаки: автономність (а відтак — спрямованість) керування інструментальної змінної; дієвість цього керування для Y; умовна незалежність інструментальної змінної від Y за фіксації причини X. Нагадаємо, що терміни «керування» і «вплив» мають на думці спрямований зв'язок, а не просто кореляцію. Стверджувати про наявність строго орієнтованого шляху від X до Y в усіх мінімальних моделях, узгоджених із даними, можна тому, що такий оршлях з необхідністю детермінується набором статистичних свідчень.

Обґрунтування принципу ідентифікації каузального впливу. Модель, зображена на рис. 6, підпадає під визначення каузального відношення (1), причому оршлях або дуга $X \to Y$ є каузальними. Структура моделі на рис. 6 є спрощенням загального випадку (тому що множина **S** — пуста) і характеризується такими статистичними свідченнями: є тільки дві умовні незалежності $\mathrm{Ind}(Z;X;Y)$, $\mathrm{Ind}(R;X;Y)$ і одна безумовна $\mathrm{Ind}(Z;\{\};R)$. Зокрема, чинна безумовна залежність $\mathrm{Dep}(Z;\{\};Y)$.

Якщо прийняти згадану вище конвенцію про можливі механізми виникнення залежностей, то залежність між X та Y може бути пояснена одним із трьох принципових припущень: змінна X каузально впливає на Y; змінна Y каузально впливає на X; X та Y обидві зазнають каузального впливу третьої змінної U (яка нерепрезентована в моделі), і, можливо, рівночасно діє один із двох перших механізмів. Для ствердження, що X має каузальний вплив на Y, достатньо спростувати (заперечити) усі альтернативні гіпотези.

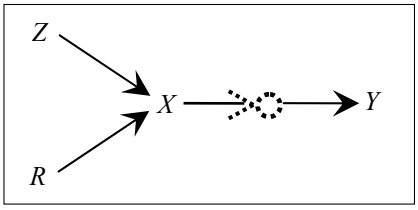

Рис. 6. Паттерн каузального зв'язку

Нехай гіпотеза H_0 — це твердження про існування оршляху від X до Y. При заданому в (1) наборі свідчень для гіпотези H_0 потенційно існують чотири основні альтернативи: H_1, H_2, H_3, H_4, які відображені відповідно на рис. 7, *а*, 7, *б*, 7, *в*, 7, *г*. Можливі й інші екзотичніші альтернативи.

Ці основні альтернативи породжуються з таких міркувань. Альтернатива каузальному впливові X на Y потребує або реверсу дуги (оршляху) $X \to Y$, або заміни її на дві дуги від прихованої змінної U, або комбінації цих конструкцій. Однак реверс дуги між X та Y у структурі рис. 6 миттєво призводить до зникнення ланцюга (відкритого шляху) між Z та Y, що суперечить залежності $\mathrm{Dep}(Z;\{\};Y)$. Відтак в альтернативних моделях доведеться вводити ребро $Z — Y$ або відповідно ребра $U — Y$ та $Z — U$, або $X — U$. Аналогічні ребра необхідні для змінної R. Отже, альтернативи поступаються з точки зору критерію простоти (додаються ребра та параметри). Однак ще вагоміше заперечення викликає порушення принципу структурно-поведінкової адекватності. Дійсно, в усіх альтернативних структурах факти $\mathrm{Ind}(Z;X;Y)$ та $\mathrm{Ind}(R;X;Y)$ графічно незумовлені. Отож ми будемо змушені накладати вимоги на співвідношення параметрів, аби «мімікрувати» ці умовні незалежності.

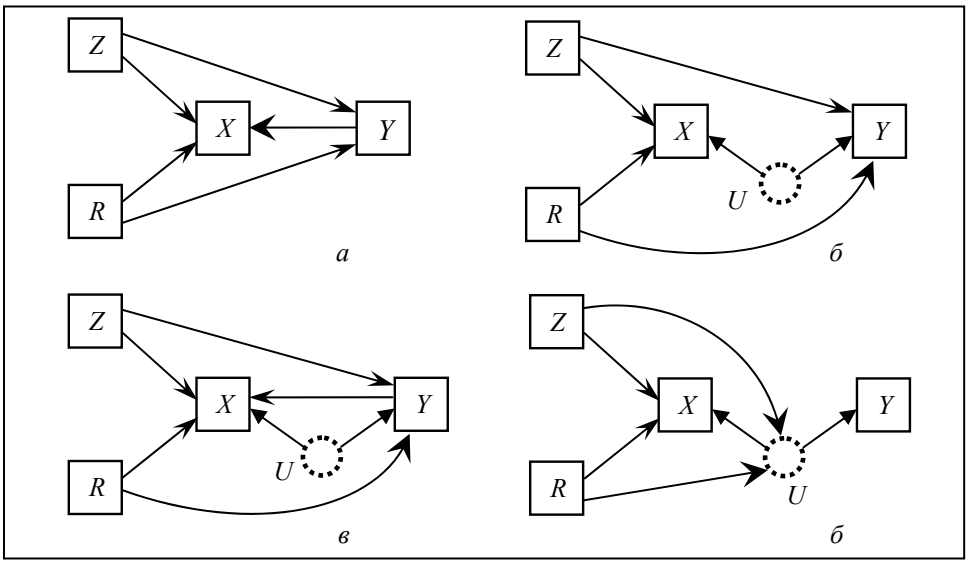

Рис. 7. Альтернативні моделі для паттерну каузального зв'язку

Розглянемо альтернативну гіпотезу H_1 (рис. 7, *а*). Дані свідчать про $\mathrm{Ind}(Z;X;Y)$, однак є ребро $Z — Y$. Для узгодження цих двох фактів маємо вдатися до такого виправдання. За кондиціювання змінної X наводиться (провокується) [13, 16] «додаткова» залежність між її «батьками» Z та Y, і більш того, ця провокована залежність між Z та Y точно нейтралізує початкову «реберну» залежність між Z та Y. Аналогічно, для виправдання факту $\mathrm{Ind}(R;X;Y)$ необхідна «анігіляція» ребра $R — Y$ з провокованою залежністю між R та Y (через X). Такі узгодження виглядають штучними (натяжка).

Для гіпотези H_2 (рис. 7, *б*) повторюються аналогічні міркування з тією відмінністю, що необхідно, аби «реберна» залежність між Z та Y точно «анігілювалася» з ланцюгом двох залежностей, який утворено з ребра $U — Y$ та з наведеної залежності між Z та U. Для гіпотези H_3 (рис. 7, *в*) для пояснення $\mathrm{Ind}(Z;X;Y)$ необхідний інший варіант «анігіляції» залежностей. Тут роль «нейтралізатора» для ребра $Z — Y$ відіграє конструкція на основі чотирьох ребер з участю прихованої змінної U. Аналогічна «анігіляція» необхідна для виправдання факту $\mathrm{Ind}(R;X;Y)$. Для гіпотези H_4 (рис. 7, *г*) повторюється аргументація, викладена для гіпотези H_1, з тією відмінністю, що замість змінної Y підставляється U.

Отже, усі моделі, що заперечують каузальний вплив X на Y, непереконливі з точки зору критичного раціоналізму, і тому мають бути відкинуті. Альтернативні гіпотези не лише потребують введення додаткових ребер, а ще й накладають жорсткі вимоги на співвідношення параметрів. Наукова методологія відкидає подібні натяжки (припущення ad hoc). Лише існування дуги $X \to Y$ (або оршляху) забезпечує задовільне пояснення статистичних відношень у даних.

Можна помітити, що визначення (1) включає в себе умови колізорного та пост-колізорного правил. У той же час не всі ребра, які орієнтовано за допомогою колізорного правила, відповідають каузальним зв'язкам. Відповідно, не всі оршляхи моделі є каузальними. Ці семантичні відмінності можна синтаксично виразити в апараті графів частково-орієнтованих індукційних шляхів (POIPG) [1]. Потрібно зауважити, що за невеликого розміру відбірки даних визначення (1) на практиці може іноді призводити до помилок. Один із головних ризиків хибної ідентифікації каузального зв'язку походить від ненадійності стандартного колізорного правила. Помилкова орієнтація ребер $Z — X — R$ може відбутися у разі слабкості транзитної залежності між Z та R. Можна підвищити надійність колізорної орієнтації ребер за допомогою інструменту провокації (реактивації) залежності [13, 16]. Відповідно, для підвищення надійності ідентифікації каузальних відношень потрібно включити (додати) у формулювання (1) умову $\mathrm{Dep}(Z; X\mathbf{S}; R)$, яка контрастує з умовою $\mathrm{Ind}(Z; \mathbf{S}; R)$. Можна сказати, що X «реактивує» залежність між Z та R (а у випадку пустої множини $\mathbf{S}$ — провокує залежність). Така корекція визначення підвищить надійність виявлення каузальних відношень.

ВИСНОВКИ

Розглянена постановка задачі — виведення каузальних моделей з даних пасивних спостережень — теоретично обтяжлива (потребує нестандартного мислення). Але саме така постановка відповідає практичним потребам і сучасним викликам. Зрозуміло, що нею не вичерпується методологія індуктивного виводу (з емпіричних даних) моделей когнітивного призначення. Можливі й інші форми збору й аналізу емпіричних даних. З одного боку, репрезентація даних у формі єдиного сумісного розподілення ймовірностей змінних є спрощенням і обмеженням. Сукупність фактів умовної (і безумовної) незалежності не вичерпує змісту аналізованих даних. Зокрема, при цьому залишається не ужитим поняття стабільності каузальних параметрів крізь різні субдомени. З іншого боку, якщо розглядати структуру АОГ-моделі лише як форму репрезентації фактів умовної незалежності в розподіленні ймовірностей, то сумісне розподілення ймовірностей містить навіть надлишкову інформацію. Наприклад, маргінальні розподілення ймовірностей кореневих змінних навряд чи несуть каузальну інформацію.

Характеристика моделей як причинно-наслідкових (у загальнонауковому розумінні цього терміна) — це питання *інтерпретації* моделі людиною. Будь-хто може не сприймати розглянені аргументи та запропонувати інші інтерпретації й інші дефініції. Та все ж видається, що наведене визначення каузального зв'язку вже самою своєю конструкцією схиляє до каузальної інтерпретації, а аргументація є концептуальною й узгодженою. Але, коли користувач збирається застосувати модель для прогнозу наслідків втручань в об'єкт, необхідно застерегти наступне. Потрібно ретельно порівняти умови збору даних та умови планованих втручань. Якщо змінна керування не була репрезентована в даних, то для прогнозу наслідків керування доведеть-

ся зробити припущення про локальний і неруйнівний (для моделі) характер керування.

Якість виведеної моделі зумовлюється повнотою набору змінних, розміром відбірки та точністю даних. Аналітик може опинитися не в змозі розпізнати реально існуюче каузальне відношення через те, що сформульований вище індуктивно-статистичний Y-паттерн (1) не відшукується. Необхідна (але недостатня) передумова для виявлення цього паттерну — відсутність двох ребер у ближньому оточенні пари «причина – наслідок». Інша передумова — наявність індикатора інструментальної змінної. Більш того, якщо (під час збирання чи обробки) відбірку даних було піддано селекції за значенням певної змінної, яка є нащадком усіх інших змінних у моделі, то правдоподібно, що жодний зв'язок моделі не буде орієнтований, тобто напрямки впливів не ідентифікуються. Також у моделі з'являться зайві ребра. Цей ефект (провокування залежностей) відомий як selection bias, а також як парадокс Берксона [1, 3, 10]. Взагалі, якщо виведена модель виявилась недостатньо інформативною, то необхідно розширити номенклатуру спостережень, збільшити обсяг відбірки даних чи підвищити точність вимірювань.

Методологія виведення каузальних моделей, що ґрунтується на систематичному аналізі статистичних відношень і на виявленні марківських властивостей, на разі видається найбільш потужною та перспективною емпірико-індуктивною методологією. Вона здатна настільки повно і точно ідентифікувати каузальні відношення, наскільки це дозволяють наявні дані спостережень. Виведення каузальних моделей із багатовимірних динамічних рядів даних ґрунтується на тих самих загальних принципах, але процедури виведення мають специфіку, зокрема, потрібне переформатування відбірки даних («кейсізація»).

Свіжий огляд літератури з каузального моделювання в економетриці на основі даних спостережень подано в [21].

ЛІТЕРАТУРА

1. *Spirtes P., Glymour C. and Scheines R.* Causation, prediction and search. — NY: MIT Press, 2001. — 543 p.
2. *Scheines R., Spirtes P., Glymour C.* et al. The TETRAD Project: Constraint based aids to causal model specification // Multivariate Behavioral Research. — 1998. — **33**. — № 1. — P. 65–118.
3. *Pearl J.* CAUSALITY: models, reasoning, and inference. — Cambridge University Press, 2000. — 526 p.
4. *Pearl J.* Graphs, causality, and structural equation models // Sociological Methods and Research. — 1998. — **27**. — № 2. — P. 226–284.
5. *Spirtes P.* A Tutorial on Causal Inference // Technical Report № CMU-PHIL-183. Dep. of Philosophy, August. — Carnegie Mellon University, PA. — 2009. — 68 p.
6. *Pearl J., Verma T.* A theory of inferred causation // Proceeding and International Conferences on Principles of Knowledge Representation and Reasoning. — Morgan Kaufmann, San Mateo, CA. — 1991. — P. 441–452.
7. *Verma T., Pearl J.* Equivalence and synthesis of causal models // The 6-th Annual Conference on Uncertainty in Artificial Intelligence. — Cambridge, MA, Elsevier Science Publishers. — 1991. — P. 220–227.

8. *Glymour C.* What s right with «Bayes Net Methods» and what is wrong with «Hunting Causes and Using Them»? // British Journal for the Philosophy of Science. — 2010. — **61**. — № 1. — P. 161–211.
9. *Pearl J.* Review of N. Cartwright «Hunting Causes and Using Them» // Economics and Philosophy. — 2010. — **26**. — № 1. — P. 69–77.
10. *Richardson T. and Spirtes P.* Causal Inference via ancestral graph Markov models // Highly Structured Stochastic Systems. — Oxford University Press, 2003. — P. 83–105.
11. *Балабанов А.С.* Минимальные сепараторы в структурах зависимостей. Свойства и идентификация // Кибернетика и системный анализ. — 2008. — № 6. — С. 17–32.
12. *Андон Ф.И., Балабанов А.С.* Структурные статистические модели: инструмент познания и моделирования // Системні дослідження та інформаційні технології. — 2007. — № 1. — С. 79–98.
13. *Балабанов А.С.* К выводу структур моделей вероятностных зависимостей из статистических данных // Кибернетика и системный анализ. — 2005. — № 5. — С. 19–31.
14. *Chow C.K., Liu C.N.* Approximating discrete probability distributions with dependence trees // Institute of Electrical and Electronics Engineers transactions on Information Theory. — 1968. — **14**. — № 3. — P. 462–467.
15. *Балабанов О.С.* Індуктивне відтворення деревовидних структур систем залежностей // Проблемы программирования. — 2001. — № 1/2. — С. 95–108.
16. *Балабанов А.С.* Реконструкция модели вероятностных зависимостей по статистическим данным. Инструментарий и алгоритм // Проблемы управления и информатики. — 2009. — № 6. — С. 90–103.
17. *Балабанов О.С.* Системи ймовірнісних залежностей: графові та статистичні властивості // Математичні машини та системи. — 2009. — № 3. — С. 80–97.
18. *Балабанов О.С.* Відкриття структур залежностей в даних: від непрямих асоціацій до каузальності // Материалы 3-й междунар. конф. «УкрПРОГ'2002». Проблемы программирования. — 2002. — № 1/2. — С. 309–316.
19. *Андон П.І., Балабанов О.С.* До відкриття латентного бінарного фактора в статистичних даних категорного типа // Доп. НАН України. — 2008. — № 9. — С. 37–43.
20. *Elidan G., Lotner N., Friedman N. and Koller D.* Discovering hidden variables: A structure-based approach // Proceedings of the 13-th Conferences on Advances in Neural Information Processing Systems (NIPS-2000). — Denver, Colorado. — 2000. — P. 479–485.
21. *Kwon D.-H., Bessler D.A.* Graphical Methods, Inductive Causal Inference, and Econometrics: A Literature Review // Computational Economics. — 2011. — **38**. — № 1. — P. 85–106.

Надійшла 22.04.2010

УДК 519.8

ПРОБЛЕМА ІСНУВАННЯ РОЗВ'ЯЗКУ В ЗАДАЧАХ РОЗПОДІЛУ ПОТОКІВ

О.Є. КІРІК

Досліджено питання існування розв'язку в задачі розподілу потоків в енергетичних мережах. Сформульовано алгоритм апріорного визначення сумісності обмежень такої задачі. Запропоновано модель з системою рівнянь неперервності, побудованої таким чином, щоб забезпечити існування допустимого розв'язку за умови зведення до мінімуму порушень замовлень споживачів.

ВСТУП

Розподіл потоків у розподільчих мережах — це складна задача, математична модель якої може містити до кількох десятків тисяч змінних. При постановці прикладних задач оптимізації розподілу потоків у таких системах доцільно спочатку дослідити питання існування розв'язку, аби за необхідності гарантувати розв'язність задачі.

Перша частина роботи присвячена дослідженням існування розв'язків задач розподілу потоків, що базуються на теоремі про максимальний потік і метод розстановки позначок [1]. Наведені викладки дозволяють заздалегідь, до здійснення процесу оптимізації, визначити, чи є сформульована задача розв'язною. У випадку дослідження реальних мереж визначення відсутності розв'язку конкретної задачі є недостатнім. Бажано певним чином переформулювати задачу, гарантувавши сумісність системи обмежень.

У другій частині роботи запропоновано модель розподілу потоків із системою рівнянь неперервності, яку побудовано таким чином, щоб забезпечити існування допустимого розв'язку. Оскільки технологічна пропускна здатність ділянок мережі є величиною фіксованою, то корегувати можна тільки замовлення споживачів та подачу речовини з витоків. При цьому порушення замовлень споживачів має бути зведене до мінімуму. Мінімізовані похибки вектора споживання отримуємо в явному вигляді, що дає простір для консультацій з експертами та є важливим з огляду на прикладну спрямованість задач, що досліджуються.

Мета роботи — дослідження питання існування розв'язку для оптимізаційної задачі зі структурою обмежень, характерною для задач розподілу потоків, побудова алгоритму визначення сумісності такої задачі, а також побудова мережевої моделі, що гарантує існування розв'язку в задачі розподілу потоків.

РІВНЯННЯ ДЛЯ ВИЗНАЧЕННЯ ДОПУСТИМОГО ПОТОКУ

Розподільчу мережу топологічно можна представити як зв'язний плоский граф [2]. Позначимо його $G = (N, V)$, де N — множина вершин, V — мно-

© *О.Є. Кірік*, 2011

жина дуг, що їх з'єднують. Кожному $v \in V$ співставлена упорядкована пара вершин (i, j), $i, j \in N$, що є, відповідно, початком і кінцем цієї дуги. Якщо споживачі розташовані у вершинах графу і для кожної вершини i відоме споживання, то кожній дузі графу v може бути приписаний деякий потік, що забезпечує доставку відповідного продукту споживачам.

Нехай $d(i)$ — функція споживання, визначена у вершинах графу, $x(i, j)$ — функція потоку, визначена на дугах графу G. Розглянемо умови, що накладаються на ці функції в класичній задачі розподілу потоків.

Основною системою рівнянь є система, що відображає закон збереження речовини у вершинах графу:

$$\sum_{j:(i,j)\in V} x(i, j) - \sum_{j:(j,i)\in V} x(i, j) = d(i), \ i \in N. \tag{1}$$

Система рівнянь (1) демонструє той факт, що кількість речовини, що втікає у вершину i мінус кількість, що витікає з цієї вершини, має дорівнювати споживанню в цій вершині. Як випливає з рівнянь (1), величини $d(i)$ мають додатне значення у вузлах — «джерелах» і від'ємне — у вузлах, де розташовані споживачі.

Виконання співвідношень (1) гарантує задоволення всіх споживачів.

Надалі зручно використовувати такі позначення. Якщо K та I підмножини множини N, то символом (K, I) позначаємо множину всіх дуг, що ведуть із $k \in K$ в $j \in I$.

Покладемо

$$d(K) = \sum_{i\in K} d(i) \text{ для всіх } K \subseteq N; \quad d^2(K) = \sum_{i\in K} d^2(i) \text{ для всіх } K \subseteq N;$$

$$x(i, K) = \sum_{\substack{j\in K \\ (i,j)\in V}} x(i, j) \text{ для всіх } K \subseteq N;$$

$$x(I, K) = \sum_{\substack{i\in I,\, j\in K \\ (i,j)\in V}} x(i, j) = \sum_{(i,j)\in (I,K)} x(i, j) \text{ для всіх } I \subseteq N, \ K \subseteq N.$$

З огляду на введені позначення система (1) буде мати вигляд:

$$x(i, N) - x(N, i) = d(i), \ i \in N. \tag{2}$$

Теорема 1. Для розв'язності системи (2) необхідно, аби виконувалося співвідношення

$$d(N) = 0. \tag{3}$$

Для доведення сумуємо (2) по $i \in N$ та отримуємо:

$$x(N, N) - x(N, N) = d(N).$$

Умова (3) є умовою збалансованості системи (2).

У реальних розподільчих мережах важливою умовою є задоволення технологічних обмежень на пропускну здатність ділянок мережі (дуг графу).

$$r^-(v) \le x(v) \le r^+(v), \tag{4}$$

де $r(v)^-$ та $r(v)^+$ — задані числа, що називаються мінімальною та максимальною пропускною здатністю дуги.

Дослідимо тепер проблему існування розв'язку системи (2) за обмежень (4). Для цього розглянемо теорему про максимальний потік та мінімальний розріз.

ЗАДАЧА ПРО МАКСИМАЛЬНИЙ ПОТІК

Сформулюємо задачу знаходження максимального потоку. Припустимо, що $r^-(v)=0$ для всіх $v\in V$. Нехай у графі G виділено дві вершини $s,t\in N$, що називаються витоком та стоком і виконуються такі співвідношення:

$$\begin{gathered} x(s,N)-x(N,s)=q, \\ x(i,N)-x(N,i)=0,\ \ i\neq s,t, \\ x(t,N)-x(N,t)=-q. \end{gathered} \tag{5}$$

Зазначимо, що ця система має в якості частинного розв'язку — нульовий (при цьому $q=0$).

Задача пошуку максимального потоку полягає в знаходженні

$$\max q \tag{6}$$

за обмежень (5).

Нехай $(K,\overline{K})$ — підмножина дуг (i,j) графу G, де $K\subset N$, $\overline{K}=N\setminus K$, $i\in K$, $j\in\overline{K}$. Розрізом графу G, що відділяє вершини s та t, називатимемо таку підмножину дуг $(K,\overline{K})$, що $s\in K$, $t\in\overline{K}$. Для цього розрізу $(K,\overline{K})$ може бути визначена його пропускна здатність, що дорівнює величині $r^+(K,\overline{K})$, тобто сумі пропускних здатностей дуг з $(K,\overline{K})$.

Якщо $(K,\overline{K})$ — розріз, то з (5), сумуючи по $i\in K$, отримаємо: $x(K,N)-x(N,K)=q$.

Враховуючи, що $K\cup\overline{K}=N$, розпишемо ліву частину цієї рівності:

$$x(K,K\cup\overline{K})-x(K\cup\overline{K},K)=x(K,K)+x(K,\overline{K})-x(K,K)-x(\overline{K},K)=q.$$

Звідси

$$x(K,\overline{K})-x(\overline{K},K)=q. \tag{7}$$

За припущенням $r^-(v)=0$ для всіх $v\in V$, тобто $x(\overline{K},K)\geq 0$. Оскільки $x(K,\overline{K})\leq r^+(K,\overline{K})$, то заміняючи в лівій частині (7) перше число на більше, а друге на менше, отримаємо: $r^+(K,\overline{K})\geq q$.

Цей результат може бути сформульований таким чином.

Теорема 2. Величина q довільного потоку з витоку s у стік t не може перевищувати пропускну здатність довільного розрізу.

Мінімальним розрізом називаємо розріз з мінімальною пропускною здатністю.

Тепер може бути сформульована основна теорема.

Теорема 3. Для довільної мережі максимальна величина потоку з s у t дорівнює пропускній здатності мінімального розрізу.

При цьому для довільного мінімального розрізу дуги з $(K,\overline{K})$ є насиченими максимальним потоком, тобто $x(v)=r^{+}(v),\ v\in(K,\overline{K}),\quad x(v)=0,$ $v\in(K,\overline{K})$.

Доведення. З огляду на теорему 2 достатньо показати, що існують потік x та розріз $(K,\overline{K})$, для яких величина потоку та пропускна здатність розрізу є рівними.

Оскільки нульовий потік задовольняє всім обмеженням задачі про максимальний потік, то задача лінійного програмування (6), (5) є сумісною. Усі розв'язки задачі є обмеженими. Лінійна функція досягає на замкненій обмеженій множині максимального значення, як наслідок, існує і максимальний потік.

Опишемо процедуру розстановки позначок для побудови допустимого та максимального потоків. Попередньо введемо деякі позначення.

Послідовність $(j_1,j_2,...,j_m),\ j_k\in N$, називається шляхом, що з'єднує вершини j_1 та j_m, якщо $(j_k,j_{k+1})\in V,\ \ k=1,...,m-1,\ j_i\neq j_l,\ i\neq l$.

Нехай шлях $J=(j_1,j_2,...,j_m)$ з'єднує вершини $j_1=a,\ j_m=b$.

Покладемо

$$\omega_J(i,j)=\begin{cases}+1, & \text{якщо } \exists k: i=j_k, j=j_{k+1},\ (j_k,j_{k+1})\in V,\ k=1,...,m-1,\\ -1, & \text{якщо } \exists k: i=j_k, j=j_{k+1},\ (j_{k+1},j_k)\in V,\ k=1,...,m-1,\\ 0 & \text{у всіх інших випадках.}\end{cases}$$

Для кожного шляху J функція ω_J визначає вектор, розмірність якого співпадає з числом дуг графу G, тобто $|V|$. Між векторами ω_J та шляхами в графі G існує взаємно однозначна відповідність.

Якщо J — шлях з вершини $a\in N$ у вершину $b\in N$, то

$$\omega_J(i,N)-\omega_J(N,i)=\begin{cases}+1 & \text{для} \quad i=a,\\ -1 & \text{для} \quad i=b, \quad i\in N,\\ 0 & \text{для} \quad i\neq a,b.\end{cases} \tag{8}$$

Для обґрунтування цього співвідношення достатньо розглянути простий шлях $J=(a,b)$, що складається лише з однієї неорієнтованої дуги. Можливі два випадки: $(a,b)\in V$ або $(b,a)\in V$. Якщо $i\neq a,b$, то всі доданки в лівій частині дорівнюють нулю, і (11) виконується. Якщо $i=a$, то в лівій частині (8) тільки один доданок $\omega_J(a,b)=1$ є відмінним від нуля, і тому ліва частина дорівнює одиниці. Якщо ж $(b,a)\in V$, то $\omega_J(b,a)=-1$, решта доданків дорівнюють нулю, і (8) знову виконується. Аналогічно для випадку $i=b$.

Для загального шляху $J=(j_1,j_2,...,j_m),\ j_1\neq j_m$, результат визначається з того факту, що $\omega_J=\sum_{k=1}^{m-1}\omega(j_k,j_{k+1})$ і крім початкової j_1 та кінцевої j_m вершин, решта вершин y_k входять у два елементарні шляхи (j_{k-1},j_k) та (j_k,j_{k+1}) один раз як кінцева, другий — як початкова.

Якщо *J*-цикл, то $\omega_J(i,N)-\omega_J(N,i)=0$ для всіх $i\in N$. Цей факт випливає із попереднього з врахуванням того, тепер усі вершини є початковими для одного елементарного шляху та кінцевими для іншого.

Нехай $x(v),\ v\in V$ — максимальний потік. Система позначок графу будується таким чином. Вершину s помічаємо парою $(+\infty,s)$. На кожному кроці вершина j отримує позначку $(\alpha(j),i)$ в одному з двох випадків.

Якщо $(i,j)\in V$, i — позначена вершина, а j — непозначена вершина та $r^+(i,j)>x(i,j)$, то $\alpha(j)=\min\{r^+(i,j)-x(i,j),\alpha(i)\}$. Якщо $(j,i)\in V$ та $x(i,j)>0$, то $\alpha(j)=\min\{x(i,j),\alpha(i)\}$.

Процес розстановки позначок є скінченим, оскільки жодна вершина не помічається двічі.

Припустимо, що ми на одному з етапів помітили вершину t і перша компонента цієї позначки є $\alpha>0$. Тоді, повертаючись назад по іншим компонентам позначок, обов'язково прийдемо до вершини s, з якої починався цей процес. Таким чином утворюється шлях від вершини s до вершини t. Позначимо його J.

Нехай новий потік з s у t дорівнює $x(v)+\alpha(t)\omega_J(v)$. За теоремою 3 він задовольняє всім обмеженням (5). Отже, потік збільшився на величину $\alpha(t)$.

Таким чином, якщо вдалося позначити вершину t, то величину потоку можна збільшити. Але оскільки, за припущенням, потік $x(v)$ був максимальним, це приводить до суперечності. Ми не можемо позначити вершину t, тому процес позначок зупиниться до того, як відбудеться помітка вершини t.

Позначимо через K множину позначених вершин і розглянемо дуги $(K,\overline{K})$ та $(\overline{K},K)$. Оскільки неможливо з множини K позначити жодну вершину з $\overline{K}$, це означає, що $x(v)=r^+(v)$ для всіх $v\in(K,\overline{K})$. Так само $x(v)=0$ для всіх $v\in(\overline{K},K)$. Тоді з (7) отримаємо, що $r^+(K,\overline{K})=q$.

Показано, що пропускна здатність деякого розрізу дорівнює потужності максимального потоку. Звідси випливає, що:

- цей розріз має мінімальну пропускну здатність;
- якщо взяти довільний інший мінімальний розріз, то таким самим чином на ньому всі дуги з $(K,\overline{K})$ є насиченими, а всі дуги з $(\overline{K},K)$ ненасиченими.

Доведення завершене.

УМОВИ РОЗВ'ЯЗНОСТІ ДЛЯ ЗАДАЧІ РОЗПОДІЛУ ПОТОКІВ

Повернемося до проблеми розв'язності системи

$$x(i,N)-x(N,i)=d(i),\ \ i\in N,$$

$$r^-(v)\le x(v)\le r^+(v),\ \ v\in V. \qquad (9)$$

Для вирішення питання щодо сумісності системи достатньо побудувати потік, що їй задовольняє.

Вважаємо, що $r^-(v)=0$. Дійсно, якщо це не так, то розглянемо нові потоки $\bar{x}(v)=x(v)-r^-(v)$. Підставивши $\bar{x}(v)$ у (9), прийдемо до нової системи:

$$x(i,N)-x(N,i)=\bar{d}(i),\ \ i\in N,\ \ r^-(v)\le\bar{x}(v)\le\bar{r}^+(v),\ \ v\in V,$$

де $\bar{r}^+(v)=r^+(v)-r^-(v)$, $\bar{d}(i)=d(i)-r^-(i,N)+r^-(i,N)$.

Це означає, що можна звести задачу, що розглядається, до попередньої, в якій нижні грані обмежень дорівнюють нулю.

Отже, розглядаємо питання про розв'язність системи (9), коли $r^-(v)=0$. Побудуємо новий граф G', додавши до графу $G(N,V)$ додаткові вершини s та t, тобто $N'=\{s\}\cup\{t\}\cup N$. Розіб'ємо N на дві таких множини S та T, що $N=S\cup T$, $S\cap T=\varnothing$ та $S=\{i\in N: d(i)>0\}$, $T=\{i\in N: d(i)\le 0\}$. До множини V додамо нові дуги (s,i), $i\in S$ та (i,t), $i\in T$. Розширену множину дуг позначимо V'. Введемо обмеження на пропускні здатності нововведених дуг: $0\le x(s,i)\le d(i)$, $i\in S$, $0\le x(i,t)\le -d(i)$, $i\in T$.

Для нового графу $G'=(N',V')$ розглянемо задачу про максимальний потік та мінімальний розріз.

Розглянемо множину $\{s\}$. Доповненням до неї буде множина $N\cup\{t\}$. Пропускна здатність розрізу $(s,N\cup\{t\})$ визначається пропускними здатностями дуг, що утворюють цей розріз і дорівнює $r^+(s,S)=d(s)$. Розглянемо тепер множину вершин $\{s\}\cup N$. Доповненням до неї буде єдина вершина t. Відповідно, пропускна здатність розрізу $(\{s\}\cup N,t)$ визначається пропускними здатностями дуг, що ведуть у t. Вона дорівнює $r^+(T,t)=d(T)$.

Насиченість усіх дуг множини (s,S) забезпечує виконання у вершинах $i\in S$ балансових співвідношень, оскільки в цьому випадку $x(i,N)-x(N,i)-x(s,i)=0$, $x(s,i)=d(i)$ для всіх $i\in S$. Тому $x(i,N)-x(N,i)=d(i)$ для всіх $i\in S$.

Аналогічно, якщо всі дуги множини (T,t) є насиченими, то виконуються умови: $x(i,N)-x(N,i)+x(i,t)=0$, $x(i,t)=-d(i)$ для всіх $i\in T$, або $x(i,N)-x(N,i)=-d(i)$ для всіх $i\in T$.

Таким чином, якщо існує потік, що насичує дуги (s,S) та (T,t) розширеного графу G', то цей потік задовольняє умовам (9) для вихідного графу G. Згідно з теоремою про максимальний потік і мінімальний розріз, вимога насиченості дуг є справедливою для мінімальних розрізів. Виведемо умови, за яких розрізи $(s,N\cup\{t\})$ і $(\{s\}\cup N,t)$ будуть мінімальними.

Розглянемо довільний розріз $K'=(\{s\}\cup K,N\setminus K)$, $K\subset N$. Випишемо значення його пропускної здатності. Вона дорівнює сумі пропускних здатностей відповідних дуг $r^+(s,S\cap\overline{K})+r^+(K,\overline{K})+r^+(T\cap K,t)$. З умови

мінімальності розрізу $(s, N \cup \{t\})$ отримаємо $d(S \cap \overline{K}) + r^+(K, \overline{K}) - d(T \cap K) \geq d(S)$. Звідси випливає, що $r^+(K, \overline{K}) \geq -d(S \cap \overline{K}) + d(S) + d(T \cap K)$. Перетворимо праву частину нерівності, враховуючи, що $N = K \cup \overline{K}$:

$$\left[d(S) - d(S \cap \overline{K})\right] + d(T \cap K) = d(S \cap K) + d(T \cap K) = d(K).$$

Ми отримали умову $r^+(K, \overline{K}) \geq d(K)$.

Аналогічно для розрізу $(\{s\} \cup N, t)$

$$d(S \cap \overline{K}) + r^+(K, \overline{K}) - d(T \cap K) \geq d(T),$$

$$r^+(K, \overline{K}) \geq -d(T) + d(T \cap K) - d(S \cap \overline{K}) =$$

$$= -d(T \cap \overline{K}) - d(S \cap \overline{K}) = -d(\overline{K}) = d(k).$$

Для задачі з ненульовою нижньою гранню обмежень ця нерівність перетвориться до вигляду:

$$r^+(K, \overline{K}) - r^-(K, \overline{K}) \geq d(K). \qquad (10)$$

Сформулюємо умови розв'язності.

Теорема 4. Необхідною і достатньою умовою розв'язності системи (9) є виконання для кожної підмножини вузлів $K \subset N$ нерівності (10).

Ця умова означає, що відтік з довільної підмножини вузлів $K \subset N$ не має перевищувати пропускну здатність дуг, які ведуть з K (або приплив у довільну підмножину вузлів $K \subset N$ не має перевищувати пропускну здатність дуг, які ведуть у K).

Сформулюємо алгоритм для визначення сумісності системи (9).

Алгоритм 1.

1. Розширюємо вихідний граф G введенням додаткового витоку та стоку.
2. Проводимо процедуру збільшення потоку, наведену в доведенні теореми 3.
3. Перший варіант. Буде отриманий максимальний потік, що означає сумісність системи.
4. Другий варіант. Буде визначена неможливість побудувати максимальний потік, що і буде означати несумісність системи. При цьому множини K та $\overline{K}$, для яких порушується (10), знаходяться автоматично, якщо через K позначити множину помічених, а через $\overline{K}$ — множину непомічених вузлів.

РОЗПОДІЛ ПОТОКІВ ІЗ МІНІМІЗАЦІЄЮ ПОХИБКИ ФУНКЦІЇ СПОЖИВАННЯ

Припустимо, що система (9) не є сумісною. Відсутність допустимого розв'язку спонукає до пошуку модифікацій моделі, які гарантували б сумісність системи обмежень.

Пропускна здатність ділянок мережі залежить від технологічних характеристик системи і не може бути за бажанням змінена. Тому залишається корегувати замовлення споживачів або подачу речовини з витоків. При цьому порушення замовлень споживачів бажано звести до мінімуму.

Для графу $G=(N,V)$ сформулюємо задачу розподілу потоків з оптимізацією доставки продукту споживачам та мінімізацією порушень замовлень. За основу візьмемо нелінійну модель розподілу потоків, що дозволяє описувати широкий спектр розподільчих мереж та докладно досліджена в [3]. Мінімізувати затрати на доставку продукту споживачам та похибку функції споживання

$$F=\frac{1}{1+\alpha}\sum_{(i,j)\in V} l(i,j)\left|x(i,j)\right|^{1+\alpha}+\frac{1}{2}\sum_{i\in N}\varepsilon^2(i) \tag{11}$$

за обмежень

$$x(i,N)-x(N,i)=d(i)+\varepsilon(i),\ \ i\in N, \tag{12}$$

$$r^-(v)\le x(v)\le r^+(v),\ \ v\in V. \tag{13}$$

Функція вартості доставки (довжина ділянки мережі) $l(i,j)$ визначена на дугах графу G. Коефіцієнт α у виразі цільової функції дозволяє враховувати фізичні закономірності для різних розподільчих мереж. При $\alpha=0$ замовлення доставляється споживачам найкоротшим шляхом, як у лінійній транспортній задачі. Функції $\varepsilon(i),\ i\in N$ відображають можливі відхилення функцій споживання від замовлень споживачів.

Необхідною умовою сумісності обмежень (12), тобто збалансованості розподільчої системи, є виконання співвідношення

$$d(N)+\varepsilon(N)=0. \tag{14}$$

Поставимо у відповідність кожному i-му співвідношенню (12) множник Лагранжа u_i, $i\in N$, а співвідношенню (14) — $\overline{u}$.

Випишемо для задачі (11)–(14) функцію Лагранжа та сформулюємо необхідні умови екстремуму.

$$L=\frac{1}{1+\alpha}\sum_{(i,j)\in V} l(i,j)\left|x(i,j)\right|^{1+\alpha}+\frac{1}{2}\sum_{i\in N}\varepsilon^2(i)+$$

$$+\sum_{i\in N}u(i)\left[x(i,N)-x(N,i)-d(i)-\varepsilon(i)\right]+\overline{u}\sum_{i\in N}\left[d(i)+\varepsilon(i)\right]=$$

$$=\sum_{(i,j)\in V}\left\{\frac{1}{1+\alpha}l(i,j)\left|x(i,j)\right|^{1+\alpha}+x(i,j)\left[u(j)-u(j)\right]\right\}+$$

$$+\sum_{i\in N}\left\{\frac{1}{2}\varepsilon^2(i)-\varepsilon(i)\left[\overline{u}-u(i)\right]-d(i)\left[\overline{u}-u(i)\right]\right\}.$$

За теоремою Куна-Такера [4], для того, аби величини $x(v)$, $v\in V$, $\varepsilon(i)$, $i\in N$ були розв'язком поставленої задачі (11)–(14), необхідно і достатньо, щоб знайшлися числа u_i, $i\in N$ та $\overline{u}$, щоб відповідні невідомі доставляли б мінімум функції Лагранжа L за обмежень (13).

З теорії двоїстості [4] розв'язування задачі (11)–(14) є еквівалентним максимізації вгнутої неперервно диференційовної функції $\Phi(u) = \min\limits_{x(i,j)\in[r^-(i,j),\, r^+(i,j)]} L$.

Позначимо $u(j,i) = \left|u(j) - u(i)\right|^{\frac{1}{\alpha}} \operatorname{sign}(u(j) - u(i))$.

З необхідних умов екстремуму отримуємо, що мінімум функції L з врахуванням простих обмежень (13) досягається в точці:

$$x^*(i,j) = \begin{cases} r^-(i,j) & \text{при} \quad u(j,i) \le r^-(i,j) l(i,j)^{\frac{1}{\alpha}}, \\ u(j,i)\, l(i,j)^{-\frac{1}{\alpha}} & \text{при} \quad u(j,i) \ge r^-(i,j) l(i,j)^{\frac{1}{\alpha}}, \\ & \quad u(j,i) \le r^+(i,j) l(i,j)^{\frac{1}{\alpha}}, \\ r^+(i,j) & \text{при} \quad u(j,i) \ge r^+(i,j) l(i,j)^{\frac{1}{\alpha}}. \end{cases} \tag{15}$$

Крім того, виконуються співвідношення

$$\varepsilon(i) = \overline{u} - u(i),\ i \in N. \tag{16}$$

Використовуючи (15), конкретизуємо вигляд функції Φ:

$$\Phi(u) = \sum_{(i,j)\in V} \phi_{(i,j)} \left[u(j) - u(i)\right] - \sum_{i\in N} \left\{ \frac{1}{2} \left[\overline{u} - u(i)\right]^2 + \left[\overline{u} - u(i)\right] d(i) \right\}, \tag{17}$$

де

$$\phi_{(i,j)} = \begin{cases} \dfrac{1}{1+\alpha} l_{ij} \left| r^-(i,j) \right|^{1+\alpha} & \text{при} \quad u(j,i) \le r^-(i,j) l(i,j)^{\frac{1}{\alpha}}, \\ -\dfrac{\alpha}{1+\alpha} u(j,i)\, l(i,j)^{-\frac{1}{\alpha}} & \text{при} \quad u(j,i) \ge r^-(i,j) l(i,j)^{\frac{1}{\alpha}}, \\ & \quad u(j,i) \le r^+(i,j) l(i,j)^{\frac{1}{\alpha}}, \\ \dfrac{1}{1+\alpha} l_{ij} \left| r^+(i,j) \right|^{1+\alpha} & \text{при} \quad u(j,i) \ge r^+(i,j) l(i,j)^{\frac{1}{\alpha}}. \end{cases}$$

Таким чином, двоїста задача зводиться до задачі безумовної оптимізації

$$\max \Phi(u), \tag{18}$$

причому її цільова функція має неперервні похідні, які ми можемо обчислити.

Знаючи оптимальні значення $u(i)$, $i \in N$, та необхідні умови екстремуму можемо за формулами (16), (17) обчислити оптимальне значення потоків x_{ij}, $(i,j) \in V$ та нові значення функцій споживання $\overline{d}(i) = d(i) + \varepsilon(i)$, $i \in N$.

АЛГОРИТМ РОЗВ'ЯЗАННЯ ДВОЇСТОЇ ЗАДАЧІ

Для розв'язання задачі (18) можна використати довільний метод нелінійного програмування [5, 6].

Можемо застосувати метод покоординатного спуску, ідея якого полягає в тому, що на кожному кроці варіюється тільки одна змінна за умови фіксації решти.

Розв'язок одновимірних задач $\max\limits_{u(i)} \Phi$ зводиться до обертання на нуль похідної цільової функції. Оскільки функція Φ є вгнутою, то похідні монотонно спадають. Для знаходження нуля неперервної монотонно спадаючої функції скористаємось методом січних, опис якого наведено у роботі [7].

Сформулюємо тепер покроковий алгоритм розв'язання задачі (18).

Алгоритм 2.

1. Вибираємо довільні початкові значення $u(i)$, $i \in N$ та $\overline{u}$.

2. Розв'язуємо задачу знаходження оптимальних значень $u(i)$, $i \in N$ максимізуючи функцію Φ методом покоординатного спуску.

3. Знаючи оптимальні значення $u(i)$, $i \in N$, і, використовуючи необхідні умови екстремуму, повертаємося до вихідних змінних $x(i, j)$, $(i, j) \in V$, що дає розподіл потоків, який гарантує мінімальне порушення замовлень споживачів $\varepsilon(i)$ у кожному вузлі $i \in N$.

ВИСНОВКИ

Запропонована математична модель розподілу потоків із мінімізацією порушень функцій споживання є, звичайно, результатом певної ідеалізації процесу планування ресурсорозподільчих процесів. Вона орієнтована на апріорний аналіз функціонування системи на наявність «вузьких місць», про які свідчитиме значне відхилення вектора споживання від планового. Ясно, що допустимі показники цих порушень будуть різними для різних продуктопроводів.

За бажанням можна накласти двосторонні обмеження на порушення функції споживання на кшталт двосторонніх технологічних обмежень (13). Завдяки цьому можна відокремити підмножину вузлів, де порушення є небажаними або взагалі недопустимими. Математична модель оптимальної доставки продукту споживачам з мінімізацією відхилень вектора споживання та вузловими технологічними обмеженнями може бути предметом подальших досліджень.

ЛІТЕРАТУРА

1. *Форд Л., Фалкерсон Д.* Потоки в сетях. — М.: Мир, 1966. — 276 с.
2. *Оре О.* Графы и их применение. — М.: Мир, 1965. — 174 с.
3. *Кірік О.Є.* Розподіл потоків в мережах складної кільцевої топології // Наук. вісті НТУУ «КПІ». — 2009. — № 2. — С. 18–26.
4. *Пшеничный Б.Н.* Выпуклый анализ и экстремальные задачи. — М.: Наука, 1980. — 320 с.
5. *Пшеничный Б.Н., Кирик Е.Е.* Методы нелинейного программирования и потоки в сетях // Кибернетика и системный анализ. — 1994. — № 6. — С. 67–77.
6. *Кірік О.Є.* Алгоритми лінеаризації та спряжених градієнтів для нелінійних задач розподілу потоків // Наук. вісті НТУУ «КПІ». — 2007. — № 3. — С. 67–73.
7. *Кірік О.Є.* Оптимізація заповнення сховищ в задачах розподілу потоків для розподільчих мереж // Наук. вісті НТУУ «КПІ». — 2010. — № 1. — С. 28–35.

Надійшла 04.09.2010

УДК 519.71

ПРИМЕНЕНИЕ СЕТЕЙ ПЕТРИ ДЛЯ АНАЛИЗА КС-ГРАММАТИК

И.Я. СПЕКТОРСКИЙ

Предложена схема использования сетей Петри для исследования некоторых свойств КС-грамматик. Метод позволяет, в частности, исследовать заданную КС-грамматику на пустоту и конечность порождаемого языка, используя дерево покрываемости соответствующей сети Петри. Кроме того, предложенный метод позволяет сформулировать необходимые условия порождения заданного слова КС-грамматикой в терминах матричного анализа соответствующей сети.

ВВЕДЕНИЕ

В работах [1–3] описан следующий метод представления формальных языков с помощью порождающих сетей Петри: каждому переходу сети сопоставляется либо один из символов терминального алфавита, либо «пустой символ» λ, и каждая последовательность запусков переходов сети, заканчивающаяся в одной из выделенных «терминальных» маркировок, определяет слово порождаемого языка. В зависимости от множества терминальных маркировок и наличия λ-переходов определяют более десяти различных классов языков, порождаемых сетями Петри [1], которые не вписываются в классическую иерархию Хомского — строго включают класс регулярных языков, строго вложены в класс контекстно-зависимых языков, и несравнимы с классом контекстно-свободных языков [4–6].

Метод анализа, предложенный в данной работе, не предполагает полного описания языка сетью Петри, и поэтому, в частности, не дает критерия порождаемости данного слова заданной формальной грамматикой, но позволяет с помощью сети Петри контролировать количество вхождений каждой буквы терминального алфавита в порождаемое слово и, как следствие, исследовать порождаемый язык на пустоту и конечность.

ОПРЕДЕЛЕНИЕ СЕТИ ПЕТРИ ДЛЯ ЗАДАННОЙ КС-ГРАММАТИКИ

Пусть $G =< N, \Sigma, P, S >$ — контекстно-свободная порождающая грамматика (КС-грамматика), где N — нетерминальный алфавит, Σ — терминальный алфавит, P — множество продукций, $S \in N$ — источник. Для грамматики G введем в рассмотрение сеть Петри N_G с множеством позиций $N \cup \Sigma$, множеством переходов P и весовой функцией W, определяемой для продукции $A \to \beta$, а также символа $\xi \in \Sigma$ соотношениями:

$$W(A \to \beta, \xi) = |\beta|_\xi \text{ (количество вхождений } \xi \text{ в } \beta),$$

$$W(\xi, A \to \beta) = \begin{cases} 1, & \text{если } \xi = A, \\ 0, & \text{если } \xi \neq A. \end{cases}$$

© *И.Я. Спекторский*, 2011

Пример 1.

Рассмотрим формальную грамматику $G =< \{S,A\},\{a,b,c\},P,S >$, где множество продукций $P = \{S \to aSab \mid A,\ A \to cA \mid \varepsilon\}$. Сеть Петри, соответствующая грамматике G, изображена на рис. 1.

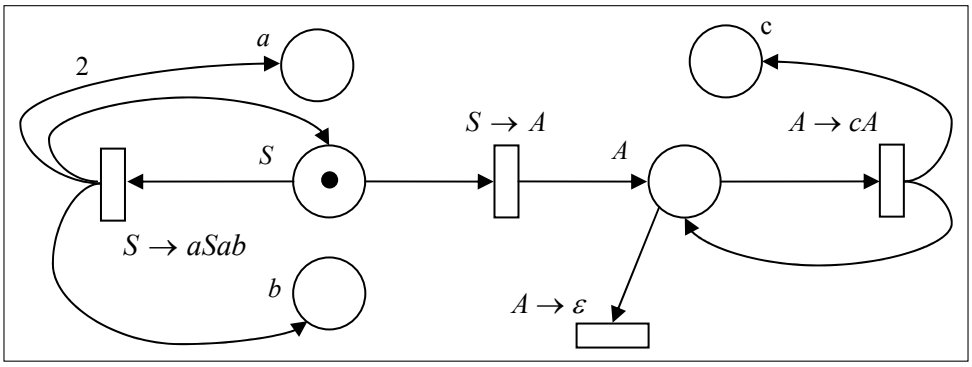

Рис. 1. Сеть Петри для грамматики с продукциями $S \to aSab \mid A,\ A \to cA \mid \varepsilon$

Легко увидеть, что данная грамматика порождает язык $\{a^n c^m (ab)^n : n \ge 0,\ m \ge 0\}$, причем каждое слово $a^n c^m (ab)^n$ порождается последовательностью применений продукций (запусков переходов) $(S \to aSab)^n (S \to A)(A \to cA)^m (A \to \varepsilon)$, что приводит к маркировке $(0,0,n,n,m)$ (предполагая порядок позиций в соответствии с перечислением в определении грамматики, т.е. S, A, a, b, c).

АНАЛИЗ С ПОМОЩЬЮ ДЕРЕВА ПОКРЫВАЕМОСТИ

Ряд свойств сети Петри N_G при фиксированной начальной маркировке ($= \{1,2,\ldots,n,\ldots\}$ — множество натуральных чисел) эффективно анализируются с помощью дерева покрываемости [1, 2].

Символ $A \in N$ называют *порождающим*, если $A \overset{*}{\Rightarrow} w$ для некоторого $w \in \Sigma$. Очевидно, что все продукции, содержащие непорождающие символы, можно исключить из P, не сужая язык, порождаемый грамматикой. Продукцию $A \to \beta$ называют *рекурсивной*, если β содержит A. Очевидно, что исключение рекурсивных продукций не влияет на порождаемость или непорождаемость нетерминальных символов.

Через μ_A ($A \in N$) обозначим такую маркировку, что

$$\mu_A = \begin{cases} 1, & \text{если } \xi = A, \\ 0, & \text{если } \xi \in (N \cup \Sigma) \setminus \{A\}. \end{cases}$$

Через T_A обозначим дерево покрываемости для начальной маркировки μ_A. Через T_A^{nr} обозначим дерево покрываемости для начальной маркировки μ_A, построенное без учета переходов, соответствующих рекурсивным продукциям.

Справедливость следующих двух теорем устанавливается непосредственно из построения сети Петри по заданной КС-грамматике.

Теорема 1. Символ $A \in N$ является порождающим тогда и только тогда, когда дерево T_A^{nr} содержит не менее одной маркировки μ, такой, что $\mu(\xi) = 0$ для всех $\xi \in N$.

Следствие. Язык, порождаемый грамматикой G, является непустым тогда и только тогда, когда дерево T_S^{nr} содержит не менее одной маркировки μ такой, что $\mu(\xi) = 0$ для всех $\xi \in N$ (т.е., когда порождающим является источник S).

Теорема 2. Грамматика G порождает бесконечный язык тогда и только тогда, когда дерево T_S содержит не менее одной маркировки μ, такой, что $\mu(\xi) = 0$ для всех $\xi \in N$, и $\mu(a) = \omega$ для некоторого $a \in \Sigma$.

Пример 2. Рассмотрим формальную грамматику $G =< \{S, A\}, \{a, b, c\}, P, S >$, где множество продукций $P = \{S \to aSab \mid cA,\ A \to c \mid \varepsilon\}$. Сеть Петри, соответствующая грамматике G, изображена на рис. 2.

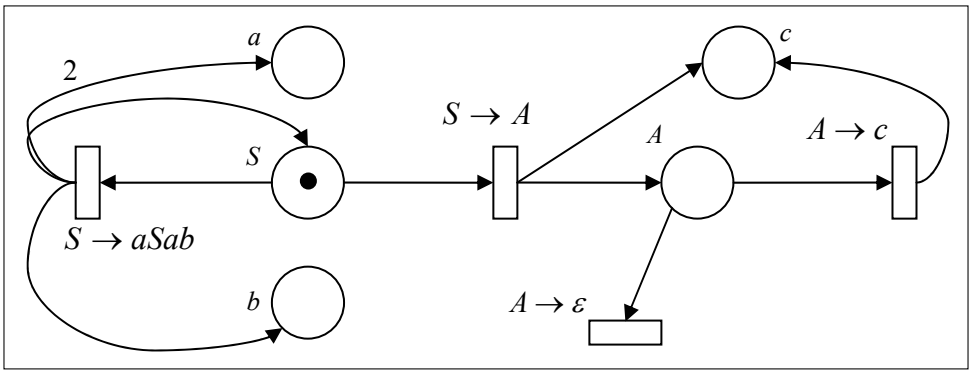

Рис. 2. Сеть Петри для грамматики с продукциями $S \to aSab \mid cA,\ A \to c \mid \varepsilon$

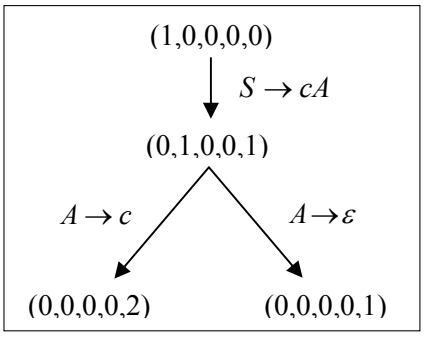

Рис. 3. Дерево T_S^{nr} для грамматики с продукциями $S \to aSab \mid cA,\ A \to c \mid \varepsilon$

Для анализа порождаемого языка на пустоту и конечность построим дерево T_S^{nr} (рис. 3) и дерево T_S (рис. 4). При построении дерева T_S^{nr}, в соответствии с определением, была исключена рекурсивная продукция $S \to aSab$. Порядок позиций при записи маркировок предполагался в соответствии с перечислением в определении грамматики, т.е. S, A, a, b, c.

Поскольку T_S^{nr} содержит маркировки μ, удовлетворяющие условию $\mu(\xi) = 0$ для всех $\xi \in \{S, A\}$ (например, маркировку (0, 0, 0, 0, 2)), заданная грамматика порождает непустой язык. Далее, дерево T_S содержит маркировки μ, удовлетворяющие условию $\mu(\xi) = 0$ для всех $\xi \in \{S, A\}$ и содержащие символ ω (маркировки $(0, 0, \omega, \omega, 1)$ и $(0, 0, \omega, \omega, 2)$), откуда следует

бесконечность языка, порождаемого данной грамматикой. Более того, поскольку символ ω находится на позициях, соответствующих символам a и b, можем сделать вывод, что порождаемый язык содержит слова со сколь угодно большим числом вхождений a и b (но не c).

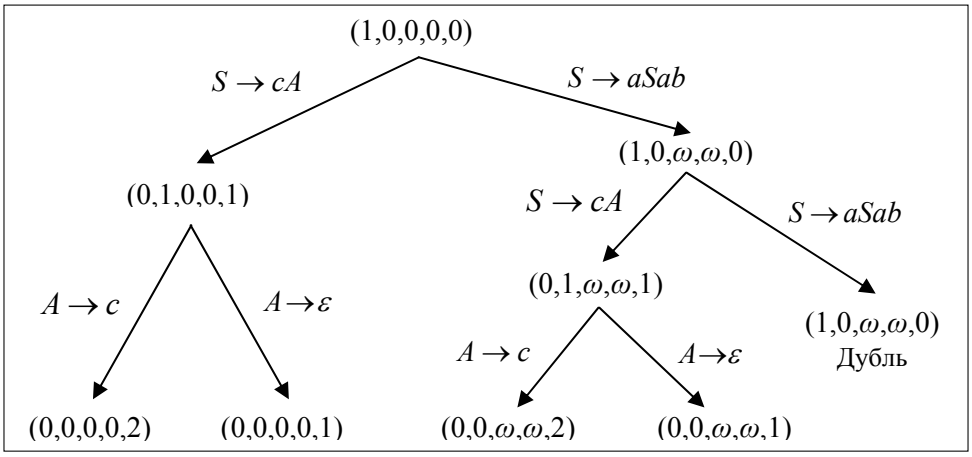

Рис. 4. Дерево T_S для грамматики с продукциями $S \to aSab \mid cA,\ A \to c \mid \varepsilon$

АНАЛИЗ С ПОМОЩЬЮ УРАВНЕНИЯ СОСТОЯНИЯ

Для анализа порождаемости грамматикой G слова $w \in \Sigma^*$ можно использовать уравнение состояний сети N_G [1, 2]. Обозначим через μ_w $(w \in \Sigma^*)$ маркировку такую, что $\mu_w(\xi) = \mid w \mid_\xi$. Очевидно, что для порождаемости слова w необходимо (но недостаточно), чтобы уравнение состояния имело хотя бы одно решение для правой части $\Delta\mu = \mu_w - \mu_S$.

Пример 3.

Рассмотрим грамматику $G = < \{S\}, \{a, b\}, \{S \to aSb \mid bSa \mid \varepsilon\} >$. Соответствующая сеть Петри N_G изображена на рис. 5.

Выпишем матрицу инцидентности для N_G, предполагая упорядоченность позиций (нетерминальных и терминальных символов) и переходов (продукций) в соответствии с перечислением в определении грамматики:

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ -1 & 0 & 0 \end{pmatrix}.$$

Уравнение состояния $A^T u = \Delta\mu$ имеет решение относительно вектора запусков $u = (u_1, u_2, u_3)$ лишь для правых частей вида $\Delta\mu = (x, y, y)$. Таким образом, из начальной маркировки $\mu_S = (1, 0, 0)$ могут достигаться только маркировки с одинаковыми второй и третьей координатами. Это означает, что грамматика G может генерировать лишь слова с одинаковым числом вхождений a и b. Однако разрешимость уравнения состояния — лишь не-

обходимое условие для порождаемости слова грамматикой; в данном примере G порождает лишь слова вида ww^r, т.е. палиндромы четной длины.

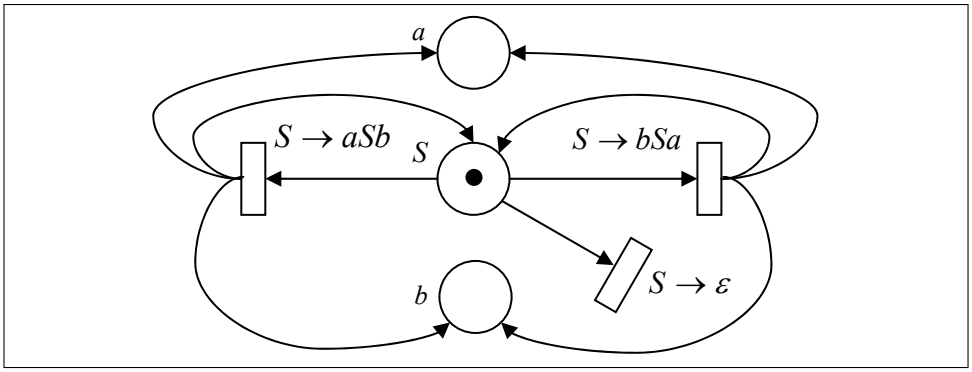

Рис. 5. Сеть Петри для грамматики с продукциями $S \to aSb \mid bSa \mid \varepsilon$

ВЫВОДЫ

Предложенный метод анализа КС-грамматик с помощью сетей Петри позволяет исследовать свойства, связанные с количеством вхождений того или иного символа терминального алфавита в слова, порождаемые данной грамматикой. В частности, с помощью сетей Петри удобно анализировать порождаемый язык на пустоту и конечность.

Используя дерево покрываемости, можно не только выявить факт бесконечности порождаемого языка, но и определить, какие именно символы могут встречаться в порождаемых словах в сколь угодно большом количестве.

Анализируя уравнение состояния сети на разрешимость, можно получить необходимые условия порождаемости слов, т.е. «оценить сверху» множество слов, которые могут порождаться данной грамматикой.

К достоинствам рассмотренного метода можно отнести простоту и наглядность. Недостаток метода, ограничивающий его применение, связан с невозможностью отслеживать порядок букв в порождаемом слове.

Направлением для дальнейших исследований предполагается обобщение рассмотренного метода на более широкий класс формальных грамматик. Кроме того, используя различные расширения сетей Петри, можно попытаться модифицировать метод с тем, чтобы получить контроль над расположением символов в словах, порождаемых данной грамматикой.

ЛИТЕРАТУРА

1. *Питерсон Дж.* Теория сетей Петри и моделирование систем. — М.: Мир, 1984. — 264 с.
2. *Котов В.Е.* Сети Петри. — М.: Наука, 1984. — 160 с.
3. *Алгоритмічні* алгебри: навч. посіб. — Київ: ІЗМН, 1997. — 480 с.
4. *Рейуорд-Смит В.Дж.* Теория формальных языков. Вводный курс. — М.: Радио и связь, 1988. — 128 с.
5. *Гросс М., Лантен А.* Теория формальных грамматик. — М.: Мир, 1971. — 296 с.
6. *Ахо А., Ульман Дж.* Теория синтаксического анализа, перевода и компиляции. У 2 т. Т. 1. — М.: Мир, 1978. — 614 с.

Поступила 24.06.2010

УДК 517.9

ДОВГОСТРОКОВІ ПРОГНОЗИ ФУНКЦІЙ СТАНУ КВАЗІЛІНІЙНИХ ГІПЕРБОЛІЧНИХ СИСТЕМ В $\mathbb{R}^n$

Н.В. ГОРБАНЬ

Розглянуто квазілінійну гіперболічну систему в просторі $\mathbb{R}^n$ з нелінійною, у загальному випадку, функцією взаємодії. Досліджено асимптотичну поведінку слабких розв'язків поставленої задачі. Доведено існування траєкторного та глобального атракторів для всіх слабких розв'язків вихідної задачі. Доведено, що потраєкторно всі слабкі розв'язки прямують до стаціонарних станів.

ВСТУП

Інтенсивні дослідження керованих п'єзоелектричних процесів і полів обумовлюють необхідність вивчення їх математичних моделей, які містять у собі квазілінійні диференціальні рівняння з частинними похідними гіперболічного типу. Більшої складності такі задачі набувають у випадках необмежених областей. У цій роботі ми зосередимось на дослідженні довгострокових прогнозів функцій стану таких задач із використанням теорії глобальних і траєкторних атракторів для м-напівпотоків, розроблених у роботах [1–10]. Зазначимо, що багатозначна динаміка розв'язків квазілінійних гіперболічних рівнянь в обмежених областях досліджувалась у роботах [1, 2, 3, 6, 9].

ПОСТАНОВКА ЗАДАЧІ

Розглядається рівняння

$$u_{tt} + \gamma u_t - \Delta u + f(x,u) = h(x), \quad (t,x) \in \mathbb{R}_+ \times \mathbb{R}^n, \tag{1}$$

де $\gamma > 0$, $n \geq 3$, $f : \mathbb{R}^{n+1} \to \mathbb{R}$ — вимірна по x та неперервна по t функція, яка задовольняє умовам:

$$h \in L^2(\mathbb{R}^n), \quad \exists\, C_1, C_2 \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n), \ C_1 \geq 0, \ \exists \alpha > 0, \ \exists c \geq 0:$$

$$|f(x,u)| \leq C_1(x) + c|u| \quad \forall\, (x,u) \in \mathbb{R}^{n+1},$$

$$\forall\, (x,u) \in \mathbb{R}^{n+1} \quad f(x,u)u \geq \alpha u^2 - C_2(x). \tag{2}$$

Надалі γ, C_1, C_2, c, α називатимемо константами задачі (1).

Покладемо

$$F(x,u) := \int_0^u f(x,s)ds, \ \ \forall (x,u) \in \mathbb{R}^{n+1}.$$

Оскільки f задовольняє умовам Каратеодорі, то $\forall (x,u) \in \mathbb{R}^{n+1}$:

© *Н.В. Горбань*, 2011

$$|F(x,u)| \le C_1(x)|u| + \frac{c}{2}|u|^2, \quad F(x,u) \ge \frac{\alpha}{2}u^2 - C_2(x)|u|. \tag{3}$$

Позначатимемо через $|\cdot|, (\cdot,\cdot)$ та $\|\cdot\|$ $((\cdot,\cdot))$ норму і скалярний добуток у $L^2(\mathbb{R}^n)$ та $H^1(\mathbb{R}^n)$ відповідно. Зауважимо, що $\forall u, v \in H^1(\mathbb{R}^n)$

$$((u,v)) = \frac{\alpha}{4}(u,v) + \sum_{i=1}^{n}\left(\frac{\partial u}{\partial x_i}, \frac{\partial v}{\partial x_i}\right).$$

Фазовим простором задачі (1) є простір $E = H^1(\mathbb{R}^n) \times L^2(\mathbb{R}^n)$.

Розв'язок задачі (1) розумітимемо в сенсі такого означення.

Означення 1. Функція $\varphi(\cdot) = (u(\cdot), u_t(\cdot))^T \in L^\infty(\tau, T; E)$ називається слабким розв'язком задачі (1) на (τ, T), якщо $\forall \psi \in H_0^1(\mathbb{R}^n)$, $\forall \eta \in \mathbb{C}_0^\infty(\tau, T)$:

$$-\int_\tau^T (u_t, \psi)\eta_t + \int_\tau^T (\gamma(u_t, \psi) + ((u,\psi)) + (f(x,u),\psi) - (h,\psi))\eta = 0. \tag{4}$$

Ставиться задача дослідження асимптотичної поведінки всіх слабких розв'язків задачі (1).

ОСНОВНІ РЕЗУЛЬТАТИ

Розглянемо клас функцій $W_\tau^T = \mathbb{C}([\tau, T]; E)$. Із умов (2), (3) для довільної функції $\varphi(\cdot) = (u(\cdot), u_t(\cdot))^T \in W_\tau^T$ коректно визначений такий функціонал:

$$V(\varphi(t)) = \frac{1}{2}|u_t(t)|^2 + \frac{1}{2}\|u(t)\|^2 + (F(\cdot, u(t)), 1) + (h, u(t)), \quad t \in [\tau, T].$$

Лема 1. Для довільного $\varphi(\cdot) = (u(\cdot), u_t(\cdot))^T \in W_\tau^T$ — слабкого розв'язку (1) на (τ, T) функція $V(\varphi(\cdot))$ є абсолютно неперервною на $[\tau, T]$ і для майже всіх $t \in [\tau, T]$

$$\frac{d}{dt}V(\varphi(t)) = -\gamma |u_t(t)|^2 .$$

Більше того, $\forall\, t \ge s$, $t, s \in [\tau, T]$: $V(\varphi(t)) \le V(\varphi(s))$.

Доведення. Нехай $\varphi(\cdot) = (u(\cdot), u_t(\cdot))^T \in W_\tau^T$ — довільний слабкий розв'язок задачі (1) на (τ, T). Тоді, внаслідок (2), $f(x,u) \in L^2(\tau, T; L^2(\mathbb{R}^n))$. Отже, із [1] функція $t \mapsto |u_t(t)|^2 + \|u(t)\|^2$ є абсолютно неперервною на $[\tau, T]$ і майже скрізь.

$$\frac{1}{2}\frac{d}{dt}\{|u_t|^2 + \|u\|^2\} = -\gamma|u_t|^2 - (f(x,u), u_t) + (h, u_t). \tag{5}$$

Для того, щоб довести, що функція $t \mapsto (F(x,u(t)),1)$ є абсолютно неперервною на $[\tau, T]$ і майже скрізь на $[\tau, T]$ виконується рівність

$$\frac{d}{dt}(F(x,u),1) = (f(x,u), u_t). \tag{6}$$

Внаслідок [1] достатньо показати її неперервність на $[\tau,T]$ і виконання (6) у сенсі скалярних розподілів на (τ,T). Доведення аналогічне [2, 3]. Таким чином, твердження леми стає очевидним.

Лему доведено.

Зауваження 1. Із твердження леми 1 випливає, що V — функція Ляпунова для задачі (1).

Оскільки $H^1(\mathbb{R}^n)$ неперервно вкладається в $L^2(\mathbb{R}^n)$, то з умов (2) для $u \in L^\infty(\tau,T;H^1(\mathbb{R}^n))$ маємо вкладення $f(x,u) \in L^2(\tau,T;L^2(\mathbb{R}^n))$. Отже, згідно з [1], для кожного слабкого розв'язку задачі (1) $\varphi(\cdot)$ на (τ,T) маємо $\varphi(\cdot) \in C([\tau,T];E)$. Це зумовлює вибір класу W_τ^T. Вкладення $\varphi(\cdot) \in C([\tau,T];E)$ дозволяє для (1) ставити задачу Коші з початковими умовами

$$u|_{t=0} = u_0 \in H^1(\mathbb{R}^n), \quad u_t|_{t=0} = v_0 \in L^2(\mathbb{R}^n). \tag{7}$$

і шукати розв'язок лише в класі $L^\infty(\tau,T;E)$.

Теорема 1. Для довільних $\varphi_0 = (u_0,v_0)^T \in E$, $T > 0$ задача Коші (1), (7) за виконання умов (2), має принаймні один слабкий розв'язок у класі W_0^T.

Доведення. З певними технічними модифікаціями повторює доведення теореми 1 із [4].

Поєднуючи результати теореми 1 та леми 1, одержимо, що для довільних $\varphi_0 = (u_0,v_0)^T \in E$ задача Коші (1), (7) за умов (2), має принаймні один розв'язок у класі $C^{\mathrm{loc}}((0,+\infty);E) \cap L^\infty_{\mathrm{loc}}((0;+\infty);E)$. Покладемо $W_0^\infty = C^{\mathrm{loc}}((0,+\infty);E)$. Для довільного $\varphi_0 \in E$ нехай $\mathcal{D}(\varphi_0)$ — сукупність усіх слабких розв'язків (визначених на $[0,+\infty)$) задачі (1) з початковими даними $\varphi(0) = \varphi_0$. Тепер для довільних $t \geq 0$, $\varphi_0 \in E$ розглянемо множину

$$G(t,\varphi_0) = \{\varphi(t) \mid \varphi(\cdot) \in \mathcal{D}(\varphi_0)\} \subset E.$$

Для непорожніх $A, B \subset E$:

$$\operatorname{dist}(A,B) = \sup_{x\in A}\inf_{y\in B} \|x-y\|_E, \quad \operatorname{dist}_H(A,B) = \max\{\operatorname{dist}(A,B), \operatorname{dist}(B,A)\},$$

$$O_\delta(A) = \{x \in E \mid \operatorname{dist}(x,A) < \delta\}, \quad B_r = \{x \in E \mid \rho(x,0) \leq r\},$$

$\overline{A} = \mathrm{cl}_E A$ — замикання A у E, $P(E)$ — сукупність усіх непорожніх підмножин E, $\beta(E)$ — сукупність усіх непорожніх обмежених підмножин E, $C(E)$ — сукупність усіх непорожніх замкнених підмножин в E, $K(E)$ — сукупність усіх непорожніх компактних підмножин E.

Зазначимо, що відображення $G: \mathbb{R}_+ \times E \mapsto P(E)$ — строгий м-напівпотік на E, тобто

- $G(0,\cdot) = I_E$ — тотожне відображення E;

- $G(t+s,x) = G(t,G(s,x)) \quad \forall\, t,s \in \mathbb{R}_+,\ \forall\, x \in E$.

Означення 2. Множина $A \subset E$ називається притягуючою множиною для м-напівпотока G, якщо для довільного $B \in \beta(E)$ і довільного околу $N(A)$ множини A в E існує $T = T(N(A),B) \in \mathbb{R}_+$ така, що $\forall\, t \geq T$ $G(t,B) \subset N(A)$ $\forall\, t \geq T$ [5].

Означення 3. Множина $\Theta \subset E$ називається глобальним атрактором для м-напівпотока G, якщо:

- Θ притягуюча множина;
- для довільної притягуючої множини Y $\Theta \subset \mathrm{cl}_E Y$ (мінімальність);
- $\Theta \subset G(t,\Theta)$, для всіх $t \geq 0$ (півінваріантність) [5].

Означення 4. М-напівпотік G називається асимптотично компактним, якщо для довільного $B \in \beta(E)$ існує $A(B) \in K(E)$ таке, що

$$\mathrm{dist}\,(G(t,B),A(B)) \to 0,\ \ t \to +\infty \quad [6].$$

Теорема 2. Нехай для задачі (1) виконано мови (2). Тоді відображення G є м-напівпотоком, для якого у фазовому просторі $E = H^1(\mathbb{R}^n) \times L^2(\mathbb{R}^n)$ існує компактний, інваріантний глобальний атрактор.

Доведення. Для доведення виконання твердження цієї теореми, беручи до уваги [6], твердження леми 1 та теореми 1, достатньо перевірити такі властивості:

- Нехай $\{\varphi^n\} \subset W_\tau^T$ — послідовність розв'язків задачі (1), причому $\varphi^n(\tau) \to \varphi_\tau$ слабо в E. Нехай задана послідовність $\{t_n\} \subset [\tau,T]$ така, що $t_n \to t_0 \in [\tau,T]$. Тоді існує $\varphi \in W_\tau^T$ — розв'язок (1) такий, що $\varphi(\tau) = \varphi_\tau$ і принаймні з точністю до підпослідовності $\varphi^n(t_n) \to \varphi(t_0)$ слабо в E. Якщо ж $\varphi^n(\tau) \to \varphi_\tau$ сильно в E, то принаймні з точністю до підпослідовності $\varphi^n(t_n) \to \varphi(t_0)$ сильно в E.
- М-напівпотік G асимптотично компактний.

Ці твердження перевіряються (із певними технічними модифікаціями) за допомогою техніки, поданої в роботах [4, 7].

Теорему доведено.

Відповідно до [6], позначимо сукупність стаціонарних станів G через $Z(G)$. Зазначимо, що

$$Z(G) = \{(z,\bar{0}) \in E \,|\, ((z,\psi)) + (f(x,z),\psi) = (h,\psi) \ \ \forall\, \psi \in H^1(\mathbb{R}^n)\}$$

є обмеженою множиною в E.

Розглянемо сімейство $K_+ = \cup_{y_0 \in H} \mathcal{D}(y_0)$ усіх слабких розв'язків рівняння (1), визначених на $[0,+\infty)$. Зазначимо, що K_+ є трансляційно інваріантним, тобто $\forall u(\cdot) \in K_+$, $\forall h \geq 0$ $u_h(\cdot) \in K_+$, де $u_h(s) = u(h+s)$, $s \geq 0$. Задамо на K_+ напівгрупу трансляцій $\{T(h)\}_{h \geq 0}$, $T(h)u(\cdot) = u_h(\cdot)$, $h \geq 0$, $u \in K_+$. Внаслідок трансляційної інваріантності K_+ отримуємо, що $T(h)K_+ \subset K_+$ при $h \geq 0$.

Побудуємо атрактор трансляційної напівгрупи $\{T(h)\}_{h\geq 0}$, що діє на K_+. На K_+ розглядатимемо топологію, індуковану з простору Фреше $C^{\mathrm{loc}}(\mathbb{R}_+;E)$. Зауважимо, що

$$f_n(\cdot)\to f(\cdot) \text{ в } C^{\mathrm{loc}}(\mathbb{R}_+;E) \Leftrightarrow \forall M>0\colon \Pi_M f_n(\cdot)\to \Pi_M f(\cdot) \text{ в } C([0,M];E),$$

де Π_M — оператор обмеження на відрізку $[0,M]$ [8]. Позначимо через Π_+ оператор обмеження на $[0,+\infty)$.

Нагадаємо, що множину $P\subset C^{\mathrm{loc}}(\mathbb{R}_+;E)\cap L_\infty(\mathbb{R}_+;E)$ називають притягуючою для простору траєкторій K_+ рівняння (1) в топології $C^{\mathrm{loc}}(\mathbb{R}_+;E)$, якщо для довільної обмеженої в $L_\infty(\mathbb{R}_+;E)$ множини $B\subset K_+$ та довільного числа $M\geq 0$ виконується співвідношення

$$\mathrm{dist}_{C([0,M];E)}(\Pi_M T(t)B, \Pi_M P)\to 0, \quad t\to+\infty.$$

Множина $U\subset K_+$ називається траєкторним атрактором у просторі траєкторій K_+ відносно топології $C^{\mathrm{loc}}(\mathbb{R}_+;E)$ [8], якщо:

- U — компактна в $C^{\mathrm{loc}}(\mathbb{R}_+;E)$ та обмежена в $L_\infty(\mathbb{R}_+;E)$;
- U — строго інваріантна відносно $\{T(h)\}_{h\geq 0}$, тобто $T(h)U=U$ $\forall h\geq 0$;
- U є притягуючою множиною для простору траєкторій K_+ у топології $C^{\mathrm{loc}}(\mathbb{R}_+;E)$.

Розглянемо рівняння (1) на всій числовій прямій. Аналогічно простору $C^{\mathrm{loc}}(\mathbb{R}_+;E)$ простір $C^{\mathrm{loc}}(\mathbb{R};E)$ оснащується топологією локальної рівномірної збіжності на кожному відрізку $[-M,M]\subset\mathbb{R}$ [8]. Функція $u\in C^{\mathrm{loc}}(\mathbb{R};E)\cap L_\infty(\mathbb{R};E)$ називається повною траєкторією рівняння (1), якщо $\forall h\in\mathbb{R}$ $\Pi_+ u_h(\cdot)\in K_+$ [8]. Нехай K — сукупність усіх повних траєкторій рівняння (1). Зазначимо, що

$$\forall h\in\mathbb{R},\ \ \forall u(\cdot)\in K:\ u_h(\cdot)\in K.$$

Теорема 3. Нехай A — глобальний атрактор з теореми 2. Тоді в просторі K_+ існує траєкторний атрактор $P\subset K_+$. При цьому істинна формула

$$P=\Pi_+ K=\Pi_+\{y\in K\mid y(t)\in A\ \ \forall t\in\mathbb{R}\}.$$

Доведення. Твердження теореми є прямим наслідком теореми 3.9 з [9].

Теорема 4. Для довільної повної траєкторії $\psi\in K$ граничні множини

$$\alpha(\psi)=\{z\in E\mid \psi(t_j)\to z \text{ для деякої послідовності } t_j\to-\infty\},$$

$$\omega(\psi)=\{z\in E\mid \psi(t_j)\to z \text{ для деякої послідовності } t_j\to+\infty\}$$

є зв'язними підмножинами $Z(G)$ на яких функція Ляпунова V стала.

Доведення. Твердження теореми є прямим наслідком теореми 2.7 з [6], леми 1 та теореми 2.

ВИСНОВКИ

Для квазілінійного гіперболічного рівняння (1) у необмеженій області з нелінійною, у загальному випадку, неперервною функцією взаємодії доведено існування глобального та траєкторного атракторів у природному фазовому та, відповідно, природному розширеному фазовому просторі. Перевірено також, що потраєкторно всі слабкі розв'язки прямують до стаціонарних положень. Результати можуть знайти своє застосування в задачах керування п'єзоелектричними процесами у формі оберненого зв'язку.

ЛІТЕРАТУРА

1. *Temam R.* Infinite-dimensional dynamical systems in mechanics and physics. — Berlin: Springer, 1988. — 500 p.
2. *Капустян О.В., Іоване Ж.* Глобальний атрактор для неавтономного хвильового рівняння без єдиності розв'язку // Системні дослідження та інформаційні технології. — 2006. — № 2. — С. 107–120.
3. *Капустян О.В.* Властивість Кнезера для неавтономного хвильового рівняння без єдиності розв'язку // Наук. вісті НТУУ «КПІ». — 2007. — № 2. — С. 107–120.
4. *Stanzhyts'kyi O.M., Horban' N.V.* Global attractor for the autonomous wave equation in with continious nonlinearity // Ukrainian Mathematical Journal. — 2008. — **60**, № 2. — P. 299–309.
5. *Melnik V.S., Valero J.* On attractors of multivalued semi-flows and differential inclusions // Set-Valued Analysis. — 1998. — **6**, № 1. — P. 83–111.
6. *Ball J.M.* Global attractors for damped semilinear wave equations // Discrete and Continuous Dynamical System. — 2004. — **10**. — P. 31–52.
7. *Stanzhitsky A.N., Gorban N.V.* On the dynamics of solutions for autonomous reaction-diffusion equation in with multivalued nonlinearity // Український математичний вісник. — 2009. — Т. 6, № 2. — С. 235–251.
8. *Вишик М.И., Чепыжов В.В.* Траекторный и глобальный аттракторы 3D системы Навье-Стокса // Математические заметки. — 2002. — Т. 71, № 2. — С. 194–213.
9. *Kapustyan O.V., Valero J.* Comparison between trajectory and global attractors for evolution systems without uniqueness of solutions // International Journal of Bifurcation and Chaos. — 2010. — **20**, Issue 9. — P. 2723–2734.
10. *Zgurovsky M.Z., Mel'nik V.S., Kasyanov P.O.* Evolution Inclusions and Variation Inequalities of Earth Data Processing II. — Heidelberg: Springer, 2011. — 274 p.

Надійшла 07.06.2011

АВТОМАТИЗОВАНІ СИСТЕМИ УПРАВЛІННЯ **AUTOMATED CONTROL SYSTEMS**

УДК 62-50

Синтез многомерных координирующих систем управления с разнотемповой дискретизацией в детерминированной среде / **Романенко В.Д., Милявский Ю.Л.** // Системні дослідження та інформаційні технології. — 2011. — № 4. — С. 7–20.

Поставлена и решена задача управления соотношениями между выходными координатами многомерного разнотемпового объекта в детерминированной среде. Предложен новый алгоритм управления, позволяющий одновременно обеспечить выполнение задачи слежения и задачи координации в переходном и установившемся режимах. Приведены результаты численного моделирования, подтверждающие практическую значимость данной разработки. Рис.: 12. Библиогр.: 6 назв.

UDC 62-50

Synthesis of multivariate coordinating multidimensional coordination control systems with multirate sampling in deterministic enviromemt / Romanenko V.D., Milyavsky Y.L. // System Research and Information Technologies. — 2011. — № 4. — P. 7–20.

The control problem of relations between the output coordinates of the multidimensional multirate objects in deterministic environment is formulated and solved. A new control algorithm that ensures simultaneous fulfillment of both monitoring and coordinating problems in transient and steady states is proposed. The results of numerical simulation, confirming the practical significance of this development are shown. Figs: 12. Refs: 6 titles.

ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ, ВИСОКОПРОДУКТИВНІ КОМП'ЮТЕРНІ СИСТЕМИ **PROGRESSIVE INFORMATION TECHNOLOGIES, HIGH-EFFICIENCY COMPUTER SYSTEMS**

УДК 004.724.2, 004.75

Алгоритмы балансирования нагрузки в Грид-системах / **Петренко А.И., Свистунов С.Я., Свирин П.В.** // Системні дослідження та інформаційні технології. — 2011. — № 4. — С. 21–36.

Проведен анализ стратегий балансирования нагрузки в Грид-системах и сравнительный анализ имеющихся алгоритмов распределения потока заданий между вычислительными ресурсами Грид-среды. Рис.: 10. Табл.: 1. Библиогр.: 52 назв.

UDC 004.724.2, 004.75

Algorithm of balancing the load in the Grid-systems / Petrenko A.I., Svistunov S.Ya., Svirin V.I. // System Research and Information Technologies. — 2011. — № 4. — P. 21–36.

The analysis of strategies of balancing load in the Grid-systems and comparative analysis of existing algorithms of the distribution of flow of the tasks between computing resources of a Grid environment are shown. Figs: 10. Tabl.: 1. Refs: 52 titles.

УДК 621.391:519.2

Неасимптотичні оцінки ефективності випадкового кодування в системі передачі інформації двійковим симетричним каналом зв'язку з відводом / **Олексійчук А.М., Гришаков С.В.** // Системні дослідження та інформаційні технології. — 2011. — № 4. — С. 37–47.

Досліджено ефективність випадкового кодування під час багаторазової передачі безнадлишкових повідомлень двійковим симетричним каналом зв'язку з відводом. Отримано неасимптотичні оцінки надійності відновлення повідомлень законним одержувачем інформації та стійкості їх захисту у відвідному каналі. Табл.: 2. Бібліогр.: 17 назв.

UDC 621.391:519.2

Unasymptotic estimates of random coding efficiency in the system of information transfer channel with take-off / Alekseychuk A.N., Gryshakov S.V. // System Research and Information Technologies. — 2011. — № 4. — P. 37–47.

The efficiency of random coding during the multiple irredundant messages transmission through a binary symmetric channel with take-off is investigated. Unasymptotic estimates of the reliability of message recovery by a legitimate recipient of information and the sustainability of their protection in take-out channel are obtained. Tabl.: 2. Refs: 17 titles.

ПРОБЛЕМИ ПРИЙНЯТТЯ РІШЕНЬ І УПРАВЛІННЯ В ЕКОНОМІЧНИХ, ТЕХНІЧНИХ, ЕКОЛОГІЧНИХ І СОЦІАЛЬНИХ СИСТЕМАХ

DECISION MAKING AND CONTROL IN ECONOMIC, TECHNICAL, ECOLOGICAL AND SOCIAL SYSTEMS

УДК 004.946

Компьютерное моделирование украинского жестового языка / **Кривонос Ю.Г., Крак Ю.В., Бармак О.В.** // Системні дослідження та інформаційні технології. — 2011. — № 4. — С. 48–60.

Предложена математическая модель и ее компьютерная реализация для получения и воспроизведения жестового языка с помощью трехмерных моделей людей. Предложена технология достаточно реалистически воспроизводить на трехмерной модели движения, полученные из видеоизображения конкретного человека — носителя жестового языка. Рис.: 7. Библиогр.: 6 назв.

UDC 004.946

Computer-based modeling of Ukrainian sigh language / Kryvonos Iu.G., Krak Iu.V., Barmak O.V. // System Research and Information Technologies. — 2011. — № 4. — P. 48–60.

A mathematical model and its software implementation for sign language speech obtaining and reproduction using 3d human model has been suggested. The technology of enough realistically reproduction of 3d model of the traffic received from the motion picture of the particular person — the bearer of the sigh language, — is proposed. Figs: 7. Refs: 6 titles.

УДК 519.816

Математичні моделі керування процесами фінансування інвестиційних проектів / **Москаленко В.В., Кондращенко В.В.** // Системні дослідження та інформаційні технології. — 2011. — № 4. — С. 61–73.

Розглянуто проблему побудови оптимальної схеми фінансування інвестиційного проекту із залученням зовнішніх джерел фінансування. Запропоновано комплекс математичних моделей, що відповідають конкретним формам фінансування, кожна з яких дозволяє розрахувати оптимальну схему фінансування та отримати оцінку її економічної ефективності з урахуванням специфіки кожної конкретної форми фінансування. Рис.: 1. Бібліогр.: 7 назв.

UDC 519.816

Mathematical models of investment projects funding processes management / Moskalenko V.V., Kondrashchenko V.V. // System Research and Information Technologies. — 2011. — № 4. — P. 61–73.

The construction of an optimal scheme of financing of the investment project with the involvement of external sources of financing are considered. A complex of mathematical models, relevant to the specific forms of financing each of which allows to calculate the optimal financing scheme and get an estimate of its cost-effectiveness taking into account the characteristics of each specific forms is proposed. Figs: 1. Refs: 7 titles.

УДК 004.9 (075)

Методы и технологии снижения пиковых нагрузок в корпоративных информационных системах с централизованным сохранением данных / Симашко В.И. // Системні дослідження та інформаційні технології. — 2011. — № 4. — С. 74–87.

Описаны пиковые нагрузки, возникающие в больших корпоративных информационных системах (КИС) с централизованным хранением данных. Предлагаются методы и технологии, позволяющие снизить или ликвидировать пиковые нагрузки на серверы и на сеть системы. Рис.: 2. Библиогр.: 3 назв.

UDC 004.9 (075)

Methods and technologies of reducing peak-loads in the corporate information systems with the centralized data preservation / Simashko V.Yo. // System Research and Information Technologies. — 2011. — № 4. — P. 74–87.

Peak overloads which arise in the big corporate information systems with the centralized data preservation are described. Methods and the technologies, which allow to reduse or liquidate peak overloads on servers and on a system network are offered. Figs: 2. Refs: 3 titles.

УДК 004.681

Математические модели экономического менеджмента информационной безопасности / Левченко Е.Г., Демчишин М.В., Рабчун А.А. // Системні дослідження та інформаційні технології. — 2011. — № 4. — С. 88–96.

Проведен сравнительный анализ количественных моделей экономического менеджмента информационной безопасности. Поставлены условия, при которых целевые функции рассмотренных моделей совпадают или достаточно близки. Выполненные расчеты иллюстрируют методику определения количества утраченной информации в зависимости от соотношения ресурсов нападения и защиты. Рис.: 3. Табл.: 2. Библиогр.: 8 назв.

UDC 004.681

The mathematical models of economic management of information security / Levchenko Y.H., Demchyshyn M.V., Rabchun A.O. // System Research and Information Technologies. — 2011. — № 4. — P. 88–96.

The comparative analysis of the quantitative models of economic management of information security is fulfilled. The conditions under which the target functions of the considered models coincide or sufficiently close are set up. The fulfilled calculations illustrate the methodology of quantity determination depending on attack and defense recourses correlation. Figs: 3. Tabl.: 2. Refs: 8 titles.

УДК 528.852:531.012

Прогнозування ерозійних процесів ґрунтів на маршрутах прокладки магістральних трубопроводів на основі геоінформаційних технологій і космічних знімків із високою роздільною здатністю зображення / Джамалов А.Т., Рагімов Р.М // Системні дослідження та інформаційні технології. — 2011. — № 4. — С. 97–103.

Досліджується можливість прогнозування ерозійних процесів у ґрунті на маршрутах прокладання магістральних трубопроводів (МТ), що дозволяє вибрати на-

йоптимальніший маршрут прокладки МТ із мінімальним ризиком. Складено Універсальне рівняння втрат ґрунту (УРВГ), визначено всі змінні, що входять в УРВГ. Отримано растрову карту моделі УРВГ для обраної тестової ділянки і таблиця оцінок інтенсивності втрат ґрунту. Рис.: 14. Табл.: 3. Бібліогр.: 4 назв.

UDC 528.852:531.012

Forecasting of soil erosion processes on the routes of the pipelines based on geoin formation technologies and high resolution satellite images / Djamalov A.T., Ragimov R.M. // System Research and Information Technologies. — 2011. — № 4. — P. 97–103.

The possibility of forecasting of erosion processes in the soil on the route of the main pipelines (MP), which allows to select the optimal route of the MP with minimal risk, is investigated. Universal Soil Loss Equation (USLE) is compiled, all the variables in USLE are defined. The raster map of USLE model for the selected test area and table estimates of the intensity of soil loss are received. Figs: 14. Tabl.: 3. Refs: 4 titles.

МАТЕМАТИЧНІ МЕТОДИ, МОДЕЛІ, ПРОБЛЕМИ І ТЕХНОЛОГІЇ ДОСЛІДЖЕННЯ СКЛАДНИХ СИСТЕМ

MATHEMATICAL METHODS, MODELS, PROBLEMS AND TECHNOLOGIES FOR COMPLEX SYSTEMS RESEARCH

УДК 681.3.00:007

От ковариаций до каузальности. Открытие структур зависимостей в данных / **Балабанов А.С.** // Системні дослідження та інформаційні технології. — 2011. — № 4. — С. 104–118.

Проанализирована современная методология вывода каузальных моделей и структур систем вероятностных зависимостей из статистических данных пассивных наблюдений. Освещены возможности, проблемы, оговорки и ограничения методов индуктивной идентификации каузальных отношений в аппарате марковских свойств и байесовых сетей. Выделены несколько ступеней каузальных моделей согласно уровню их обоснованности и адекватности источнику данных. Сформулирован статистический паттерн, который сводит обоснование вывода о каузальном характере связи двух переменных к тестированию набора статистических фактов (не)зависимости. Рис.: 7. Библиогр.: 21 назв.

UDC 681.3.00:007

From covariation to causation. Discovery of structures of dependency in data / Balabanov O.S. // System Research and Information Technologies. — 2011. — № 4. — P. 104–118.

The current methodology of output casual models and structures of systems of probabilistic dependencies of stafistical data of passive observation is analysed. The problems, features, traps and limitations of the methods of the inductive identification of casual relation in the unit of marcov properties and bayesias nets are highlighted. Several stages of casual models according to the level of their validity and adequacy of the data source are emphasized. The statistical pattern, which brings the justification of a finding about casual nature of the connections between two variables to the test of a set of statistical facts of (in)dependency is formulated. Figs: 7. Refs: 21 titles.

УДК 519.8

Проблема существования решения в задачах распределения потоков / **Кирик Е.Е.** // Системні дослідження та інформаційні технології. — 2011. — № 4. — С. 119–128.

Исследованы вопросы существования решения в задаче распределения потоков в энергетических сетях. Сформулирован алгоритм априорного определения совместности ограничений такой задачи. Предложена модель с системой уравнений

непрерывности, построенной таким образом, чтобы обеспечить существование допустимого решения при условии сведения к минимуму нарушений заказов потребителей. Библиогр.: 7 назв.

UDC 519.8

The problem of solution existence in the flow distribution problems / **Kirik O.E.** // System Research and Information Technologies. — 2011. — № 4. — P. 119–128.

Questions of the existence of a solution for the problem of flow distribution in power networks are investigated. The algorithm of a priori definition of restrictions compatibility of this problem is formulated. A model with the system of equations of continuity, built in such a way as to ensure the existence of a feasible solution provided to minimize disruption customer orders is constructed. Refs: 7 titles.

УДК 519.71

Застосування мереж Петрі для аналізу КВ-граматик / **Спекторський І.Я.** // Системні дослідження та інформаційні технології. — 2011. — № 4. — С. 129–133.

Запропоновано схему використання мереж Петрі для дослідження деяких властивостей КВ-граматик. Метод дозволяє, зокрема, досліджувати задану КВ-граматику на порожність та скінченність породжуваної мови, використовуючи дерево покриття відповідної мережі Петрі. Крім того, запропонований метод дозволяє сформулювати необхідні умови породження заданого слова КВ-граматикою в термінах матричного аналізу відповідної мережі. Рис.: 5. Бібліогр.: 6 назв.

UDC 519.71

Application of Petri nets for the analysis of CF-grammars / Spectorsky I.Ya. // System Research and Information Technologies. — 2011. — № 4. — P. 129–133.

The scheme of the use of Petri nets for the study of some properties of the CF-grammars is proposed. This method, enables, in particular, to investigate the emptiness and finiteness of language, generated by given CF-grammar, using tree cover of the relevant Petri net. Additionally, the proposed method allows to formulate the necessary conditions for the generation of a given word by CF-grammar in terms of a matrix analysis of the relevant network. Figs: 5. Refs: 6 titles.

УДК 517.9

Долгосрочные прогнозы функций состояния квазилинейных гиперболических систем в $\mathbb{R}^n$ / **Горбань Н.В.** // Системні дослідження та інформаційні технології. — 2011. — № 4. — С. 134–139.

Рассматрена квазилинейная гиперболическая система в пространстве $\mathbb{R}^n$ с нелинейной, в общем случае, функцией взаимодействия. Исследовано асимптотическое поведение слабых решений поставленной задачи. Доказано существование траекторного и глобального аттракторов рассматриваемой задачи. Доказано, что потраекторно все слабые решения стремятся к стационарным состояниям. 10 назв.

UDC 517.9

Long-time forecasts for state functions of quasilinear hyperbolic systems in $\mathbb{R}^n$ / **Gorban N.V.** // System Research and Information Technologies. — 2011. — № 4. — P. 134–139.

The quasilinear hyperbolic system in $\mathbb{R}^n$ with non-linear in general case interaction function is considered. The asymptotic behaviour of weak solutions for considered problem is investigated. The existing of trajection and global attractors for all weak solutions of investigated problem is proved. It is proved that the trajectories of all weak solutions passing to the stationary states. Refs: 10 titles.

ВІДОМОСТІ ПРО АВТОРІВ

Балабанов Олександр Степанович,
кандидат технічних наук, старший науковий співробітник Інституту програмних систем НАН України, Київ

Бармак Олександр Володимирович,
доцент, кандидат технічних наук, старший науковий співробітник Інституту кібернетики ім. В.М. Глушкова НАН України, Київ

Горбань Наталія Володимирівна,
молодший науковий співробітник Навчально-наукового комплексу «Інститут прикладного системного аналізу» НТУУ «КПІ» МОН та НАН України, Київ

Гришаков Сергій Володимирович,
здобувач Інституту спеціального зв'язку та захисту інформації Національного технічного університету «КПІ», Україна, Київ

Демчишин Мирослав Володимирович,
аспірант кафедри засобів захисту інформації Національного авіаційного університету, Україна, Київ

Джамалов Алісейн Талиб,
кандидат технічних наук, заступник генерального директора Національного аерокосмічного агенства Науково-дослідного інституту аерокосмічної інформатики, Азербайджан, Баку

Кірік Олена Євстафіївна,
старший науковий співробітник, кандидат фізико-математичних наук, вчений секретар Навчально-наукового комплексу «Інститут прикладного системного аналізу» НТУУ «КПІ» МОН та НАН України, Київ

Кондращенко Володимир Валерійович,
аспірант кафедри автоматизованих систем управління Національного технічного університету «Харківський політехнічний інститут», Україна, Харків

Крак Юрій Васильович,
доктор фізико-математичних наук, професор кафедри моделювання складних систем Київського національного університету ім. Тараса Шевченка, Україна, Київ

Кривонос Юрій Георгійович,
академік НАН України, доктор фізико-математичних наук, заступник директора Інституту кібернетики ім. В.М. Глушкова НАН України, Київ

Левченко Євген Григорович,
доцент, кандидат фізико-математичних, доцент кафедри засобів захисту інформації Національного авіаційного університету, Україна, Київ

Мілявський Юрій Леонідович,
аспірант Навчально-наукового комплексу «Інститут прикладного системного аналізу» НТУУ «КПІ» МОН та НАН України, Київ

Москаленко Валентина Володимирівна,
кандидат технічних наук, доцент кафедри автоматизованих систем управління Національного технічного університету «Харківський політехнічний інститут», Україна, Харків

Олексійчук Антон Миколайович,
доктор технічних наук, професор кафедри Інституту спеціального зв'язку та захисту інформації Національного технічного університету України «КПІ», Україна, Київ

Петренко Анатолій Іванович,
професор, доктор технічних наук, завідувач кафедри системного проектування Навчально-наукового комплексу «Інститут прикладного системного аналізу» НТУУ «КПІ» МОН та НАН України, Київ

Рабчун Андрій Олександрович,
співробітник Державного університету інформаційно-комунікаційних технологій, Україна, Київ

Рагімов Рауф Махмуд,
кандидат технічних наук, заступник директора Національного аерокосмічного агенства Науково-дослідного інституту аерокосмічної інформатики, Азербайджан, Баку

Романенко Віктор Демидович,
професор, доктор технічних наук, заступник директора з науково-педагогічної роботи Навчально-наукового комплексу «Інститут прикладного системного аналізу» НТУУ «КПІ» МОН та НАН України, Київ

Свірін Павло Володимирович,
асистент Навчально-наукового комплексу «Інститут прикладного системного аналізу» НТУУ «КПІ» МОН та НАН України, Київ

Свістунов Сергій Якович,
кандидат технічних наук, старший науковий співробітник Інституту теоретичної фізики ім. М.М. Боголюбова НАН України, Київ

Сімашко Володимир Йосипович,
кандидат технічних наук, доцент кафедри економічної кібернетики Рівненського державного гуманітарного університету, Україна, Рівне

Спекторський Ігор Якович,
доцент, кандидат фізико-математичних наук, доцент кафедри математичних методів системного аналізу Навчально-наукового комплексу «Інститут прикладного системного аналізу» НТУУ «КПІ» МОН та НАН України, Київ

Зміст журналу

«Системні дослідження та інформаційні технології»

за 2011 р.

ЗМІСТ № 1

Маслянко П.П., Землянський Ю.Р., Рябушенко А.В. Системний аналіз методів прогнозування для фінансово-інвестиційної діяльності 7

Кузнєцова Н.В., Бідюк П.І. Інформаційна технологія аналізу фінансових даних на основі інтегрованого методу 22

Песчанский А.И. Календарное техническое обслуживание элементов монотонной системы с учетом их минимального аварийного восстановления 34

Лопатін О.К. «Alert»-технології, що ґрунтуються на теорії динамічних систем в економічних задачах 50

Гороховатский В.А. Оценка сходства структурных объектов как множеств компонент 57

Говорущенко Т.О. Процесс повторного тестування під час експертизи програмного забезпечення 71

Фаттахова М.И. Приближенный анализ моделей моносервисных беспроводных сетей микросотовой структуры 87

Бачинский В.А., Гиоргизова-Гай В.Ш. Выбор протокола динамической маршрутизации в корпоративной IP-сети 99

Науменко І.Я., Кизима В.І. Ехолокація у хвилеводах за відсутності реперів та апріорних даних про швидкість звуку 111

Алхімова С.М., Яценко В.П. Огляд методу перфузійної комп'ютерної томографії та його застосування в онкології 119

Середний С.С. Оценивание вероятности дефолта по кредитным операциям с использованием логистической регрессии и кластерного анализа 126

Перверза Е.В. Сценарный подход в задачах анализа сложных социальных систем 133

Реферати 144

Відомості про авторів 149

ЗМІСТ № 2

Романенко В.Д., Мілявський Ю.Л. Координуюче керування багатовимірним об'єктом із різнотемповою дискретизацією в стохастичному середовищі 7

Редько Е.В., Подладчикова Т.В., Подладчиков В.Н. Повышение эффективности оценивания параметров распределения в условиях ограниченной экспериментальной информации 21

Четырбок П.В. Инвариантность по скалярному критерию распознавания образов 28

Бессараб В.И., Зайцева Э.Е., Коваленко Е.Г. Использование аппарата идемпотентных алгебр для анализа процессов в сети с предоставлением интегрированных услуг 35

Замаруєва І.В., Бадьоріна Л.М. Метод кількісного оцінювання відповідей у системах тестування знань відкритого типу 41

Зак Ю.А. Количественные методы обработки мнений экспертов при оценке качества и перспективности представленных для финансирования инновационных проектов 47

Крючковский В.В., Ходаков Д.В. Реализация принятых решений 62

Махорт А.П. Про вибір стратегії оподаткування в економічній системі за наявності монополістів 76

Руденко С.С., Буждиган О.Я. Оцінка стійкості екосистем на основі аналізу рольових трофічних мереж 88

Левченко Є.Г., Прус Р.Б. Оптимізаційні економічні задачі в системах захисту інформації 98

Редько І.В., Снігур Н.М. Алгебраїчна характеристика класу графових перетворень 104

Кисленко Ю.І. Збереження цілісності відтворюваного на шляху вербалізації окремого фрагмента дійсності 115

Шаховська Н.Б. Формальне подання простору даних у вигляді алгебраїчної системи 128

Круглый стол по обсуждению концепции международной научно-исследовательской программы 141

Реферати 143

Відомості про авторів 149

ЗМІСТ № 3

Редько И.В., Татариков А.О., Семенченко М.П. Открыто-замкнутые основания сред интеграции. Часть II 7
Савченко І.О. Методологічне і математичне забезпечення розв'язання задач передбачення на основі модифікованого методу морфологічного аналізу 18
Максимей И.В., Шевченко Д.Н. Технология имитационного моделирования параметрических отказов технических систем 29
Осіпенко Д.В. Підхід до інтеграції системи кредитного скорингу та моделі керування активами та пасивами комерційного банку 38
Каргин А.А., Пятикоп Е.Е. Модели представления ключевых элементов в задаче покомпонентного распознавания символов текста 48
Кондратенко Н.Р., Лужицький С.В., Чеборака О.В. Нечіткі моделі в задачах підбору персоналу при формуванні соціальних груп 56
Зайченко Ю.П., Ови Нафас Агаи Аг Гамиш. Исследование двойственной задачи оптимизации инвестиционного портфеля в нечетких условиях 63
Михалевич В.М. К параметрической форме моделирования ситуации в общей задаче принятия решения 77
Мазурок Т.Л., Тодорцев Ю.К. Нейромережева реалізація інтелектуальної підтримки прийняття рішень в автоматизованому управлінні навчанням 88
Панченко Б.Е., Писанко И.Н. О топологии путей нормализации в реляционном каркасе 102
Марценюк В.П., Вакуленко Д.В., Андрущак І.Є. Оптимальне керування режимами медикаментозної терапії та фізіотерапії в задачі реконструкції кісткової тканини 108
Данилов В.Я., Науменко І.Я., Кизима В.І., Клименко С.М. Згасання звуку в міжтрубному просторі свердловин 123
Бохонов Ю.Є. Про періодичні розв'язки квазілінійного звичайного диференціального рівняння другого порядку 133
Арсеньев Д.Г., Амбражей А.Н., Головин Н.М. Таратухин В.В. Сотрудничество Академического центра компетенции SAP СПбГПУ и высших учебных заведений стран СНГ в рамках программы «Университетский Альянс SAP» 138
Реферати 143
Відомості про авторів 149

ЗМІСТ № 4

Романенко В.Д., Мілявський Ю.Л. Синтез багатовимірних координуючих систем керування з різнотемповою дискретизацією в детермінованому середовищі 7
Петренко А.І., Свістунов С.Я., Свірін П.В. Алгоритми балансування навантаження в Грід-системах 21
Алексейчук А.Н., Гришаков С.В. Неасимптотические оценки эффективности случайного кодирования в системе передачи информации по двоичному симметричному каналу связи с отводом 37
Кривонос Ю.Г., Крак Ю.В., Бармак О.В. Комп'ютерне моделювання української жестової мови 48
Москаленко В.В., Кондращенко В.В. Математические модели управления процессами финансирования инвестиционных проектов 61
Сімашко В.Й. Методи і технології зниження пікових навантажень у корпоративних інформаційних системах із централізованим збереженням даних 74
Левченко Є.Г., Демчишин М.В., Рабчун А.О. Математичні моделі економічного менеджменту інформаційної безпеки 88
Джамалов А.Т., Рагимов Р.М. Прогнозирование эрозионных процессов почв на маршрутах прокладки магистральных трубопроводов на основе геоинформационных технологий и космических снимков высокого разрешения 97
Балабанов О.С. Від коваріацій до каузальності. Відкриття структур залежностей у даних 104
Кірік О.Є. Проблема існування розв'язку в задачах розподілу потоків 119
Спекторский И.Я. Применение сетей Петри для анализа КС-грамматик 129
Горбань Н.В. Довгострокові прогнози функцій стану квазілінійних гіперболічних систем в $\mathbb{R}^n$ 134
Реферати 140
Відомості про авторів 145
Зміст журналу «Системні дослідження та інформаційні технології» за 2011 р. 147
Автори статей за 2011 р. 149

АВТОРИ СТАТЕЙ ЗА 2011 р.

Алхімова Світлана Миколаївна, № 1
Амбражей Антон Миколайович, № 3
Андрущак Ігор Євгенович, № 3
Арсеньєв Дмитро Германович, № 3
Бадьорина Любов Миколаївна, № 2
Балабанов Олександр Степанович, № 4
Бармак Олександр Володимирович, № 4
Бачинський Владислав Анатолійович, № 1
Бессараб Володимир Іванович, № 2
Бідюк Петро Іванович, № 1
Бохонов Юрій Євгенович, № 3
Буждиган Оксана Ярославівна, № 2
Вакуленко Дмитро Вікторович, № 3
Гіоргізова-Гай Вікторія Шалвівна, № 1
Говорущенко Тетяна Олександрівна, № 1
Головін Микита Михайлович, № 3
Горбань Наталія Володимирівна, № 4
Гороховатський Володимир Олексійович, № 1
Гришаков Сергій Володимирович, № 4
Данилов Валерій Якович, № 3
Демчишин Мирослав Володимирович, № 4
Джамалов Алісейн Талиб, № 4
Зайцева Еліна Євгенівна, № 2
Зайченко Юрій Петрович, № 3
Зак Юрій Олександрович, № 2
Замаруєва Ірина Вікторівна, № 2
Землянський Юрій Романович, № 1
Каргін Анатолій Олексійович, № 3
Кірік Олена Євстафіївна, № 4
Кизима Володимир Іванович, № 1, 3
Кисленко Юрій Іванович, № 2
Клименко Сергій Миколайович, № 3
Коваленко Євгенія Геннадіївна, № 2
Кондратенко Наталія Романівна, № 3
Кондращенко Володимир Валерійович, № 4
Крак Юрій Васильович, № 4
Кривонос Юрій Георгійович, № 4
Крючковський Віктор Володимирович, № 2
Кузнєцова Наталія Володимирівна, № 1
Левченко Євген Григорович, № 2, 4
Лопатін Олексій Константинович, № 1
Лужецький Сергій Володимирович, № 3
Мазурок Тетяна Леонідівна, № 3
Максімей Іван Васильович, № 3
Марценюк Василь Петрович, № 3
Маслянко Павло Павлович, № 1
Махорт Андрій Пилипович, № 2
Михалевич Вадим Михайлович, № 3
Мілявський Юрій Леонідович, № 2, 4
Москаленко Валентина Володимирівна, № 4
Науменко Ігор Якович, № 1, 3
Ові Нафас Агаі Аг Гаміш, № 3
Олексійчук Антон Миколайович, № 4
Осіпенко Денис В'ячеславович, № 3
Панченко Борис Євгенович, № 3
Переверза Катерина Володимирівна, № 1
Песчанський Олексій Іванович, № 1
Петренко Анатолій Іванович, № 4
Писанко Іван Миколайович, № 3
Подладчіков Володимир Миколайович, № 2
Подладчікова Тетяна Володимирівна, № 2
Прус Руслана Богданівна, № 2
П'ятикоп Олена Євгенівна, № 3
Рабчун Андрій Олександрович, № 4
Рагімов Рауф Махмуд, № 4
Редько Євген Вікторович, № 2
Редько Ігор Володимирович, № 2, 3
Романенко Віктор Демидович, № 2, 4
Руденко Світлана Степанівна, № 2
Рябушенко Андрій Віталійович, № 1
Савченко Ілля Олександрович, № 3
Свірін Павло Володимирович, № 4
Свістунов Сергій Якович, № 4
Семенченко Михайло Петрович, № 3
Середній Сергій Сергійович, № 1
Сімашко Володимир Йосипович, № 4
Снігур Наталія Миколаївна, № 2
Спекторский Ігор Якович, № 4
Таратухін Віктор Володимирович, № 3
Татаріков Олександр Олегович, № 3
Тодорцев Юрій Костянтинович, № 3
Фаттахова Мехрибан Іса кизи, № 1
Ходаков Даніїл Вікторович, № 2
Чеборака Олександр Валерійович, № 3
Четирбок Петро Васильович, № 2
Шаховська Наталя Богданівна, № 2
Шевченко Дмитро Миколайович, № 3
Яценко Валентин Порфирович, № 1

ПРАВИЛА ОФОРМЛЕННЯ СТАТЕЙ

для журналу

«СИСТЕМНІ ДОСЛІДЖЕННЯ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ»

1. Стаття з анотаціями подається у двох варіантах: **роздрукована** (на папері формату А4 через 1,5 інтервали у двох примірниках, розмір шрифту 12 пт., поля зліва та справа по 3 см) та в **електронному** вигляді.

Назва статті, прізвище та ініціали автора (авторів), а також анотація (від 700 до 1000 символів) подаються українською, російською та англійською мовами.

2. Текст статті набирається в редакторі Microsoft Word 2003 шрифтом Times New Roman із вирівнюванням по ширині.

3. На початку статті вказується індекс УДК, назва статті, ініціали та прізвище автора (авторів).

4. Список літератури подається в порядку посилання після тексту статті.

5. Набір формул здійснюється в редакторі формул «Equations».

6. Символи у формулах (крім особливих випадків) набираються курсивом. Допускаються прямі символи, але їх треба розмітити або включити до списку спеціальних символів.

7. Нумерувати необхідно тільки ті формули (винесені на середину рядків), на які є посилання в тексті.

8. Таблиці та рисунки виконуються в одному стилі, їм надається номер, назва і розташовуються після посилань у тексті. Не слід повторювати в статті підписи до рисунків та таблиць.

9. Стаття разом із таблицями, рисунками і списком літератури має бути не більше 15 сторінок. Оглядові статті обсягом до 25 сторінок приймаються за домовленістю із редколегією.

10. До рукопису додаються відомості про кожного автора: країна, місто, місце роботи, службова посада, вчений ступінь та звання, сфера наукових досліджень, контактні телефони, адреса та e-mail.