УДК 004.056.5:519.6

А.А. Кобозєва, канд. фіз.-мат. наук, доц.,
Одес. нац. політехн. ун-т

АНАЛІЗ ІНФОРМАЦІЙНОГО ПРОЦЕСУ НА ОСНОВІ ЙОГО УНІВЕРСАЛЬНОЇ ФОРМАЛІЗАЦІЇ

А.А.Кобозєва. **Аналіз інформаційного процесу на основі його універсальної формалізації.** Запропонована загальна формалізація інформаційного процесу, на основі якої проводиться аналіз його структури й деяких властивостей, що відкриває можливості для розпаралелювання процесу обробки.

А.А.Кобозева. **Анализ информационного процесса на основе его универсальной формализации.** Предложена общая формализация информационного процесса, на основе которой проводится анализ его структуры и некоторых свойств, открывающий возможности для распараллеливания процесса обработки.

A.A.Kobozeva. **Analysis of informational process on the basis of its universal formalization.** Universal formalization of informational process is proposed. Analysis of structure and characteristics of informational process on the basis of its formalization is carried out, with the aim of paralleling of processing.

Процес впровадження нових інформаційних технологій в усі сфери життя сучасного суспільства, що вступає в постіндустріальний період свого розвитку, який можна назвати інформаційним, є неможливим без вирішення питань інформаційної безпеки в різних сферах: політичній, військовій, екологічній, природничо-науковій, технічній, соціальній, нормативно-правовій, економічній, фінансовій.

Процес математизації інформаційної безпеки відстає від її потреб. Проявом цього відставання є відсутність універсальної формалізації довільної інформаційної системи, інформаційного процесу (ІП), підходу до питань аналізу й обробки даних про стан системи, заснованого на загальному математичному базисі, що значною мірою гальмує процес створення загального наукового базису захисту інформації, про необхідність якого не один раз говорилося в сучасних відкритих джерелах [1]. Тому важливою є побудова загальної формалізації ІП, дослідження його структури й деяких властивостей на основі отриманого формального представлення, підсумками якого повинні стати

вирішення питання про чутливість ІП до збурних дій;

визначення шляхів і можливостей для розпаралелювання процесу обробки даних про ІП.

Під *інформаційною системою* (ІС) розуміється сукупність інформаційних ресурсів і взаємозалежних засобів, які здійснюють зберігання, обробку й передачу інформації; ІС забезпечує протікання ІП, тобто процесів, пов'язаних зі збереженням, обробкою й передачею інформації [1].

Під *структурою ІП* розуміється сукупність інформаційних зв'язків на рівні окремих операцій по перетворенню параметрів ІС.

ІП характеризується зміною параметрів, що його визначають, або приведенням одних параметрів (вихідних) у відповідність з іншими (вхідними) за деяким законом, тобто може бути формально представлений у вигляді неперервної вектор-функції скінченної кількості змінних:

$$\Phi(x_1,\ldots,x_n)=\begin{pmatrix}\varphi_1(x_1,\ldots,x_n)\\ \varphi_2(x_1,\ldots,x_n)\\ \vdots\\ \varphi_m(x_1,\ldots,x_n)\end{pmatrix}=\begin{pmatrix}\Phi_1\\ \Phi_2\\ \vdots\\ \Phi_m\end{pmatrix}, \qquad (1)$$

де $(\Phi_1,\Phi_2,\ldots,\Phi_m)\in R^m$ — вихідні,

$(x_1,\ldots,x_n)\in D\subseteq R^n$ — вхідні параметри,

D — область визначення $\Phi(x_1,\ldots,x_n)$.

Функція (1) породжує на D m дійсних функцій [2]

$$\varphi_i(x_1,\ldots,x_n)=\Phi_i,\ i=\overline{1,m}\ . \qquad (2)$$

Має місце таке твердження [3].

Твердження 1. ІП (ІС) може бути формально представлений у вигляді скінченної множини дійсних функцій (2), а аналіз процесу зведений до аналізу отриманих функцій.

Одним з найбільш важливих питань аналізу ІП є встановлення міри його чутливості до збурних дій (похибок вхідних даних). При формалізації процесу у вигляді (1) його результат можна розглянути як результат задачі про обчислення функції (1), тобто сукупності функцій (2). Задача називається *чутливою* до збурних дій, якщо навіть малі збурні дії можуть привести до значної похибки результату, і *нечутливою* в протилежному випадку. *Чутливість ІП* буде визначатися чутливістю задачі обчислення сукупності функцій (2) [3].

Нехай $\varphi_i\in C^1(D),\ i=\overline{1,m}$, тобто мають усі часткові похідні, неперервні на D. Припустимо, що значення одного з вихідних параметрів Φ_j однозначно визначається сукупністю значень інших $\Phi_1,\ldots,\Phi_{j-1},\Phi_{j+1}\ldots,\Phi_m$, тобто якщо $\Omega_0\subseteq R^{m-1}$ є множина точок, що відповідають усіляким точкам $(x_1,x_2,\ldots,x_n)\in D$, то в Ω_0 буде мати місце функціональна залежність:

$$\Phi_j=f(\Phi_1,\ldots,\Phi_{j-1},\Phi_{j+1},\ldots,\Phi_m) \qquad (3)$$

причому $f\in C^1(\Omega),\ \Omega\subseteq R^{m-1},\ \Omega\supseteq\Omega_0$, Ω — відкрита множина, а при підстановці (2) в (3) виходить тотожність відносно $(x_1,x_2,\ldots,x_n)\in D$:

$$\varphi_j(x_1,\ldots,x_n)\equiv f(\varphi_1(x_1,\ldots,x_n),\ldots,\varphi_{j-1}(x_1,\ldots,x_n),\varphi_{j+1}(x_1,\ldots,x_n),\ldots,\varphi_m(x_1,\ldots,x_n))\ . \qquad (4)$$

В цьому випадку функція φ_j залежить від функцій $\varphi_1,\ldots,\varphi_{j-1},\varphi_{j+1},\ldots\varphi_m$ в області D. Функції $\varphi_1,\varphi_2,\ldots\varphi_m$ називаються *залежними* в області D, якщо одна з них залежить від інших. Якщо ні в D, ні в будь-якій області $E\subseteq D$ не має місця тотожність (4), то функції $\varphi_1,\varphi_2,\ldots\varphi_m$ називаються *незалежними* в D [2].

Визначення. Вихідні параметри ІП будуть *незалежними (залежними*), якщо незалежними (залежними) в області D будуть функції (2), які їх визначають.

Зауваження 1. У випадку незалежності вихідних параметрів їх визначення відповідно до (2) і аналіз може проводитися одночасно (паралельно), що значно скорочує при необхідності час реалізації і аналізу ІП. *Цей процес можна представити як сукупність не зв'язаних між собою "простих" процесів*, результатом кожного з яких є одержання лише одного параметра Φ_i, а дослідження поданого ІП зведеться до досліджень скінченної сукупності "простих", які можна проводити одночасно.

Таким чином, для того, щоб урахувати можливість зменшення часу обробки даних про ІП, спочатку необхідно встановити, чи є вихідні параметри (функції (2)) незалежними (залежними).

Відповідь на це питання дає матриця Якобі для функцій $\varphi_i,\ \ i=\overline{1,m}$:

$$\begin{pmatrix}\frac{\partial\varphi_1}{\partial x_1} & \ldots & \frac{\partial\varphi_1}{\partial x_n}\\ \ldots\ldots\ldots\ldots & & \\ \frac{\partial\varphi_m}{\partial x_1} & \ldots & \frac{\partial\varphi_m}{\partial x_n}\end{pmatrix}\ . \qquad (5)$$

Нехай $n\geq m$.

Теорема 1. Нехай ранг матриці Якобі (5) ІП (2) в області D є r, при цьому $r=m$. Тоді в D знайдеться така область зміни вхідних параметрів ІП, у якій вихідні будуть незалежними, а початковий процес може бути представлений у вигляді сукупності "простих".

Доказ. Якщо хоч один визначник m-го порядку, складений з елементів матриці (5), відрізняється від нуля в області D, то в цій області функції φ_i, $i=\overline{1,m}$, а тому і вихідні параметри ІП, незалежні. Ранг матриці Якобі (5) в області D — це найвищий з порядків визначників, утворених з елементів цієї матриці, що не дорівнюють тотожньо нулю в D. Нехай ранг матриці Якобі ІП в області D є r і досягається в точці $M^0(x_1^0, x_2^0, ..., x_n^0) \in D$, тобто визначник r-го порядку в цій точці відрізняється від нуля, при цьому $r=m$. Тоді в D знайдеться така область зміни вхідних параметрів ІП, в якій вихідні будуть незалежними.

Нехай в області D вихідні параметри ІП можуть бути як залежними, так і незалежними.

Для однаковості викладу всі параметри далі позначаються u_i, де $i=\overline{1,N}$, $N=n+m$ (для $i=\overline{1,n}$ u_i відповідають вхідним параметрам). Тоді реалізацію процесу можна формалізувати таким чином:

$$u_k = F_k(u_{k_1},\ldots,u_{k_{s_k}}), \quad n<k\le N, \quad k_1,\ldots,k_{s_k}<k, \tag{6}$$

де всі F_k є досить гладкими функціями своїх аргументів. Як результат процесу розглядається сукупність величин u_k (вихідні параметри). Не накладаючи серйозних обмежень, можна припустити, що результат — це величина u_N.

Співвідношення (6) задають процес обчислення функції (1) у нових позначеннях

$$\Phi(u_1,\ldots,u_n) = (u_{n+1}, u_{n+2},\ldots,u_N)^T. \tag{7}$$

Якщо обчислення (7) спочатку задані за допомогою (6), то при великих N одержати явний вираз функції Φ (тобто функцій φ_i, $i=\overline{1,m}$) через вхідні дані ($u_i, i=\overline{1,n}$) важко й не завжди можливо. Буде отримано достатню умову такого представлення.

Кожне рівняння системи (6) еквівалентне рівнянню

$$F_k(u_{k_1},\ldots,u_{k_{s_k}}) - u_k = F_k(u_{k_1},\ldots,u_{k_{s_k}}) - u_k + 0\sum_{i\ne k,k_1,\ldots,k_{s_k}} u_i = G_k(u_1,u_2,\ldots,u_N) =$$
$$= G_k(x_1,\ldots,x_n,\Phi_1,\ldots,\Phi_m) = 0,$$

що приводить (6) до еквівалентної системи рівнянь, у загальному випадку — нелінійної:

$$\begin{cases} G_1(x_1,\ldots,x_n;\Phi_1,\ldots,\Phi_m)=0, \\ G_2(x_1,\ldots,x_n;\Phi_1,\ldots,\Phi_m)=0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ G_m(x_1,\ldots,x_n,\Phi_1,\ldots,\Phi_m)=0 \end{cases}. \tag{8}$$

Якщо $G_i \in C^1(D)$, $i=\overline{1,m}$, де D — $n+m$-вимірний прямокутний паралелепіпед

$$D=[x_1^0-\Delta_1, x_1^0+\Delta_1]\times\ldots\times[x_n^0-\Delta_n, x_n^0+\Delta_n]\times[\Phi_1^0-\overline{\Delta}_1, \Phi_1^0+\overline{\Delta}_1]\times\ldots\times[\Phi_m^0-\overline{\Delta}_m, \Phi_m^0+\overline{\Delta}_m]$$

з центром в точці $(x_1^0,\ldots,x_n^0,\Phi_1^0,\ldots,\Phi_m^0)$, координати якої задовольняють системі (8), і визначник матриці Якобі для функцій G_i, $i=\overline{1,m}$, по змінним $\Phi_1,\ldots,\Phi_m$ відрізняється від нуля, тобто

$$\det\begin{pmatrix} \frac{\partial G_1}{\partial \Phi_1} & \frac{\partial G_1}{\partial \Phi_2} & \dots & \frac{\partial G_1}{\partial \Phi_m} \\ \frac{\partial G_2}{\partial \Phi_1} & \frac{\partial G_2}{\partial \Phi_2} & \dots & \frac{\partial G_2}{\partial \Phi_m} \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ \frac{\partial G_m}{\partial \Phi_1} & \frac{\partial G_m}{\partial \Phi_2} & \dots & \frac{\partial G_m}{\partial \Phi_m} \end{pmatrix} \neq 0,$$

тоді в деякому околі точки $\langle x_1^0,\dots,x_n^0,\Phi_1^0,\dots,\Phi_m^0\rangle$ система (8) визначає $\Phi_1,\dots,\Phi_m$ у вигляді (2), до того $\varphi_i \in C^1(D), \quad i=\overline{1,m}$.

Тому ІП, який визначається за допомогою Φ, можна досліджувати через рекурентні співвідношення (6). Для цього система (6) перетворюється до еквівалентного вигляду:

$$F_k\left(u_{k_1},\dots,u_{k_{s_k}}\right)-u_k=0\ ,\quad n<k\le N;\ \ k_1,\dots,k_{s_k}<k. \tag{9}$$

Задача про встановлення міри чутливості ІП (6) зводиться до аналізу чутливості системи (9). Для цього розглядається збурена система

$$F_k\left(u_{k_1}+\Delta u_{k_1},\dots,u_{k_{s_k}}+\Delta u_{k_{s_k}}\right)-\left(u_k+\Delta u_k\right)=0 \quad n<k\le N;\ \ k_1,\dots,k_{s_k}<k, \tag{10}$$

де збурення Δu_k малі. Після віднімання (9) з (10), використовуючи для функції F_k вираз [3]

$$F_k\left(u_{k_1}+\Delta u_{k_1},\dots,u_{k_{s_k}}+\Delta u_{k_{s_k}}\right)=F_k\left(u_{k_1},\dots,u_{k_{s_k}}\right)+\frac{\partial F_1\left(u_{k_1},\dots,u_{k_{s_k}}\right)}{\partial u_{k_1}}\ \Delta u_{k_1}+$$
$$+\dots+\frac{\partial F_k\left(u_{k_1},\dots,u_{k_{s_k}}\right)}{\partial u_{k_{s_k}}}\ \Delta u_{k_{s_k}}+\bar{o}\left(\sqrt{\Delta u_{k_1}^2+\dots+\Delta u_{k_{s_k}}^2}\right),\ \text{коли}\ \sqrt{\Delta u_{k_1}^2+\dots+\Delta u_{k_{s_k}}^2}\to 0, \tag{11}$$

з точністю до нескінченно малих другого порядку, отримується система лінійних алгебраїчних рівнянь щодо збурень Δu_k [4]

$$\sum_{i=1}^{s_k}\frac{\partial F_k\left(u_{k_1},\dots,u_{k_{s_k}}\right)}{\partial u_{k_i}}\ \Delta u_{k_i}-\Delta u_k=0\ ,\quad n<k\le N;\ \ k_1,\dots,k_{s_k}<k. \tag{12}$$

Матриця системи (12) позначається далі **Ψ** і називається *варіаційною матрицею інформаційного процесу* (6) (ВМІП). ЇЇ елементи ψ_{ij} в загальному випадку визначаються як

$$\Psi_{ij}=\begin{cases} -1, & \text{якщо}\quad j=i+n, \\ \dfrac{\partial F_{i+n}}{\partial u_j}, & \text{якщо}\quad j\ \text{є одним з чисел}\ (i+n)_1,\dots,(i+n)_{s_{i+n}} \\ 0, & \text{в інших випадках} \end{cases}.$$

Матриця **Ψ** — це матриця Якобі функцій

$$F_k\left(u_{k_1},\ldots,u_{k_{s_k}}\right)-u_k=f_k\left(u_1,\ldots,u_N\right)$$

по змінним $u_1,\ldots,u_N$ розміру $m\times N$. Вона, як правило, сильно розріджена та, як випливає з (12), має повний ранг.

ВМІП відіграє важливу роль при аналізі структури ІП та дослідженні його чутливості до збурних дій.

Процес поширення похибок [4] у ході ІП (6), реалізація якого зводиться до обчислення функції (7), при реальних обчисленнях точних формул (6) дає

$$\overline{u}_k=\overline{F}_k\left(\overline{u}_{k_1},\ldots,\overline{u}_{k_{s_k}}\right)\ \Leftrightarrow\ \overline{u}_k=F_k\left(\overline{u}_{k_1},\ldots,\overline{u}_{k_{s_k}}\right)+\eta_k,\quad n<k\le N,\ k_1,\ldots,k_{s_k}<k, \tag{13}$$

де $\overline{F}_k$ — збурена "близька" до F_k функція, реальне обчислення якої здійснюється при реалізації (6),

$\overline{u}_k$ — реально задана або обчислена величина u_k,

η_k — еквівалентна абсолютна похибка (підсумковий результат збурної дії), яка вноситься в результат обчислення F_k.

Рівняння (13) можна представити у вигляді:

$$\overline{u}_k+\varepsilon_k=F_k\left(\overline{u}_{k_1}+\varepsilon_{k_1},\ldots,\overline{u}_{k_{s_k}}+\varepsilon_{k_{s_k}}\right),\quad n<k\le N,\ k_1,\ldots,k_{s_k}<k, \tag{14}$$

де ε_k — збурення $\overline{u}_k$ (ці збурення вводяться й у вхідні дані).

Якщо взяти збурені вхідні дані $u_1+\varepsilon_1,\ldots,u_n+\varepsilon_n$ і провести з ними точний процес (6), то на кожному кроці цього процесу як точний результат буде $\overline{u}_k+\varepsilon_k$.

Враховуючи (13) і (14),

$$\varepsilon_k=F_k\left(\overline{u}_{k_1}+\varepsilon_{k_1},\ldots,\overline{u}_{k_{s_k}}+\varepsilon_{k_{s_k}}\right)-F_k\left(\overline{u}_{k_1},\ldots,\overline{u}_{k_{s_k}}\right)-\eta_k,\quad n<k\le N,\ k_1,\ldots,k_{s_k}<k, \tag{15}$$

Якщо збурення вхідних даних $\varepsilon_1,\ldots,\varepsilon_n$ відомі, то за допомогою (15) можна визначити інші збурення $\varepsilon_k, k>n$ (значення u_k, збурення η_k визначаються реалізацією ІП (13)), що дасть можливість для встановлення міри чутливості ІП до збурних дій. Через це є необхідність розглянути (15) докладно.

Система (15) у загальному випадку є нелінійною відносно ε_k. Враховуючи рівність (11), (15) замінюється лінійною системою

$$\varepsilon_k=\sum_{i=1}^{s_k}\frac{\partial F_k\left(\overline{u}_{k_1},\ldots,\overline{u}_{k_{s_k}}\right)}{\partial u_{k_i}}\varepsilon_{k_i}-\eta_k,\quad n<k\le N,\ k_1,\ldots,k_{s_k}<k.$$

Реально обчислені величини $\overline{u}_i, i=k_1,\ldots,k_{sk}$, замінюються на точні:

$$\sum_{i=1}^{s_k}\frac{\partial F_k\left(u_{k_1},\ldots,u_{k_{s_k}}\right)}{\partial u_{k_i}}\varepsilon_{k_i}-\varepsilon_k=\eta_k,\quad n<k\le N,\ k_1,\ldots,k_{s_k}<k. \tag{16}$$

Матриця системи (16) — це ВМІП (6), тому система завжди сумісна. Таким чином, величини збурень ε_k визначаються як розв'язок системи лінійних алгебраїчних рівнянь (16), властивості якої визначаються властивостями ВМІП.

Зауваження 2. Оскільки ε_k використовуються для аналізу ІП з погляду встановлення його чутливості до збурних дій, важливо, щоб ці значення були отримані як можна точніше, для чого матриця системи (16), яка є ВМІП (6), повинна бути добре обумовленою.

Для подальшого аналізу ІП йому у відповідність можна поставити орграф у такий спосіб. Зіставляється k-й вершині графа одержання величини u_k. Перші n вершин символізують введення початкових даних $u_1,...,u_n$ і називаються вхідними, а інші вершини — обчислення u_k як значень функцій F_k з (6). Вважається, що дуга йде з i-ї вершини в j-у в тому й тільки тому випадку, коли при обчисленні величини u_j величина u_i використовується як аргумент. Відповідно до (6) дуги не будуть входити в k-у вершину, якщо $k \le n$. Якщо $k > n$, то в k-у вершину будуть входити дуги з вершин з номерами $k_1,..., k_{sk}$.

Виходячи зі способу побудови графа, стає очевидною істинність твердження 2 та теореми 2.

Твердження 2. Граф ІП є ациклічним.

Теорема 2. Вихідні параметри ІП незалежні тоді й тільки тоді, коли граф ІП дводольний.

У графі ІП наочно представлені відомості про те, як окремі перетворення у процесі пов'язані між собою інформаційно, які перетворення можуть виконуватися одночасно, які необхідно виконуються пізніше або раніше, ніж інші і т.д. *Граф ІП описує всю картину поширення інформації під час його протікання.*

З отриманим графом можна пов'язати матрицю **Ф** розмірами $m \times N$ з елементами φ_{ij}:

$$\varphi_{ij} = \begin{cases} -1, & \text{якщо } j = i + n, \\ 1, & \text{якщо } j \text{ є одним з чисел } (i+n)_1,\ldots,(i+n)_{s_{i+n}} \\ 0, & \text{в інших випадках} \end{cases}.$$

Очевидно, k-й стовпець матриці **Ф** відповідає параметру u_k, а k-й рядок — параметру u_{k+n}. В k-му рядку елемент –1 стоїть в тому стовпці, номер якого відповідає номеру параметра u_{k+n}, що обчислюється. Елементи +1 стоять в тих стовпцях, номери яких відповідають номерам аргументів параметра u_{k+n}, що обчислюється. Матриця **Ф** описує зв'язок параметрів u_k між собою й називається *матрицею інформаційної зв'язності ІП* (6) (МІЗІП). Очевидним є зв'язок МІЗІП з ВМІП: структури ненульових елементів обох матриць повністю співпадають.

Для графа ІП по МІЗІП заміною ненульових елементів якимись числами можна одержати нескінченну сукупність матриць, кожна з яких є *зваженою* МІЗІП. Будь-яка зважена МІЗІП, частковим випадком якої є і ВМІП, дозволяє однозначно відновити граф ІП, а тому разом з ВМІП може використовуватися для його аналізу.

Для орграфа, що відповідає ІП, стандартним чином визначається $N \times N$-матриця суміжності **В** з елементами b_{ij} [5]. Матриця **В** тісно пов'язана з МІЗІП **Ф**: **Ф** — підматриця, що складається з останніх m рядків матриці $\mathbf{B}^{\mathbf{T}}$-**I**, де **I** — одинична матриця відповідного розміру. У зв'язку з цим для приведення МІЗІП до більш "зручного" з погляду можливостей її обробки вигляду за рахунок перенумерації рядків і стовпців можна здійснювати відповідну перенумерацію для матриці суміжності.

Зауваження 3. До значного зменшення часу, необхідного для аналізу ІП, приводить виявлення й наступне використання його внутрішнього паралелізму на основі відповідного графа, тобто можливостей паралельного (одночасного) виконання (аналізу) вхідних в ІП операцій [4]. Із цією метою можуть використовуватися топологічні сортування графа.

Таким чином, запропонована загальна формалізація довільного ІП, універсальність якої дає можливість для аналізу структури процесу і його властивостей, які не обмежуються розглянутими в роботі, дозволяє зменшити час обробки даних про ІП шляхом виявлення й використання його внутрішнього паралелізму, що, як можна судити по публікаціях з відкритих джерел, ніколи не

робилося раніше, є черговим кроком на шляху створення єдиного наукового базису інформаційної безпеки.

Література

1. Ленков, С.В. Методы и средства защиты информации: в 2 т. / С.В. Ленков, Д.А. Перегудов, В.А. Хорошко. — К.: Арий, 2008. —
   Т.2: Информационная безопасность. — 344 с.
2. Фихтенгольц, Г.М. Курс дифференциального и интегрального исчисления. Т.1 / Г.М. Фихтенгольц. — М.: Наука, 1969. — 608 с.
3. Кобозева, А.А. Общий подход к анализу состояния информационных объектов, основанный на теории возмущений / А.А. Кобозева // Вісн. Східноукр. нац. ун-ту ім. В. Даля. — 2008. — № 8(126), ч.1. — С. 72 — 81.
4. Воеводин, В.В. Параллельные вычисления / В.В. Воеводин, Вл.В. Воеводин. — СПб.: БХВ-Петербург, 2002. — 608 с.
5. Харари, Ф. Теория графов / Ф. Харари; пер.с англ. В.П.Козырева. — М.: Мир, 1973. — 300 с.

Труды Одесского политехнического университета, 2009, №1(31), с.128-133.