МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Наукові праці
Донецького національного технічного університету

Серія: “Інформатика, кібернетика та обчислювальна техніка”

Всеукраїнський науковий збірник

Заснований у травні 1996 року

Виходить 2 рази на рік

№ 2(37)’ 2023

Луцьк – 2024

УДК 004+519.6+519.7

Публікується згідно з рішенням Вченої ради ДВНЗ «Донецький національний технічний університет» (протокол № 2 від 29.02.2024).

Збірник містить наукові статті співробітників ДонНТУ та інших навчальних і наукових закладів України, які є науковими партнерами ДонНТУ. Публікації висвітлюють результати наукових досліджень і розробок в таких напрямках, як інформатика, чисельні методи, паралельні обчислення, програмування, розробка засобів обчислювальної техніки, дослідження комп'ютерних мереж, машинна графіка і обробка зображень, математичне моделювання в різних галузях. Матеріали збірника призначені для наукових співробітників, викладачів, інженерно-технічних працівників, аспірантів та студентів.

Засновник та видавець – Донецький національний технічний університет (ДонНТУ)

Адреса редакції:
Юридична адреса: 85300, Україна, Донецька область, м. Покровськ, пл. Шибанкова, 2, ДВНЗ «ДонНТУ».
Фактична адреса: 43003, Україна, Волинська область, м. Луцьк, вул. Потебні, 56, ДВНЗ «ДонНТУ».
E-mail: yevhen.bashkov@donntu.edu.ua

Збірник зареєстровано в Державному комітеті інформаційної політики, телебачення та радіомовлення України. Свідоцтво: серія КВ, №7374 від 03.06.2003.

Збірник включено до переліку наукових фахових видань України, в яких можуть публікуватися результати дисертаційних робіт на здобуття наукових ступенів доктора наук, кандидата наук та ступеня доктора філософії за спеціальностями 121 Інженерія програмного забезпечення, 122 Комп'ютерні науки, 123 Комп'ютерна інженерія (наказ Міністерства освіти і науки України №409 від 17 березня 2020 р.)

Збірник "Наукові праці ДонНТУ, серія "Інформатика, кібернетика та обчислювальна техніка" за наказом № 409 МОНУ від 17. 03.2020 отримав категорію Б.

ISSN 1996-1588

ЗМІСТ

УДК 004.222 + 004.315

[1]О.В. Самощенко, канд. техн. наук, доц.
[2]Б.С. Гусєв, канд. техн. наук, доц.
В.В. Лапко, канд. техн. наук, доц.
[1]Київський національний університет імені Тараса Шевченка, м. Київ, Україна
aleksandr.samoshchenko@gmail.com
[2]Національний університет біоресурсів і природокористування України, м. Київ, Україна
gusevbs@gmail.com

Ділення цілих чисел в доповняльному коді із знакозмінним нулем залишку діленого

Доповняльні коди знакових цілих операндів вказуються єдиною формулою для додатних і від'ємних чисел та обчислюються як залишок по модулю, що дорівнює кількості двійкових наборів доповняльного коду числа. Аналітично обґрунтовані оригінальні способи фіксації переповнення розрядної сітки частки при діленні цілих чисел з використанням вихідного переносу і ознаки нульового значення часткового залишку діленого при одночасному отриманні значення старшого біту частки. Показано, що з метою уніфікації алгоритму обчислення бінарних значень цифрових розрядів частки, нульові часткові залишки діленого при визначенні цифр частки та обчисленні наступного часткового залишку діленого необхідно перетворити у код нуля відповідно до формату додатного або від'ємного двійкового коду в залежності від знаку діленого. Доведена коректність застосування коду перетвореного нуля при обчисленні бінарних значень розрядів частки та залишку. Математично підтверджено правило формування залишку у відповідності до знаку діленого.

Ключові слова: доповняльний код, негативний нуль, позитивний нуль, модуль доповняльного коду, залишок по модулю, переповнення

DOI: 10.31474/1996-1588-2023-2-37-4-14

Вступ

Практична реалізація операції ділення цілих чисел в доповняльному коді за допустимих значень операндів передбачає контроль переповнення розрядної сітки частки [1-3]. Спрощення контролю переповнення частки та зменшення апаратних витрат досягаються при використанні стандартних інформаційних сигналів, що формуються [4-6]. При діленні доречно застосовувати, наприклад, вихідний перенос та ознаку наявності нульового значення залишку діленого.

Особливості подання операндів у доповняльному коді призводять до нетотожних дій при різних співвідношеннях знаків операндів за причиною, у тому числі, вимоги неоднакового опрацювання коду нуля [7-10]. Для формування коректного коду частки нульові значення часткових залишків діленого пропонується розглядати у форматі позитивного або негативного нуля у відповідності до знаку діленого.

Загальна методика ділення цілих чисел

Обчислення частки від ділення цілих чисел визначено співвідношенням:

$$a = bd + c, \tag{1}$$

де a, b, c, d – знакові цілі числа, що вказують на значення діленого, дільнику, залишку та частки відповідно.

Однозначність розкладання діленого на складові та цілочисловий тип ділення забезпечуються виконанням умов:

$$b \neq 0, c < b \tag{2}$$

При діленні знакових цілих чисел задача (1) зводиться до розв'язку рівняння

$$A = B \cdot D + C, \tag{3}$$

де $A = \pm|A|, B = \pm|B|, D = \pm|D|, C = \pm|C|$ – коди діленого, дільника, частки та залишку відповідно, де додатність (+) і від'ємність (–) складових операції ділення позначено як NA, NB, ND, NC та закодовано нулем і одиницею:

$$NX = \begin{cases} 0, if\ X \geq 0; \\ 1, if\ X < 0, \end{cases}$$

де $X = A, B, D, C$. Загалом, таке комп'ютерне зображення чисел відповідає терміну "прямий

код". При практичному розв'язанні задачі (3) знакові операнди відображаються в доповняльних кодах.

Знак частки, відповідно до правил алгебри, визначатиметься логічним виразом

$$ND = NA \oplus NB . \quad (4)$$

Однозначність розкладання знакового діленого на складові забезпечується при обмеженнях (2) на модулі чисел та збереженні співвідношення (1) між модулями чисел:

$$|A| = |B| \cdot |D| + |C| .$$

При довільних знаках діленого і дільника коректне визначення частки забезпечується при виконанні наступного співвідношення між знаковими розрядами числових даних:

$$NB \oplus ND = NC ,$$

що вказує на знак доданку в рівнянні (3), як результат від формування добутку дільника і частки. Враховуючи (4), утворюється вираз для визначення знаку коректного коду залишку діленого

$$NC = NB \oplus (NA \oplus NB) = NA . \quad (5)$$

В задачі (1) існує $a \ge b \cdot d$. З урахуванням властивостей операції множення невід'ємних цілих чисел, розрядність числових даних співвідноситься за виразом

$$l_a \ge l_b + l_d , \quad (6)$$

де l_a , l_b , l_d – максимальна кількість значущих розрядів діленого (довжина двійкового коду модулю діленого), дільника і частки відповідно. При вказанні модулів дільника та частки $(n-1)$-розрядними кодами: $l_b = l_d = n-1$, нерівність (6) приймає вигляд

$$l_a \ge 2n-2 . \quad (7)$$

З урахуванням знакового розряду розрядність змінних задачі (3) збільшується

$$l_B = l_D = n ,$$

де l_B, l_D – розрядність знакових числових значень дільника та частки відповідно.

При однорідній пам'яті комп'ютера, що складається з комірок розрядністю n біт, більш довге ділене (7) розташовується в двох суміжних комірках пам'яті. Тому розрядність l_A комп'ютерного відображення знакового діленого в задачі (3) визначається як $l_A = 2n$.

Згідно з (2), для комп'ютерного відображення знакового залишку діленого задачі (3) визначається розрядністю

$$l_C = l_B = n .$$

Відображення частки у доповняльному коді вказується n-розрядним поліномом

$$D^T(n,1) = [ND] D^T(n-1,1) = \\ = (V + D)_{mV} = V signD + D, \quad (8)$$

де $V = 2^n$ – параметр відображення доповняльного коду розрядністю n біт, $(X)_{mV}$ – позначення процедури обчислення залишку за модулем V коду (X), $D^T(n-1,1)$ - $(n-1)$-розрядний доповняльний код частки (3), $signX = 0$ або $signX = 1$ - знаковий біт додатного або від'ємного числа X відповідно, $ND = signD$.

Область значень функції $D^T(D)$ при $D \ge 0$ згідно з (8), відповідає інтервалу:

$$D^T(n,1) = [0] D^T(n-1,1) \in [0; V/2-1],$$

де $[0; V/2-1]$ – діапазон значень двійкового коду в полі $(n-1,1)$. З $signB(if\ B \ge 0) = 0$ випливає:

$$D^T(n,1) = V signD + D = D .$$

При $D < 0$ інтервал зміни значень функції $D^T(D)$, згідно з (8), відповідає інтервалу:

$$D^T(n,1) = [1] D^T(n-1,1) \in [V/2; V-1],$$ отже

$$D^T(n,1) = V signD + D = V + D .$$

Таким чином, область визначення n-розрядної функції $D^T(D)$ дорівнює:

$$D \in \begin{cases} [0; V/2-1], if\ D \ge 0; \\ [-V/2; -1], if\ D < 0. \end{cases} \quad (9)$$

Частка D в області визначення функції (8) на різних ділянках зміни частки відображається двійковими кодами різного виду, зіставленими додатним та від'ємним числам:

$$D \in \begin{cases} +0 d_{n-1} d_{n-2} ... d_2 d_1, if\ D \in [0; V/2-1]; \\ -d_n d_{n-1} d_{n-2} ... d_2 d_1, if\ D \in [-V/2; -1], \end{cases} \quad (10)$$

де d_i – цифрові розряди модуля частки при $i = 1,2,...,n-1,n$.

Таким чином, згідно з (8-10), відображення додатних значень частки D є цілочисловим двійковим кодом

$$D^T(n,1)=(V+D)_{mV}=$$
$$=\left(V+\left(+0d_{n-1}d_{n-2}...d_2d_1\right)\right)_{mV}=$$
$$=[0]d_{n-1}d_{n-2}...d_2d_1,$$

де $d_{n-1}d_{n-2}...d_2d_1$ – цифрові розряди модуля додатної частки; $[0]$ – знаковий розряд додатної частки $D\ge 0$.

Відображення від'ємних числових значень частки D в доповняльному коді, згідно з (8-10), є цілочисловим двійковим кодом:

$$D^T(n,1)=(V+D)_{mV}=\left(V+\left(-d_nd_{n-1}...d_2d_1\right)\right)_{mV}=$$
$$=V-d_nd_{n-1}...d_2d_1=\left(2^n-1\right)-d_nd_{n-1}...d_2d_1+1=$$
$$=\left(1_n1_{n-1}...1_21_1\right)-d_nd_{n-1}...d_2d_1+1=\left(1_n-d_n\right)\cdot 2^{n-1}+$$
$$+\left(1_{n-1}-d_{n-1}\right)\cdot 2^{n-2}+...+\left(1_2-d_2\right)\cdot 2^1+$$
$$+\left(1_1-d_1\right)\cdot 2^0+1=\overline{d_n}\cdot 2^{n-1}+\overline{d_{n-1}}\cdot 2^{n-2}+$$
$$+...+\overline{d_2}\cdot 2^1+\overline{d_1}\cdot 2^0+1=\left[\overline{d_n}\right]\overline{d_{n-1}}\,\overline{d_{n-2}}...\overline{d_2}\,\overline{d_1}+1,$$

де $\overline{d_{n-1}}\,\overline{d_{n-2}}...\overline{d_2}\,\overline{d_1}$ – інверсні значення цифрових розрядів модуля від'ємної частки, $\left[\overline{d_n}\right]$ – інверсне значення в позиції знакового розряду доповняльного коду частки $D^T(n,1)$.

Таким чином, процедура формування доповняльного коду частки визначається системою:

$$D^T(n,1)=\begin{cases}[d_n]d_{n-1}d_{n-2}...d_2d_1,\\ \quad if\ D\in[0;V/2-1];\\ \left[\overline{d_n}\right]\overline{d_{n-1}}\,\overline{d_{n-2}}...\overline{d_2}\,\overline{d_1}+1,\\ \quad if\ D\in[-V/2;-1],\end{cases}\qquad(11)$$

де $d_{n-1}d_{n-2}...d_2d_1$ – бінарні значення цифрових розрядів прямого коду модуля частки, які визначаються розв'язкою задачі (1), $[d_n]$ та $\left[\overline{d_n}\right]$ – бінарні значення в знаковій позиції доповняльного коду частки.

Ділення додатних чисел в доповняльному коді

При $A>0$, $B>0$, згідно з (11), значення модуля додатної частки обмежено $|D|<2^{n-1}$, тому переповнення розрядної сітки частки при діленні додатних чисел визначається нерівністю

$$A<B\cdot 2^{n-1},$$

де $A=+|A|, |A|\in[0;H/2-1], H=2^{2n}$ - параметр відображення доповняльного коду розрядністю $2n$ біт, $B=+|B|, |B|\in[1;V/2-1]$.

Переповнення відсутнє при $A-B\cdot 2^{n-1}<0$, рівнозначне

$$2A-B\cdot 2^n<0,$$

що відповідає представленню зміщеного дільника $B(n,1)\cdot 2^n$ відповідно до розрядності подвоєного діленого.

Таким чином,

$$OVPP=\begin{cases}0, if\ A_n<0;\\ 1, if\ A_n=0;\\ 1, if\ A_n>0,\end{cases}\qquad(12)$$

де $A_n=2A-B\cdot 2^n$, $OVPP$ – ознака переповнення розрядної сітки частки при діленні додатних чисел.

Обчислення залишку діленого A_n в доповняльному коді визначається виразом

$$A_n^T(2n,1)=\left(2A^T(2n,1)-B^T(n,1)\cdot 2^n\right)_{mH},\qquad(13)$$

де $A^T(2n,1)=[NA]A^T(2n-1,1)=(H+A)_{mH}=$
$=HsignA+A=A$, $B^T(n,1)=[NB]B^T(n-1,1)=$
$=(V+B)_{mV}=VsignB+B=B$.

Вираз (13) подається у вигляді:

$$A_n^T(2n,1)=\left(H+2A^T(2n,1)-B^T(n,1)\cdot 2^n\right)_{mH}=$$
$$=\left(\left(2A^T(2n,1)\right)_{mH}+\left(\left(2^n-1\right)-B^T(n,1)+1\right)_{mV}\cdot 2^n\right)_{mH}$$
$$=\left(\left(2A^T(2n,1)\right)_{mH}+\left(KE(n,1)-B^T(n,1)+1\right)_{mV}\cdot 2^n\right)_{mH}=\qquad(14)$$
$$=\left(\left(2A^T(2n,1)\right)_{mH}+\left(\overline{B^T(n,1)}+1\right)_{mV}\cdot 2^n\right)_{mH}=$$
$$=\left(\left(2A^T(2n,1)\right)_{mH}+B_{TR}^T(n,1)\cdot 2^n\right)_{mH},$$

де $A^T(2n,1)=[0]|A|(2n-1,1)$, $\overline{B^T(n,1)}$ – порозрядна n-розрядна інверсія доповняльного коду дільника:

$$B^T(n,1)=[0]|B|(n-1,1),$$
$$B_{TR}^T(n,1)=\overline{B^T(n,1)}+1$$

– перетворений доповняльний код дільника.

Згідно з (14), формується двійковий код залишку діленого A_n та переносу $E\cdot 2^{2n}$, де E – вихідний перенос повної суми залишку діленого

$$S(2n+1,1)=2A^T(2n,1)+B_{TR}^T(n,1)\cdot 2^n=$$
$$=2A^T(2n,1)+H-B^T(n,1)\cdot 2^n=$$
$$=E\cdot 2^{2n}+2A-B\cdot 2^n=E\cdot 2^{2n}+A_n=$$
$$=\begin{cases}>2^{2n}, if\ A_n>0, E=1;\\ 2^{2n}, if\ A_n=0, E=1;\\ <2^{2n}, if\ A_n<0, E=0.\end{cases}$$

Отже,

$$E=\begin{cases}1, if\ A_n>0, N_n=0;\\ 1, if\ A_n=0, N_n=0;\\ 0, if\ A_n<0, N_n=1,\end{cases} \quad (15)$$

де N_n – знаковий біт доповняльного коду залишку діленого $A_n^T(2n,1)=[N_n]A_n^T(2n-1,1)$.

Згідно з (12-15), формування ознаки переповнення частки при діленні додатних чисел описується виразом:

$$OVPP=E.$$

За відсутності переповнення $OVPP$, бінарні значення розрядів додатної частки, згідно з (3) та (11), визначаються розв'язкою рівняння

$$A=B\cdot\left(d_n\cdot 2^{n-1}+...+d_2\cdot 2+d_1\right)+C.$$

З урахуванням (12), за відсутності переповнення $OVPP$, бінарне значення в позиції знакового розряду d_n додатної частки при діленні додатних чисел визначається $d_n=0$.

В загальному випадку значення біту d_{n-1} визначається системою нерівностей:

$$d_{n-1}=\begin{cases}1, if\ A_{n-1}\ge 0;\\ 0, if\ A_{n-1}<0,\end{cases} \quad (16)$$

де

$$A_{n-1}=\begin{cases}A-B\cdot 2^{n-2}, if\ d_n=0;\\ A-B\cdot 2^{n-1}-B\cdot 2^{n-2}, if\ d_n=1.\end{cases} \quad (17)$$

При представленні зміщеного дільника $B(n,1)\cdot 2^{n-1}$ у відповідності до розрядності діленого $A(2n,1)$ вираз (17) приймає вигляд:

$$A_{n-1}=\begin{cases}2A_n+B\cdot 2^n, if\ d_n=0;\\ 2A_n-B\cdot 2^n, if\ d_n=1.\end{cases}$$

В системі доповняльного коду залишок діленого A_{n-1} утворюється відповідно до системи виразів

$$A_{n-1}=\begin{cases}2A_n^T(2n,1)+B^T(n,1)\cdot 2^n, if\ d_n=0;\\ 2A_n^T(2n,1)+B_{TR}^T(n,1)\cdot 2^n, if\ d_n=1.\end{cases} \quad (18)$$

Отже, згідно з (16), (18), при діленні додатних чисел існує

$$d_{n-1}=\begin{cases}1, if\left(N_{n-1}=0\right)\wedge\left(NB=0\right);\\ 0, if\left(N_{n-1}=1\right)\wedge\left(NB=0\right),\end{cases} \quad (19)$$

де N_{n-1} – бінарне значення у знаковому розряді доповняльного коду залишку діленого $A_{n-1}^T(2n,1)=[N_{n-1}]A_{n-1}^T(2n-1,1)$.

Таким чином, при діленні додатних чисел бінарне значення d_{n-1} визначається співвідношенням:

$$d_{n-1}=\overline{N_{n-1}\oplus NB}. \quad (20)$$

При змінах індексів вирази (18-20) вказують на правила формування усіх наступних цифрових розрядів додатної частки.

Кінцеве значення залишку діленого формується у відповідності до значення часткового залишку діленого $A_1^T(2n,1)$, який, згідно з (5), при $N_1\ne NA$ коригується:

$$C^T(n,1)=\begin{cases}A_1^T(2n,n+1), if\ N_1=0;\\ A_0^T(2n,n+1), if\ N_1=1,\end{cases}$$

де $C^T(n,1)$ – цілочисловий залишок діленого при діленні додатних чисел, N_1 - значення у знаковому біті доповняльного коду залишку діленого:

$$A_1^T(2n,1)=[N_1]A_1^T(2n-1,1),$$
$$A_1^T(2n,n+1)=A_1^T(2n,1)\cdot 2^{-n},$$
$$A_0^T(2n,n+1)=\left(A_1^T(2n,n+1)+B^T(n,1)\right)_{mV}.$$

Ділення цілих чисел в доповняльному коді при додатному діленому та від'ємному дільнику

При $A>0$, $B<0$, згідно з (11), значення модуля від'ємної частки обмежено $|D|\le 2^{n-1}$, тому переповнення розрядної сітки частки визначається нерівністю

$$A\le|B|\cdot 2^{n-1},$$

де $A=+|A|, |A|\in[0;H/2-1]$, $|B|\in[1;V/2]$.

Переповнення відсутнє при

$$A+B\cdot 2^{n-1}\le 0,$$

рівнозначне

$$2A+B\cdot 2^n\le 0,$$

що відповідає представленню зміщеного дільника $B(n,1)\cdot 2^n$ відповідно до розрядності подвоєного діленого, $B=-|B|$.

Таким чином,

$$OVPN=\begin{cases}0, if\ A_n<0;\\ 0, if\ A_n=0;\\ 1, if\ A_n>0,\end{cases} \quad (21)$$

де $A_n=2A+B\cdot 2^n$, $OVPN$ – ознака переповнення розрядної сітки частки при діленні цілих чисел за наявності додатного діленого і від'ємного дільника.

Обчислення залишку діленого A_n в доповняльному коді визначається виразом

$$A_n^T(2n,1)=\left(2A^T(2n,1)+B^T(n,1)\cdot 2^n\right)_{mH}, \quad (22)$$

де $A^T(2n,1)=[NA]A^T(2n-1,1)=(H+A)_{mH}=$
$=HsignA+A=A$, $B^T(n,1)=[NB]B^T(n-1,1)=$
$=(V+B)_{mV}=VsignB+B=2^n+B=2^n-|B|$.

Згідно з (22), формується двійковий код залишку діленого A_n, ознаки ZA про наявність залишку, що дорівнює нулю та переносу $E\cdot 2^{2n}$, де E – вихідний перенос повної суми залишку діленого:

$$S(2n+1,1)=2A^T(2n,1)+B^T(n,1)\cdot 2^n=$$
$$=2A+\left(2^n-|B|\right)\cdot 2^n=2^{2n}+2A-|B|\cdot 2^n=$$
$$=E\cdot 2^{2n}+2A+B\cdot 2^n=E\cdot 2^{2n}+A_n=$$
$$=\begin{cases}>2^{2n}, if\ A_n>0, E=1;\\ 2^{2n}, if\ A_n=0, E=1;\\ <2^{2n}, if\ A_n<0, E=0.\end{cases}$$

Отже,

$$E=\begin{cases}1, if\ ZA=1;\\ 1, if\ (N_n=0)\wedge(ZA=0);\\ 0, if\ (N_n=1)\wedge(ZA=0),\end{cases} \quad (23)$$

де N_n – знаковий біт доповняльного коду залишку діленого $A_n^T(2n,1)=[N_n]A_n^T(2n-1,1)$,

$$ZA=\begin{cases}0, if\ A_n\neq 0;\\ 1, if\ A_n=0.\end{cases}$$

Згідно з (21-23), формування ознаки переповнення частки при діленні цілих чисел за наявності додатного діленого і від'ємного дільника описується виразом:

$$OVPN=E\oplus ZA.$$

За відсутності переповнення $OVPN$, бінарні значення розрядів від'ємної частки, згідно з (3) та (10), визначаються розв'язкою рівняння

$$A=-B\cdot\left(d_n\cdot 2^{n-1}+d_{n-1}\cdot 2^{n-2}...+d_2\cdot 2^1+d_1\cdot 2^0\right)+C\cdot$$

З урахуванням (22), за відсутності переповнення $OVPN$, значення біту в позиції знакового розряду d_n частки при діленні цілих чисел за наявності додатного діленого і від'ємного дільника визначається системою:

$$d_n=\begin{cases}0, if\ A_n<0;\\ 1, if\ A_n=0.\end{cases} \quad (24)$$

При діленні додатного діленого і від'ємного дільника, згідно з (11), доповняльний код від'ємної частки визначається системою виразів:

$$D^T(n,1)=\begin{cases}\left[\overline{0}\right]\overline{d_{n-1}}\,\overline{d_{n-2}}...\overline{d_2}\,\overline{d_1}+1, if\ |D|<2^{n-1};\\ \left[\overline{1}\right]\overline{0_{n-1}}\,\overline{0_{n-2}}...\overline{0_2}\,\overline{0_1}+1, if\ D=-2^{n-1},\end{cases} \quad (25)$$

де $\left[\overline{0}\right], \left[\overline{1}\right]$ – бінарні інверсні значення в позиції знакового розряду d_n від'ємної частки.

Отже, згідно з (24-25), при діленні цілих чисел за наявності додатного діленого і від'ємного дільника існує

$$d_n=\begin{cases}1, if\ (N_n=0)\wedge(NB=1);\\ 0, if\ (N_n=1)\wedge(NB=1),\end{cases} \quad (26)$$

де N_n - знаковий біт доповняльного коду залишку діленого $A_n^T(2n,1)=[N_n]A_n^T(2n-1,1)$.

Таким чином, з урахуванням (26), бінарне значення в позиції знакового розряду доповняльного коду від'ємної частки визначається логічним виразом

$$d_n=N_n\oplus NB. \quad (27)$$

В загальному випадку значення біту d_{n-1} доповняльного коду від'ємної частки визначається системою нерівностей:

$$d_{n-1}=\begin{cases}1, if\ A_{n-1}\geq 0;\\ 0, if\ A_{n-1}<0,\end{cases} \quad (28)$$

де

$$A_{n-1}=\begin{cases}A+B\cdot 2^{n-2}, if\ d_n=0;\\ A+B\cdot 2^{n-1}+B\cdot 2^{n-2}, if\ d_n=1.\end{cases} \quad (29)$$

При представленні зміщеного дільника $B(n,1)\cdot 2^{n-1}$ у відповідності до розрядності діленого $A(2n,1)$ вираз (29), згідно з (27), приймає вигляд:

$$A_{n-1}=\begin{cases}2A_n-B\cdot 2^n, if\ d_n=0;\\ 2A_n+B\cdot 2^n, if\ d_n=1.\end{cases}$$

В системі доповняльного коду залишок діленого A_{n-1} утворюється відповідно до системи виразів

$$A_{n-1}=\begin{cases}2A_n^T(2n,1)+B_{TR}^T(n,1)\cdot 2^n, if\ d_n=0;\\ 2A_n^T(2n,1)+B^T(n,1)\cdot 2^n, if\ d_n=1.\end{cases} \quad (30)$$

Згідно з (28), (30), існує

$$d_{n-1}=\begin{cases}1, if\ (N_{n-1}=0)\wedge(NB=1);\\ 0, if\ (N_{n-1}=1)\wedge(NB=1),\end{cases} \quad (31)$$

де N_{n-1} – бінарне значення у знаковому розряді доповняльного коду залишку діленого $A_{n-1}^T(2n,1)=[N_{n-1}]A_{n-1}^T(2n-1,1)$.

Таким чином, згідно з (31), значення біту d_{n-1} доповняльного коду від'ємної частки визначається логічним виразом

$$d_{n-1}=N_{n-1}\oplus NB.$$

В загальному випадку, при діленні цілих чисел за наявності від'ємного діленого і додатного дільника доповняльний код від'ємної частки визначається інверсним кодом відповідно до (11), (25):

$$\overline{d_{n-1}}=\overline{N_{n-1}\oplus NB}. \quad (32)$$

При змінах індексів вирази (30-32) вказують на правила формування усіх наступних цифрових розрядів від'ємної частки.

Кінцеве значення залишку діленого формується у відповідності до значення часткового залишку діленого $A_1^T(2n,1)$, який, згідно з (5), при $N_1\neq NA$ коригується:

$$C^T(n,1)=\begin{cases}A_1^T(2n,n+1), if\ N_1=0;\\ A_0^T(2n,n+1), if\ N_1=1,\end{cases}$$

де $C^T(n,1)$ – цілочисловий залишок діленого, N_1 – значення у знаковому біті доповняльного коду залишку діленого

$$A_1^T(2n,1)=[N_1]A_1^T(2n-1,1),$$
$$A_1^T(2n,n+1)=A_1^T(2n,1)\cdot 2^{-n},$$
$$A_0^T(2n,n+1)=\left(A_1^T(2n,n+1)+B_{TR}^T(n,1)\right)_{mV}.$$

Доповняльний код частки, згідно з (11), формується інкрементуванням інверсного коду частки

$$D^T(n,1)=\left[\overline{d_n}\right]\overline{d_{n-1}}\,\overline{d_{n-2}}\ldots\overline{d_2}\,\overline{d_1}+1.$$

Ділення цілих чисел в доповняльному коді при від'ємному діленому та додатному дільнику

При $A<0$, $B>0$, згідно з (11), значення модуля від'ємної частки обмежено $|D|\le 2^{n-1}$, тому переповнення розрядної сітки частки визначається нерівністю

$$|A|\le B\cdot 2^{n-1},$$

де $B=+|B|$, $B<2^{n-1}$, $|A|\le 2^{2n-2}$.

Переповнення відсутнє при

$$A+B\cdot 2^{n-1}\ge 0,$$

рівнозначне

$$2A+B\cdot 2^n\ge 0,$$

що відповідає представленню зміщеного дільника $B(n,1)\cdot 2^n$ відповідно до розрядності подвоєного діленого, $A=-|A|$.

Таким чином,

$$OVNP=\begin{cases}0, if\ A_n>0;\\ 0, if\ A_n=0;\\ 1, if\ A_n<0,\end{cases} \quad (33)$$

де $A_n=2A+B\cdot 2^n$, $OVNP$ – ознака переповнення розрядної сітки частки при діленні цілих чисел за наявності від'ємного діленого і додатного дільника.

Обчислення залишку діленого A_n в доповняльному коді визначається виразом

$$A_n^T(2n,1)=\left(2A^T(2n,1)+B^T(n,1)\cdot 2^n\right)_{mH}, \quad (34)$$

де $A^T(2n,1)=[NA]A^T(2n-1,1)=$
$=(H+A)_{mH}=HsignA+A=H+A=2^{2n}-|A|$,
$B^T(n,1)=[NB]B^T(n-1,1)=(V+B)_{mV}=B$.

Згідно з (34), з урахуванням обмеження $|A|<2^{2n-1}$, формується двійковий код залишку діленого A_n, ознаки ZA_n про наявність залишку, що дорівнює нулеві, та переносу $E\cdot 2^{2n}$, де E – вихідний перенос повної суми залишку діленого

$$S(2n+1,1)=2A^T(2n,1)+B^T(n,1)\cdot 2^n=$$
$$=2\left(2^{2n-1}-|A|\right)+B\cdot 2^n=$$
$$=2^{2n}+2A+B\cdot 2^n=E\cdot 2^{2n}+A_n=$$
$$=\begin{cases}>2^{2n}, if\ A_n>0, E=1;\\ 2^{2n}, if\ A_n=0, E=1;\\ <2^{2n}, if\ A_n<0, E=0.\end{cases}$$

Отже,

$$E=\begin{cases}1, if\ A_n>0, N_n=0;\\ 1, if\ A_n=0, N_n=0;\\ 0, if\ A_n<0, N_n=1,\end{cases} \quad (35)$$

де N_n – знаковий біт доповняльного коду залишку діленого $A_n^T(2n,1)=[N_n]A_n^T(2n-1,1)$.

Згідно з (33),(35), формування ознаки переповнення частки при діленні цілих чисел за наявності від'ємного діленого і додатного дільника вказується виразом:

$$OVNP=\overline{E}.$$

За відсутності переповнення $OVNP$, бінарні значення розрядів від'ємної частки, згідно з (3) та (10), визначаються розв'язкою рівняння

$$A=-B\cdot\left(d_n\cdot 2^{n-1}+d_{n-1}\cdot 2^{n-2}...+d_2\cdot 2^1+d_1\cdot 2^0\right)+C.$$

З урахуванням (34), за відсутності переповнення $OVNP$, значення біту в позиції знакового розряду d_n частки при діленні цілих чисел за наявності від'ємного діленого і додатного дільника визначається системою

$$d_n=\begin{cases}0, if\ A_n>0;\\ 1, if\ A_n=0.\end{cases} \quad (36)$$

Звідки, згідно з (33-34), з урахуванням $NB=0$, випливає

$$d_n=\begin{cases}0, if\ (ZA=0)\wedge(N_n=0);\\ 1, if\ (ZA=1)\wedge(N_n=0).\end{cases} \quad (37)$$

З урахуванням знаку діленого $NA=1$,

$$N_n^{NZ}=ZA_n\oplus N_n,$$

де N_n^{NZ} – ознака знакового біту коду від'ємного нуля в полі часткового залишку діленого $A_n^T(2n,1)$,

$$ZA_n=\begin{cases}0, if\ A_n\neq 0;\\ 1, if\ A_n=0.\end{cases}$$

З урахуванням (36-37) та ознаки N_n^{NZ}, існує

$$d_n=N_n^{NZ}\oplus NB. \quad (38)$$

В загальному випадку, бінарне значення d_{n-1} визначається системою нерівностей

$$d_{n-1}=\begin{cases}0, if\ A_{n-1}>0;\\ 1, if\ A_{n-1}\le 0,\end{cases} \quad (39)$$

де $A_{n-1}=\begin{cases}A+B\cdot 2^{n-2}, if\ d_n=0;\\ A+B\cdot 2^{n-1}+B\cdot 2^{n-2}, if\ d_n=1.\end{cases}$

При представленні зміщеного дільника $B(n,1)\cdot 2^{n-1}$ у відповідності до розрядності діленого $A(2n,1)$ вираз (39) приймає вигляд:

$$A_{n-1}=\begin{cases}2A_n-B\cdot 2^n, if\ d_n=0;\\ 2A_n+B\cdot 2^n, if\ d_n=1.\end{cases}$$

В системі доповняльного коду залишок діленого A_{n-1} утворюється відповідно до системи виразів:

$$A_{n-1}=\begin{cases}2A_n^T(2n,1)+B_{TR}^T(n,1)\cdot 2^n, if\ d_n=0;\\ 2A_n^T(2n,1)+B^T(n,1)\cdot 2^n, if\ d_n=1.\end{cases} \quad (40)$$

Згідно з (38), з урахуванням $NB=0$, існує

$$d_{n-1}=\begin{cases}0, if\ (ZA_{n-1}=0)\wedge(N_{n-1}=0);\\ 1, if\ (ZA_{n-1}=0)\wedge(N_{n-1}=1);\\ 1, if\ (ZA_{n-1}=1)\wedge(N_{n-1}=0),\end{cases} \quad (41)$$

де N_{n-1} – бінарне значення у знаковому розряді доповняльного коду залишку діленого $A_{n-1}^T(2n,1)=[N_{n-1}]A_{n-1}^T(2n-1,1)$,

$$ZA_{n-1}=\begin{cases}0, if\ A_{n-1}\neq 0;\\ 1, if\ A_{n-1}=0.\end{cases}$$

Значення біту d_{n-1} доповняльного коду від'ємної частки визначається виразом

$$d_{n-1}=N_{n-1}^{NZ}\oplus NB,$$

де $N_{n-1}^{NZ}=ZA_{n-1}\oplus N_{n-1}$ – ознака знакового біту коду від'ємного нуля в полі часткового залишку діленого $A_{n-1}^T(2n,1)$.

В загальному випадку, за наявності від'ємного діленого і додатного дільника доповняльний код від'ємної частки визначається інверсним кодом відповідно до (11), (25):

$$\overline{d_{n-1}}=\overline{N_{n-1}^{NZ}\oplus NB}. \quad (42)$$

При змінах індексів вирази (40-42) вказують на правила формування усіх наступних цифрових розрядів від'ємної частки.

Кінцеве значення залишку діленого формується у відповідності до значення часткового залишку діленого $A_1^T(2n,1)$, який, згідно з (5), при $N_1\neq NA$ коригується:

$$C^T(n,1)=\begin{cases}A_1^T(2n,n+1), if\ N_1^{NZ}=1;\\ A_0^T(2n,n+1), if\ N_1^{NZ}=0,\end{cases}$$

де $C^T(n,1)$ – цілочисловий залишок діленого, $N_1^{NZ} = ZA_1 \oplus N_1$ – ознака знакового біту коду від'ємного нуля в полі залишку діленого $A_1^T(2n,1)$, ZA_1 – ознака наявність залишку діленого A_1, що дорівнює нулеві, N_1 – бінарне значення у знаковому біті доповняльного коду залишку діленого

$$A_1^T(2n,1) = [N_1]A_1^T(2n-1,1),$$

$$A_1^T(2n,n+1) = A_1^T(2n,1)\cdot 2^{-n},$$

$$A_0^T(2n,n+1) = \left(A_1^T(2n,n+1) + B_{TR}^T(n,1)\right)_{mV}.$$

Доповняльний код частки, згідно з (11), формується інкрементуванням інверсного коду частки

$$D^T(n,1) = \left[\overline{d_n}\right]\overline{d_{n-1}}\,\overline{d_{n-2}}\ldots\overline{d_2}\,\overline{d_1} + 1.$$

Ділення цілих від'ємних чисел в доповняльному коді

При $A<0$, $B<0$, згідно з (11), значення модуля додатної частки обмежено $|D| < 2^{n-1}$, тому переповнення розрядної сітки частки визначається нерівністю:

$$|A| < |B|\cdot 2^{n-1},$$

де $|B| \le 2^{n-1}$, $|A| \le 2^{2n-2}$.

Переповнення відсутнє при $A - B\cdot 2^{n-1} > 0$, рівнозначне $2A - B\cdot 2^n > 0$, що відповідає представленню зміщеного дільника $B(n,1)\cdot 2^n$ відповідно до розрядності подвоєного діленого, $A = -|A|$, $B = -|B|$.

Таким чином,

$$OVNN = \begin{cases} 0, if\ A_n > 0; \\ 1, if\ A_n = 0; \\ 1, if\ A_n < 0, \end{cases} \quad (43)$$

де $A_n = 2A - B\cdot 2^n$, $OVNN$ – ознака переповнення розрядної сітки частки при діленні від'ємних чисел.

Обчислення залишку діленого A_n в доповняльному коді визначається виразом

$$A_n^T(2n,1) = \left(2A^T(2n,1) - B^T(n,1)\cdot 2^n\right)_{mH}, \quad (44)$$

де $A^T(2n,1) = [NA]A^T(2n-1,1) =$

$$= (H + A)_{mH} = HsignA + A = H + A = 2^{2n} - |A|,$$

$$B^T(n,1) = [NB]B^T(n-1,1) = 2^n - |B|.$$

Вираз (44) подається у вигляді

$$A_n^T(2n,1) = \left(H + 2A^T(2n,1) - B^T(n,1)\cdot 2^n\right)_{mH} = \left(\left(2A^T(2n,1)\right)_{mH} + B_{TR}^T(n,1)\cdot 2^n\right)_{mH}, \quad (45)$$

де $B_{TR}^T(n,1)$ – перетворений доповняльний код дільника відповідно до (14).

Згідно з (44,45), з урахуванням обмеження $|A| < 2^{2n-1}$, формується двійковий код залишку діленого A_n, ознаки ZA_n про наявність залишку, що дорівнює нулеві, та переносу $E\cdot 2^{2n}$, де E – вихідний перенос повної суми залишку діленого

$$S(2n+1,1) = 2A^T(2n,1) + B_{TR}^T(n,1)\cdot 2^n =$$
$$= 2\left(2^{2n-1} - |A|\right) + \left(2^n - 2^n + |B|\right)\cdot 2^n =$$
$$= 2^n + 2A - B\cdot 2^n = E\cdot 2^{2n} + A_n =$$
$$= \begin{cases} > 2^{2n}, if\ A_n > 0, E = 1; \\ 2^{2n}, if\ A_n = 0, E = 1; \\ < 2^{2n}, if\ A_n < 0, E = 0. \end{cases}$$

Отже,

$$E = \begin{cases} 1, if\ A_n > 0, N_n = 0; \\ 1, if\ A_n = 0, N_n = 0; \\ 0, if\ A_n < 0, N_n = 1, \end{cases} \quad (46)$$

де N_n – знаковий біт доповняльного коду залишку діленого:

$$A_n^T(2n,1) = [N_n]A_n^T(2n-1,1).$$

Згідно з (43-46), формування ознаки переповнення частки при діленні від'ємних чисел описується виразом:

$$OVNN = \overline{E \oplus ZA_n}, \text{ де } ZA_n = \begin{cases} 0, if\ A_n \ne 0; \\ 1, if\ A_n = 0. \end{cases}$$

За відсутності переповнення $OVNN$, бінарні значення розрядів додатної частки, згідно з (3) та (11), визначаються розв'язкою рівняння

$$A = B\cdot\left(d_n\cdot 2^{n-1} + \ldots + d_2\cdot 2 + d_1\right) + C.$$

З урахуванням (43), за відсутності переповнення $OVNN$, бінарне значення в позиції знакового розряду d_n додатної частки при діленні від'ємних чисел визначається $d_n = 0$.

В загальному випадку, бінарне значення розряду d_{n-1} визначається системою

$$d_{n-1} = \begin{cases} 0, if\ A_{n-1} > 0; \\ 1, if\ A_{n-1} = 0; \\ 1, if\ A_{n-1} < 0, \end{cases}$$

де

$$A_{n-1}=\begin{cases}A-B\cdot 2^{n-2},if\ d_n=0;\\ A-B\cdot 2^{n-1}-B\cdot 2^{n-2},if\ d_n=1.\end{cases} \quad (47)$$

При представленні зміщеного дільника $B(n,1)\cdot 2^{n-1}$ у відповідності до розрядності діленого $A(2n,1)$ вираз (47) приймає вигляд:

$$A_{n-1}=\begin{cases}2A_n+B\cdot 2^n,if\ d_n=0;\\ 2A_n-B\cdot 2^n,if\ d_n=1.\end{cases}$$

В системі доповняльного коду залишок діленого A_{n-1} утворюється відповідно до системи виразів:

$$A_{n-1}=\begin{cases}2A_n^T(2n,1)+B^T(n,1)\cdot 2^n,if\ d_n=0;\\ 2A_n^T(2n,1)+B_{TR}^T(n,1)\cdot 2^n,if\ d_n=1.\end{cases} \quad (48)$$

З урахуванням $NB=1$, існує

$$d_{n-1}=\begin{cases}0,if\ (ZA_{n-1}=0)\wedge(N_{n-1}=0);\\ 1,if\ (ZA_{n-1}=0)\wedge(N_{n-1}=1);\\ 1,if\ (ZA_{n-1}=1)\wedge(N_{n-1}=0),\end{cases} \quad (49)$$

де N_{n-1} – бінарне значення у знаковому розряді доповняльного коду залишку діленого:

$$A_{n-1}^T(2n,1)=[N_{n-1}]A_{n-1}^T(2n-1,1),$$

$$ZA_{n-1}=\begin{cases}0,if\ A_{n-1}\neq 0;\\ 1,if\ A_{n-1}=0.\end{cases}$$

Таким чином, значення біту d_{n-1} доповняльного коду додатної частки визначається виразом:

$$d_{n-1}=\overline{N_{n-1}^{NZ}\oplus NB}, \quad (50)$$

де $N_{n-1}^{NZ}=ZA_{n-1}\oplus N_{n-1}$ – ознака знакового біту коду від'ємного нуля в полі часткового залишку діленого $A_{n-1}^T(2n,1)$.

При змінах індексів вирази (48-50) вказують на правила формування усіх наступних цифрових розрядів додатної частки.

Кінцеве значення залишку діленого формується у відповідності до значення часткового залишку діленого $A_1^T(2n,1)$, який, згідно з (5), при $N_1\neq NA$ коригується наступним чином:

$$C^T(n,1)=\begin{cases}A_1^T(2n,n+1),if\ N_1^{NZ}=1;\\ A_0^T(2n,n+1),if\ N_1^{NZ}=0,\end{cases}$$

де $C^T(n,1)$ – цілочисловий залишок діленого, $N_1^{NZ}=ZA_1\oplus N_1$ – ознака знакового біту коду від'ємного нуля в полі залишку діленого $A_1^T(2n,1)$, ZA_1 – ознака про наявність залишку діленого A_1, що дорівнює нулеві, N_1 – бінарне значення у знаковому біті доповняльного коду залишку діленого.

$$A_1^T(2n,1)=[N_1]A_1^T(2n-1,1),$$

$$A_1^T(2n,n+1)=A_1^T(2n,1)\cdot 2^{-n},$$

$$A_0^T(2n,n+1)=\left(A_1^T(2n,n+1)+B^T(n,1)\right)_{mV}.$$

Висновок

Запропонований формалізований опис операції ділення цілих операндів, що базується на використанні зміщеного кодування і виконанні дій за пріоритетним використанням модулярності даних, показав методологічну повноту, чітку математичну послідовність висновків та коректність отриманих результатів.

Оригінальний метод комп'ютерного подання цілих чисел у цифровій формі передбачає кодування додатних та від'ємних цілих чисел відповідно до однакової математичної залежності, фіксацію переповнення розрядної сітки частки одночасно з обчисленням знакового розряду. Перетворення поточних кодів нульових значень залишків діленого у відповідності до формату негативного нуля створює умови уніфікації виразу обчислення бінарних значень цифрових розрядів частки при будь-яких співвідношеннях знаків діленого і дільника. За відсутності переповнення знаковий та цифрові розряди доповняльного коду частки обчислюються однаково.

Значущість та перспективність результатів дослідження, що проведено, грунтується на доведених математичних відношеннях, нескладності запропонованого алгоритму операції ділення та можливості створення відповідних конкурентоспроможних пристроїв ділення цілих чисел в доповняльних кодах.

Список літератури

1. Intel 64 and IA-32 Architectures Software Developer's Manual. Volume 2. Order Number: 325383-078US, December 2022. URL: http://www.intel.com (дата звернення: 25.12.2022).
2. Patankar U., Koel A. Review of Basic Classes of Dividers Based on Division Algorithm. Thomas Johann

Seebeck Department of Electronics, Tallinn University of Technology, 19086 Tallinn, Estonia. January, 2021. Digital Object Identifier 10.1109/ACCESS.2021.3055735.
3. Самощенко О. В., Гусєв Б. С., Лапко В. В. Математичний опис операції ділення цілих чисел. Наукові праці Донецького національного технічного університету. Серія "Інформатика, кібернетика та обчислювальна техніка". 2023. Випуск 2(35), 2022. Випуск 1(36), 2023. С. 85-96.
4. Самощенко О. В., Лапко В. В. Контроль переповнення при додаванні та відніманні цілих операндів в системі доповняльних кодів. Наукові праці Донецького національного технічного університету. Серія "Інформатика, кібернетика та обчислювальна техніка". 2020. Випуск 2(31). С. 30-38, DOI: 10.31474/1996-1588-2020-2-31-30-38.
5. Samoshchenko O., Marhiiev H., Miroshkin O. Comparison of Floating-Point Numbers Absolute Values when Representation of the Exponents with Sign-Inverse Two's Complementary Code. 2020 IEEE 2nd International Conference on Advanced Trends in Information Theory (ATIT), Kyiv, Ukraine, 2020. P. 23-28. URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9349292&isnumber=9349252.
6. Ugurdag H.F., Dinechin F., Gener Y.S., Gören S., Didier L.-S. Hardware division by small integer constants. IEEE Transactions on Computers, Vol. 66, No.12, December 2017. P. 2097-2110.
7. Ebergen J., Jamadagni N. Radix-2 Division Algorithms with an Over-Redundant Digit Set. IEEE Transactions on computers. Vol. 64, No. 9, September 2015. P. 2652-2663.
8. Subha S. A Modified Synthetic Division Algorithm. International Journal of Computational and Applied Mathematics. ISSN 1819-4966. Volume 12, Number 3 (2017). P. 691-697.
9. Narendra K., ShabirAhmed B.J., Swaroop Kumar K., Asha G.H. FPGA Implementation of Fixed point Integer Divider Using Iterative Array Structure. International Journal of Engineering and Technical Research (IJETR). ISSN: 2321-0869. Volume-3, Issue-4, April 2015. P. 170-179.
10. Matthews E., Lu A., Fang Z., Shannon L. Rethinking Integer Divider Design for FPGA-Based Soft-Processors. 2019 IEEE 27th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM). Electronic ISSN: 2576-2621. URL: https://ieeexplore.ieee.org/document/8735506.

References

1. Intel 64 and IA-32 Architectures Software Developer's Manual. Volume 2. Order Number: 325383-078US, December 2022, available at: http://www.intel.com.
2. Patankar, U., Koel, A. (2021), "Review of Basic Classes of Dividers Based on Division Algorithm", *Thomas Johann Seebeck Department of Electronics, Tallinn University of Technology,* 19086 Tallinn, Estonia. Digital Object Identifier 10.1109/ACCESS.2021.3055735.
3. Samoshchenko, O. V., Gusev, B. S., Lapko, V. V. (2023), "Mathematical description the divide of integer numbers", [Matematichniy opis operatsiy dilennya tsilykh chisel], *Naukovi pratsi Donets'koho natsional'noho tekhnichnoho universytetu. Seriya "Informatyka, Kybernetyka ta obchysliuvalna tekhnika",* No. 2(35)-1(36), pp. 85-96.
4. Samoshchenko, O. V., Lapko, V. V. (2020), "Overflow verification of integer addition and subtraction in the two`s- complement code", [Kontrol perepovnennya pri dodavanni ta vidnimanni tsilykh operandiv v systemi dopovnyalnikh kodiv], *Naukovi pratsi Donets'koho natsional'noho tekhnichnoho universytetu. Seriya "Informatyka, Kybernetyka ta obchysliuvalna tekhnika",* No. 2(31), pp. 30-38.
5. Samoshchenko, O., Marhiiev, H., Miroshkin, O. (2020), "Comparison of Floating-Point Numbers Absolute Values when Representation of the Exponents with Sign-Inverse Two's Complementary Code," *2020 IEEE 2nd International Conference on Advanced Trends in Information Theory (ATIT), Kyiv, Ukraine,* 2020, pp. 23-28, available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9349292&isnumber=9349252.
6. Ugurdag, H.F., Dinechin, F., Gener, Y.S., Gören, S., Didier, L.-S. (2017), "Hardware division by small integer constants", *IEEE Transactions on Computers, Vol. 66, No.12,* pp. 2097-2110.
7. Ebergen, J., Jamadagni, N. (2015), "Radix-2 Division Algorithms with an Over-Redundant Digit Set", *IEEE Transactions on computers,* Vol. 64, No. 9, pp. 2652-2663.
8. Subha, S. (2017), "A Modified Synthetic Division Algorithm", *International Journal of Computational and Applied Mathematics*. ISSN 1819-4966 Volume 12, Number 3, pp. 691-697.
9. Narendra, K., ShabirAhmed, B.J., Swaroop Kumar, K, Asha, G.H. (2015), "FPGA Implementation of Fixed point Integer Divider Using Iterative Array Structure", *International Journal of Engineering and Technical Research (IJETR).* ISSN: 2321-0869, Volume-3, Issue-4, pp. 170-179.
10. Matthews, E., Lu, A., Fang, Z., Shannon, L. (2019), "Rethinking Integer Divider Design for FPGA-Based Soft-Processors", *2019 IEEE 27th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM).* Electronic ISSN: 2576-2621, available at: https://ieeexplore.ieee.org/document/8735506.

Надійшла до редакції 12.03.2023

[1]O. V. SAMOSHCHENKO, [2]B. S. GUSEV, V. V. LAPKO
[1]Taras Shevchenko National University of Kyiv, Kyiv, Ukraine
[2]National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine
aleksandr.samoshchenko@gmail.com, gusevbs@gmail.com

DIVISION OF INTEGERS IN COMPLEMENTARY CODE WITH SIGN-CHANGING ZERO REMAINDER

The complementary codes of the signed integer operands of the division operation are indicated by a single formula for positive and negative numbers and are calculated as a remainder modulo the number of binary sets of the complementary code of the number. The task of dividing integers is reduced to the operation of dividing polynomials of complementary codes. For the mathematical description of the complementary code of integers, the form of a shifted and modulo-truncated special code is introduced, which is adequately indicated by a polynomial of positive and negative numbers.

The functional dependence of the definition area of the additional code and the bit rate of the corresponding polynomial is established. In the complementary code system, an algorithm for calculating a special polynomial fraction, invariant to the signs of the operands, is proposed. Analytically substantiated original methods of fixing the overflow of the bit grid of the fraction when dividing integers using the output carry and the sign of the zero value of the partial remainder of the divided while simultaneously obtaining the value of the most significant bit of the fraction.

In order to unify the algorithm for calculating the binary values of the digital digits of the quotient, zero partial remainders of the divisor when determining the digits of the quotient and calculating the next partial remainder of the quotient are proposed to be converted into a code of zero according to the format of a positive or negative binary code depending on the sign of the quotient. The correctness of the application of the transformed zero code when calculating the binary values of the digits of the quotient and the remainder is proven. A proven algorithm for calculating the additional code of the fraction based on the value of the polynomial of the fraction. The proposed algorithm for calculating the correct complementary code of the remainder divided by all ratios of the signs of the operands. The rules for the formation of the remainder in accordance with the division sign are specified and mathematically confirmed.

Key words: complementary code, negative zero, positive zero, module of the complementary code, remainder after the module, overflow

УДК 004.056.63

А.О. Нікітенко, аспірант
Донецький національний технічний університет, м. Луцьк, Україна
andrii.nikitenko@donntu.edu.ua

Системи виявлення мережевих вторгнень на основі нейронних мереж глибокого навчання

Статтю присвячено огляду систем виявлення мережевих вторгнень на основі нейронних мереж глибокого навчання та шляхів їх створення. Проведено огляд предметної області, визначено класифікацію кіберзагроз та користь систем виявлення мережевих вторгнень при запобіганні основним типам кіберзагроз. Розглянуто публічні набори даних для створення систем виявлення мережевих вторгнень та визначені їх характеристики. Проаналізовано підходи до створення систем виявлення мережевих вторгнень з використанням рекурентних нейронних мереж, в тому числі, з використанням вентильного рекурентного вузла в комбінації з механізмом уваги. Визначено основні завдання для подальших досліджень.

Ключові слова: нейронні мережі, системи виявлення мережевих вторгнень, кібербезпека, публічні набори даних, механізм уваги

DOI: 10.31474/1996-1588-2023-2-37-15-21

Вступ

З відкриттям інтернету та поширенням цифрових технологій кібербезпека стала однією з найважливіших та найбільш актуальних проблем нашого часу. В мережі, що ніколи не спить, вороги постійно шукають шляхи проникнення в системи та мережі, зламуючи паролі та використовуючи різноманітні атаки для отримання конфіденційної інформації. Захист від їх загроз вимагає надзвичайної уваги та інноваційних рішень.

Однак, разом зі зростанням кількості загроз, розвиваються і засоби для їх виявлення та запобігання, тому однією з найважливіших складових сучасної кібербезпеки є системи виявлення мережевих вторгнень (NIDS – Network Intrusion Detection System). Ці системи відіграють важливу роль у спробі виявити та відсіяти аномальну активність в мережах, що може свідчити про можливий вторгнення чи атаку.

За останнє десятиліття, системи виявлення мережевих вторгнень зазнали значних трансформацій завдяки впровадженню глибокого навчання та нейронних мереж. Ці інноваційні технології стали ключовими інструментами у підвищенні ефективності таких систем та забезпеченні надійного захисту від великої кількості загроз, навіть тих, що розвиваються з експоненційною швидкістю.

При розробці ефективних і гнучких NIDS для невідомих майбутніх атак виникають, в основному, дві проблеми. По-перше, складно правильно вибрати ознаки з набору даних мережевого трафіку для виявлення аномалій. Ознаки, вибрані для одного класу атак, можуть не працювати для інших категорій атак через постійні зміни та еволюцію сценаріїв атак. По-друге, недоступність маркованого трафіку з реальних мереж для розробки NIDS. Потрібні величезні зусилля, щоб створити такий маркований набір даних з необробленого мережевого трафіку, зібраного за певний період або в режимі реального часу. Крім того, щоб зберегти конфіденційність внутрішньої організаційної структури мережі, а також конфіденційність різних користувачів, адміністратори мереж неохоче повідомляють про будь-які вторгнення, які могли статися в їхніх мережах [1].

Метою роботи є розгляд ролі нейронних мереж глибокого навчання в сучасних NIDS та докладний аналіз публічних наборів даних для тренування систем виявлення мережевих вторгнень для підвищення точності та швидкості виявлення загроз та дослідження теперішніх рішень які використовують глибоке навчання (DL – deep learning) для створення NIDS.

Огляд предметної області

Під кібератакою можна розуміти будь-які дії, спрямовані на порушення конфіденційності, цілісності або доступності інформаційних систем, даних або мереж. До них належать використання шкідливого програмного забезпечення, шкідливих вебсайтів, атаки із застосуванням соціальної інженерії та інші методи.

Існують такі основні види кібертатак [2]: атаки з використанням ресурсів, компрометація доступу користувача, порушення кореневого доступу, компрометація вебдоступу, шкідливі атаки, відмова в обслуговуванні. Нижче на

рисунку 1 приведено класифікацію основних кіберзагроз.

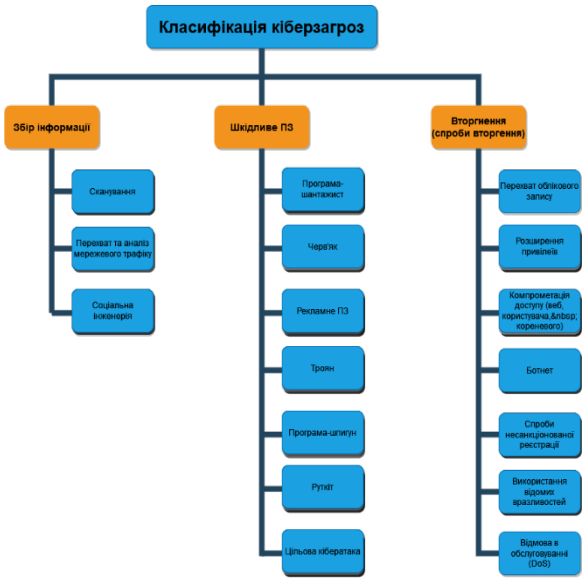


Рисунок 1 – Класифікація кіберзагроз
Джерело: Авторський рисунок

Під NIDS можна розуміти систему або інструмент, який працює з мережею і відстежує трафік на предмет підозрілої або незвичайної шкідливої активності або порушень політик і повідомляє про це адміністратору за допомогою системи управління інформацією про безпеку і подіями [4]. Проникнення в мережу зловмисників може призвести до втрати потенційно важливої інформації, витоку даних і втрати довіри користувачів. Загалом NIDS можна розділити на три категорії: IDS на основі хоста (HIDS – Host-Based IDS), NIDS та гібридна IDS.

NIDS може бути корисною для виявлення і реагування на перераховані види кібератак:

1) атаки з використанням ресурсів: NIDS може виявити аномальну активність в мережі, таку як перехоплення сеансів або надмірну кількість запитів між довіреними клієнтами і серверами, вона може реагувати на такі атаки і сповіщати адміністраторів про можливі порушення конфіденційності та цілісності;

2) компрометація доступу користувача: NIDS може виявити спроби несанкціонованого доступу до облікових записів користувачів шляхом аналізу змін у звичайних моделях поведінки користувачів та сповіщати адміністраторів про ці спроби;

3) порушення кореневого доступу: NIDS може виявляти спроби отримання несанкціонованого доступу до облікового запису адміністратора шляхом аналізу несправедливих дій в системі та надсилання сповіщень;

4) компрометація вебдоступу: NIDS може виявити спроби атак на вебсайти, такі як SQL-ін'єкції або міжсайтовий скриптинг, і сповіщати адміністраторів про ці аномалії в трафіку;

5) шкідливі атаки: NIDS може виявити надходження шкідливого програмного забезпечення до системи на основі аналізу сигнатур і аномальної поведінки, тим самим запобігаючи атакам від програм, таких як віруси, черв'яки, троянські коні і шпигунські програми;

6) відмова в обслуговуванні: NIDS може виявити аномальну активність, яка вказує на атаки відмови в обслуговуванні, такі як масові запити або перевантаження мережі і серверів, і приймати заходи для зменшення їх впливу на доступність системи.

Отже, NIDS може допомогти запобігти багатьом типам кібератак, надаючи раннє виявлення та сповіщення про можливі загрози, що дозволяє адміністраторам приймати відповідні заходи для захисту інформаційних систем, даних і мереж.

Набори даних для створення NIDS

Ефективність системи виявлення мережевих вторгнень значною мірою залежить від якості та кількості даних, які використовуються для навчання та оцінки системи. З розвитком машинного навчання та нейронних мереж набори даних стали ключовим компонентом при побудові точних та надійних систем виявлення вторгнень. Обраний набір даних використовується не тільки для навчання моделі IDS, але й для оцінки ефективності запропонованої IDS-моделі [5].

Оскільки важко збирати дані про атаки в реальному часі безпосередньо, дослідники можуть використовувати загальнодоступні стандартні набори даних на основі мережевого трафіку, такі як DARPA 1998, KDD Cup 99, KYOTO, UNSW-NB15, DEFCON, 105 CAIDA, CDX, NSL-KDD, TWENTE, CSC DoS, CICIDS2017, CSE-CIC-IDS2018, ISCX, ADFA2013, CTU-13 тощо.

Для подальших досліджень обрано такі набори даних: NSL-KDD, CICIDS2017 та UNSW-NB15. Ці три набори даних широко використовуються в сучасних моделях, завдяки чому експериментальні результати краще порівнюються з найсучаснішими моделями.

Набір даних NSL-KDD [12] широко використовується в експериментах з виявлення вторгнень. Більшість дослідників використовують набір даних NSL-KDD як еталонний набір даних. NSL-KDD не тільки ефективно розв`язує проблему передискретизації, притаманну набору даних KDD Cup 1999. Категорії набору даних зроблено більш збалансованими шляхом їх розумного налаштування.

Таким чином, модель класифікатора трафіку не зміщена в бік більш часто використовуваних категорій. Набір даних NSL-KDD включає навчальний набір (KDDTrain+) та тестовий набір (KDDTest+). Ці набори даних

фіксують нормальний трафік і чотири типи атак: DoS, Probe, R2L, U2R.

Таблиця 1 – Розподіл типів атак у наборі даних NSL-KDD

Назва атаки	Частота
Normal	77054
DoS	53385
Probe	14077
R2L	3749
U2R	252

Як видно з таблиці 1, набір даних містить мітки даних трафіку категорій нормального трафіку та чотирьох типів трафіку атак відповідно: DoS (атака на відмову в обслуговуванні), R2L (несанкціонований доступ з віддаленої машини), U2R (несанкціонований доступ до локальних привілеїв суперкористувача (root)) та Probe (спостереження та інші дослідження). Після того, як кожен трафік чисельно охарактеризовано, отримано власний вектор трафіку. Всього існує 41 ознака, включаючи базові ознаки, ознаки вмісту та ознаки комунікації. У тестовому наборі є деякі унікальні типи атак, але в навчальному наборі їх немає, тому модель краще відображає реальну здатність моделі виявляти зловмисний трафік на цьому тестовому наборі [6].

Набір даних CIC-IDS-2017 [13] містить дані про загальний трафік атаки, саме в реальному фоновому трафіку (звичайному трафіку) для запуску змодельованої хакерської атаки, а також через монітор для збору мережевого трафіку даних. Цей набір даних охоплює дуже широкий спектр трафіку. Наприклад, він містить повну топологію мережі, включаючи модеми, брандмауери, комутатори, маршрутизатори, різні операційні системи (Windows, Ubuntu і Mac OS) і різноманітні атаки, ймовірно, включаючи веб-атаки, злом грубою силою, DoS, DDoS, звичайні атаки на проникнення, серцеві кровотечі (heartbleed), ботнети, сканування мережі [6]. Крім того, тип даних трафіку атак відкалібрований відповідно до атак у кожному періоді, як показано в таблиці 2, де відображено розподіл різних зразків атак у наборі даних, оскільки нормальний потік більший за вибірку атакуючого трафіку. Тому для забезпечення узагальнюючої здатності моделі необхідне балансування даних. Набори даних CIC -IDS2017 і CSE-CIC-IDS2018 перетворюють дані трафіку в числову векторну інформацію шляхом обробки функцій, і характеристики трафіку можуть досягати 79 елементів. Це більше, ніж кількість ознак NSL-KDD, що дозволяє підвищити точність моделі виявлення зловмисного трафіку.

Набір даних UNSW-NB15 [14] створений в лабораторії кіберполігону UNSW для створення гібриду реальної сучасної звичайної діяльності та синтетичної сучасної поведінки.

Сирі мережеві пакети з набору даних UNSW-NB15 були створені за допомогою інструменту IXIA PerfectStorm в Лабораторії кіберполігону UNSW в Канберрі для створення гібриду реальної сучасної звичайної діяльності та синтетичної сучасної поведінки атаки.

Таблиця 2 – Розподіл типів атак у наборі даних CIC-IDS2017

Назва атаки	Частота	Назва атаки	Частота
SSH-Patator	5897	Web Attack and Brute Force	1507
FTP-Patator	7938	Web Attack and XSS	652
DoS slowloris	5796	Web Attack and Sql Injection	21
DoS GoldenEye	10293	Infiltration	36
Heartbleed	11	Bot	1966
DoS Slowhttptest	5499	PortScan	158930
DoS Hulk	231073	DDoS	128027

Інструмент tcpdump був використаний для захоплення 100 ГБ необробленого трафіку (наприклад, Pcap-файлів). Цей набір даних містить дев'ять типів атак, а саме: Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode та Worms [7]. Розподіл атак наведено в таблиці 3. Використовуються інструменти Argus, Bro-IDS та дванадцять алгоритмів, розроблених для генерації 49 ознак з міткою класу.

Таблиця 3 – Розподіл типів атак у наборі даних UNSW-NB15

Назва атаки	Частота
Normal	2,218,761
Fuzzers	24,246
Analysis	2,677
Backdoors	2,329
DoS	16,353
Exploits	44,525
Generic	215,481
Reconnaissance	13,987
Shellcode	1,511
Worms	174

Глибоке навчання при створенні NIDS

Машинне навчання відіграє важливу роль у забезпеченні кібербезпеки, надаючи потужні інструменти для виявлення і запобігання кібератакам. Однак слід зазначити, що машинне навчання потребує суттєвої роботи експертів

(наприклад, інженерів-функціоналістів), які можуть виконати важливе завдання з визначення релевантних ознак даних перед запуском алгоритмів поверхневого навчання. DL базується на багаторівневому представленні вхідних даних і може автономно визначати ознаки за допомогою певного процесу навчання на основі представлення [3]. Здатність DL систем аналізувати величезні обсяги даних і виявляти приховані патерни робить його невід'ємним компонентом сучасних систем безпеки. Таким чином, машинне навчання є фундаментальною частиною кіберзахисту, але DL піднімає цю сферу на новий рівень, забезпечуючи вищий рівень захисту і реакції на загрози. Тому нейронні мережі глибокого навчання будуть і надалі розглядатися як один із сучасних та ефективних методів протидії кіберзагрозам.

На сьогодні в дослідженнях [8], [9], [10], в яких отримані високі показники точності моделей, за основу береться рекурентна нейронна мережа (RNN – Recurrent Neural Network) через свою здатність до виділення часових характеристик з вхідних даних. Циклічна структура RNN може зберігати історичну інформацію та надавати можливості моделювання послідовностей. В момент часу t мережевий шар приймає вхідні дані x_t поточної часової мітки та прихований стан попередньої часової мітки $h_t - 1$, тому поточний стан h_t може бути визначений наступним чином:

$$h_t = \sigma(W_{xh}x_t + W_{hh}h_{t-1} + b_h), \quad (1)$$

де $W_{xh}x_t$ та W_{hh} - ваги матриць;
b – зміщення;
σ – функція активації.

Однак рекурентні нейронні мережі схильні до градієнтного зникнення та градієнтного вибуху, тому для запобігання цьому зазвичай обирають архітектуру нейронної мережі типу довга короткочасна пам`ять (LSTM – Long Short Term Memory) або вентильний рекурентний вузол (GRU – Gated Recurrent Unit). Отримано експериментальні результати [11] які показали, що GRU більше підходить для створення NIDS, ніж LSTM. В структурі GRU, (рисунок 2) існує два типи вузлів: вузол скидання r_t (reset gate) та вузол уточнення z_t (update gate). Вони працюють разом, щоб визначити процес оновлення інформації.

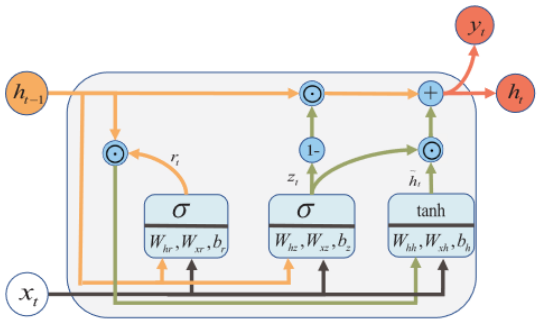


Рисунок 2 – Структура GRU [10]

Припустимо, що поточний вхід x_t, а новий стан h_t через час t складається з двох частин: стану-кандидата $\tilde{h}_t$ і минулий стан h_{t-1}.

$$h_t = (1 - z_t)h_{t-1} + z_t\tilde{h}_t. \quad (2)$$

Вузол скидання r_t працює у процесі отримання стану-кандидата. Спосіб отримання стану-кандидата подібний до традиційного RNN, за винятком механізму вузлів.

$$\tilde{h}_t = \tanh(x_tW_{xh} + W_{hh}(r_i \odot h_{t-1}) + b_h), \quad (3)$$

де $\odot$ позначає добуток Адамара;
W - вагова матриця;
b - зміщення.

Тут r_t допомагає контролювати, скільки інформації з минулого стану може бути додано до стану-кандидата. Вузол скидання r_t оновлюється наступним чином:

$$r_t = \sigma(x_tW_{xr} + h_{t-1}W_{hr} + b_r). \quad (4)$$

Відповідно до рівняння (4) для отримання нового стану h_t, вузол уточнення z_t відіграє роль для збалансування попереднього стану h_{t-1} і поточного стану-кандидата $\tilde{h}_t$. z_t можна розглядати як вузол для розподілу минулої інформації та нової інформації. Оновлення z_t відбувається подібно до оновлення r_t :

$$z_t = \sigma(x_tW_{xz} + h_{t-1}W_{hz} + b_z). \quad (5)$$

Експерименти [11] показали, що BiGRU та BiLSTM працюють ефективніше ніж GRU та LSTM при створенні систем виявлення мережевих вторгнень, в цьому випадку обрана двонаправлений GRU. Двонаправлений вентильний рекурентний вузол (BiGRU – Bidirectional Gated Recurrent Unit) – це вдосконалена версія GRU, яка працює в двох напрямках. Вона підсумовує пряму інформацію $\vec{h}$ і зворотну інформацію $\overleftarrow{h}$ для покращення можливостей вилучення ознак.

$$\vec{h}_t = \overrightarrow{GRU}(x_t), t \in [1, T],$$
$$\overleftarrow{h}_t = \overleftarrow{GRU}(x_t), t \in [1, T]. \quad (6)$$

Все більше у дослідженнях набирає обертів використання різноманітних механізмів уваги в моделях NIDS. У дослідженні [8] запропоновано гібридну модель виявлення вторгнень на основі BiGRU з механізмом уваги. Застосовано стековий розріджений автокодувальник (SSAE – Stacked Sparse Autoencoder) для вивчення високорівневого представлення вихідних даних трафіку, а витягнуті низьковимірні ознаки використовувались як вхідні дані для класифікатора.

Результати експерименту показують, що використання SSAE та механізму уваги на основі зрізів може значно прискорити процес класифікації. Запропонована модель може ефективно виявляти мережеві вторгнення, точність

досягає понад 98,68%, а рівень помилкових тривог нижче 1,32%, що перевершує попередні відносні методи. Для експериментів використано набір даних UNSW-NB15.

В роботі [9] запропоновано двошарову двонаправлену мережу довготривалої короткочасної пам'яті з механізмом уваги для розрізнення даних трафіку, враховуючи, що дані трафіку залежать від часу. Вона досягає найвищої точності 99,05%, найвищої швидкості виявлення 99,36% і найвищого показника f1 99,154% на наборі даних UNSW-NB15. У кожному випадку запропонована методика перевершує традиційні класифікатори машинного навчання. Однак, через високу обчислювальну вартість складних структур DNN, вони мережі не навчалися на інших еталонних наборах даних IDS в цьому дослідженні.

Обидва дослідження використовують так звану увагу на основі зрізів (Slice-Based Attention). На основі їх попередніх досліджень виявлено, що дані про трафік залежать від часу. Інформація про трафік за кілька суміжних моментів часу дуже корисна для визначення поточного типу трафіку. Тому запропоновано об'єднати кілька фрагментів даних про трафік у вигляді зрізу трафіку. Для того, щоб зменшити обсяг обчислень під час оптимізованого множення матриць, в цьому процесі використовується скалярнодобуткова увага (dot-product attention), яка зображена на рисунку 3.

В дослідженні [10] представлена модель виявлення вторгнень з ієрархічним механізмом уваги. Кілька даних про трафік об'єднано в певному порядку, а також досліджено вплив різної кількості попереднього трафіку на продуктивність системи. Було продемонстровано, що запропонована модель досягла задовільної продуктивності на наборі даних UNSW-NB15, з точністю понад 98,76%. В роботі використана комбінація уваги на основі зрізів та уваги на основі особливостей (Feature-Based Attention), яка зосереджена на тому, що не всі функції мають однакову важливість для представлення інформації про окремий трафік. Таким чином, щоб повністю вивільнити потенціал деяких ознак і зафіксувати ознаки, які є дійсно важливими для представлення трафіку, використовується механізм уваги, щоб визначити, на якій ознаці слід зосередитися. Крім того, механізм уваги на основі місцезнаходження не має додаткових об'єктів інтересу і є релевантним лише для кожного входу в самому джерелі даних. Таким чином, він більш підходить для роботи з вхідними характеристиками.

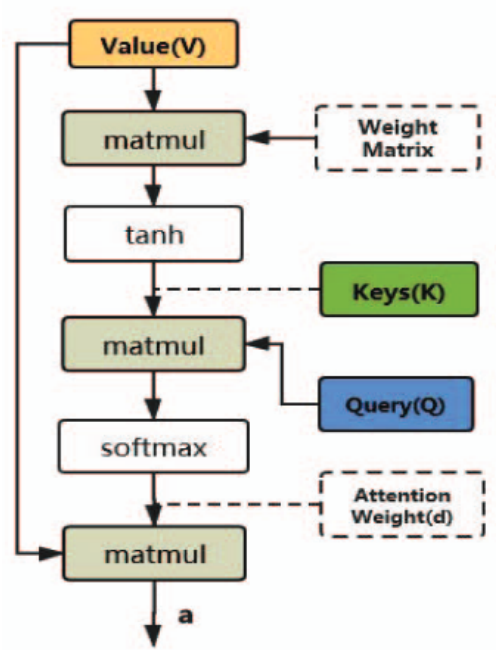


Рисунок 3 – Ілюстрація скалярнодобуткової уваги [8]

Висновки

В рамках дослідження проведено огляд предметної області, визначено та розглянуто основні публічні набори даних для створення NIDS та їх характеристики. Підкреслено, що на сьогодні машинне навчання є фундаментальною частиною кіберзахисту, при цьому DL підіймає цю сферу на новий рівень, забезпечуючи вищий рівень захисту і реакції на загрози. Тому потрібно і на далі розвивати напрям з використанням глибокого навчання. Розглянуто новітні підходи до створення NIDS з використанням рекурентних нейронних мереж (RNN), а конкретно вентильний рекурентний вузол (GRU) в комбінації з механізмом уваги.

В подальших дослідженнях планується удосконалення архітектури для NIDS з використанням рекурентних нейронних мереж та розгляд можливості використання механізму уваги. Одночасно планується дослідити можливість використання автокодувальників для попереднього вилучення ознак з наборів даних для тренування.

Список літератури

1. A Deep Learning Approach for Network Intrusion Detection System / A. Javaid et al. 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS), New York City, United States, 3–5 December 2015. 2016. URL: https://doi.org/10.4108/eai.3-12-2015.2262516 (дата звернення: 18.09.2023).

2. Нікітенко А.О. Використання нейронних мереж глибокого навчання в галузі кібербезпеки – Луцьк, Київ: Збірник матеріалів Восьмої міжнародної науково-технічної конференції "Моделювання і комп'ютерна графіка", 2023. С. 17-23.
3. On the effectiveness of machine and deep learning for cyber security / G. Apruzzese et al. 2018 10th International Conference on Cyber Conflict (CyCon), Tallinn, 29 May – 1 June 2018. 2018. URL: https://doi.org/10.23919/cycon.2018.8405026 (дата звернення: 18.09.2023).
4. Verma J., Bhandari A., Singh G. Review of existing data sets for network intrusion detection system. Advances in Mathematics: Scientific Journal. 2020. Vol. 9, no. 6. P. 3849–3854. URL: https://doi.org/10.37418/amsj.9.6.64 (дата звернення: 18.09.2023).
5. Нікітенко А.О. Набори даних для створення систем виявлення мережевих вторгнень з використанням нейронних мереж. Київ: Збірник матеріалів XLI науково-технічної конференції молодих вчених та спеціалістів Інституту проблем моделювання в Енергетиці ім. Г.Є. Пухова НАН України, 2023. С. 104–106.
6. Liu X., Liu J. Malicious traffic detection combined deep neural network with hierarchical attention mechanism. Scientific Reports. 2021. Vol. 11, no. 1. URL: https://doi.org/10.1038/s41598-021-91805-z (дата звернення: 18.09.2023).
7. Moustafa N., Slay J. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). 2015 Military Communications and Information Systems Conference (MilCIS), Canberra, Australia, 10–12 November 2015. 2015. URL: https://doi.org/10.1109/milcis.2015.7348942 (дата звернення: 18.09.2023).
8. An Efficient Intrusion Detection Model Combined Bidirectional Gated Recurrent Units With Attention Mechanism / J. Wang et al. 2020 7th International Conference on Behavioural and Social Computing (BESC), Bournemouth, United Kingdom, 5–7 November 2020. 2020. URL: https://doi.org/10.1109/besc51023.2020.9348310 (дата звернення: 18.09.2023).
9. Intrusion Detection Based on Bidirectional Long Short-Term Memory with Attention Mechanism / Y. Yang et al. Computers, Materials & Continua. 2023. Vol. 74, no. 1. P. 801–815. URL: https://doi.org/10.32604/cmc.2023.031907 (дата звернення: 18.09.2023).
10. An Intrusion Detection Model With Hierarchical Attention Mechanism / C. Liu et al. *IEEE Access*. 2020. Vol. 8. P. 67542–67554. URL: https://doi.org/10.1109/access.2020.2983568 (дата звернення: 18.09.2023).
11. Network Intrusion Detection Technology Based on Convolutional Neural Network and BiGRU / B. Cao et al. Computational Intelligence and Neuroscience. 2022. Vol. 2022. P. 1–20. URL: https://doi.org/10.1155/2022/1942847 (дата звернення: 18.09.2023).
12. NSL-KDD dataset. URL: https://www.unb.ca/cic/datasets/nsl.html
13. Intrusion Detection Evaluation Dataset (CIC-IDS2017). URL: https://www.unb.ca/cic/datasets/ids-2017.html
14. The UNSW-NB15 Dataset. URL: https://research.unsw.edu.au/projects/unsw-nb15-dataset

References

1. Javaid A. et al. (2016), "A Deep Learning Approach for Network Intrusion Detection System", *9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS)*, New York City, United States, pp. 3–5, available at: https://doi.org/10.4108/eai.3-12-2015.2262516 (date of access: 18.09.2023).
2. Nikitenko, A. (2023), "The use of deep learning neural networks in the field of cybersecurity", *Lutsk, Kyiv: Proceedings of the Eighth International Scientific and Technical Conference "Modeling and Computer Graphics."* ["Vykorystannia neironnykh merezh hlybokoho navchannia v haluzi kiberbezpeky", Lutsk, Kyiv: Zbirnyk materialiv Vosmoi mizhnarodnoi naukovo-tekhnichnoi konferentsii "Modeliuvannia i komp'iuterna hrafika"], pp. 17-24.
3. Apruzzese, G., et al. (2018), "On the effectiveness of machine and deep learning for cyber security" *2018 10th International Conference on Cyber Conflict (CyCon)*, Tallinn, 29 May – 1 June 2018, available at: https://doi.org/10.23919/cycon.2018.8405026 (date of access: 18.09.2023).
4. Verma, J., Bhandari, A., Singh, G. (2020), "Review of existing data sets for network intrusion detection system", *Advances in Mathematics: Scientific Journal*. 2020. Vol. 9, no. 6, pp. 3849–3854, available at: https://doi.org/10.37418/amsj.9.6.64 (date of access: 18.09.2023).
5. Nikitenko, A. (2023), "Datasets for creating network intrusion detection systems using neural networks", *Kyiv: Proceedings of the XLI Scientific and Technical Conference of Young Scientists and Specialists of the Pukhov Institute for Modeling Problems in Energy of the National Academy of Sciences of Ukraine*, [Nabory danykh dlia stvorennia system vyiavlennia merezhevykh vtorhnen z vykorystanniam neironnykh merezh. Kyiv: Zbirnyk

materialiv XLI naukovo-tekhnichnoi konferentsii molodykh vchenykh ta spetsialistiv Instytutu problem modeliuvannia v Enerhetytsi im. H.Ye. Pukhova NAN Ukrainy], pp. 104-106.
6. Liu, X., Liu, J. (2021), "Malicious traffic detection combined deep neural network with hierarchical attention mechanism", *Scientific Reports*. Vol. 11, no. 1., available at: https://doi.org/10.1038/s41598-021-91805-z (date of access: 18.09.2023).
7. Moustafa, N., Slay, J. (2015), "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)", *2015 Military Communications and Information Systems Conference (MilCIS)*, Canberra, Australia, 10–12 November 2015, available at: https://doi.org/10.1109/milcis.2015.7348942 (date of access: 18.09.2023).
8. Wang, J., et al. (2020), "An Efficient Intrusion Detection Model Combined Bidirectional Gated Recurrent Units With Attention Mechanism" *2020 7th International Conference on Behavioural and Social Computing (BESC)*, Bournemouth, United Kingdom, 5–7 November 2020, available at: https://doi.org/10.1109/besc51023.2020.9348310 (date of access: 18.09.2023).
9. Yang, Y., et al. (2023), "Intrusion Detection Based on Bidirectional Long Short-Term Memory with Attention Mechanism", *Computers, Materials & Continua*. Vol. 74, no. 1, pp. 801–815, available at: https://doi.org/10.32604/cmc.2023.031907 (date of access: 18.09.2023).
10. Liu, C., et al. (2020), "An Intrusion Detection Model With Hierarchical Attention Mechanism", *IEEE Access*. Vol. 8. pp. 67542–67554, available at: https://doi.org/10.1109/access.2020.2983568 (date of access: 18.09.2023).
11. Cao, B., et al. (2022), "Network Intrusion Detection Technology Based on Convolutional Neural Network and BiGRU", *Computational Intelligence and Neuroscience*. Vol. 2022, pp. 1–20, available at: https://doi.org/10.1155/2022/1942847 (date of access: 18.09.2023).
12. "NSL-KDD dataset", available at: https://www.unb.ca/cic/datasets/nsl.html
13. "Intrusion Detection Evaluation Dataset (CIC-IDS2017)", available at: https://www.unb.ca/cic/datasets/ids-2017.html
14. "The UNSW-NB15 Dataset", available at: https://research.unsw.edu.au/projects/unsw-nb15-dataset

Надійшла до редакції 7.10.2023

A. NIKITENKO
Donetsk National Technical University, Lutsk, Ukraine
andrii.nikitenko@donntu.edu.ua

NETWORK INTRUSION DETECTION SYSTEMS BASED ON DEEP LEARNING NEURAL NETWORKS

The article is devoted to an overview of network intrusion detection systems based on deep learning neural networks and ways to create them. An overview of the subject area is carried out, the classification of cyber threats and the usefulness of network intrusion detection systems for preventing the main types of cyber threats are determined. Public datasets for creating network intrusion detection systems are considered and their characteristics are determined. Approaches to creating network intrusion detection systems using deep learning neural networks are analyzed.

Today, a variety of architectures are widely used to detect network intrusions: deep neural networks, autoencoders, convolutional neural networks, and recurrent neural networks. In particular, the use of recurrent neural networks, such as long short-term memory and gated recurrent unit in combination with various attention mechanisms, is relevant. In the process of writing this article, the main tasks for further research were identified.

The purpose of this paper is to study the role of deep learning neural networks in modern network intrusion detection systems and to analyze in detail public datasets for training networked intrusion detection systems to improve the accuracy and speed of threat detection, as well as to study modern solutions that use deep learning to create network intrusion detection system for real-time intrusion detection.

Keywords: neural networks, network intrusion detection systems, cybersecurity, public datasets, attention mechanism

UDC 004.89

[1]Y. E. Horichenko, PhD student,
[2]O. A. Pozdnyakov, PhD student,
[3]A. V. Tulenkov, Assistant,
[4]A. V. Parkhomenko, PhD, Associate Professor
National University "Zaporizhzhia Polytechnic", Zaporizhzhia, Ukraine
[1]gorichenko@zp.edu.ua
[2]olegpozdnyakovua@gmail.com
[3]aetulenkov@gmail.com
[4]parhom@zntu.edu.ua

Research and practical implementation of intelligent methods and models for controlling power generation from alternative energy sources

The problem of predicting power generation from alternative energy sources in HAS was investigated in the paper. Its relevance is caused by the inability to forecast and monitor power generation from alternative energy sources, which always depends on numerous factors within open-source HAS platforms. This drawback significantly limits the efficient use of clean energy resources and makes it difficult for homeowners to become independent of external power grid. It leaves them without the ability to make informed choices about their energy usage and appliance scheduling. The object of the study is a process of predicting power generation for alternative energy sources within HAS. The objective of the study is to develop a predictive model capable of forecasting power generation from alternative energy sources within HAS. The constructed model is designed to predict power generation using historical weather and power generation data. The structure of the LSTM model involves two sequential LSTM layers, followed by fully connected layers and two hidden fully connected layers, activated by ReLU activation function. An experimental investigation was conducted using a dataset collected from a real HAS. The dataset covers the period from August 9, 2022, to July 18, 2023, and includes a total of 8256 instances. The developed LSTM model demonstrated performance with an RMSE of 287.67 watts and a coefficient of determination (R^2) of 0.74. There is a notable drop from the initial 466,15 watts to 287.67 watts after the implementation of a new predictive model, which indicates a good accuracy enhancement of approximately 38.3%. The results obtained are based on the creation and fine-tuning of a specialized LSTM model, which effectively predicts power generation. These findings from the investigation suggest integration of the proposed predictive model into real HAS that depends on alternative energy sources. Such integration has the potential to optimize energy usage efficiency while reducing reliance on traditional power grids.

Key words: power generation, forecasting, alternative energy sources, home automation system, long short-term memory, OpenHAB

DOI: 10.31474/1996-1588-2023-2-37-22-33

Introduction

In today's world people are becoming increasingly concerned about the environment and looking for sustainable ways to produce energy. Renewable energy sources have become a ray of hope for a greener future.

More and more countries are realizing the huge potential of renewable energy sources. These sources, like solar, wind, hydro and geothermal energy, have many benefits compared to the old-fashioned methods that rely on fossil fuels. One of the main reasons renewable energy is so attractive is because it is kind to the environment. Unlike traditional ways of making energy that create a lot of air pollution from burning fossil fuels, renewable sources don't produce much or any harmful gases that contribute to climate change. This helps to fight against global warming and it also means cleaner air for everyone to breathe [1].

The incorporation of alternative energy sources into home automation systems (HAS) has become a significant trend that is changing how energy is produced and used in households. By integrating clean energy technologies as part of HAS, two key advantages are realized: the utilization of clean energy and the attainment of independence from the external power grid. However, this decentralized approach to power generation necessitates a comprehensive strategy for effectively predicting and monitoring the power generation processes of these renewable sources [2].

Today, a variety of HAS platforms exists, including popular open-source software such as Open Home Automation Bus (OpenHAB), Home Assistant, Domoticz, OpenMotics and HomeGenie. These

platforms are designed to simplify home automation, allowing homeowners to manage and monitor various Internet of things (IoT) devices [3]. However, a common drawback is the inability to predict and track power generation from alternative energy sources which depends on various factors such as weather conditions and the time of day. This underscores the relevance of the study.

The object of the study is the process of predicting power generation for alternative energy sources within HAS.

In today's digital age, machine learning (ML) methods have emerged as powerful tools for predicting data across various facets of life. However, it is important to note that the effectiveness of each ML method is not universal and depends on the specific characteristics of the input data and the results of experimental analyses. This study is dedicated to selecting and adapting an appropriate ML method aimed at accurately predicting power generation within HAS. The development and integration of a predictive model into HAS will play a crucial role in the future by enabling homeowners to make well-informed choices about their energy usage and creating an optimal schedule for electric appliances. As a result, it will promote energy independence and maximize the utilization of clean energy sources.

The subject of the study are ML methods aimed at predicting power generation based on weather forecasts. These ML methods act as a bridge between meteorological data and energy output prediction, facilitating an understanding of how external environmental variables influence power generation.

The objective of the study is to develop a predictive model capable of forecasting power generation from alternative energy sources within HAS.

Problem statement

The real HAS [4] was investigated, which integrates solar panels and wind turbines as sources for producing clean energy. This system has OpenHAB installed as home automation software, which allows homeowners to control different IoT devices and create automation rules for them. However, a key limitation of this platform lies in its inability to predict and monitor power generation from alternative energy sources which depends on various factors. This drawback highlights the necessity for the development of a predictive model that can provide accurate forecasts of power generation based on weather conditions. Integrating such a development into the real HAS would enable to estimate the available amount of electricity for future use and efficiently optimize the use of electrical appliances from alternative sources, thereby reducing reliance on electricity from the external power grid.

In the study [4] a novel strategy for managing energy consumption has been developed, focusing on an intelligent methodology. This approach comprises two main steps: predicting power generation using weather forecasts and optimizing electrical appliance schedules to maximize the utilization of clean energy, while also considering the residents' habits. Different types of artificial intelligence architectures, including feedforward neural network (FNN), recurrent neural networks (RNN) and neural networks with long short-term memory (LSTM), were developed to predict power generation. The FNN proved to be the most successful, producing results with a root mean square error (RMSE) of 466.15 watts, which is not sufficiently accurate for predicting power generation.

Enhancing the accuracy of power generation forecasts requires the development of a new predictive model using advanced ML methods.

Review of the literature

Over the past few years, numerous ML methods have been investigated. To develop a predictive model according to the objective of the study, it is necessary to begin by analyzing existing solutions.

In study [5], researchers explored gated recurrent units (GRU) and LSTM neural networks for predicting hourly solar radiation, comparing their performance with and without additional weather variables. The dataset encompassed two years' worth of hourly solar radiation data along with corresponding weather data. The findings demonstrated that incorporating weather variables and cloud cover information into both the GRU and LSTM models led to significant improvements in prediction accuracy. The LSTM model exhibited a slightly lower mean absolute percentage error (MAPE) of 23.79%, whereas the GRU model demonstrated a MAPE of 25.67%. Notably, the GRU model demonstrated faster learning and prediction capabilities compared to the LSTM model. To conclude, the study indicated that the difference in prediction accuracy between the GRU and LSTM models was not particularly substantial.

In the research [6] the focus lies on using LSTM, GRU and Drop-GRU neural networks to predict short-term power consumption. The Drop-GRU model includes inserting a dropout layer between consecutive GRU layers. This process effectively reduces the complexity of the neural network and helps prevent overfitting. The results of the study highlight that the Drop-GRU neural network stands out by being more efficient and delivering better performance than both the GRU and LSTM models. The key performance metrics such as RMSE and mean squared error (MAE) exhibit low values, while the coefficient of determination (R^2) is

almost perfect (near +1) in comparison to other models.

The study described in [7] focuses on creating a neural network model to predict short-term household power consumption. To do this, the researchers used data from smart meters in individual households, which recorded the amount of electricity used in kWh for each hour of the day. They built and compared three types of models: LSTM, GRU and autoregressive integrated moving average (ARIMA). Interestingly, LSTM and GRU showed very similar performance, with LSTM performing slightly better according to metrics like MSE, RMSE, R^2, MAE and MAPE. An intriguing finding was that the GRU model achieved similar accuracy with fewer training epochs, indicating its greater efficiency in terms of memory and training requirements. On the other hand, the ARIMA model, which is more oriented towards statistics, performed less accurately than both LSTM and GRU models, which was to be expected given that ARIMA is better suited for longer-term forecasts rather than short-term ones.

The paper [8] investigates power consumption prediction using three approaches: support vector machine (SVM), multilayer perceptron (MLP) based artificial neural network (ANN) and random forest (RF). Data collected from a Supervisory Control and Data Acquisition System spanning January to December 2017, with no missing values, informs the study. The dataset of 52417 instances encompasses attributes like date, time, temperature, humidity, wind speed, general diffuse flows, diffuse flows and power consumption. Among the methods, RF emerges as the most effective, showcasing lower errors (MAE: 0.2161, MSE: 0.0930, RMSE: 0.3049) and the highest R^2 (0.9069). RF also excels in efficiency, requiring less time for training and testing than SVM and ANN. The study confirms RF's prowess in providing accurate real-time power consumption predictions, supported by robust metrics and prediction plots.

The study [9] aimed to improve the accuracy of short-term wind power generation forecasts in Turkish wind farms by comparing five regression ML methods: RF, gradient boosting machine (GBM), k-nearest neighbors (kNN), decision tree (DT) and Extra Trees regression. These methods were tested using historical wind speed and direction data collected from January 2018 to December 2018 at 10-minute intervals. The results highlighted the effectiveness of RF, kNN, GBM and Extra Trees regression in enhancing short-term wind power predictions. Notably, the GBM-based ensemble algorithm demonstrated the best performance with a MAE value of 0.0277, MAPE value of 0.3310, RMSE value of 0.0672, MSE value of 0.0045 and an R^2 value of 0.9651. On the other hand, the DT algorithm performed less satisfactorily, with a MAE of 0.0336, MAPE of 0.3309, RMSE of 0.0884 and MSE of 0.0078.

The studies conducted have explored various approaches for forecasting power generation and consumption with a focus on enhancing accuracy. The GRU and LSTM models have emerged as standout performers, especially when combined with dropout techniques [5, 6, 7]. Furthermore, other ML methods such as RF, GBM, kNN and Extra Trees regression have demonstrated their efficiency in achieving precise predictions [8, 9].

The success of these methods forms the basis for the next study, which aims to develop, compare and select the best predictive model with the highest level of accuracy in power generation prediction.

Materials and methods

In the field of predicting power generation, it is crucial to carry out a series of computer experiments and select an approach that offers the most accurate forecasts for the investigated HAS in work [4].

Before delving into forecasting, it is essential to focus on analyzing and preparing the data. By carefully studying and getting the data into the right shape, this will contribute to the creation of a more accurate predictive model. The first essential step is to conduct a comprehensive data analysis.

A software described in [4] captures and stores hourly data, including historical weather forecasts and power generation in watts, was used. The dataset does not contain any missing values and covers the period from August 9, 2022, to July 18, 2023, and includes a total of 8256 instances. Within this dataset, various attributes are included, such as date, time, temperature, thermal index, dew point, humidity, precipitation, precipitation chance, precipitation type, snow amount, snow depth, wind gust, wind speed, wind direction, sea level pressure, cloud cover, visibility, solar radiation, solar energy, ultraviolet index, severe risk, weather conditions and the amount of power generated by solar panels and wind turbines.

Figure 1 shows hourly power generation in watts over a day, while Figure 2 displays power generation over the entire period.

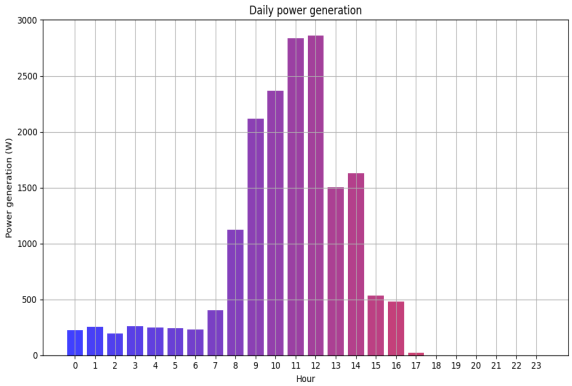


Figure 1 – Hourly power generation throughout a day

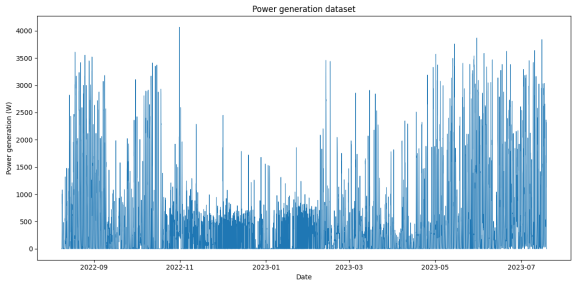


Figure 2 – Power generation throughout the entire period

The dataset has been preprocessed to create five new attributes: hour, quarter of the year, day of the month, day of the year and day of the week. Figure 3 depicts the correlation between various pairs of attributes including these additional features. As observed from the figure, attributes such as solar radiation, ultraviolet index, solar energy, temperature, thermal index (feels like), wind gust, wind speed, dew point and hour show moderate correlation with power generation while the remaining attributes exhibit only a minimal correlation with power generation.

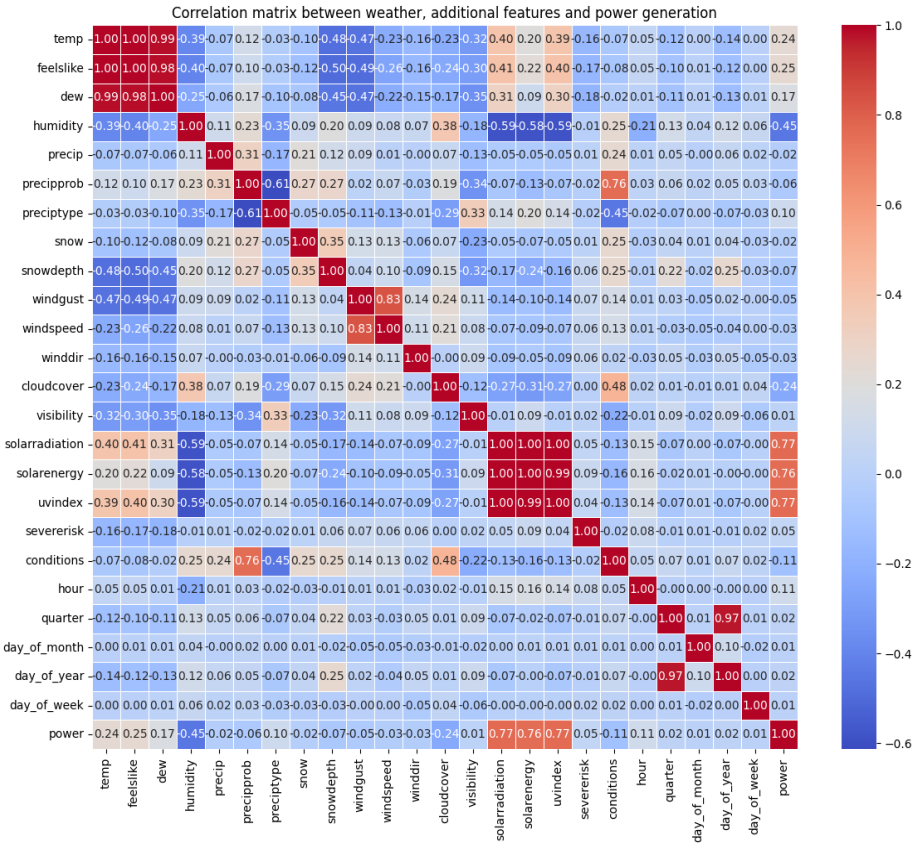


Figure 3 – Correlation matrix for attribute pairs in the dataset with additional features

Given the high correlation between temperature and the thermal index (feels like), as well as the strong connection among solar radiation, solar energy and the UV index, it becomes apparent that including all of these factors as features in the prediction model is unnecessary. The presence of highly correlated features within a dataset can lead to multicollinearity, which, in turn, has the potential to disrupt the model's stability and interpretability. To address this, it is important to adopt a strategic approach in feature selection, considering both practical relevance and domain knowledge. In this study, a set of features was chosen for the future prediction model: temperature, dew point, wind gust, solar radiation and hour of the day. These chosen features exhibit good correlations with power generation among the available options, making them the most suitable candidates for inclusion.

Following data analysis, the second crucial step is data pre-processing.

The dataset comprises features of diverse data types, highlighting the need for data normalization. In order to address this variance and bring coherence to the dataset, Min-max normalization was chosen as the normalization technique, ensuring that the data is brought onto a common scale without sacrificing information. The Min-max normalization formula, defined as:

$$\bar{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

where x represents the actual data, $x_{\min}$ is the minimum value of the dataset and $x_{\max}$ signifies the maximum value, ensures that the resulting normalized

value $\bar{x}$ remains confined within the inclusive range of [0, 1]. By undergoing this transformation, the predictive model can harness features that have been rendered consistently scaled and comparable. This enhancement enables the model to make more precise predictions and informed decisions based on the uniformized dataset [10].

After data pre-processing, the third important step is to split the data. In this study, the dataset was divided into a substantial 80% training subset and a 20% test subset, as illustrated in Figure 4. Due to the limited size of the dataset, creating a validation subset is not appropriate here. The primary concern is that allocating samples for validation could significantly reduce the already constrained training data, impacting the robustness of the model. In this scenario, the focus is wisely directed toward maximizing the available training data to construct a stronger model.

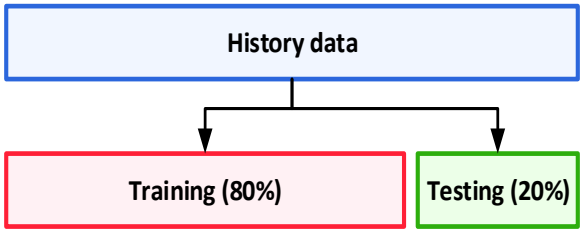


Figure 4 – Dataset splitting scheme

Experiments

This research paper delves into the field of power generation forecasting by investigating seven different approaches. They include LSTM, GRU, GRU with Dropout, as well as RF, GBM, kNN and Extra Trees regression. These methods were selected based on good performance and widespread use in previous research. Each algorithm has a unique theoretical foundation in the context of forecasting, consistently delivering promising results. Additionally, some of these algorithms have adjustable settings known as hyper-parameters, which impact factors such as speed, adaptability, stability and accuracy in predictions. To fine-tune these settings and optimize algorithm performance, a technique provided by the Python library "GridSearchCV", was employed. This approach aids in identifying the optimal configuration, resulting in heightened precision and predictive capabilities.

In this study, predictive modeling was initiated with the first implemented model, which is LSTM. It is a powerful RNN architecture known for its ability to capture intricate temporal patterns in sequential data. This versatile technology has been widely employed in various fields, including language modeling, machine translation and image captioning [11]. In the context of forecasting power generation, LSTM's adeptness at analyzing historical time series data makes it a valuable tool. By assimilating factors like weather conditions and past power generation records, LSTM can provide precise predictions, enhancing the optimization of energy production and distribution systems for greater sustainability.

The constructed LSTM model is designed to forecast power generation based on historical weather and power data. The model consists of two sequential LSTM layers followed by fully connected layers. The first LSTM layer, comprising 200 units with a ReLU activation function, takes input sequences of the weather features. The second LSTM layer, with 100 units and a ReLU activation, refines the extracted features. Further, the architecture includes two hidden fully connected layers of 50 and 25 units, respectively, also activated by ReLU. The output layer, consisting of a single unit, aims to predict power generation.

The ReLU activation function was chosen due to its ability to mitigate the vanishing gradient problem, promoting effective training. ReLU's property of allowing positive values and zeroing out negatives introduces non-linearity, crucial for capturing complex relationships in power generation prediction [12]. This makes ReLU a good choice for enhancing the LSTM model's accuracy in forecasting power generation based on historical data.

The model was trained over 100 epochs using the Adam optimizer with a learning rate of 0.001 and the MSE was utilized as the loss function. The training process involved mini-batches of size 24 to update the model's parameters efficiently.

Following LSTM, the second model introduced is the GRU. It represents a type of RNN architecture akin to the LSTM network. Engineered to capture intricate temporal patterns in sequential data, the GRU offers a simpler structure with fewer parameters, rendering it computationally efficient and faster to train. Its core distinction lies in its handling of memory cell states – replacing them with a candidate activation vector governed by reset and update gates. This architecture permits selective retention and incorporation of information over time, empowering the network to discern meaningful patterns within sequences [13].

The constructed GRU model was designed to forecast power generation, mirroring the architecture of the LSTM network for later performance comparison. The GRU model, like LSTM, incorporated a similar structure, including the number of layers, units and activation functions. The proposed model was trained over a specified number of epochs using the chosen optimizer. This similarity in architecture enabled an equitable assessment of the two models' predictive capabilities during subsequent stages of the experiment, fostering insightful performance evaluation.

To address overfitting concerns, the third model implemented is GRU with Dropout Regularization. Dropout is a regularization technique in neural networks aimed at preventing overfitting by randomly deactivating a portion of the neurons during

each training iteration. The fundamental idea behind dropout is to simulate training a large number of neural networks with different architectures by temporarily removing neurons along with their connections. This process helps to create a diverse ensemble of networks that can work together to make predictions. During training, dropout is applied to hidden units in a layer with a certain probability, effectively forcing the network to learn robust and distributed representations of the data, as neurons must work independently and adapt to various contexts. This prevents co-adaptations and overfitting, making the network more resilient and capable of generalizing well to unseen data [14].

To enhance the performance of the GRU model and mitigate overfitting, dropout layers were integrated into the architecture. These dropout layers were added after each GRU layer, employing a dropout rate of 0.2. Rigorous experimentation demonstrated that this particular dropout rate delivered optimal outcomes, effectively enhancing the model's ability to generalize and curbing overfitting tendencies. The core architecture of the model remained unaltered, encompassing two GRU layers with 200 and 100 units respectively, followed by dense layers of sizes 50, 25 and 1. The addition of dropout layers contributed to better regularization and improved the model's accuracy.

Transitioning to ensemble methods, the fourth model in the research is RF. It is an ensemble learning method widely employed for classification, regression and predictive tasks, offering enhanced accuracy and robustness through the integration of multiple decision trees. The fundamental concept revolves around constructing a "forest" of decision trees, each trained on distinct subsets of the data and features. The ensemble effect mitigates overfitting and enhances generalization by aggregating the predictions from individual trees [15].

Achieving optimal performance with RF depends on careful hyperparameter tuning. Hyperparameters are pre-learning settings that significantly influence model behavior. The RF model was constructed, employing an auto fine-tuning approach to optimize its hyperparameters for the precise prediction of power generation. This strategic tuning process ensures that the algorithm adapts to the unique characteristics of the power generation data, ultimately enhancing its performance and predictive accuracy.

The fifth model applied is GBM, a boosting algorithm known for its capacity to construct powerful predictive models by iteratively improving model weaknesses. GBM is a powerful ML algorithm that belongs to the ensemble learning family. The main idea behind GBM is to sequentially combine a set of weak learners, typically decision trees, to create a robust predictive model. It addresses both bias and variance trade-offs, making it highly effective for various applications, including power generation forecasting [16].

The process of tuning hyperparameters in GBM is of paramount importance to achieve optimal model performance. Hyperparameters control various aspects of the model's behavior, allowing customization for specific tasks. The GBM model was constructed using an automated approach to finely adjust its hyperparameters, aiming to optimize the algorithm's performance for predicting power generation.

The sixth model employed is kNN, a non-parametric algorithm that relies on the similarity of data points to make predictions. kNN is a ML algorithm used for classification, regression and clustering tasks. Its main idea is to predict the label or value of a new data point based on the majority class or average value of its k nearest neighbors in the feature space. The algorithm operates on the assumption that similar data points are likely to have similar labels or values [17].

Tuning hyperparameters is crucial to optimize the performance of the kNN algorithm, as the choice of hyperparameters significantly influences its predictive accuracy and computational efficiency. The kNN model was constructed with careful focus on optimizing algorithm performance for precise power generation prediction. The hyperparameters were fine-tuned using an automated approach. This adaptive strategy, coupled with systematic experimentation, ensured that the kNN model achieved its highest predictive potential.

Lastly, the seventh model is Extra Trees, another ensemble method similar to RF, but with additional randomness, which can lead to improved model diversity and performance. Extra Trees regression is an ensemble ML algorithm that harnesses the power of decision trees for predictive modeling. Its fundamental principle revolves around the aggregation of multiple unpruned decision trees, each grown independently from the entirety of the training dataset. Unlike traditional decision trees, Extra Trees deliberately introduces randomness at critical junctures, such as the selection of split points. This randomness aims to mitigate overfitting and enhance the diversity of the constituent trees. The algorithm's efficiency lies in forecasting complex relationships has prompted its application in various fields, including power generation forecasting [18].

Achieving optimal performance requires careful hyperparameter tuning. Each hyperparameter plays a pivotal role in influencing the behavior and generalization capability of the model. The Extra Trees model was meticulously constructed by employing an automated process to fine-tune its crucial hyperparameters. This approach aimed to meticulously optimize the algorithm's performance in predicting power generation.

Results

In the final phase of the study, the conclusions derived from evaluating various power generation predictive models are investigated. These models were developed using a suite of tools, including:

– Python 3.10.4;

– TensorFlow and Keras libraries for predictive model creation;

– NumPy, pandas and scikit-learn libraries for data manipulation and analysis;

– Matplotlib library for creating visualizations.

The computations were executed on a machine equipped with 16 GB of RAM and powered by an Intel Xeon E3-1230 v2 3.70 GHz processor.

The study included seven different predictive models: LSTM, GRU, GRU with Dropout Regularization, RF, GBM, kNN and Extra Trees regression. Each model was evaluated using a set of metrics. Ideally, the goal was to minimize the MAE, MAPE and RMSE. The coefficient of determination (R^2) was expected to approach values closer to 1, indicating a strong correlation between predicted and actual values.

MAE calculates the total of absolute differences between predicted and actual values, while MAPE offers insights into accuracy on a relative scale. RMSE is a critical metric, measures the standard deviation of prediction errors. Smaller RMSE values indicate stronger model performance, particularly when errors are squared and averaged. Among these metrics, RMSE stands out due to its sensitivity to larger errors after squaring and averaging. When RMSE is minimized, it corresponds to lower error rates in real-world applications, emphasizing the model's precision. Additionally, the closeness between RMSE and MAE values highlights consistent error magnitudes, confirming the model's effectiveness and wide applicability.

Table 1 presents a comprehensive breakdown of results, featuring metrics such as MAE, MAPE, RMSE, MSE and R^2 for both training and testing datasets used in power generation forecasting. The examination of errors within the training dataset illustrates the model's adaptability, while errors within the testing dataset reveal the model's capacity to generalize across new, unseen data. This analysis provides valuable insights into the model's suitability and its potential in real-world scenarios.

Table 1 – Model performances based on the MAE, MAPE, RMSE, MSE and R^2 metrics

Model	Performance Evaluation on Training Dataset					Performance Evaluation on Testing Dataset					Training Time (s)
	MAE	MAPE	RMSE	MSE	R^2	MAE	MAPE	RMSE	MSE	R^2	
LSTM	139.10	41.18	208.85	43619.40	0.87	178.56	56.71	287.67	82756.29	0.74	626.90
GRU	133.47	40.32	195.90	38378.53	0.88	197.83	57.83	323.05	104361.24	0.71	533.29
GRU with Dropout	179.29	53.14	256.63	65858.90	0.81	210.97	65.39	308.07	94906.90	0.66	563.32
Random Forest	156.75	46.81	227.05	51552.56	0.85	224.94	68.01	357.19	127582.54	0.62	3.76
GBM	188.89	56.41	268.32	71997.32	0.78	230.85	69.79	358.13	128258.29	0.62	1.62
kNN	0.00	0.00	0.00	0.00	1.00	227.53	68.79	353.00	124607.38	0.63	0.01
Extra Trees	154.97	46.28	222.76	49621.88	0.85	228.78	69.17	350.25	122675.27	0.63	2.07

The evaluation of various predictive models can be visualized through overlapping scatter plots, providing insight into the relationships between predicted and actual power generation. Additionally, a 24-hour chart comparing forecasted and actual power generation offers a comprehensive perspective on model accuracy. These graphical representations are demonstrated in Figures 5-11.

In Figure 5a, the scatter plot represents the correlation between predicted power generation values for the LSTM model and the corresponding actual values from the testing dataset. The plot includes a red linear regression line, visually representing the relationship between predicted and original power generation. To quantify the correlation, Pearson's correlation coefficient is also displayed within the same figure. Figure 5b extends this analysis by presenting a 24-hour comparison between measured and predicted power generation using the LSTM model during the testing phase.

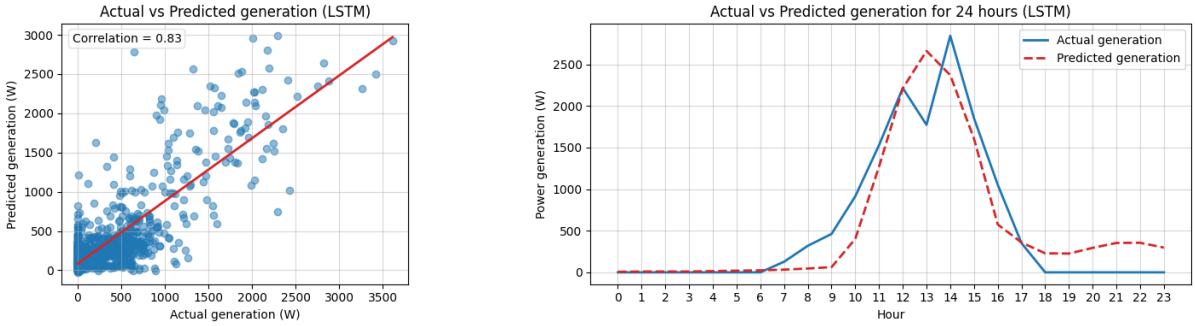
Continuing this pattern, the next figures provide similar insights into the performance of different predictive models. Each figure delves into the respective ML technique: GRU (Fig. 6), GRU with Dropout Regularization (Fig. 7), RF (Fig. 8), GBM (Fig. 9), kNN (Fig. 10) and Extra Trees regression (Fig. 11).

These graphical representations offer a clear visual interpretation of the models' abilities to capture underlying patterns and trends in power generation data, providing a valuable perspective on their effectiveness.

The analysis of Figures 5-11 provides valuable insights into the performance of various predictive models for power generation forecasting. Among these models, the LSTM model stands out with the highest accuracy in predicting power generation, achieving the highest correlation coefficient of 0.83. During the training phase, the LSTM's predictions closely align with the measured data, as can be seen on the 24-hour chart (Fig 5b). However, it is worth noting that there is a moderate spread between some predicted and actual values, as observed in the scatter plot (Fig. 5a).

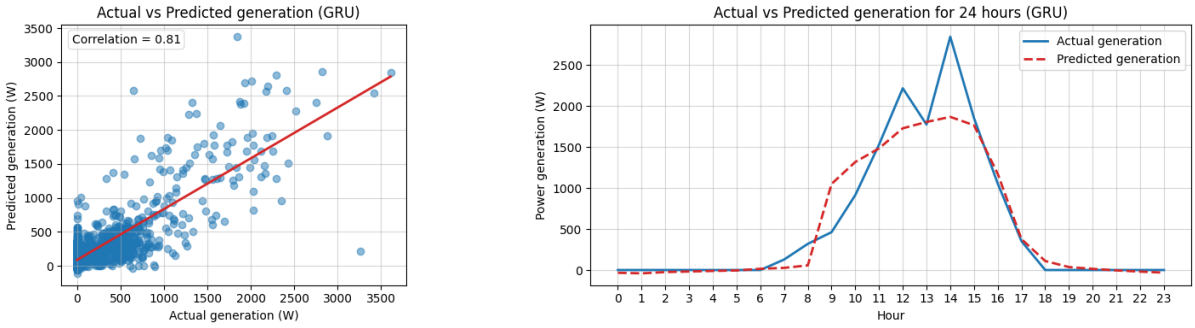
In terms of performance, both the GRU and GRU with Dropout Regularization models also demonstrate promising results, achieving correlation coefficients of 0.81 and 0.82, respectively. These models demonstrate similar predictive capabilities to the LSTM, though with a slightly decreased performance.

Conversely, the other predictive models, including RF, GBM, kNN and Extra Trees regression, show significantly lower accuracy in power generation prediction. Consequently, they may not be suitable choices for tasks involving power generation forecasting, given their low performance when compared to the LSTM, GRU and GRU with Dropout Regularization models.



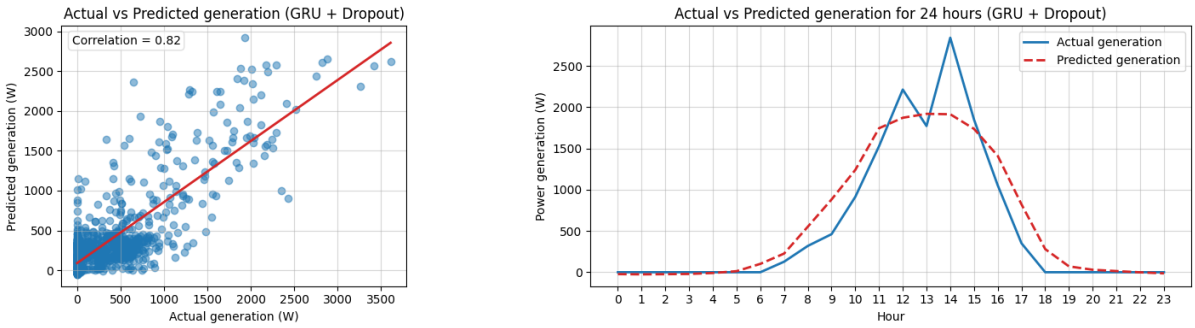
(a) (b)

Figure 5 – Scatter plot performance power generation forecasting by the LSTM (a) and a 24-hour comparison between predicted and actual power generation (b)



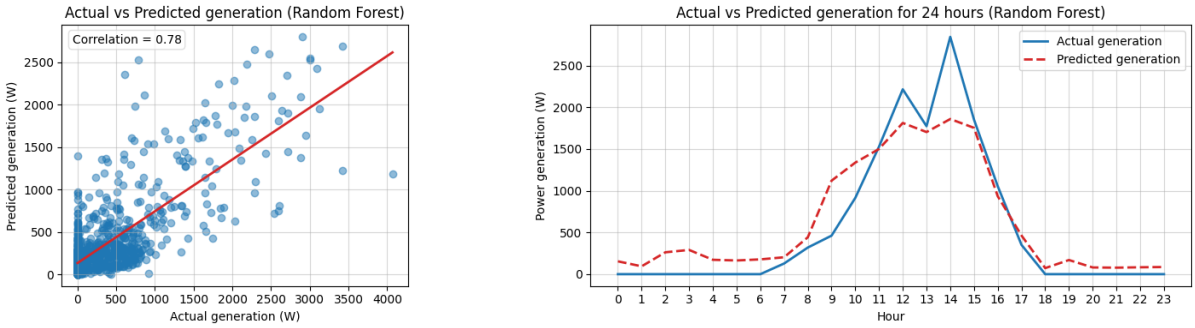
(a) (b)

Figure 6 – Scatter plot performance power generation forecasting by the GRU (a) and a 24-hour comparison between predicted and actual power generation (b)



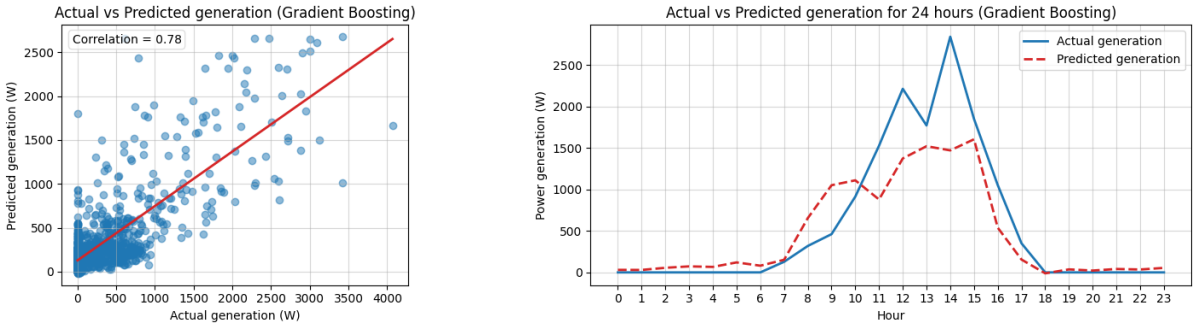
(a) (b)

Figure 7 – Scatter plot performance power generation forecasting by the GRU with Dropout Regularization (a) and a 24-hour comparison between predicted and actual power generation (b)



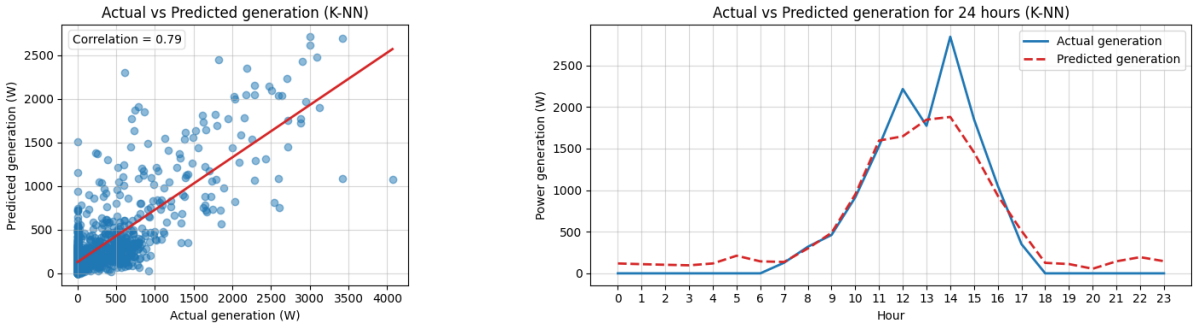
(a) (b)

Figure 8 – Scatter plot performance power generation forecasting by the RF (a) and a 24-hour comparison between predicted and actual power generation (b)



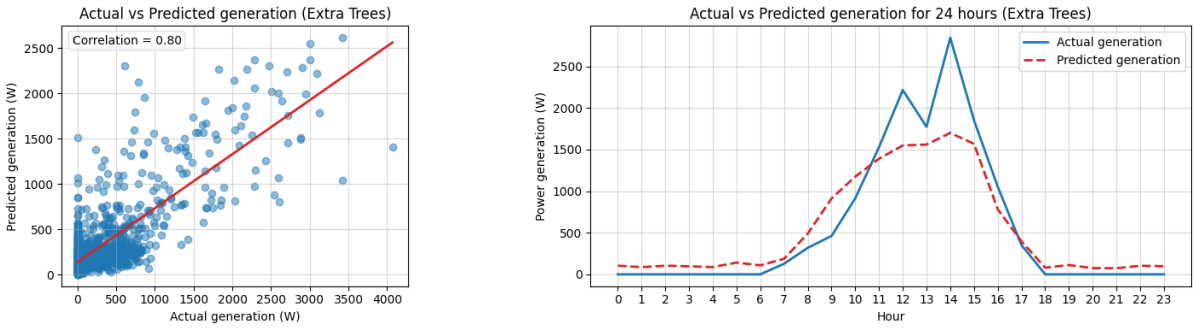
(a) (b)

Figure 9 – Scatter plot performance power generation forecasting by the GBM (a) and a 24-hour comparison between predicted and actual power generation (b)



(a) (b)

Figure 10 – Scatter plot performance power generation forecasting by the kNN (a) and a 24-hour comparison between predicted and actual power generation (b)



(a) (b)

Figure 11 – Scatter plot performance power generation forecasting by the Extra Trees regression (a) and a 24-hour comparison between predicted and actual power generation (b)

Discussion

A comprehensive study (Table 1) compared seven models for accurate power generation forecasting using a real HAS [4] dataset from August 9, 2022, to July 18, 2023.

Within the models that were implemented, several stand out for their good predictive performance in power generation forecasting. The LSTM model outperformed others, with a competitive MAE of 178.56, MAPE of 56.71%, RMSE of 287.67, and R2 of 0.74. However, its drawback is the lengthy training time of 626.90 seconds. The GRU model showed promise, achieving a MAE of 197.83, MAPE of 57.83%, RMSE of 323.05, and R2 of 0.71, with a shorter training time (533.29 seconds) than LSTM. The GRU model with Dropout Regularization had slightly worse results (MAE: 210.97, MAPE: 65.39%, RMSE: 308.07, R2: 0.66, training time: 563.32 seconds). Other methods (RF, GBM, kNN, Extra Trees regression) performed less impressively.

LSTM emerged as the most accurate predictive model, offering potential integration into a real HAS. This would provide homeowners with precise power generation forecasts, supporting informed decisions on energy consumption and optimal schedules for electrical appliances, contributing to enhanced energy independence and reduced electricity costs.

Conclusions

The problem of predicting power generation from alternative energy sources in HAS was investigated in the paper. It was considered a regression problem. The existing open-source HAS platforms mainly concentrate on simplifying home automation tasks and lack the ability to predict and manage power generation which depends on various factors such as weather conditions and the time of day. This drawback significantly limits the efficient use of clean energy resources and makes it difficult for homeowners to become independent of external power grid. To address this gap, the study suggests selecting and adapting an appropriate ML method to accurately forecast power generation within HAS.

The model predicts power generation using historical weather and power data. It comprises two LSTM layers (200 and 100 units) for processing weather features. Two fully connected layers (50 and 25 units) are included, both activated by ReLU. The output layer aims to predict power generation with a single unit. An experimental investigation was conducted using a dataset collected from a real HAS. The study compares the accuracy of seven different predictive models through experimental evaluation. The results demonstrate that the LSTM model outperforms other models with an RMSE of 287.67 and a coefficient of determination (R^2) of 0.74.

The developed model provides an interface for integration into real HAS scenarios dependent on alternative energy sources. This interface enables the prediction of power generation within the HAS, facilitating optimal energy utilization and promoting energy independence.

The scientific novelty of the obtained results lies in the implementation and adaption of a specialized LSTM model, which accurately forecasts power generation using historical weather and power data.

The practical significance of the obtained results lies in the improvement of power generation forecast accuracy. Comparing RMSE values, there is a notable drop from the initial 466,15 watts to 287.67 watts after the implementation of a new predictive model. This indicates a good accuracy enhancement of approximately 38.3%. This improvement enables the integration of the developed predictive model into real HAS that depends on alternative energy sources. Such integration has the potential to optimize energy usage efficiency while reducing reliance on traditional power grids.

The prospects for the further research involve integrating the developed predictive model into real HAS. While the LSTM model demonstrates good predictive capabilities in power generation forecasting, there is still room for model enhancement. One crucial aspect to consider is the dataset size, as a more extensive dataset could potentially enhance the model's accuracy. Furthermore, refining the architecture through optimizations like adjusting the model's depth and width holds promise for improving results.

Bibliography

1. Modern Trends in the Development of Renewable Energy: The Experience of the EU and Leading Countries of the World / Melnyk L. et al.; Mechanism of an economic regulation. 2020. Vol. 3 (89). P. 117-133. doi: 10.21272/mer.2020.89.09.
2. Yonghong Ma., Baixuan Li. Hybridized Intelligent Home Renewable Energy Management System for Smart Grids. Sustainability. 2020. Vol. 12 (5). P. 1-14. doi: 10.3390/su12052117.
3. Amos Z. The Best Open-Source Home Automation Platforms for 2023, 2023. URL: https://rehack.com/iot/smart-home/open-source-home-automation.
4. Research and Software Implementation of Intelligent Method of Energy Consumption Control / Horichenko Y. et al.; Proceedings of the 6th International Workshop on Computer Modeling and Intelligent Systems (CMIS-2023), Zaporizhzhia, Ukraine. 2023. P. 1-13.

5. Hour-Ahead Solar Irradiance Forecasting Using Multivariate Gated Recurrent Units / Wojtkiewicz J., Hosseini M., Gottumukkala R., Chambers T.; Energies. 2019. Vol. 12 (21). P. 1-13. doi: 10.3390/en12214055.
6. Predicting Energy Consumption Using LSTM, Multi-Layer GRU and Drop-GRU Neural Networks / Mahjoub S., Chrifi-Alaoui L., Marhic B., Delahoche L.; Sensors. 2022. Vol. 22 (11). P. 1-20. doi: 10.3390/s22114062.
7. Emshagin S., Halim W., Kashef, R. Short-term Prediction of Household Electricity Consumption Using Customized LSTM and GRU Models. arXiv preprint. 2022. Vol. 1. P. 1-11. doi: 10.48550/arXiv.2212.08757.
8. Zogaan W. Power Consumption prediction using Random Forest model. International Journal of Mechanical Engineering. 2022. Vol. 7. P. 329-341.
9. A Machine Learning-Based Gradient Boosting Regression Approach for Wind Power Production Forecasting: A Step towards Smart Grid Environments / Singh U., Rizwan M., Alaraj M., Alsaidan I.; Energies. 2021. Vol. 14 (16). P. 1-21. doi: 10.3390/en14165196.
10. Mejbah A. Min-Max Normalization, 2022. URL: https://www.kaggle.com/code/mejbahahammad/min-max-normalization.
11. Sugandhi A. What is Long Short-Term Memory (LSTM) – Complete Guide, 2023. URL: https://www.knowledgehut.com/blog/web-development/long-short-term-memory.
12. Brownlee J. How to Fix the Vanishing Gradients Problem Using the ReLU, 2019. URL: https://machinelearningmastery.com/how-to-fix-vanishing-gradients-using-the-rectified-linear-activation-function.
13. Nama A. Understanding Gated Recurrent Unit (GRU) in Deep Learning, 2023. URL: https://medium.com/@anishnama20/understanding-gated-recurrent-unit-gru-in-deep-learning-2e54923f3e2.
14. Adrian G. A review of Dropout as applied to RNNs, 2018. URL: https://adriangcoder.medium.com/a-review-of-dropout-as-applied-to-rnns-72e79ecd5b7b.
15. Nisha A. Tuning Random Forest Hyperparameters, 2022. URL: https://www.kdnuggets.com/2022/08/tuning-random-forest-hyperparameters.html.
16. Jain A. Complete Machine Learning Guide to Parameter Tuning in Gradient Boosting (GBM) in Python, 2016. URL: https://www.analyticsvidhya.com/blog/2016/02/complete-guide-parameter-tuning-gradient-boosting-gbm-python.
17. Joos K. The k-Nearest Neighbors (kNN) Algorithm in Python, 2021. URL: https://realpython.com/knn-python.
18. Brownlee J. How to Develop an Extra Trees Ensemble with Python, 2021. URL: https://machinelearningmastery.com/extra-trees-ensemble-with-python.

References

1. Melnyk, L., Derykolenko, O., Mazin, Yu., Matsenko, O., Piven, V. (2020), "Modern Trends in the Development of Renewable Energy: The Experience of the EU and Leading Countries of the World", *Mechanism of an economic regulation*, vol. 3 (89), pp. 117-133.
2. Yonghong, Ma., Baixuan, Li. (2020), "Hybridized Intelligent Home Renewable Energy Management System for Smart Grids", *Sustainability*, vol. 12 (5), pp. 1-14.
3. Amos, Z. (2023), "The Best Open-Source Home Automation Platforms for 2023", available at: https://rehack.com/iot/smart-home/open-source-home-automation.
4. Horichenko, Y., Parkhomenko, A., Pozdnyakov, O., Tulenkov, A., Zalyubovskiy, Y., Gladkova, O., Parkhomenko, A. (2023), "Research and Software Implementation of Intelligent Method of Energy Consumption Control", *Proceedings of the 6th International Workshop on Computer Modeling and Intelligent Systems (CMIS-2023)*, Zaporizhzhia, Ukraine, pp. 1-13.
5. Wojtkiewicz, J., Hosseini, M., Gottumukkala, R., Chambers, T. (2019), "Hour-Ahead Solar Irradiance Forecasting Using Multivariate Gated Recurrent Units", *Energies*, vol. 12 (21), pp. 1-13.
6. Mahjoub, S., Chrifi-Alaoui, L., Marhic, B., Delahoche, L. (2022), "Predicting Energy Consumption Using LSTM, Multi-Layer GRU and Drop-GRU Neural Networks", *Sensors*, vol. 22 (11), pp. 1-20.
7. Emshagin, S., Halim, W., Kashef, R. (2022), "Short-term Prediction of Household Electricity Consumption Using Customized LSTM and GRU Models", *arXiv preprint*, vol. 1, pp. 1-11.
8. Zogaan, W. (2022), "Power Consumption prediction using Random Forest model", *International Journal of Mechanical Engineering*, vol. 7, pp. 329-341.
9. Singh, U., Rizwan, M., Alaraj, M., Alsaidan, I. (2021), "A Machine Learning-Based Gradient Boosting Regression Approach for Wind Power Production Forecasting: A Step towards Smart Grid Environments", *Energies*, vol. 14 (16), pp. 1-21.
10. Mejbah, A. (2022), "Min-Max Normalization", available at: https://www.kaggle.com/code/mejbahahammad/min-max-normalization.

11. Sugandhi, A. (2023), "What is Long Short-Term Memory (LSTM) – Complete Guide", available at: https://www.knowledgehut.com/blog/web-development/long-short-term-memory.
12. Brownlee, J. (2019), "How to Fix the Vanishing Gradients Problem Using the ReLU", available at: https://machinelearningmastery.com/how-to-fix-vanishing-gradients-using-the-rectified-linear-activation-function.
13. Nama, A. (2023), "Understanding Gated Recurrent Unit (GRU) in Deep Learning", available at: https://medium.com/@anishnama20/understanding-gated-recurrent-unit-gru-in-deep-learning-2e54923f3e2.
14. Adrian, G. (2018), "A review of Dropout as applied to RNNs", available at: https://adriangcoder.medium.com/a-review-of-dropout-as-applied-to-rnns-72e79ecd5b7b.
15. Nisha, A. (2022), "Tuning Random Forest Hyperparameters", available at: https://www.kdnuggets.com/2022/08/tuning-random-forest-hyperparameters.html.
16. Jain, A. (2016), "Complete Machine Learning Guide to Parameter Tuning in Gradient Boosting (GBM) in Python", available at: https://www.analyticsvidhya.com/blog/2016/02/complete-guide-parameter-tuning-gradient-boosting-gbm-python.
17. Joos, K. (2021), "The k-Nearest Neighbors (kNN) Algorithm in Python", available at: https://realpython.com/knn-python.
18. Brownlee, J. (2021), "How to Develop an Extra Trees Ensemble with Python", available at: https://machinelearningmastery.com/extra-trees-ensemble-with-python.

Надійшла до редакції 06.12.2023

[1]Ю. Є. ГОРІЧЕНКО, [2]О. А. ПОЗДНЯКОВ, [3]А. В. ТУЛЕНКОВ, [4]А. В. ПАРХОМЕНКО
Національний університет «Запорізька політехніка», м. Запоріжжя, Україна
[1]gorichenko@zp.edu.ua
[2]olegpozdnyakovua@gmail.com
[3]aetulenkov@gmail.com
[4]parhom@zntu.edu.ua

ДОСЛІДЖЕННЯ ТА ПРАКТИЧНА РЕАЛІЗАЦІЯ ІНТЕЛЕКТУАЛЬНИХ МЕТОДІВ І МОДЕЛЕЙ КОНТРОЛЮ ГЕНЕРАЦІЇ ЕЛЕКТРОЕНЕРГІЇ З АЛЬТЕРНАТИВНИХ ДЖЕРЕЛ

У даній статті було досліджено проблему прогнозування генерації електроенергії з альтернативних джерел в системах домашньої автоматизації. Актуальність проблеми обумовлена неможливістю прогнозувати генерацію електроенергії з альтернативних джерел енергії, яка залежить від різноманітних факторів, в рамках платформ систем домашньої автоматизації з відкритим вихідним кодом. Цей недолік значно обмежує ефективне використання екологічно чистих енергетичних ресурсів домовласниками, позбавляючи їх можливості робити усвідомлений вибір щодо кількості енергоспоживання та планувати розклад використання електроприладів. Об'єктом роботи є процес прогнозування генерації електроенергії з альтернативних джерел в системах домашньої автоматизації.

Метою роботи є розробка прогностичної моделі, здатної прогнозувати генерацію електроенергії з альтернативних джерел енергії в системі домашньої автоматизації.

Розроблена нейромережева модель призначена для прогнозування генерації електроенергії, використовуючи історичні дані про прогноз погоди та генерацію електроенергії. Структура моделі LSTM включає два послідовних шари LSTM, за якими слідують повністю підключені шари, а також два приховані повністю підключені шари, що активуються функцією активації ReLU.

Експериментальне дослідження проводилося з використанням набору даних, зібраних з реальної системи домашньої автоматизації. Цей набір охоплює період з 9 серпня 2022 року по 18 липня 2023 року і складається з 8256 записів. Розроблена нейромережева модель LSTM продемонструвала продуктивність з середньоквадратичною помилкою (RMSE), значення якої становить 287,67 Вт та коефіцієнтом детермінації (R^2), що дорівнює 0,74. В результаті чого спостерігається помітне зниження середньоквадратичної помилки з 466,15 Вт до 287,67 Вт після реалізації нової нейромережевої моделі, що свідчить про підвищення точності прогнозування приблизно на 38,3%. Отримані результати дослідження базуються на створенні та налаштуванні спеціалізованої моделі LSTM, яка ефективно прогнозує генерацію електроенергії. Ці результати пропонують інтегрувати запропоновану прогностичну модель у реальну систему домашньої автоматизації, яка використовує альтернативні джерела енергії. Така інтеграція потенційно може оптимізувати ефективність енергоспоживання та знизити залежність від традиційних електромереж.

Ключові слова: генерація електроенергії, прогнозування, альтернативні джерела енергії, система домашньої автоматизації, довга короткочасна пам'ять, OpenHAB

УДК 004.021:622.412

В.І. Голінько[1], д-р техн. наук, проф.,
О.В. Голінько[2], аспірант
Національний технічний університет «Дніпровська політехніка», м. Дніпро, Україна.
[1]golinkongu@gmail.com; [2]HolinkoAV@gmail.com

Теоретико-методологічні засади комп'ютерного моніторингу систем вибухозахисту

Розглянуто напрями вдосконалення систем вибухозахисту вугільних шахт. Показано, що суттєво підвищити надійність вибухозахисту можна за рахунок доповнення існуючих систем моніторингу вибухозахисту додатковою системою комп'ютерного моніторингу засобів контролю вибухонебезпечності з функціями безперервного стеження за працездатністю засобів контролю вмісту метану в режимі реального часу, прогнозування її зміни в часі та автоматичного відновлення працездатності шляхом регулювання парламентів засобів, а також прийняття рішень стосовно видів та рівнів втручання в роботу технологічних систем на основі оцінки ризиків на всіх рівнях управління. Наведено алгоритм роботи «інтелектуального» аналізатора метану реалізованого на базі мікроконтролера ATmega8.

Ключові слова: комп'ютерний моніторинг, вибухозахист, аналізатори метану, методи, алгоритми, мікроконтролер

DOI: 10.31474/1996-1588-2023-2-37-34-41

Вступ

Моніторинг, як постійне або регулярне спостереження за яким-небудь об'єктом або процесом з метою оцінки, порівняння та прогнозування застосовується повсюдно, в різних областях, як у науці, так і в сфері управління. Будь-які управлінські інформаційні процеси включають в себе процедури реєстрації, збору, передачі, зберігання, обробки, видачі інформації та прийняття рішень на основі зібраної інформації [1]. Так існує моніторинг навколишнього середовища з метою його охорони і контролю впливу господарської діяльності на природу. Спочатку з'явився моніторинг в ґрунтознавстві, екології, а потім він поширився і в інші галузі практичної діяльності. Ведуться спостереження за динамікою стану навколишнього середовища, щоб при виявленні негативних змін вжити заходів щодо їх мінімізації або усунення.

В соціології моніторинг відстежує зміни в суспільстві, вивчає соціально-політичні явища, дає прогнози. В психології застосовується психолого-педагогічний моніторинг, його мета – контроль педагогічного впливу на всебічний розвиток дитини та попередження несприятливих факторів. У медицині відстежується загальний вплив навколишнього середовища на здоров'я людини. У сфері виробництва і господарства моніторинг також важливий. У будівництві це контроль та спостереження за будівництвом будівлі та її стану, стану ґрунтів і навколишньої забудови. При цьому використовується спеціальне обладнання, що дозволяє виявити можливі негативні тенденції і вжити заходів щодо їх виправлення на початковому етапі.

Моніторинг широко застосовується в сучасному бізнесі, зокрема, для відстеження ситуації на ринку і конкуренції. Для цього здійснюється спостереження за діями конкурентів, пропозицією продуктів, їх рекламою, відгуки покупців про товари тощо [1].

Інформаційні технології являють собою ті засоби і методи, за допомогою яких реалізуються процедури реєстрації, збору, передачі, зберігання, обробки, видачі інформації та прийняття рішень на основі зібраної інформації в різних інформаційних системах. З розвитком інформаційних технологій нині практично всі моніторингові системи будуються на базі комп'ютерів та інформаційних систем. В цьому разі моніторингові системи трансформується в комп'ютерні моніторингові системи.

Існують і моніторингові системи, які спостерігають за роботою самих мереж. Моніторингом мережі називають роботу системи, яка виконує постійне спостереження за комп'ютерною мережею у пошуках повільних чи несправних систем та яка при виявленні збоїв повідомляє про них за допомогою засобів оповіщення [2].

Моніторинг комп'ютерної мережі це безперервний процес, який здійснюється за таким принципом: спеціальні програми з певною періодичністю надсилають запити пристроям у мережі. У разі некоректного або невдалого результату такого запиту система моніторингу посилає сигнал про несправність або автоматично

проводить певні дії, регламентовані протоколом для виникнення аварійних ситуацій.

Моніторинг можна поділити на різні типи, до найпоширеніших типів моніторингу можна віднести такі як:

- моніторинг об'єкту, призначений для надання інформації, необхідної для постійного спостереження за станом об'єкта, виявлення та усунення проблем, прийняття рішень;
- моніторинг процесу, призначений для надання інформації, необхідної для постійного планування та аналізу роботу, оцінки успіху або в іншому випадку реалізації проектів та програми, виявлення та вирішення проблем і викликів, використання можливості по мірі їх виникнення;
- моніторинг результатів чи впливу, спрямований на оцінку змін, які викликані проектом чи програмою.

Основою систем комп'ютерного моніторингу являються моніторингові програми, які мають багато цілей, залежно від типу моніторингу та контексту. Однак основною метою є регулярний збір інформації про процес, а потім її використання для внесення постійних коригувань в залежності від стану системи для підтримки її ефективної працездатності.

Постановка задачі

Підземний видобуток вугілля завжди пов'язаний з комплексом шкідливих та небезпечних виробничих чинників, вплив яких зумовлений, насамперед, гірничо-геологічними умовами, використанням високопродуктивних машин і механізмів, значною інтенсивністю робіт очисних та підготовчих робіт [4]. Регулярність виникнення надзвичайних ситуацій та аварій, високий рівень травматизму обумовлюють необхідність використання прогресивних підходів до забезпечення безпеки працівників вугледобувних шахт.

Однією з розповсюджених причин виникнення аварій на вугледобувних підприємствах є вибухи метану [5]. На ризик виникнення вибухів метану на вугільних шахтах впливає безліч як зовнішніх, так і внутрішніх чинників. Перші значною мірою визначаються впливом природного середовища: гірничо-геологічними умовами видобутку вугілля, глибиною розробки, властивостями вугілля та вміщуючи порід, гірським тиском, метановістю родовищ, наявністю суфлярів та раптових викидів метану тощо. Зовнішні чинники можуть мати суттєве значення для формування безпечного стану виробничого середовища. До внутрішніх чинників на гірничих підприємствах відносяться стан виробничого обладнання, технологія видобутку, організація робіт, стан протиаварійної готовності підприємств, наявність та стан систем та засобів колективного і індивідуального захисту працюючих тощо.

Важливим заходом з попередження вибухів метану є моніторинг систем вибухозахисту, одним із основних елементів якого є система контролю концентрації метану в атмосфері гірничих виробок шахт. Останнім часом на вугільних шахтах в основному використовуються системи моніторингу вибухонебезпеки на основі комплексу аерогазового інформаційного КАГІ та уніфікованої телекомунікаційної системи диспетчерського контролю та автоматизованого керування гірничими машинами й технологічними комплексами (УТАС).

Існуючі системи моніторингу вибухонебезпеки здійснюють контроль, передачу, зберігання, видачу інформації про вміст метану в рудниковій атмосфері та видачу управляючого сигналу на відключення електроенергії, як основного джерела запалення метаноповітряної суміші.

Основними і найменш надійними елементами системи моніторингу вибухонебезпеки є газоаналізатори. Складні умови їх експлуатації (випадкові удари, заливання водою, дія екстремальних значень факторів тощо), а також несанкціоноване втручання в роботу аналізаторів можуть впливати на працездатність та стабільність їх роботи, а це, в свою чергу, може призвести до неспрацьовування захисного вимкнення електроенергії, що є однією із причин виникнення аварій на шахтах. Прикладом цього є аварія з чисельними смертельними наслідками спричинена вибухом метаноповітряної суміші, яка трапилася 2 березня 2017 року на шахті «Степова» ДП Львіввугілля, де, незважаючи на наявність досить коштовної системи УТАС, засоби вибухозахисту не забезпечили захисного відключення електричної мережі при наявності вибухонебезпечної концентрації метану, що стало однією із причин вибуху.

Важливим заходом з підвищення надійності роботи системи моніторингу є пробні перевірки її спрацювання. Згідно чинних інструкцій по експлуатації існуючих стаціонарних засобів контролю вмісту метану необхідно здійснювати щоденну перевірку їх працездатності та періодично перевіряти їх в шахтних умовах з використанням атестованих газових сумішей, що обумовлює значні експлуатаційні витрати. Встановлено [6], що щоденна перевірка справності апаратури контролю метану у порівнянні зі щомісячною перевіркою дозволяє знизити ймовірність вибуху у 300 разів, а за умови зменшення терміну між перевірками апаратури до 4 годин ймовірність вибуху за рік може бути меншою допустимого значення 10^{-4} .

Зважаючи на викладене, актуальною задачею є підвищення надійності системи моніторингу вибухозахисту шляхом впровадження сучасних інформаційних технологій та комп'ютерної техніки.

Методологічною основою системи комп'ютерного моніторингу вибухозахисту повинно стати наступне положення: в умовах мінімальних витрат на створення та експлуатацію системи комп'ютерного моніторингу вибухозахисту, забезпечити зниження ймовірності вибуху метану до допустимого значення.

Основні результати досліджень

Суттєво підвищити надійність вибухозахисту вугільних шахт можна лише за умови доповнення існуючих систем моніторингу вибухозахисту додатковою системою комп'ютерного моніторингу засобів контролю вибухонебезпечності з функціями безперервного стеження за працездатністю засобів контролю вмісту метану в режимі реального часу, прогнозування її зміни в часі та автоматичного відновлення працездатності шляхом регулювання парламентів засобів, а також прийняття рішень стосовно видів та рівнів втручання в роботу технологічних систем на основі оцінки ризиків на всіх рівнях управління.

Фактично, це передбачає створення дворівневої системи комп'ютерного моніторингу вибухозахисту, тобто доповнення існуючих систем комп'ютерного моніторингу вибухозахисту додатковою підсистемою комп'ютерного моніторингу засобів контролю вмісту метану, та суттєвого розширення функцій системи на всіх рівнях. Структура такої системи комп'ютерного моніторингу вибухозахисту наведена на рис. 1.

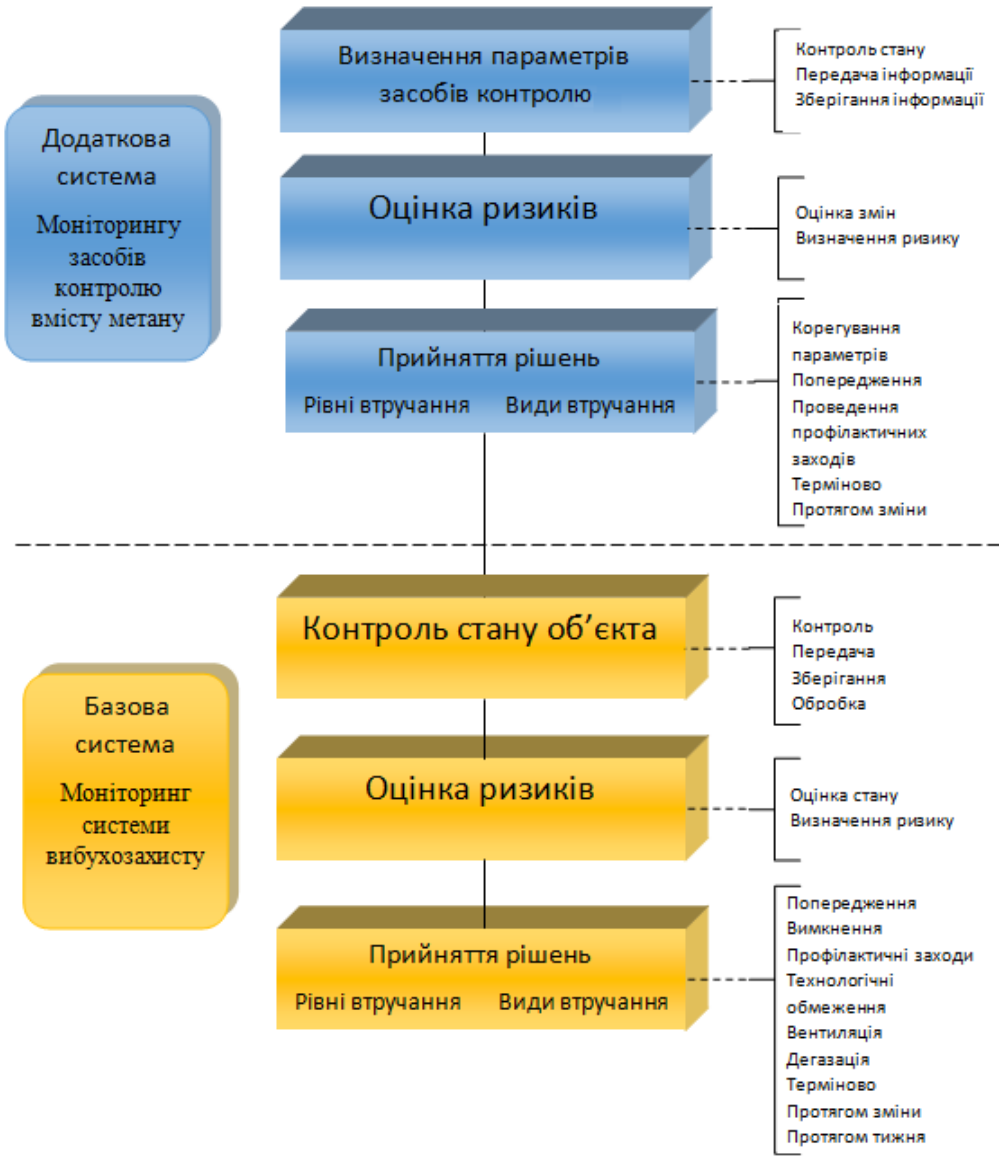


Рисунок 1 – Структура ссистеми комп'ютерного моніторингу вибухозахисту

Нині в існуючих засобах контролю вмісту метану в основному використовують термокаталітичний метод. Виконані дослідження цього методу та заснованих на ньому засобів контролю, дозволили істотно підвищити надійність та стабільність роботи газоаналізаторів [7,8]. Однак складні умови експлуатації газоаналізаторів можуть впливати на працездатність та стабільність їх роботи. Тому, згідно чинних інструкцій з експлуатації існуючих стаціонарних засобів контролю вмісту метану, необхідно здійснювати щоденну перевірку їх працездатності та періодично перевіряти їх в шахтних умовах з використанням атестованих газових сумішей, що обумовлює значні експлуатаційні витрати.

Питанням автоматичної діагностики стану аналізаторів метану приділяється значна увага. Виконано теоретичний аналіз та дослідження методу виявлення зміщення «нуля» аналізаторів метану шляхом зниження напруги живлення термогрупи до величини, при якій не протікає реакція окислення метану та аналізу зміни вольтамперних характеристик термоелементів, запропоновано алгоритм моніторингу зміщення «нуля» та відновлення працездатності аналізаторів шляхом його корегування [9].

Запропоновано метод контролю чутливості первинних перетворювачів та корегування показань аналізаторів метану, що полягає в аналізі зміни вихідних характеристики перетворювачів на дільниці виходу вихідної напруги вимірювального моста на горизонтальну частину (плато) характеристики при зміні величини струму через чутливі елементи [10].

Виконано теоретичний аналіз та дослідження методу виявлення випадків забруднення газодифузійних фільтрів датчиків метану, шляхом аналізу перехідних процесів, що виникають в датчиках при зміні режиму живлення термоелементів. Запропоновано алгоритм виявлення випадків забруднення [11].

Розроблено метод підвищення швидкодії та надійності систем автоматичного газового захисту, шляхом одночасного використання в аналізаторах метану двох датчиків, один з яких є високо стабільний але відносно інерційний - термокаталітичний, а інший - малоінерційний оптичний. Обґрунтовано метод виявлення вибухонебезпечності середовища, який полягає в тому, що за відсутності суттєвих збурень, викликаних газодинамічними явищами, інформація про концентрацію метану та формування сигналу на відключення електрообладнання здійснюється на основі вимірювання концентрації термокаталітичним датчиком, а малоінерційний оптичний датчик служить лише для визначення величини приросту концентрації метану за короткий проміжок часу при газодинамічних явищах. Запропоновано алгоритм для реалізації методу [12].

Приведені розробки дають змогу реалізувати на практиці систему комп'ютерного моніторингу засобів контролю вибухонебезпечності. Основним елементом такої системи є «інтелектуальний» аналізатор метану, виконаний на базі мікроконтролера ATmega8.

Зазвичай для об'єднання декількох алгоритмів в одному контролері застосовують методи дієвої керованості. Так, при застосуванні мікроконтролера ATmega8 для цього використовують підсистему переривань за внутрішніми запитами, що формуються відповідно до програми роботи аналізатора метану в атмосфері гірничих виробок. В такому випадку все програмне забезпечення аналізатора складається з наступних компонентів.

Налагодження (повірка) аналізатора: установка нуля, чутливості, підготовка до роботи, початкові установки; використання програмою регістрів та виконання підготовчих послідовностей операцій для калібрування.

Основний алгоритм – являє собою неперервний цикл вимірювання вихідної напруги вимірювального моста, виведення результатів вимірювання на індикатор, передачу інформації про концентрацію метану в систему автоматичної сигналізації, накопичення інформації про поточні середні значення концентрації метану і величину їх середньоквадратичного відхилення, періодичне запам'ятовування поточних середніх величин, порівняння поточних середніх величин зі значеннями цих величин в попередні періоди часу, формування запитів на діагностику стану блоків і виявлення несанкціонованого втручання в роботу датчика метану або його аварійного стану.

Алгоритм вектору переривань – від внутрішнього пристрою мікроконтролера, з якого прийшов запит на переривання основного алгоритму, виклик алгоритму обробки події і установка пріоритету в черговості обробки (в разі одночасних запитів).

В аналізаторі метану, побудованому з використанням сучасних мікропроцесорних засобів обробки інформації, в якому реалізована функція автоматичної діагностики нульових показань, чутливості та стану газодифузійного фільтру відсутні елементи для регулювання «нуля» і чутливості аналізатора. Тому аналізатор має два режими роботи: режим налагоджування і робочий режим. Всі операції налагодження і перевірки здійснюються без втручання оператора згідно розробленої програми роботи аналізатора.

В режимі налагодження виносний датчик аналізатора метану встановлюється в малогабаритну камеру, яка за командою з дисплею аналізатора, почергово заповнюється атмосферним повітрям чи атестованою газовою сумішшю з вмістом метану 1,0%.

При подачі чистого повітря аналізатором встановлюються задані величини струму через чутливі елементи, визначаються та заноситься в пам'ять аналізатора величини напруги на елементах та їх різниця, визначаються значення коефіцієнтів математичної моделі вольтамперних характеристики термоелементів датчика метану.

При подачі атестованої газової суміші до виносного датчика аналізатора визначається та встановлюється попередньо визначена для чистого повітря величина напруги на порівняльному елементі, визначається напруга на робочому елементі, розраховується чутливість датчика аналізатора у мВ/%CH_4, встановлюються початкові значення параметрів первинних перетворювачів, які використовуються при діагностиці стану аналізатора та корегуванні його показань [9-11].

У робочому режимі безперервно виконуються такі операції:

- вимірювання напруги на елементах;
- підтримка визначеної при регулюванні величини напруги на порівняльному елементі;
- визначення різниці напруги на елементах;
- розрахунок та передача інформації про вміст метану.

За зовнішнім запитом оператора чи періодично, відповідно до визначеного програмою роботи аналізатора часу та за умови стабільної й безпечної газової ситуації в місці контролю, здійснюється перевірка й коректування нульових показань аналізатора перевірка й коректування чутливості та перевірка стану газодифузійного фільтра аналізатора. При цьому послідовно реалізуються відповідні алгоритми перевірки й коректування, викладені в [9-11]

Блок-схема алгоритму роботи аналізатору наведена на рис. 2 (налагодження та основний режим) та рис.3 (перевірка і коригування параматрів аналізатора).

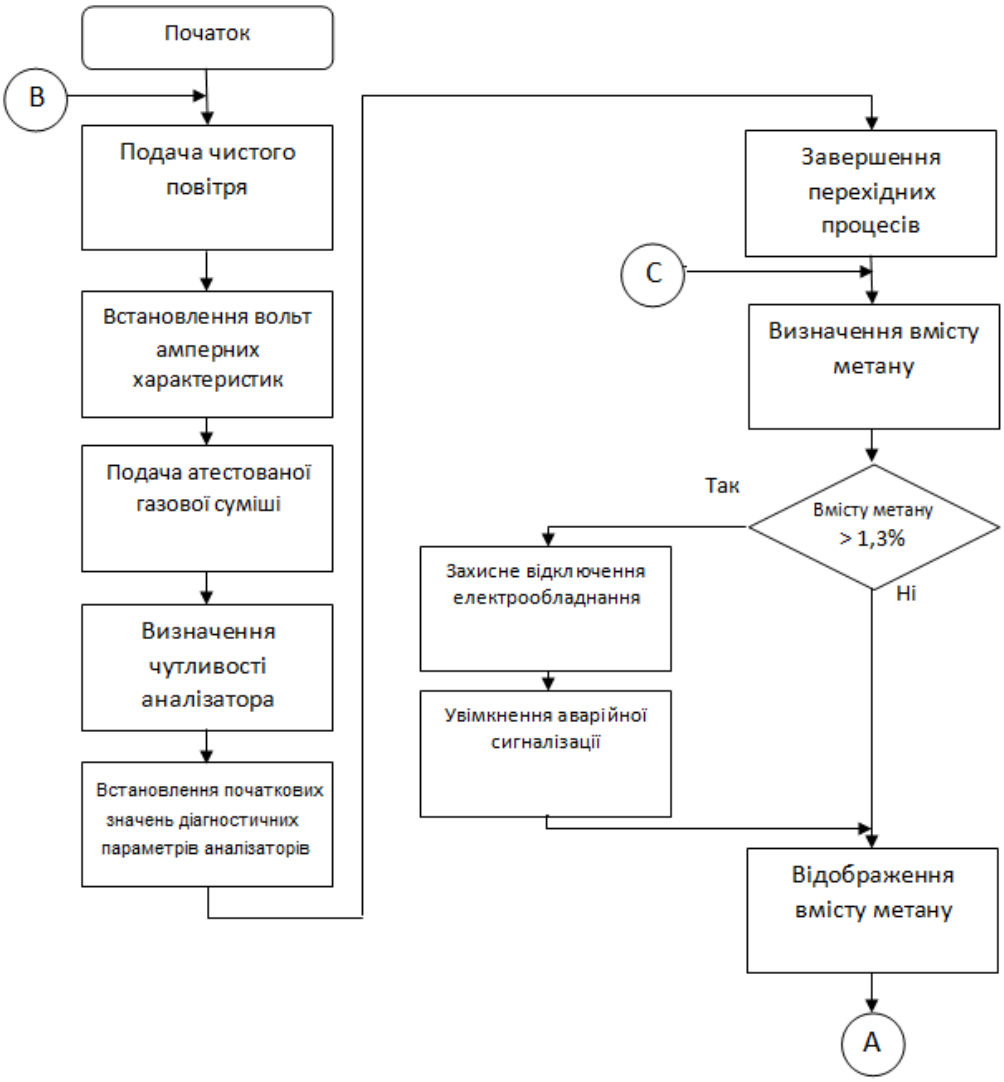


Рисунок 2 – Блок-схема алгоритму роботи програми (налагодження та основний режим)

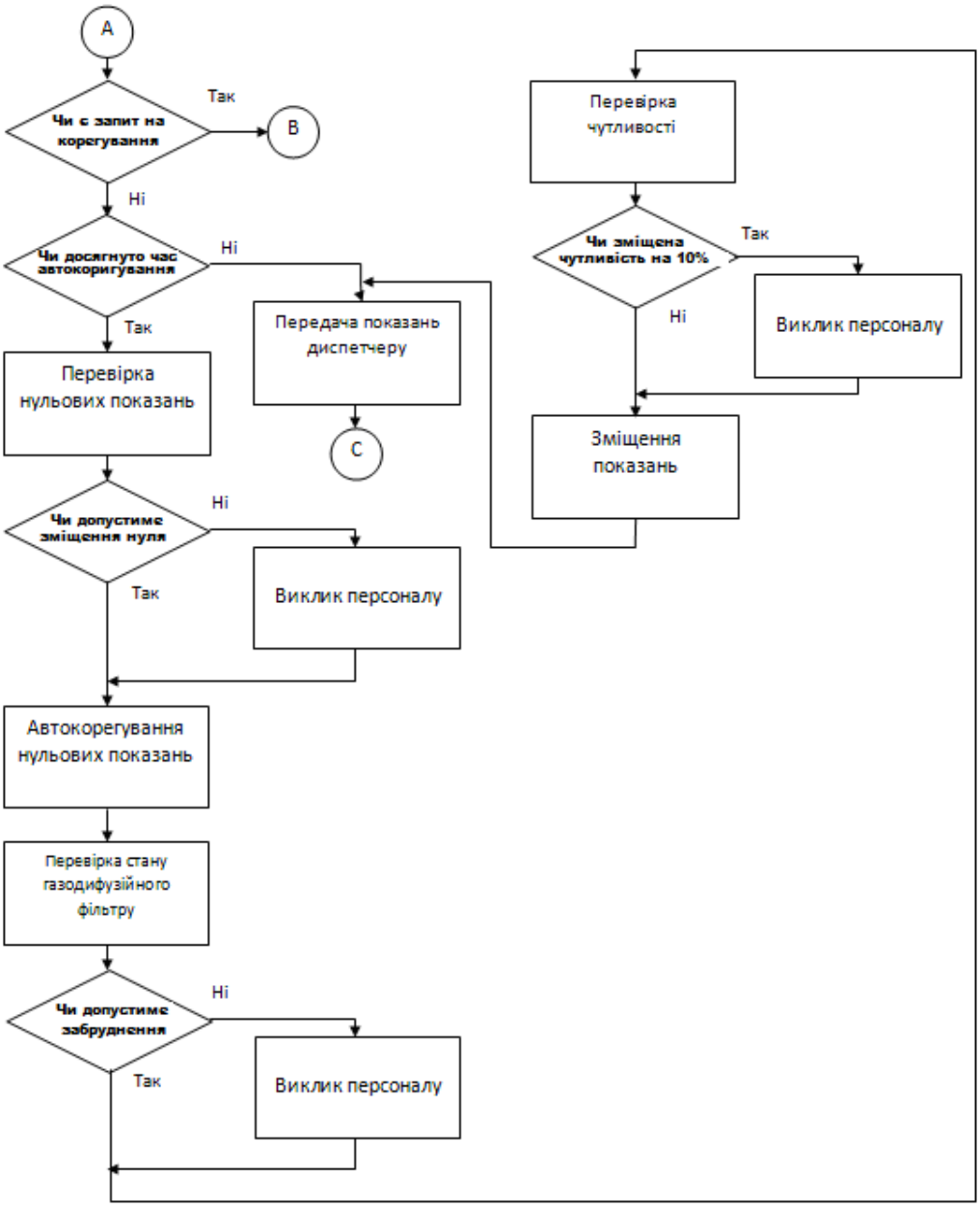


Рисунок 3 – Блок-схема алгоритму роботи програми (перевірка і коригування параметрів)

Комп'ютерний моніторинг засобів контролю вмісту метану в даній системі включає перевірку працездатності та діагностику стану засобів контролю вибухонебезпечності, передачу та зберігання інформації, оцінку змін стану засобів контролю, визначення ризику виникнення надзвичайних ситуацій, обумовлених цими змінами, прийняття рішень стосовно видів втручання в роботу системи контролю (здійснення програмного коригування результатів вимірювання, попередження про недопустимий характер змін, повідомлення про необхідність проведення профілактичних заходів) та рівнів втручання (терміново, з дозволеною затримкою в час).

Базовий рівень системи комп'ютерного моніторингу вибухозахисту крім зазначених раніше задач та функцій (контроль, передачу, зберігання, видачу інформації про вміст метану в рудниковій атмосфері та видачу управляючого сигналу на відключення електроенергії) потребує доповнення функціями оцінки ризику виникнення надзвичайних ситуацій, обумовлених газодинамічними процесами в гірничих виробках шахт, прийняття рішень стосовно видів та рівнів втручання в роботу технологічних систем, систем вентиляції, дегазації, системи контролю, введення обмежень, проведення профілактичних заходів тощо.

Висновок та напрямок подальших досліджень

Підвищити надійність систем вибухозахисту вугільних шахт можна лише за

умови доповнення існуючих систем вибухозахисту додатковою підсистемою комп'ютерного моніторингу засобів контролю вмісту метану, яка здійснює безперервне стеження за працездатністю засобів контролю вмісту метану в режимі реального часу, прогнозування її зміни в часі та автоматичне відновлення працездатності шляхом регулювання парламентів засобів, а також забезпечує прийняття рішень стосовно видів та рівнів втручання в роботу технологічних систем на основі оцінки ризиків на всіх рівнях управління.

Виконані дослідження термокаталітичного методу контролю вмісту метану, та заснованих на ньому засобів контролю, дозволили розробити методи та алгоритми автоматичної діагностики стану аналізаторів метану в частині контролю їх нульових показань, чутливості, стану газодифузійних фільтрів, що створює передумови для розробки та впровадження системи комп'ютерного моніторингу засобів контролю вмісту метану. Основним елементом такої системи є «інтелектуальний» аналізатор метану, виконаний на базі мікроконтролера ATmega8.

Використання сучасних мікропроцесорних засобів обробки інформації дає можливість не тільки автоматично перевіряти працездатність аналізаторів, а і здійснити корегування показань газоаналізатора при зміні чутливості, зміщенні нульових показань та обмеженні доступу метаноповітряної суміші до чутливих елементів.

В перспективі система комп'ютерного моніторингу вибухозахисту повинна інтегруватися в загальну для шахти систему моніторингу умов праці – як систему спостережень, збору, обробки, передачі, збереження й аналізу інформації про стан умов праці, прогнозування їхніх змін та розробки обґрунтованих рекомендацій для прийняття управлінських рішень з врахуванням ризику та соціально-економічних наслідків впливу надзвичайних ситуацій, шкідливих та небезпечних виробничих чинників на працівників та ефективність виробництва.

Список літератури

1. Simister N., Monitoring. M&E Training & Consultancy. Intrac, 2017.
2. Sonny S. Ariss, Computer monitoring: benefits and pitfalls facing management. Information & Management 39, 2002. P. 553–558.
3. Дорожинський О.Л. Геоматика в моніторингу довкілля та оцінці загрозливих ситуацій. монографія. Львів: Вид-во Львів. політехніки, 2016. 399 с.
4. Tyuleneva T., Kabanov E., Moldazhanov M., Plotnikov E. Improving the Professional Risk Management System for Methane and Coal Dust Explosions Using a Risk-based Approach. E3S Web of Conferences 278, 2021. 01027.
5. Aldhafeeri T., Tran M., Vrolyk R., Pope M., Fowler M. A Review of Methane Gas Detection Sensors: Recent Developments and Future Perspectives. Inventions 2020, 2020. Vol. 5, 28 p.
6. Колосюк А.В., Колосюк В.П. Вибухонебезпека систем електропостачання гірничих машин на шахтах, небезпечних за раптовими викидами вугілля і газу. Монографія. Кременчук: КрНУ імені Михайла Остроградського, 2017. 178 с.
7. Голінько В.І., Котляров А.К. Контроль вибухонебезпечності середовища в гірничих виробках та обладнанні вугільних шахт. Монографія / Днепр: вид-во «Ліра», 2010. 368 с.
8. Голінько В.І., Белоножко А.В., Белоножко В.В. Контроль вибухонебезпечності газових сумішей під час аварійних загазувань гірничих виробок шахт. Монографія. Днепропетровск: НГУ, 2014. 209 с.
9. Алексеєв М.О., Голінько О.В. Автоматична діагностика стану стаціонарних термокаталітичних газоаналізаторів. Збірник наукових праць НГУ. Д.: НГУ, 2018. № 53. С. 223–229
10. Алексеєв М.О., Голінько О.В. Автоматичний контроль чутливості датчиків стаціонарних термокаталітичних аналізаторів метану. Гірничий вісник. 2020. Вип. 107. С. 16–22.
11. Голінько В.І., Голінько О.В. Методи та алгоритми контролю стану газодифузійного фільтра аналізаторів метану. Наукові праці ДонНТУ. Серія: «Інформатика, кібернетика та обчислювальна техніка», №2-35, 2022-№1-36, 2023. С. 22-29.
12. Голінько В.І., Голінько О.В. Методи та алгоритми підвищення швидкодії систем автоматичного газового захисту. Наукові праці ДонНТУ. Серія: «Проблеми моделювання та автоматизації проектування», 2021. №1(17). С. 79-87.

References

1. Simister, N. (2017), "Monitoring" M&E Training & Consultancy. Intrac.
2. Sonny, S. (2002), Ariss, Computer monitoring: benefits and pitfalls facing management. Information & Management 39, pp. 553–558.
3. Dorozhynskyi, O.L. (2016), Geomatics in environmental monitoring and assessment of threatening situations. Monograph [Heomatyka v monitorynhu dovkillya ta otsintsi zahrozlyvykh sytuatsiy. Monografiya] – L′viv: Vyd-vo L′viv. politekhniky. 399 p.
4. Tyuleneva, T., Kabanov, E., Moldazhanov, M., Plotnikov, E. (2021), "Improving the Professional Risk

Management System for Methane and Coal Dust Explosions Using a Risk-based Approach" *E3S Web of Conferences 278*, 01027.
5. Aldhafeeri, T., Tran, M., Vrolyk, R., Pope, M., Fowler, M. (2020), "A Review of Methane Gas Detection Sensors: Recent Developments and Future Perspectives" Inventions 2020, 5, 28 p.
6. Kolosyuk, A.V., Kolosyuk, V.P. (2017), "Explosion hazard of power supply systems of mining machines in mines, dangerous for sudden emissions of coal and gas", *Monograph.* [Vybukhonebezpeka system elektropostachannya hirnychykh mashyn na shakhtakh, nebezpechnykh za raptovymy vykydamy vuhillya i hazu. Monohrafiya]. Kremenchuk: Mykhailo Ostrogradsky KrNU. 178 p.
7. Golinko, V.I., Kotlyarov, A.K. (2010), "Control of explosiveness of the environment in mine workings and equipment of coal mines", *Monograph.* [Kontrol' vzryvoopasnosti sredy v gornykh vyrabotkakh i oborudovanii ugol'nykh shakht. Monografiya]. Izdatel'stvovo «Lira». Dnepr, 368 p.
8. Golinko, V.I., Kotlyarov, A.K. Belonozhko, V.V. (2004), "Explosion hazard control of mine workings", *Monograph* [Kontrol vzryvoopasnosti gornykh vyrabotok shakht. Monografiya]. Nauka i obrazovaniye. Dnepr, 207 p.
9. Alekseev, M.O., Golinko, O.V. (2018), "Automatic diagnostics of the state of stationary thermocatalytic gas analyzers", [Avtomatychna diahnostyka stanu statsionarnykh termokatalitychnykh hazoanalizatoriv*]. Zbirnyk naukovykh prats' NHU. D. NHU*. No 53, pp. 223-229
10. Alekseev, M.O., Golinko, O.V. (2020), "Automatic sensitivity control of sensors of stationary thermocatalytic methane analyzers", [Avtomatychnyy kontrol′ chutlyvosti datchykiv statsionarnykh termokatalitychnykh analizatoriv metanu]. *Hirnychyy visnyk.* 107. pp. 16–22.
11. Golinko, V.I., Golinko, O.V. (2023), "Methods and algorithms for monitoring the state of the gas diffusion filter of methane analyzers", [Metody ta alhorytmy kontrolyu stanu hazodyfuziynoho fil′tra analizatoriv metanu. Naukovi pratsi DonNTU. Seriya: «Informatyka, kibernetyka ta obchyslyuval′na tekhnika] Scientific works of DonNTU. Series: "Informatics, Cybernetics and Computer Science", №2-35, 2022-№1-36, pp. 22-29.
12. Golinko, V.I., Golinko, O.V. (2021), "Methods and algorithms for increasing the speed of automatic gas protection systems", [Metody ta alhorytmy pidvyshchennya shvydkodiyi system avtomatychnoho hazovoho zakhystu. Naukovi pratsi DonNTU. Seriya: «Problemy modelyuvannya ta avtomatyzatsiyi proektuvannya»], Scientific works of DonNTU. Series: "Problems of modeling and design automation". №1 (17), pp. 79-87.

Надійшла до редакції 22.05.2023

V. HOLINKO[1], O. HOLINKO
Dnipro University of Technology, Dnipro, Ukraine
[1]golinkongu@gmail.com; [2]HolinkoAV@gmail.com

THEORETICAL AND METHODOLOGICAL PRINCIPLES OF COMPUTER MONITORING OF EXPLOSION PROTECTION SYSTEMS

The areas of improvement of the explosion protection systems of coal mines through the implementation of computer monitoring systems are considered. The following provision is adopted as the methodological basis of such systems: in conditions of minimal costs for the creation and operation of the computer explosion protection monitoring system, to ensure the reduction of the probability of a methane explosion to an acceptable value.
It is shown that it is possible to significantly increase the reliability of explosion protection by supplementing the existing explosion protection systems with an additional computer monitoring system of means of explosive control. Such a system provides automatic performance testing and diagnostics of the state of explosives control devices, information transfer and storage, assessment of changes in the state of control devices, determination of the risk of emergency situations caused by these changes, decision-making regarding types of intervention in the control system.
The basic level of the computer monitoring system of explosion protection needs to be supplemented with the functions of assessing the risk of emergency situations caused by gas-dynamic processes in the mine workings, making decisions regarding the types and levels of intervention in the work of technological equipment, ventilation and degassing systems, introducing restrictions and carrying out preventive measures.
The algorithm of operation of the computer monitoring system of means of explosiveness control is presented.

Key words: computer monitoring, explosion protection, methane analyzers, methods, algorithms, microcontroller

УДК 004.056.63

Д.Ю. Хома, аспірант,
Донецький національний технічний університет, м. Луцьк, Україна
dmytro.khoma.asp@donntu.edu.ua

Стеганографічні алгоритми та стегоаналіз на основі класичних методів і нейронних мереж

Статтю присвячено огляду алгоритмів і моделей створення стегозображень на основі класичних стеганоргафічних методів, а також методів з використанням штучного інтелекту. Проведено огляд предметної області, визначено основні метрики для порівняння стеганографічних алгоритмів. Виконано порівняльний аналіз зображень створених за допомогою різних стеганографічних методів при різному корисному навантаженні. Наведено результати роботи різних стегоаналізаторів на основі статистичних методів, і методів глибокого навчання. Визначено основні завдання для подальших досліджень.

Ключові слова: стеганографія, метрики, генеративна змагальна мережа, методи стеганографічного аналізу, нейронна мережа, стегодетектор, стійкість, подібність

DOI: 10.31474/1996-1588-2023-2-37-42-53

Вступ

В умовах постійно змінюючогося інформаційного середовища та зростаючої потреби у захисті конфіденційної інформації від сторонніх очей та зловмисників виникли дві передові технології, що відкривають нові можливості в області конфіденційності: стеганографія та генеративно-змагальні мережі (далі GAN).

Мистецтво стеганографії виникло в давні часи і служило прихованим засобом спілкування під час конфліктів і шпигунства. За століття стеганографія еволюціонувала від приховування секретних повідомлень у текстах і зображеннях до цифрового віку, де вона знаходить нові застосування в великому обсязі цифрового контенту. Залежність від цифрового зв'язку і конфіденційності даних викликала потребу в ефективних стеганографічних методах, які можуть ефективно приховувати інформацію, уникаючи виявлення.

Стеганографія зображень – це метод приховування повідомлень всередині зображень. У той час як інші методи, такі як криптографія, спрямовані на те, щоб забезпечити конфіденційність самого повідомлення, стеганографія має на меті приховати факт існування вкладеного повідомлення.

Генеративно-змагальні мережі є революційним проривом у галузі штучного інтелекту та машинного навчання. Такі мережі складаються з пари нейронних мереж, які змагаються між собою: генератор створює синтетичні дані, схожі на реальні, а дискримінатор відрізняє справжні дані від підроблених. GAN проявили виняткову майстерність у створенні реалістичних зображень, відео та аудіо, розширюючи можливості творчого створення контенту та комп'ютерного зору.

Одним із прикладів поєднання стеганографії та GAN є створення реалістичних зображень із прихованою інформацією, такою як текст чи інше зображення. Зазвичай такі зображення називають стегозображеннями.

Оцінка якості стегозображень включає в себе використання різних метрик, які враховують різні аспекти прихованої інформації та впливу стеганографічних процесів на якість зображення. Обрання метрик залежить від конкретного застосування стеганографії та вимог до якості та безпеки. Комбінація різних метрик часто використовується для отримання більш об'єктивної оцінки.

Метою роботи є огляд методів і метрик створення та порівняння стегозображень, використання загальнодоступних стего-аналізаторів на стегозображеннях, порівняльний аналіз якості і стійкості до виявлення створених стегозображень за допомогою класичних методів цифрової стеганографії зображень, а також створених за допомогою моделі штучного інтелекту на основі GAN.

Огляд предметної області

Стеганографія – це наука, що вивчає методи та способи приховування конфіденційної інформації. Її розвиток охоплює великий проміжок історії, починаючи від використання невидимих чорнил та мікроточок і закінчуючи сучасними системами вбудовування повідомлень у цифрові файли.

Сучасна стеганографія застосовується в різних галузях, включаючи введення водяних знаків у цифрові файли для захисту від

несанкціонованого копіювання, вбудовування додаткової інформації у телевізійні та радіопрограми, приховану анотацію документів у медичних закладах та маркування ДНК-послідовностей у генетиці [1]. Крім того, цифрова стеганографія використовується зловмисниками для створення прихованих каналів передачі конфіденційних даних, розповсюдження неприпустимого контенту та несанкціонованої передачі промислової і військової інформації [2].

Зображення є одним із найпопулярніших об'єктів для застосування стеганографії через їхню велику розповсюдженість і об'єм даних, який може бути використаний для приховування інформації. Стеганографія зображень використовує різноманітні методи та техніки для вбудовування конфіденційної інформації в самі зображення так, щоб це було практично непомітно для людей.

Одним із поширених методів стеганографії зображень є LSB (Least Significant Bit) метод [3]. Цей підхід використовує найменш важливі біти пікселів зображення для вбудовування додаткової інформації без значущих змін у вигляді зображення. Такі зміни непомітні для людей при звичайному перегляді зображення.

Додатково, стеганографія зображень може використовувати методи, які адаптовані до конкретних особливостей графічних файлів, таких як JPEG чи PNG. Наприклад, для формату JPEG може бути використана стеганографія на основі коефіцієнтів дискретного косинусного перетворення (DCT), щоб вбудувати інформацію в частини зображення, які менше впливають на його якість.

Застосування стеганографії зображень розповсюджується в різних сферах. Наприклад, в цифровій обробці зображень стеганографія може використовуватися для водяних знаків, які вказують на авторство чи автентичність зображення. У медичній галузі, вбудовування конфіденційної інформації у медичні зображення може допомагати забезпечити конфіденційність пацієнтів.

Такі методи застосовуються не лише до зображень, але і до інших мультимедійних файлів, таких як аудіо та відео, що дозволяє стеганографії виявити широкий спектр застосувань у цифровому світі.

Штучний інтелект (далі ШІ) в останні роки значно розширив можливості стеганографії зображень, додаючи нові рівні складності та ефективності до процесу приховування інформації [4]. Існує кілька напрямків, де ШІ і стеганографія зображень взаємодіють для створення більш ефективних та непростих методів:

1) автоматичний вибір місця для вбудовування: застосування алгоритмів машинного навчання дозволяє автоматично визначити оптимальні регіони зображення для вбудовування інформації, такі моделі можуть аналізувати структуру та властивості пікселів, шукаючи місця, де можна вставити дані без значущого порушення візуальної якості;

2) глибоке вбудовування інформації: використання глибоких нейронних мереж для створення нових методів вбудовування, які можуть ураховувати більш сложні аспекти зображення, наприклад, GAN можуть допомагати створювати стеганографічні методи, що приховують інформацію в непомітних деталях зображень;

3) використання контексту та семантики: алгоритми ШІ можуть аналізувати контекст та семантичні зв'язки на зображенні, визначаючи місця, де вбудовування буде менш помітним, наприклад, врахування особливостей обличчя на фотографії може допомагати створювати методи стеганографії, які уникають цих областей для збереження візуальної непомітності;

4) реакція на зовнішні зміни: ШІ може надавати механізми, які дозволяють адаптувати вбудовування інформації до змін у зовнішньому середовищі, наприклад, система може виявляти різні рівні компресії зображень та вибирати оптимальні методи вбудовування для кожного випадку;

5) стійкість до атак: використання ШІ дозволяє розробляти стеганографічні методи, які є стійкими до спроб виявлення чи зламування, та можуть навчатися адаптуватися до різноманітних методів аналізу зображень, ускладнювати завдання виявлення прихованої інформації.

Отже, інтеграція штучного інтелекту у стеганографію зображень відкриває нові перспективи для розвитку більш ефективних та стійких систем приховування інформації в цифрових медіафайлах. Однак, разом із цим, виникають нові виклики, такі як розробка більш прогресивних заходів безпеки, так званих стегоаналізаторів, для ідентифікації стего-зображень.

Стегоаналізатори – це програми або алгоритми, спеціально призначені для виявлення або вилучення прихованої інформації в мультимедійних файлах, включаючи зображення. Їхнє завдання полягає в аналізі цифрового контенту з метою виявлення слідів стеганографії.

Основні методи, які використовуються стегоаналізаторами для виявлення прихованої інформації в зображеннях:

1) аналіз статистичних параметрів: стегоаналізатори можуть аналізувати статистичні параметри зображення, такі як розподіл яскравості, частотні характерис-тики та інші параметри, де аномальності у цих параметрах можуть вказувати на наявність вбудованої інформації;

2) перевірка відмінностей у бітах: порівнюють оригінальне та змінене зображення, визначаючи різницю у значеннях бітів, де суттєві

відмінності можуть свідчити про наявність стеганографії;

3) аналіз коефіцієнтів стиснення: приховування інформації можуть впливати на стиснення зображення, тому стегоаналізатори використо-вують методи аналізу різниці між оригінальними та стиснутими даними для виявлення можливих стеганографічних слідів;

4) виявлення аномальних шаблонів: стегоаналізатори можуть виявляти аномалії в шаблонах або структурах зображення, що можуть вказувати на вбудовану інформацію.

5) використання машинного навчання: деякі стегоаналізатори використовують техніки машинного навчання для автоматизації процесу виявлення стеганографії, такі моделі навчаються розпізнавати патерни, характерні для стеганографічних вбудовувань.

Сучасні стеганографічні методи намагаються пристосуватися до викликів, які представляють стегоаналізатори, застосовуючи технології штучного інтелекту та адаптивні алгоритми, що роблять виявлення прихованої інформації складнішим завданням. Це приводить до постійної еволюції обох сторін – розробників стеганографічних методів та стегоаналізаторів – у пошуках нових технологічних рішень та стратегій.

В цілому стає зрозуміло, що з плином часу та розвитком штучного інтелекту, все більше і більше методів створення стегозображень, а також стегоаналізаторів в своїй основі використовують нейронні мережі.

Для цих цілей популярності набула модель штучного інтелекту GAN, що підходить як для створення стегозображень, так і для стегоаналізу, оскільки містить в собі дві мережі штучного інтелекту, одна з яких вчиться генерувати стегозображення, а інша ідентифікувати їх.

Генеративно-змагальні мережі – це клас штучних інтелектуальних моделей, які були вперше представлені Яном Гудфеллоу та його колегами у 2014 році [5]. Ці моделі внесли значний вклад у галузь машинного навчання, зокрема у генеративне моделювання. Вони надають ефективний метод для створення синтетичних даних, які максимально схожі до реальн розподілу даних.

Основна ідея GAN полягає в тому, що ця модель використовує дві нейронні мережі, які знаходяться у конкурентному відношенні одна до одної. Одна мережа, відома як генератор, старається створити дані, які найкраще імітують реальний розподіл. Інша мережа, відома як дискримінатор, намагається відрізнити справжні дані від синтетичних. Ця конкурентна взаємодія між мережами призводить до вражаючих результатів у створенні зображень, аудіо, тексту та інших творчих областях.

Генеративно-змагальні мережі можна розглядати через метафору: генератор – це майстерні фальсифікатори, які намагаються створити фальшиву валюту, тоді як дискримінатор виступає як експерт-інспектор, що намагається відрізнити справжні гроші від підроблених. Обидві мережі постійно покращуються через конкурентний процес, що веде до удосконалення їхніх здібностей з часом.

Архітектура GAN (рисунок 1) в загальному вигляді включає дві основні частини: генератор і дискримінатор.

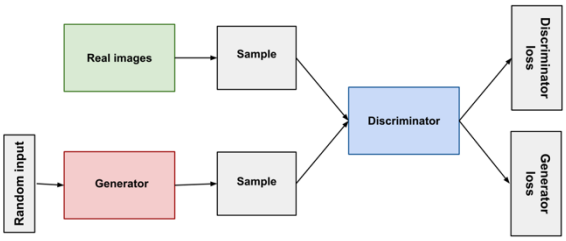


Рисунок 1 – Загальна архітектура GAN [1]

Генератор приймає на вхід випадковий вектор z (зазвичай з розподілу шуму, такого як нормальний розподіл) і намагається згенерувати дані, які схожі на навчальний набір. Функція втрати для генератора враховує те, наскільки добре згенеровані дані проходять через дискримінатор, тобто його можливість визначати, чи є зображення реальним.

Формула для обчислення втрат генератора:

$$G_{loss} = -log\left(D\big(G(z)\big)\right), \quad (1)$$

де $D\big(G(z)\big)$ – ймовірність класифікації дискримінатором згенерованих даних як реальних, $G(z)$ – функція генерації даних,

Логарифм ймовірності того, що дискримінатор визначить згенеровані дані як реальні, взятий зі знаком мінус, введено для того, щоб максимізувати цю втрату під час оптимізації генератора. Максимізація цієї втрати вказує на те, що генератор стає кращим у створенні даних, які дискримінатор вважає реальними.

Отже, головна ідея формули втрат генератора полягає в тому, що генератор намагається максимізувати ймовірність того, що його згенеровані дані будуть визнані як реальні дискримінатором, що приводить до покращення якості створених даних.

Дискримінатор приймає на вхід як згенеровані генератором дані, так і реальні дані з навчального набору і намагається відрізнити одне від іншого. Функція втрат для дискримінатора враховує його здатність правильно класифікувати дані як реальні x чи згенеровані $G(z)$.

Формула для обчислення функції втрат дискримінатора:

$$D_{loss} = -log\big(D(x)\big) - log\left(1 - \left(D\big(G(z)\big)\right)\right), \quad (2)$$

де $D(x)$ – ймовірність правильної класифікації реальних даних дискримінатором.

Логарифм ймовірності правильної класифікації реальних даних взятий зі знаком мінус для того, щоб максимізувати цю втрату під час оптимізації дискримінатора. Максимізація цієї втрати вказує на те, що дискримінатор стає кращим у правильному визначенні реальних даних.

В той же час, інша частина функції втрат дискримінатора вказує на ймовірність того, що дискримінатор правильно класифікує згенеровані генератором дані як сгенеровані. Логарифм від 1 мінус ймовірність цього сценарію також береться зі знаком мінус для того, щоб максимізувати цю втрату під час оптимізації дискримінатора. Максимізація цієї втрати вказує на те, що дискримінатор стає кращим у визначенні згенерованих даних як справжніх.

Отже, функція втрат для дискримінатора спрямована на те, щоб максимізувати ймовірність правильної класифікації як реальних, так і згенерованих даних, що веде до поліпшення його здатностей відрізняти справжні дані від синтетичних.

Наприклад, для створення зображень генератор генерує зображення з випадкових шумових векторів, тоді як дискримінатор навчається розрізняти між реальними та створеними даними. Така сама ідея лежить і в основі генерації стегозображень, тільки замість звичайних зображень, генератор створює стегозображення, вбудовуючи в момент його створення повідомлення, а дискримінатор намагається розрізнити, чи є подане йому на вхід зображення таким що містить приховане повідомлення, чи ні.

Ці дві мережі вступають у конкурентну взаємодію, що призводить до постійного удосконалення якості генерації і розпізнавання даних.

Алгоритми створення стегозображень

Для створення і подальшого порівняння стегозображень було обрано використовувати такі класичні алгоритми кодування повідомлення в зображення-обкладинку, такі як: LSB (Least Significant Bit), DCT (Discrete Cosine Transform) і PVD (Pixel Value Differencing).

Метод LSB (Least Significant Bit) – це техніка стеганографії, що використовує найменш вагомі біти (найменш значущі біти) в бінарному представленні даних для приховування (англ. encode) іншої інформації без помітних змін у вигляді оригінальних даних [6]. Основна ідея полягає в тому, щоб внести невеликі зміни в найменш вагомі біти, які зазвичай менше впливають на сприйняття людьми. Приклад такого вбудовування інформації показано на рисунку 2.

Вибір конкретних бітів для приховування залежить від специфіки застосовуваного методу.

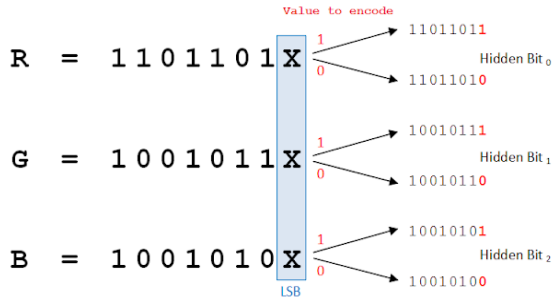


Рисунок 2 – Приклад кодування одного пікселя за допомогою LSB [7]

Зазвичай обираються найменш вагомі біти, які визначають найменш значущі біти в бінарному представленні пікселів зображення. У випадку зображень RGB (Red, Green, Blue), часто використовують останній біт кожного каналу, оскільки зміни в ньому менше помітні для людського ока. Нехай $I(x,y)$ значення пікселя одного з каналів, а bit_i – i-й біт повідомлення , тоді формула вбудовування бітів в найменший значущий біт будя-якого з каналів пікселя буде виглядати так:

$$I'_{x,y} = \left\lfloor \frac{I_{x,y}}{2} \right\rfloor \cdot 2 + bit_i, \qquad (3)$$

де $I'_{x,y}$ – нове значення пікселя одного з каналів після вбудовування;

$\lfloor \cdot \rfloor$ – функція округлення вниз.

Важливо враховувати, що при надто великій зміні в найменш вагомих бітах може виникнути помітна деградація якості зображення або зміна текстового контенту.

Метод LSB може бути використаний для різних цілей, таких як приховування інформації, водяний знак, або навіть атаки на безпеку, тому важливо ретельно вибирати метод та контролювати розмір внесених змін для забезпечення непомітності.

Для декодування (англ. decode) прихованої інформації застосовують той самий метод LSB, витягаючи найменш вагомі біти з області де припускається, що інформація була вбудована. Тоді функція видобуття i-того біту з $I'_{x,y}$ пікселю буде:

$$bit'_i = I'_{x,y} \; mod \; 2. \qquad (2)$$

Важливо зазначити, що для ефективного видобуття інформації, отримання доступу до оригінальних даних є ключовим, оскільки тільки на їхній основі можна точно визначити, які біти були модифіковані. Алгоритм видобуття полягає у зборі найменш вагомих бітів з відповідних пікселів чи символів і їх об'єднанні в бінарний потік. Далі проводиться конвертація бінарного потоку в текстовий або графічний формат, в залежності від того, яка інформація була прихована. Якщо стеганографічний алгоритм не дуже складний або

захищений, інформацію може бути відновлено досить легко. Тому важливо застосовувати додаткові заходи безпеки для ускладнення видобуття прихованої інформації без належного ключа чи доступу до оригінальних даних.

Метод DCT (Discrete Cosine Transform) – дискретне косинусне перетворення в стеганографії використовується для приховування інформації в частотному просторі зображення [8]. Для цього зображення розбивається на блоки, зазвичай розміром 8x8 пікселів, і кожен блок піддається перестворенню.

$$F(u,v) = \\ = C(u)C(v)\sum_{x=0}^{N-1}\sum_{y=0}^{M-1} f(x,y) \\ \cdot \cos\left[\frac{(2x+1)u\pi}{2N}\right]\cos\left[\frac{(2y+1)v\pi}{2M}\right], \quad (4)$$

де N і M – ширина і висота блоку відповідно,

$F(u,v)$ – значення DCT в пікселі з координатами (u,v) у частотному просторі,

$f(x,y)$ – значення пікселя в просторі зображення,

$C(u)$ і $C(v)$ – коефіцієнти, що забезпечують ортогональність перетворення.

Коефіцієнти $C(u)$ і $C(v)$ визначаються:

$$\mathrm{C}(u) = \begin{cases} \sqrt{1/N}, & u = 0 \\ \sqrt{2/N}, & \text{інакше} \end{cases}, \quad (5)$$

$$\mathrm{C}(v) = \begin{cases} \sqrt{1/M}, & v = 0 \\ \sqrt{2/M}, & \text{інакше} \end{cases}. \quad (6)$$

DCT перетворює просторові інтенсивності пікселів в частотний домен, представляючи їх як суму косинусних функцій різної частоти. Зазвичай, для приховування використовуються низькочастотні компоненти DCT, оскільки вони менше помітні для людського зору.

Після DCT перетворення вибираються деякі коефіцієнти, зазвичай в лівому верхньому куті блоку, які вважаються менш важливими для сприйняття людським оком. Інформація для приховування кодується в ці вибрані коефіцієнти. Зазвичай, це виконується шляхом заміни менш значущих бітів коефіцієнтів інформацією, яку необхідно приховати. Для зменшення помітності внесених змін, зазвичай обираються невеликі значення для кодування, які не суттєво впливають на оригінальну інтенсивність пікселів. Після внесення змін, блоки піддаються зворотньому DCT перетворенню для отримання зображення з прихованою інформацією.

Метод DCT в стеганографії дозволяє приховувати інформацію в частотному просторі, що може бути ефективним для непомітного вбудовування даних у зображення.

Для видобуття прихованої інформації спочатку зображення розбивається на блоки, а кожен блок піддається DCT перетворенню. За допомогою заздалегідь встановленого ключа чи параметрів, вибираються ті ж самі коефіцієнти, які використовувалися для приховування інформації. Видобуті біти з цих коефіцієнтів об'єднуються для відновлення прихованої інформації.

Далі застосовуються зворотні операції DCT перетворення для отримання оригінального блоку пікселів з внесеними прихованими даними. Процес повторюється для всіх блоків на зображенні, і отримані дані об'єднуються для отримання повної прихованої інформації.

Застосування методу DCT в стеганографії вимагає врахування того, що прихована інформація може бути вразливою до атак, якщо атакуючий знає методологію вбудовування і володіє ключем чи параметрами. Тому, безпека методу часто пов'язана з ефективністю використання ключа для приховування та видобуття інформації.

Метод PVD (Pixel Value Differencing) – одним з методів стеганографії, який може бути використаний для непомітного вбудовування інформації у зображення, заснований на відмінностях у значеннях пікселів [9].

Для приховування інформації за допомогою PVD зображення розбивається на пари сусідніх пікселів, і різниця між їхніми значеннями використовується для внесення прихованої інформації. Далі приховування даних здійснюється шляхом зміни значення менш важливого пікселя в парі (зазвичай меншого за абсолютною інтенсивністю) в залежності від біта інформації, яку потрібно приховати. Зміни виконуються так, щоб вони були менше помітні для спостерігача, і для цього може використовуватися різноманітний набір правил та параметрів. Крім того, важливо враховувати, що зміни не повинні бути великими, щоб уникнути помітних артефактів на зображенні.

Нехай $I_{x,y}$ та $I_{x+1,y}$ – сусідні пікселі, тоді формула для вбудовування буде виглядати так:

$$I'_{x,y} = I_{x,y} + \beta \cdot sign\left(I_{x+1,y} - I_{x,y}\right) \cdot 2^{bit}, \quad (7)$$

де $I'_{x,y}$ – нове значення пікселя,

$sign$ – функція знаку (позначається як 1, 0 або -1),

β – контрольований параметр,

bit – вбудовуваний біт (0 або 1).

Для видобуття прихованої інформації використовуються ті ж пари сусідніх пікселів, які були використані при приховуванні. Зміни в інтенсивності менш важливого пікселя, які були внесені під час приховування, аналізуються для відновлення бітової інформації. Застосовуються визначені параметри і правила для визначення, які

зміни в інтенсивності пікселів вказують на наявність прихованої інформації. Використовуючи ключ або параметри, що використовувалися при приховуванні, виконується обернений процес для витягнення бітової інформації. Така процедура повторюється для всіх пар сусідніх пікселів на зображенні, і зібрані біти об'єднуються для отримання повної прихованої інформації. Виходячи з усього вищесказаного, формула для вилучення вбудованого біту буде виглядати так:

$$bit' = round\left(\frac{I'_{x,y} - I_{x,y}}{2^{bit}}\right), \quad (8)$$

де $round$ – функція округлення.

Метод PVD є одним з прикладів стеганографії, який може бути використаний для непомітного вбудовування інформації у зображення, заснований на відмінностях у значеннях пікселів. Також цей метод відзначається своєю простотою та непомітністю, але важливим є факт того, що безпека від виявлення може залежати від вибору параметрів та правил, використовуваних для внесення та видобуття змін.

Крім того, було вирішено також використати для порівняння один з методів цифрової стеганографії зображень на основі GAN – SteganoGAN. Автори пропонують рішення з використанням згорток на основі DCGAN, для отримання високоємної стеганографії зображень, яку неможливо виявити.

SteganoGAN складається з трьох модулів (рисунок 3): Encoder, який приймає зображення обкладинки та тензор даних, або повідомлення, і створює стеганографічне зображення; Decoder, який приймає стеганографічне зображення та намагається відновити тензор даних, і Critic, який оцінює якість обкладинки та стеганографічних зображень. Крім того, Encoder передбачає використання однієї з трьох різних архітектур на вибір: Basic, Residual або Dence.

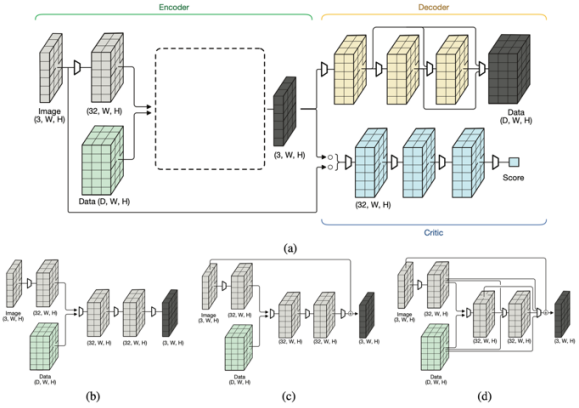


Рисунок 3 – Архітектура SteganoGAN [10]

Щоб оцінити продуктивність стеганографічної систему штучного інтелекту, використовується метрика Акуратності вилучення секретних даних (Accuracy) і «біт на піксель» (BPP), що означає міру інформації, що виміряє кількість прихованої інформації на одиницю пікселя у зображенні. Однак замість метрики BPP автори використовують більш надійну метрику RS-BPP, що базується на використанні кодування Ріда-Соломона.

Також вимірюють якість зображення за допомогою пікового співвідношення сигнал/шум (PSNR) і структурної подібності (SSIM). Експериментальні результати демонструють ефективність різних варіантів «SteganoGAN», причому варіант Dence демонструє кращу продуктивність щодо відносного корисного навантаження та якості зображення.

Крім того, у документі досліджується здатність системи уникати інструментів стеганалізу, які використовуються для виявлення прихованих даних. Наведені результати показують, що створені в моделі штучного інтелекту «SteganoGAN» стеганографічні зображення важко виявити, оскільки вони відповідають мінімальним вимогам для життєздатного алгоритму стеганографії.

Також обговорюються порівняння з традиційними методами стеганографії, такими як HUGO та JSteg, підкреслюючи переваги та потенціал запропонованого підходу SteganoGAN на основі глибокого навчання.

Дослідження представляє інноваційний і гнучкий підхід до стеганографії зображень, який підтримує різні розміри зображень і довільні двійкові дані, досягаючи вищих відносних корисних навантажень порівняно з існуючими методами, залишаючись невиявленими. SteganoGAN відкриває нові можливості в галузі стеганографії, де глибоке навчання та GAN відіграють вирішальну роль у розширенні меж методів приховування даних.

Метрики оцінювання стегозображень

Для оцінки стеганографічних алгоритмів в загальному використовуються три метрики: обсяг даних, який можна приховати в зображенні, тобто ємність, подібність між обкладинкою та стеганографічним зображенням, яка називається спотворенням, і здатність уникнути виявлення інструментами стеганалізу, яка називається секретністю [11]. У цьому розділі описано деякі показники для оцінки продуктивності стеганографічних алгоритмів.

Метрика *BPP* (Bits Per Pixel) – це метрика, що вказує на кількість бітів інформації, яку припадає на кожен піксель у зображенні. Вона важлива для оцінки роздільної здатності та якості зображень, оскільки більше бітів на піксель дозволяє зберігати більше деталей та інформації, але може вимагати більше простору для зберігання або передачі.

Формула BPP обчислюється як відношення обсягу інформації $bitsn$ (зазвичай у бітах) до кількості пікселів у зображенні розміром $W \cdot H$:

$$BPP = \frac{bitsn}{W \cdot H}. \quad (8)$$

Використовується для оцінки обсягу інформації, яку можна вбудувати в кожен піксель зображення без помітних змін в якості. Окрім вимірювання відносного корисного навантаження, нам також потрібно виміряти якість стеганографічного зображення. Одним із широко використовуваних показників для вимірювання якості зображення є пікове відношення сигнал/шум (PSNR).

Метрика $PSNR$ (Peak Signal-to-Noise Ratio) – враховує відношення сигнал-шум і вимірює, наскільки добре спотворене (сгенероване) зображення відповідає оригіналу. PSNR виражається у децибелах (dB) і може бути інтерпретований таким чином:

– 30-40 dB: дуже хороша якість;

– 20-30 dB: прийнятна якість;

– Менше 20 dB: низька якість.

Чим вище значення PSNR, тим менше помилок відносно оригіналу і, отже, якість спотвореного зображення вважається вищою.

За наявності двох зображень X і Y розміром $W \cdot H$ і коефіцієнтом масштабування sc (для 8-бітних зображень дорівнює 255), який представляє максимально можливу різницю в числовому представленні кожного пікселя, PSNR визначається через функцію середньої квадратичної помилки (MSE):

$$MSE = \frac{1}{WH}\sum_{i=1}^{W}\sum_{j=1}^{H}\left(X_{i,j} - Y_{i,j}\right)^2, \quad (9)$$

$$PSNR = 20 \cdot \log_{10}(sc) - 10 \cdot \log_{10}(MSE), \quad (10)$$

Хоча $PSNR$ часто використовується для оцінки спотворень, які виникають внаслідок застосування алгоритмів стеганографії, вважається, що він не є ідеальним для порівняння різних типів таких алгоритмів. Тому використовується ще один показник, а саме індекс структурної подібності (SSIM), який допомогає оцінювати якість зображень.

Метрика $SSIM$ (Structural Similarity Index) – широко використовується в індустрії мовлення для вимірювання якості зображення та відео. В стеганографії ця метрика використовується для вимірювання структурної схожості між двома зображеннями. $SSIM$ розроблений для того, щоб враховувати сприйнятливість людського зору до змін в контрасті, яскравості та структурі зображення, і він часто використовується для оцінки якості стиснених, оброблених чи сгенерованих зображень.

Основні компоненти індексу структурної подібності включають:

1) яскравість (англ. luminance): оцінює, наскільки зображення яскравості відповідає реальним відчуттям яскравості в людському зорі;

2) контраст (англ. contrast): визначає, як добре контраст відображається на зображенні порівняно з реальним контрастом;

3) структурна подібність (англ. structural similarity): вимірює, наскільки добре структурні елементи, такі як текстури та деталі, відтворені на зображенні.

За наявності двох зображень X і Y $SSIM$ можна розрахувати за формулою:

$$SSIM_{X,Y} = \frac{(2\mu_X\mu_Y + c_1)\cdot(2\sigma_{XY} + c_2)}{(\mu_X^2 + \mu_Y^2 + c_1)\cdot(\sigma_X^2 + \sigma_Y^2 + c_2)}, \quad (11)$$

де μ_X – середнє значення X, μ_Y – середнє значення Y, σ_X^2 – дисперсія X, σ_Y^2 – дисперсія Y, σ_{XY} – коваріантність X і Y, $c1 = (k_1 L)^2$, $c2 = (k_2 L)^2$ – дві змінні, в яких $k_1 = 0.01$, $k_1 = 0.03$, L – динамічний діапазон значень пікселів, що при використанні 8-бітного зображення дорівнює 255.

Значення метрики $SSIM$ варіюється в межі [-1, 1], де 1 вказує на повну ідентичність між зображеннями, а -1 навпаки, говорить про те, що два зображення повністю відрізняються. Високий показник SSIM вказує на високу схожість між зображеннями, тоді як низький SSIM вказує на відмінності в якості.

Порівняння стеганографічних алгоритмів

Для порівняння на практиці основних метрик кожним з наведених вище класичних стеганографічних алгоритмів, а також моделлю ШІ SteganoGAN було створено стегозображення з різним значенням BPP. Для методів LSB, DCT, PVD та SteganoGAN граничні значення досить великі, було вирішено провести експеримент з створенням стегозображень з:

$$BPP \in [0.01, 0.1, 1, 1.5, 2, 2.5, 3, 3.5, 4]. \quad (12)$$

В ході експерименту виявилось, що для методу DCT найбільшим значенням BPP є 0.1, тому було вирішено залишити цей метод, і продовжити експеримент з більш великими значеннями BPP для методів LSB, PVD, SteganoGAN. Крім того, метод LSB показав, що при зростанні BPP метод перставав працювати через брак вільного місця в зображенні, але алгоритм досить гнучкий, і дозволяє використання більше ніж одного найменшого значущого біта в зображенні, тому

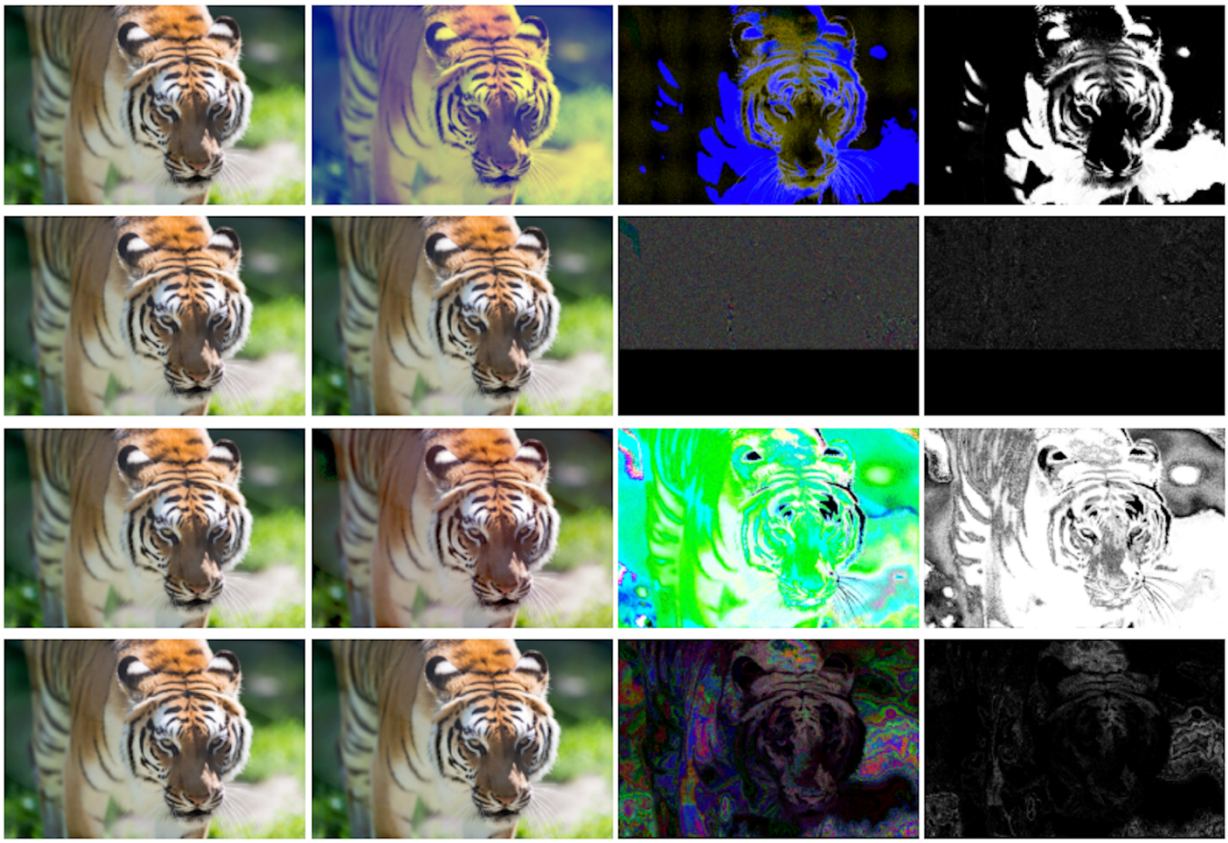


Рисунок 4 – Порівняння оригінального зображення із стегозображенням
Джерело: Авторський рисунок

було прийнято рішення використовувати 2 найменших значущих бітів зображення, що для експеременту було більше ніж достатньо для кодування повідомлення в зображення.

На рисунку 4 наведено порівняння оригінального зображення із стегозображеннями, створених різними методами. Перша колонка містить в собі оригінальні зображення, друга – стегозображення що сгенероване одним з чотирьох вищенаведених алгоритмів, третя колонка містить бінарізовану різницю між оригінальним і стегозображенням за трьома каналами, четверта колонка містить бінарізовану різницю між оригінальним і стегозображенням в сірому тоні. Перший рядок містить стегозображення, створене за допомогою методу DCT з найбільшим можливим $BPP = 0.1$. Другий, третій і четвертий рядок містять стегозображення створені за допомогою методів LSB, SteganoGAN і PVD відповідно, з використанням $BPP = 4$.

На наведеному вище зображенні в першому рядку можна побачити різницю між оригінальним і стегозображенням створеного за допомогою методу DCT, і сказати, що дане зображення є стегозображенням. Проте, візуальне порівняння стегозображеннь інших методів з оригінальним вже не таке однозначне, тому скористуємось метриками $BPP, MSE, PSNR$ та $SSIM$ для їх порівняння. На рисунку 5 наведено графіки залежності метрики MSE від BPP для методів DCT, LSB, SteganoGAN та PVD. З рисунку видно що навіть при малому значенні BPP, метод DCT видає набагато більшу помилку, в інтервалі $MSE \in [1300, 1400]$, ніж інші стеганографічні методи. В той же час, два інші класичні методи видають помилку близьку до нуля в інтервалі $MSE \in [0, 1]$, тоді як оцінки SteganoGAN знаходяться в інтервалі $MSE \in [1300, 1400]$.

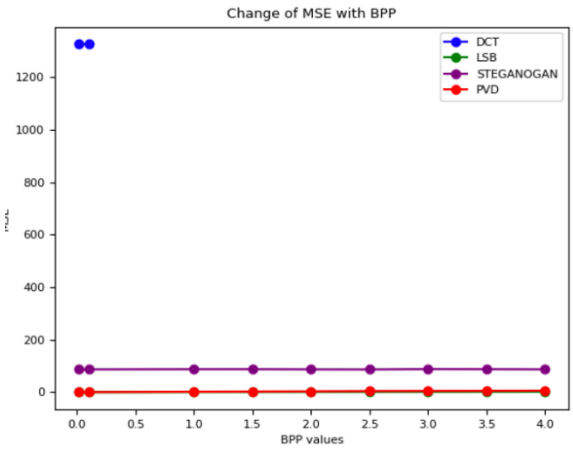


Рисунок 5 – Графік залежності метрики MSE від BPP
Джерело: Авторський рисунок

На рисунку 6 наведено графік залежності метрики PSNR від BPP.

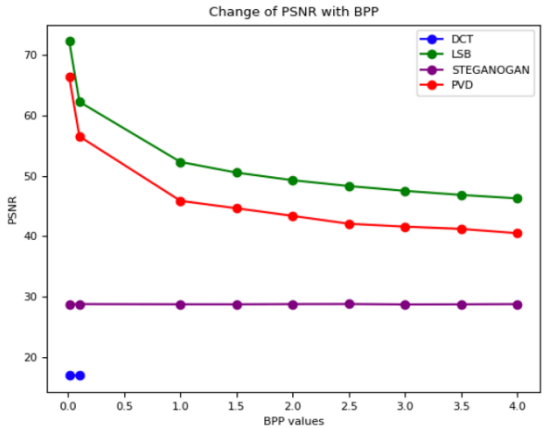


Рисунок 6 – Графік залежності метрики $PSNR$ від BPP
Джерело: Авторський рисунок

З графіку видно, що при малому значенні метрики BPP методи LSB і PVD дають дуже хороший результат в 50-80 dB метрики PSNR, але при збільшенні $BPP, PSNR$ ціх двох методів поступово зменшується до 40-50 dB. Метод StaganoGAN дає стабільний результат в 29 dB метрики $PSNR$ при будь-якому значенні BPP. Також бачимо, що метод DCT навіть при низькому значенні BPP дає поганий результат в 10 dB метрики $PSNR$. На рисунку 7 наведено графік залежності метрики $SSIM$ від BPP.

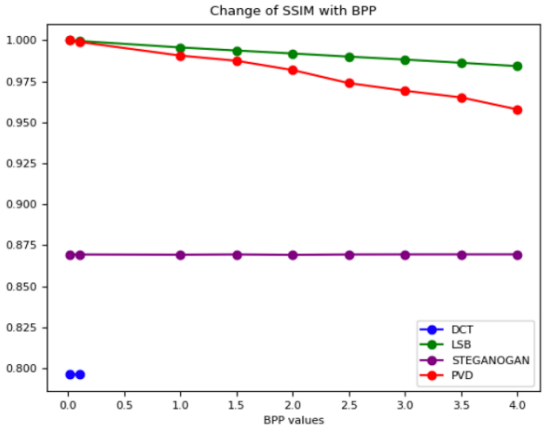


Рисунок 7 – Графік залежності метрики $SSIM$ від BPP
Джерело: Авторський рисунок

Як і з іншими метриками, методи LSB і PVD надають кращий результат метрики $SSIM$, близький до 100% структурної ідентичності при дуже малих значеннях BPP, однак при збільшенні BPP цей результат поступово зменшується. В той же час, метод SteganoGAN дає стабільно гарний результат в 87% метрики $SSIM$ при будь-якому значенні BPP. Метод DCT надає найгірший результат з усіх розглянутих стеганографічних методів в 79% структурної ідентичності, навіть при малих занченнях BPP.

Загалом, виходячи з результатів розрахунку метрик для кожного з розглянутих стеганографічних алгоритмів, можна сказати, що найгірше показав себе метод DCT, який видавав найгірший з усіх результат по всім метрикам навіть при найменших значеннях $BPP \in [0.01, 0.1]$. Найкраще себе показали методо LSB та PVD, хоча за усіма метриками, при збільшенні BPP, їх якість поступово падала. В той же час метод SteganoGAN видавав стабільно добрий результат за всіма метриками, при будь-якому значенні BPP.

Використання стегодетекторів

Для порівняння методів LSB, DCT, PVD та SteganoGAN до стійкості до стегоаналізу було принято рішення використати як статистичні стегоаналізатори так і стегоаналізатори засновані на методах глибокого навчання.

В якості статистичного стегоаналізатору було обрано використати популярний інструмент стеганалізу з відкритим вихідним кодом під назвою StegExpose [12], який поєднує в собі кілька існуючих методів стеганалізу, включаючи:

– RS Analysis;
– Chi Squared Attack;
– Primary Sets;
– Sample Pairs.

Щоб оцінити ефективність кожного з розглянутих стеганографічних методів уникнення виявлення, було використано всі сгенеровані стегозображення цих методів з різними значеннями BPP, а результати було перевірено за допомогою StegExpose і приведено в таблиці 1, де на перехресті алгоритму, за допомогою якого було створне стегозображення, і стегоаналізатору цифра 0 означає, що таке зображення з відповідним BPP було розпізнане як таке що не містить призованих даних, а цифра 1 означає що зображення є стегозображенням.

Недавні дослідження [13-15] показали багатообіцяючі результати у виявленні стеганографічних зображень за допомогою підходів на основі глибокого навчання. Тому для вивчення того, чи можуть стеганографічні методи протистояти інструментам стегоаналізу на основі глибокого навчання, було використано запропоновані для стеганалізу моделі детекторів на базі ШІ: YeNet, XuNet та SrNet. Результат використання цих моделей ШІ для стегоаналізу також приведено в таблиці 1. Слід зауважити, що інструменти стего-аналізу StegExpose та YeNet були використані в роботі [10] для перевірки стійкості стегозображень до стегоаналізу, тому очікувалось, що нові перевірки моделі SteganoGAN підтвердять описані результати перевірки, тоді як інші XuNet та SrNet покажуть результати не гірше. Крім того, використана в дослідженні модель для стегоаналізу

Таблиця 1 – Виявлення сетегозображень за допомогою різних стегоаналізаторів

BPP	Алгоритм	StegExpose	SrNet	YeNet	XuNet
0.01	LSB	0	1	0	0
	PVD	0	1	0	0
	DCT	0	1	0	0
	SteganoGAN	0	0	0	1
0.1	LSB	0	1	0	0
	PVD	0	1	0	0
	DCT	0	0	0	0
	SteganoGAN	0	0	0	0
1	LSB	1	1	0	0
	PVD	0	1	0	0
	DCT	0	0	0	0
	SteganoGAN	0	0	0	1
1.5	LSB	1	1	0	0
	PVD	0	1	0	0
	SteganoGAN	0	0	0	1
2	LSB	1	1	0	0
	PVD	0	1	0	0
	SteganoGAN	0	0	0	1
2.5	LSB	1	1	0	0
	PVD	0	0	0	0
	SteganoGAN	0	0	0	1
3	LSB	1	1	0	0
	PVD	0	1	0	0
	SteganoGAN	0	0	0	1
3.5	LSB	1	1	0	0
	PVD	0	0	0	0
	SteganoGAN	0	0	0	1
4	LSB	1	0	0	0
	PVD	0	1	0	0
	SteganoGAN	0	0	0	1

YeNet показала себе як така, що не підходить для виявлення прихованих даних в зображеннях, а StegExpose як було сказано в роботі [12] більш придатний для виявлення стегозображень зтворених зо допомогою методу LSB. Модель SrNet досить непогано спрацювала з всіма класичними методами стеганографії LSB, PVD та DCT, але не змогла виявити стегозобарження що були створені за допомогою SteganoGAN. В той же час, як можна побачити з результатів, модель XuNet виявила 8 із 9 стегозображень, тобто майже всі стегозображення сгенеровані за допомогою SteganoGAN, що повністю спростовує наданий авторами результат стійкості до стегоаналізу в 99%, а також каже про те, що саме модель XuNet найбільше з усіх підходить для стегоаналізу зображень сгенерованих за допомогою стеганографічних моделей на основі GAN.

Висновки

В рамках дослідження проведено огляд предметної області, визначено основні моделі і алгоритми створення стегозображень, виділено їх особливості. Підкреслено, що на сьогодні машинне навчання є домінуючою технологією в сфері стеганографії зображень. Особливу увагу в цьому напрямку надано генеративно-змагальним мережам (GAN), чия архітектура підходить для створення якісних моделей для стеганорафії, так і для стегоаналізу. Ретельно оглянуто архітектуру і алгоритми використання трьоїх класичних методів створення стегозображень: LSB, PVD та DCT, а також одиного з методів заснованих на GAN: SteganoGAN. Наведено основні метрики оцінювання стегозображень, їх формули і значення.

Результати експериментів із створення стегозображень та їх стеганоаналізу свідчать про те, що моделі засновані на штучному інтелекту мають більший потенціал для стеганографії зображень ніж класичні алгоритми, оскільки мають стабільні результати метрик незалежно від значення BPP. Крім того, виявлено, що зазначена властивість стійкості моделі SteganoGAN [10] до стегоаналізу не відповідає дійсності, оскільки модель XuNet виявила 88,8% сгенерованих SteganoGAN стегозображень.

В подальших дослідженнях планується удосконалення моделі StaganoGAN за допомогою використання згорткових нейронних мереж, інших

методів оптимізації градієнтного спуску, а також використання новітніх стегодетекторів на основі штучного інтелекту для визначення стійкості стегозображень.

Список літератури

1. Image Steganography: Recent Trends and Techniques / Sana P.K. et al. 2021. P. 37–45. URL: https://doi.org/10.4018/978-1-7998-7160-6.ch002 (date of access: 10.12.2023).
2. Conway M. Code wars: Steganography, signals intelligence, and terrorism. Knowledge, Technology & Policy, Jun 2003. Vol. 16, no. 2. P. 45–62. URL: https://doi.org/10.1007/s12130-003-1026-4 (date of access: 10.12.2023).
3. Fridrich J., Goljan M., & Du R. Reliable detection of LSB steganography in color and grayscale images. MM&Sec '01: P Proceedings of the 5th Workshop on Information Hiding and Multimedia Security, 2001. P. 27–30. URL: https://doi.org/10.1145/1232454.1232466 (date of access: 10.12.2023).
4. Hermassi H. Blind image steganalysis using features extraction and machine learning. 18th International Multi-Conference on Systems, Signals & Devices (SSD), 22-25 March 2021. P. 673–680. URL: https://doi.org/10.1109/SSD52085.2021.9429334 (date of access: 10.12.2023).
5. Generative Adversarial Networks / Goodfellow I. et al. Proc. of 27th International Conference on Neural Information Processing Systems, 2014. P. 2672–2680.
6. Fridrich J. Steganography in Digital Media: Principles, Algorithms, and Applications. 1st Edition. – New York: Cambridge University Press, 2009. P. 437.
7. LSB-Steganography - Python program to steganography files into images using the Least Significant Bit. URL: https://www.kitploit.com/2018/02/lsb-steganography-python-program-to.html
8. Bhattacharya T., Dey N., Chaudhuri S. A session based multiple image hiding technique using DWT and DCT. International Journal of Computer Application (IJCA), 2012. Vol. 38, no. 5. P. 398–409.
9. Pradhan A., Sekhar R.K., Swain G. Adaptive PVD Steganography Using Horizontal, Vertical, and Diagonal Edges in Six-Pixel Blocks. Security and Communication Networks, 2 August 2017. Vol. 2017. URL: https://doi.org/10.1155/2017/1924618 (date of access: 10.12.2023).
10. SteganoGAN: High Capacity Image Steganography with GANs / Zhang K., Alfredo C., Lei Xu, Kalyan V. 2019. URL: https://doi.org/10.48550/arXiv.1901.03892 (date of access: 10.12.2023).
11. Performance evaluation parameters of image steganography techniques / Pradhan A., Sahu A., Swain G., Sekhar R.K. International Conference on Research Advances in Integrated Navigation Systems, RAINS, 2016. URL: https://doi.org/10.1109/RAINS.2016.7764399 (date of access: 10.12.2023).
12. Boehm B., StegExpose-a tool for detecting lsb steganography, 2014. URL: https://arxiv.org/abs/1410.6656 (date of access: 10.12.2023).
13. Ye J., Ni J., Yi Y. Deep Learning Hierarchical Representations for Image Steganalysis. IEEE Transactions on Information Forensics and Security, November 2017. Vol. 12, no. 11. P. 2545–2557. URL: https://doi.org/10.1109/TIFS.2017.2710946 (date of access: 10.12.2023).
14. Xu G., Wu H., Shi Y. Structural Design of Convolutional Neural Networks for Steganalysis. IEEE Signal Processing Letters, May 2016. Vol. 23, no. 5. P. 708–712. URL: https://doi.org/10.1109/LSP.2016.2548421 (date of access: 10.12.2023).
15. Boroumand M., Chen M., Fridrich J., Deep Residual Network for Steganalysis of Digital Images. IEEE Transactions on Information Forensics and Security, May 2019. Vol. 14, no. 5. P. 1181–1193, URL: https://doi.org/10.1109/TIFS.2018.2871749 (date of access: 10.12.2023).

References

1. Sana, P.K., et al. (2021), "Image Steganography: Recent Trends and Techniques", pp. 37–45, available at: https://doi.org/10.4018/978-1-7998-7160-6.ch002 (date of access: 10.12.2023).
2. Conway, M. (2003), "Code wars: Steganography, signals intelligence, and terrorism" *Knowledge, Technology & Policy*, Jun 2003. Vol. 16, no. 2, pp. 45–62, available at: https://doi.org/10.1007/s12130-003-1026-4 (date of access: 10.12.2023).
3. Fridrich, J., Goljan, M., & Du R. (2001), "Reliable detection of LSB steganography in color and grayscale images", *MM&Sec '01: P Proceedings of the 5th Workshop on Information Hiding and Multimedia Security*, pp. 27–30, available at: https://doi.org/10.1145/1232454.1232466 (date of access: 10.12.2023).
4. Hermassi, H. (2021), "Blind image steganalysis using features extraction and machine learning", *18th International Multi-Conference on Systems, Signals & Devices (SSD),* 22-25 March 2021. pp. 673–680, available at: https://doi.org/10.1109/SSD52085.2021.9429334 (date of access: 10.12.2023).

5. Goodfellow, I. et al. (2014), "Generative Adversarial Networks", *Proc. of 27th International Conference on Neural Information Processing Systems*, pp. 2672–2680.
6. Fridrich, J. (2009), "Steganography in Digital Media: Principles, Algorithms, and Applications", 1st Edition. *– New York: Cambridge University Press*, pp. 437.
7. "LSB-Steganography - Python program to steganography files into images using the Least Significant Bit", available at: https://www.kitploit.com/2018/02/lsb-steganography-python-program-to.html
8. Bhattacharya, T., Dey, N., Chaudhuri, S. (2012), "A session based multiple image hiding technique using DWT and DCT", *International Journal of Computer Application (IJCA)*, vol. 38, no. 5, pp. 398–409.
9. Pradhan, A., Sekhar, R.K., Swain, G. (2017) "Adaptive PVD Steganography Using Horizontal, Vertical, and Diagonal Edges in Six-Pixel Blocks", *Security and Communication Networks,* vol. 2017, available at: https://doi.org/10.1155/2017/1924618 (date of access: 10.12.2023).
10. Zhang, K., Alfredo, C., Lei, Xu, Kalyan, V. (2019), "SteganoGAN: High Capacity Image Steganography with GANs", available at: https://doi.org/10.48550/arXiv.1901.03892 (date of access: 10.12.2023).
11. Pradhan, A., Sahu, A., Swain, G., Sekhar, R.K. (2016), "Performance evaluation parameters of image steganography techniques", *International Conference on Research Advances in Integrated Navigation Systems, RAINS*, available at: https://doi.org/10.1109/RAINS.2016.7764399 (date of access: 10.12.2023).
12. Boehm, B. (2014), "StegExpose-a tool for detecting lsb steganography", available at: https://arxiv.org/abs/1410.6656 (date of access: 10.12.2023).
13. Ye, J., Ni, J., Yi, Y. (2017), "Deep Learning Hierarchical Representations for Image Steganalysis", *IEEE Transactions on Information Forensics and Security*, November 2017. Vol. 12, no. 11, pp. 2545–2557, available at: https://doi.org/10.1109/TIFS.2017.2710946 (date of access: 10.12.2023).
14. Xu, G., Wu, H., Shi, Y. (2016), "Structural Design of Convolutional Neural Networks for Steganalysis", *IEEE Signal Processing Letters*, vol. 23, no. 5. pp. 708–712, available at: https://doi.org/10.1109/LSP.2016.2548421 (date of access: 10.12.2023).
15. Boroumand, M., Chen, M., Fridrich, J. (2019), "Deep Residual Network for Steganalysis of Digital Images", *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1181–1193, available at: https://doi.org/10.1109/TIFS.2018.2871749 (date of access: 10.12.2023).

Надійшла до редакції 13.11.2023

D. KHOMA
Donetsk National Technical University, Lutsk, Ukraine
dmytro.khoma.asp@donntu.edu.ua

STEGANOGRAPHIC ALGORITHMS AND STEGOANALYSIS BASED ON CLASSICAL METHODS AND NEURAL NETWORKS

The article is devoted to an overview of algorithms and models for creating steganographic images based on classical steganographic methods, as well as methods using artificial intelligence. An overview of the subject area is given, and the main metrics for comparing steganographic algorithms are identified. A comparative analysis of images created using different steganographic methods with different payloads is performed. The results of the work of various stegoanalyzers based on statistical methods and deep learning methods are presented. The main tasks for further research are identified.

The purpose of this paper is to provide an overview of algorithms and models for generating steganographic images, a comparative analysis various steganographic techniques, and presenting results from diverse stegoanalyzers utilizing statistical and deep learning methods.

Keywords: steganography, metrics, generative adversarial network, methods of steganographic analysis, neural network, stegodetector, robustness, similarity

УДК 004.056

М.С. Кондратенко, аспірант
Інститут проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України,
м. Київ, Україна
nikolay.ns95@gmail.com

Узгодження математичних та юридичних аспектів при використанні смарт-контрактів

В статті розглянуто питання правового статусу смарт-контрактів в Україні, досліджено проблеми та перешкоди, які виникають при узгодженні розумних контрактів з законодавством України. В роботі наводяться основні відомості про смарт-контракти, їх характеристики та особливості застосування, відбувається порівняння смарт-контрактів з традиційними договорами з урахуванням переваг та недоліків кожного з них, проводиться аналіз правового досвіду країн, в яких активно використовуються смарт-контракти та пропонуються шляхи вирішення проблем, які при цьому виникають

Ключові слова: блокчейн, смарт-контракт, розумний договір, смарт-договір, криптовалюта, Bitcoin, Ethereum, правове регулювання смарт-контрактів

DOI: 10.31474/1996-1588-2023-2-37-54-66

Вступ

Розвиток цифрових технологій суттєво змінює принципи та інструменти, на яких базується функціонування фінансових ринків розвинутих країн. Однією з проривних технологій, яка зумовила напрямки подальших змін у багатьох галузях, стала блокчейн-технологія.

Саме завдяки виникненню і швидкому розвитку блокчейн-платформ виник новий тип договорів – так звані смарт-контракти, тобто «розумні угоди». Смарт-контракт - це формальна угода, яка створена у вигляді комп'ютерного коду і може бути укладена, змінена або припинена лише за допомогою конкретної комп'ютерної програми. Така угода виконується автоматично та безперервно, і не потребує участі третьої довіреної сторони. Вона також може функціонувати в умовах повної недовіри між сторонами.. Ці корисні риси смарт-контрактів зумовлюють постійне зростання їх популярності. Хоч існує багато юридичних питань, пов'язаних зі смарт-контрактами, проте вони все ж таки досить успішно працюють на різних блокчейн-платформах, зокрема на платформі Ethereum.

Смарт-контракти на блокчейн-платформі для операцій в електроенергетиці, на відміну від традиційних контрактів, мають кілька вагомих переваг:

1) швидкі розрахунки, проводяться негайно після факту споживання електроенергії;

2) зручний вибір параметрів тарифікації та управління трафіком;

3) відсутність проблем з боржниками завдяки автоматизованій системі розрахунків;

4) відмова від посередників, що призводить до зниження ціни на електроенергію на 5-10%.

Ці смарт-контракти ініціюються однією із сторін (споживачем, мережею або генеруючою компанією) і укладаються після узгодження всіх деталей.

Можна навести такі потенційні очікування від впровадження смарт-контрактів в енергетиці:

- зменшення витрат завдяки відмові від посередників;
- збільшення гнучкості у виборі тарифу та системи оплати;
- посилення ролі звичайних споживачів у плануванні обсягів споживання;
- можливість безпосередньо закуповувати електроенергію від виробників;
- захист учасників від недобросовісних маніпуляцій з показниками обліку;
- захист мережевих та генеруючих компаній від ризику банкрутства компаній-постачальників;
- захист споживачів від недостовірної інформації щодо відключень.

На даний час єдиним можливим середовищем для повноцінного функціонування смарт-контрактів є блокчейн, оскільки він дозволяє зберігати історію всіх транзакцій та захищає цю історію від «зловмисних» змін. Тому сторонам, які бажають укласти смарт-контракт замість «звичайного» договору, потрібно обрати середовище для укладення відповідної угоди. Зазвичай таким середовищем виступає Ethereum, трохи менш поширеними є Polkadot, Tron та Ripple [1]. Останнім часом з'явились також нові платформи Side Chains і NXT. Кожен сайдчейн - це повноцінний блокчейн зі своїми правилами, алгоритмами, користувачами, способами отримання винагороди, механізмами захисту та іншими базовими складовими. Однак це не автономна платформа, оскільки будь-який

сайдчейн завжди так чи інакше пов'язаний з основним блокчейном. Деякі сайдчейни (Side Chains) сумісні з EVM (Ethereum Virtual Machine) і можуть виконувати контракти, розроблені для віртуальної машини Ethereum. EVM-сумісні сайдчейни підтримують смарт-контракти, написані на мові Solidity, а також на інших мовах смарт-контрактів EVM, що означає, що смарт-контракти, написані для Ethereum Mainnet, також працюватимуть на EVM-сумісних сайдчейнах [2].

В свою чергу, NXT є більш розширеною платформою блокчейну. Вона була випущена в 2013 році і спроектована для створення децентралізованих додатків і послуг. Ця платформа використовує власну криптовалюту NXT і надає можливість створювати власні токени на основі цієї блокчейн мережі. Стійкість платформи забезпечується за рахунок алгоритму на основі принципу Proof-of-stake (PoS). Окрім цього до ключових особливостей NXT можна віднести наступні:

- NXT не є криптовалютою, це криптоплатформа з вбудованою підтримкою додатково створюваних для неї додатків (у тому числі і криптовалют);
- NXT автоматично повертає суму транзакції назад на рахунок користувача, якщо не було жодного підтвердження транзакції;
- NXT не зберігає гаманець користувача на фізичному носії, а використовує технологію Brain Wallet;
- NXT підтримує роботу з токенами, що дозволяють сторонньому ресурсу ідентифікувати їх.

власника рахунку NXT без введення ключової фрази [3].

Якщо ж потрібно реалізувати використання смарт-контрактів на рівні державних установ, то швидше за все потрібно буде створювати новий блокчейн, з особливими властивостями, пристосований для використання саме смарт-контрактів певного типу і який майже напевно буде частково централізованим. Зазначимо, що ідея створення частково централізованого блокчейну на базі протоколу консенсусу PoP було розглянуто у роботах [4], [5].

Концепція смарт-контрактів з першими належними згадками виникла у 1994 році, коли криптограф та правовий експерт Нік Сабо висловив думку, що застосовуючи електронний децентралізований реєстр, можна створювати контракти, які виконуються автоматично. Але на практиці ця ідея була повноцінно реалізована лише у 2008 році завдяки появі технології блокчейн. Зараз використання смарт-контрактів набуває суттєвого практичного значення. Особливо вважаємо за потрібне виділити такі сфери їх застосування як ведення каскадних державних реєстрів [4], [5] та укладання різних угод щодо продажу «зеленої» електроенергії [11]. Але при цьому на шляху їх використання виникає багато юридичних питань і проблем. Деякі з них є реальними, і пов'язані в основному з тим, що смарт-контракти є досить новим інструментом, який ще не знайшов відображення у законодавствах переважної більшості країн. Деякі інші проблеми можна вважати більш надуманими, ніж реальними, вони пояснюються настороженим ставленням більшості людей до нових технологій, які є досить складними для розуміння звичайного громадянина, що не є спеціалістом у відповідній галузі.

Основною метою цієї роботи є: аналіз юридичних аспектів і проблем, пов'язаних з використанням смарт-контрактів; огляд та аналіз відповідних рішень, прийнятих іншими країнами у цьому питанні; аналіз можливих шляхів вирішення таких проблем, з урахуванням досвіду інших країн.

Актуальність роботи зумовлена як стрімким поширенням використання смарт-контрактів у різних галузях, так і їх спроможністю вирішити широкий спектр нагальних питань (ведення державних реєстрів, захист від рейдерства, спрощення p2p-продажу електроенергії між домогосподарствами, тощо), за умови визначення їх юридичного статусу.

Робота має наступну структуру. У першому розділі ми надаємо основні відомості про смарт-контракти, їх характеристики та особливості застосування. У другому розділі ми аналізуємо законодавчий досвід країн, в яких активно використовуються смарт-контракти, та визначаємо проблеми, які при цьому залишаються відкритими. У третьому розділі ми пропонуємо шляхи вирішення деяких із цих проблем. У висновках визначаємо, чого нам вдалося досягти у цій роботі та якими можуть бути подальші дослідження за цим напрямком.

Визначення та основні властивості смарт-контрактів

Концепція блокчейн, яка використовується в криптовалютах, передбачає наявність в кожному новому блоці значення геш-функції від попереднього, тобто кожен наступний блок залежить від усіх попередніх блоків. Крім того, за рахунок криптографічних властивостей геш-функції, обчислювально неможливо (за прийнятний час) підробити чи змінити якийсь блок таким чином, щоб значення геш-функції від цього блоку залишилось незмінним (властивість стійкості криптографічної геш-функції до колізій [7]). Тому якщо зловмисник бажає підмінити якусь інформацію, що міститься у блоці з номером n, за умови, що вже були опубліковані блоки з номерами $n+1,\dots,n+m$, то йому потрібно буде змінити всі і ці наступні блоки, оскільки зміниться значення геш-функції у кожному з наступних блоків. Зазначимо, що у «класичному»

децентралізованому блокчейні інформація про всі проведені раніше операції вноситься до певного реєстру, який зберігається не на одному центральному сервері, а у кожного активного учасника мережі, якого називають «нодом».

Смартконтракти, що функціонують у блокчейні, також є частиною цієї інформації. Все це робить смарт-контракти максимально інформативними (наприклад, можна подивитися інформацію про всіх власників нерухомості з моменту її будівництва) і захищеними від шахрайства, за умови, що більшість нодів є чесними учасниками (не можна зламати або підробити сервер з даними, так як інформація про угоди та їх умови є у всіх активних користувачів системи).

У криптовалютних блокчейнах ми можемо визначити смарт-контракт як додаток або програму, що працює на блокчейні. Як правило, він працює як цифрова угода, дотримання якої забезпечується певним набором правил. Ці правила визначені комп'ютерним кодом, який реплікується та виконується всіма нодами мережі.

Смарт-контракти блокчейну дозволяють створювати протоколи, які не потребують довіри сторін. Це означає, що дві сторони можуть брати на себе зобов'язання через блокчейн, не знаючи один одного та не довіряючи один одному. Вони можуть бути впевнені, якщо умови не будуть виконані, контракт не буде виконаний. Крім того, використання смарт-контрактів може усунути потребу в посередниках, що значно знижує операційні витрати [7].

Зазначимо, що повністю реалізувати концепцію смарт-контрактів в найпершому блокчейні Bitcoin неможливо внаслідок певних обмежень, закладених його розробником у його генезис-блок з метою безпеки. Тому активний розвиток smart-contracts отримали лише в 2013 році з появою цифрової валюти Ethereum та відповідного блокчейну [8]. Її засновник Віталій Бутерін розширив можливості створення смарт-контрактів в середовищі Ethereum, створивши універсальну децентралізовану блокчейн-платформу з можливістю програмування різних систем зберігання і обробки даних. При цьому не будь-яка угода може бути укладена у формі смартконтракту, а лише така, умови якої можна описані як математичні правила.

Смарт-контракт працює як детермінована програма, що виконує певне завдання, умови виконання якого описані у цьому контракті спеціальними математичними функціями на певній мові програмування. Зокрема, смарт-контракт часто використовує так звані умовні оператори "якщо… то…". Формально кажучи, смарт-контракти не є ні юридичними контрактами, ні розумними (смарт). Це просто програмний код, що записаний у розподіленому блокчейн-середовищі і виконання якого активується за певними правилами.

У мережі Ethereum смарт-контракти відповідають за виконання блокчейн-операцій та керування ними, коли користувачі (адреси) взаємодіють один з одною. Будь-яка адреса, яка не є смарт-контрактом, називається зовнішнім акаунтом (Externally-Owned Account - EOA). Таким чином, смарт-контракти контролюються комп'ютерним кодом, а EOA – користувачами. По суті, смарт-контракти Ethereum складаються з коду контракту та двох публічних ключів. Перший публічний ключ надається автором контракту. Інший ключ відповідає самому контракту, і є цифровим ідентифікатором, унікальним для кожного смарт-контракту. Активація будь-якого смарт-контракту можлива лише через транзакцію в блокчейні, при чому в цій транзакції обов'язково має бути присутнім цифровий підпис, виконаний на приватному ключі, що відповідає вказаному у контракті відкритому ключу. Смарт-контракт може бути активований або відповідним користувачем (відкритий ключ якого вказано у контракті), або іншим смартконтрактом (який, в свою чергу, було активовано легітимним користувачем). Тобто першим тригер для активації смарт-контракту завжди викликає є користувач.

Смарт-контракти на Ethereum як правило мають наступні характеристики.

Розподіленість. Смарт-контракти реплікуються та розподіляються у всіх нодах мережі Ethereum. Це одна з основних відмінностей від тих рішень, що використовуються на централізованих серверах.

Детермінованість. Смарт-контракти виконують лише ті дії, для яких вони були розроблені, і лише за умови виконання вимог, визначених контрактом. Крім того, результат завжди буде однаковим незалежно від того, хто їх виконує. Це унеможливлює маніпулювання контрактом та виключає «людський фактор» з його виконання.

Автономність. Смарт-контракти автоматизують виконання всіх видів завдань, які можуть бути задані програмно. Тому після активації вони працюють як автономні програми. До активації смарт-контракт знаходиться в "бездіяльності" і не виконує жодних дій.

Незмінність. Смарт-контракти не можна змінити після його розміщення у блокчейні. Але його можна "видалити" – правда, лише в тому випадку, якщо у ньому передбачена відповідна функція його видалення. Тому можна стверджувати, що смарт-контракти є захищеними від несанкціованого доступу до їх коду, в тому числі до його зміни.

Можливість налаштування. Хоча для програмування смарт-контрактів використовуються специфічні мови програмування (наприклад, Solidity [10]), отримані

з «класичних» шляхом суттєвого спрощення, їх можливості все одно дозволяють широкий вибір налаштувань для смарт-контракту. Ці налаштування стосуються дій, які смарт-контракт виконує, умов його активації, умов його видалення, умов виклику інших смарт-контрактів, тощо. Зокрема, у середовищі Eteriumсмарт-контракти можна використовувати для створення багатьох типів децентралізованих додатків (DApps).

Відсутність необхідності у третій довіреній стороні та взагалі у довірі між сторонами. Дві або більше сторін можуть взаємодіяти через смарт-контракти, не знаючи один одного та не довіряючи одна одній (тобто в умовах повної недовіри), оскільки блокчейн-технологія гарантує непідмінність всіх операцій та даних, що використовуються в смарт-контракті.

Прозорість. Оскільки смарт-контракти засновані на публічному блокчейні, їхній вихідний код не тільки незмінний, але й доступний всім [8].

Об'єкти смарт-контрактів.

Елементами «розумного» контракту є:

- сторони угоди, що мають цифровий підпис, які погоджуються або відмовляються від відповідності товару або послуги висунутим раніше вимогам;
- предмет договору - товар або послуги, які будуть відправлені в обмін на грошові кошти;
- умови, при дотриманні яких буде проведений автоматичний обмін благами, наприклад, відповідність поставленого товару стандартам якості, повинні мати повний математичний опис;
- децентралізована платформа, в якій написаний алгоритм (програмний код) самого смарт-контракту.

Види смарт-контрактів.

Залежно від рівня автоматизації, смарт-контракти можуть бути:

1) повністю автоматизованими;
2) переважно на паперовому носії, однак частина пунктів договору перенесені в смарт-контракт, наприклад, проведення транзакцій;
3) з копією на паперовому носії.

Для кращого розуміння специфіки функціонування та використання смарт-контрактів, проведемо їх порівняння зі стандартними «паперовими» договорами, які використовуються повсюдно.

Носій інформації:

- смарт-контракт: комп'ютерний алгоритм на платформі блокчейн;
- стандартний паперовий договір: папір.

На чому ґрунтується документ:

- смарт-контракт: програмний код;
- стандартний паперовий договір: норми права.

Можливість змінити умови:

- смарт-контракт: не можна змінити умови чинного контракту;
- стандартний паперовий договір: можна переписати, змінити (доп. угода), по-різному інтерпретувати.

Складнощі в укладанні контракту:

- смарт-контракт: висока, часто потрібен програміст;
- стандартний паперовий договір: середня, іноді потрібен юрист.

Виконання умов договору:

- смарт-контракт: виконуються автоматично усіма учасниками;
- стандартний паперовий договір: можуть бути не виконані сторонами, або робота може бути зроблена неякісно.

Застосування покарання:

- смарт-контракт: автоматично при настанні певних умов (виключено «людський фактор»);
- стандартний паперовий договір: суперечки зазвичай вирішуються через суд.

Наявність посередників:

- смарт-контракт: угоди проводяться без посередників;
- стандартний паперовий договір: часто необхідна допомога юриста, нотаріуса, участь державних установ.

Валюта розрахунків:

- смарт-контракт: зазвичай – криптовалюта, але можна налаштувати смарт-контракт для розрахунку «звичайними» фіатними валютами;
- стандартний паперовий договір: реальні гроші.

Час проведення операції:

- смарт-контракт: практично миттєво;
- стандартний паперовий договір: вимагає часу, особливо для міжнародних операцій.

Місцезнаходження сторін:

- смарт-контракт: контракт може бути підписаний без особистої присутності сторін, які можуть перебувати в будь-якій точці світу;
- стандартний паперовий договір: часто необхідна особиста зустріч представників сторін.

Ризик шахрайських операцій:

- смарт-контракт: практично виключений;
- стандартний паперовий договір: невеликий [6].

Обмеження при використанні смарт-контрактів

Смарт-контракти складаються із комп'ютерного коду, написаного людиною. Це пов'язано з численними ризиками, оскільки код потенційно може містити певні помилки. В ідеалі вони мають бути написані та розгорнуті досвідченими програмістами, особливо коли йдеться про конфіденційну інформацію або великі суми грошей. Крім цього, внаслідок обмеженості

мови програмування, що використовується для написання смарт-контрактів, централізовані системи можуть запропонувати більше рішень та функцій, ніж смарт-контракти. Основна відмінність полягає в тому, що смарт-контракти працюють у розподіленій P2P-мережі, а не на централізованому сервері. І оскільки вони засновані на блокчейн-системі, вони зазвичай або незмінні, або їх дуже важко змінити. Бути незмінним може бути добре в одних ситуаціях, але дуже погано в інших. Наприклад, коли в 2016 році зламали децентралізовану автономну організацію (DAO) під назвою The DAO, мільйони ETH були вкрадені через недоліки в коді смарт-контракту. Оскільки смарт-контракт був незмінним, розробники не могли виправити код. Зрештою це призвело до хард-форку, тобто до створення альтернативного блокчейну для Ethereum (Ethereum Classic) [9]. Простіше кажучи, в одному блокчейні результати атаки було скасовано і гроші повернуто законним власникам (це частина поточного блокчейну Ethereum), а в іншому результати атаки залишились незмінними (тепер цей блокчейн називається Ethereum Classic).

Важливо відзначити, що проблема виникла не через вразливість самого блокчейну Ethereum. Натомість це було викликано неправильною реалізацією смарт-контракту.

Ще одне обмеження у використанні смарт-контрактів пов'язане з їх невизначеним правовим статусом. Причина полягає не лише в тому, що блокчейн знаходиться в «сірій правовій зоні» для більшості країн, але й тому, що смарт-контракти не відповідають чинній правовій базі. Наприклад, часто для укладання контрактів необхідно, щоб обидві сторони були належним чином ідентифіковані та старші за 18 років. Псевдоанонімність, що забезпечується блокчейн-технологією, у поєднанні з відсутністю посередників, не дозволяє належним чином виконувати ці вимоги. Хоча існують потенційні шляхи вирішення цього питання, юридична сила смарт-контрактів є реальною проблемою, особливо коли йдеться про розподілені мережі, в яких відсутнє будь-яке централізоване регулювання.

Необхідність гарантій стійкості блокчейну до основних атак для врегулювання юридичного статусу смарт-контрактів

Оскільки смарт-контракти, що функціонують у блокчейні, мають ряд особливостей, які не притаманні їх «звичайним» аналогам, то природно, що ці властивості впливатимуть і на юридичні аспекти функціонування смарт-контрактів. У цьому розділі ми надамо необхідні умови, що висуваються до блокчейну, а також визначимо, коли смарт-контракт може вважатись зареєстрованим і таким, що не може бути відміненим.

Основними атаками на блокчейн вважається дві атаки – атака подвійної витрати [12], [13] та атака розгалуження [12], [14]. Вони суттєво відрізняються і за метою, і за технікою виконання, і за аналітичними оцінками імовірності успіху, але одна їх риса є спільною – наявність альтернативного ланцюжка достатньо великої довжини, який конкурує з основним і може суттєво змінити історію блокчейну. Це означає, що смарт-контракт може зникнути з цієї історії після свого укладання та/або після виконання. При цьому великою проблемою є те, що деякі дії, які виконував смарт-контракт і які виконувались on-chain, також зникнуть з історії, а ті які виконувались off-chain можуть, взагалі кажучи, не змінитись при зміні історії блокчейну.

Наприклад, якщо смарт-контракт переводив якість кошти з одного криптовалютного гаманця в інший, і це саме та криптовалюта, яка функціонує у цьому ж блокчейні, то ця дія буде відмінена; а якщо переводились банкові кошти, то зміна історії блокчейну навпаки не вплине на таку транзакцію. Або інший приклад: нехай у смарт-контракті було встановлено або змінено право власності на деякий об'єкт нерухомості. Якщо історію блокчейну буде змінено, то і ці зміни стосовно власності на нерухомість теж не збережуться.

Способи вирішення такої проблеми були запропоновані Накамото у [13] і розглядались з того часу у величезній кількості сучасних робіт (для ознайомлення можна скористатись [14], [15] та бібліографією в них). Основним таким способом є отримання так званих «блоків підтвердження», тобто очікування певної (інколи досить великої) кількості блоків підтвердження, які були створені після цього блоку у тому ж самому ланцюжку.

З наведених прикладів зрозуміло, що юридичне обґрунтування смарт-контрактів неможливе без їх захищеності, і в першу чергу від атак на блокчейн, який є його середовищем. Тому для набуття чинності смарт-контракту, аналогічно до вимоги до криптовалютних транзакцій, необхідна також певна кількість блоків підтвердження. Ця кількість суттєво залежить від структури блокчейну та алгоритму консенсусу, який в ньому використовується. Далі наведемо формули для визначення кількості блоків підтвердження, в залежності від типу консенсусу, на якому базується блокчейн.

Нижче наведемо аналітичні вирази, які, в залежності від алгоритму консенсусу у блокчейні, дозволяють не тільки оцінити імовірність атаки, що змінить історію блокчейну, а й визначити необхідну кількість блоків підтвердження, після виходу яких імовірність створити альтернативний ланцюжок блоків є нехтувано малою.

Імовірність атаки для протоколу консенсусу Proof-of-Work (враховуючи час затримки) [12].

Імовірність зміни історії блокчейну зловмисником, у якого доля обчислювальних потужностей становить p_M, після z блоків підтвердження дорівнює

$$P(z) = \begin{cases} 1, \quad if\ p'_M \geq p'_H; \\ 1 - \sum_{k=0}^{z} P_z(k)\left(1 - \left(\frac{p'_M}{p'_H}\right)^{z-k}\right), \text{інакше,} \end{cases}$$

де

$$P_n(k) = \frac{p_H^n}{(n-1)!} \cdot \frac{e^{-\alpha_M n D_H} \cdot (\alpha_M n D_H)^k}{k!} \cdot \sum_{i=0}^{k} \frac{(n-i+1)! \cdot C_k^i}{(\alpha n D_H)^i},$$

час затримки дорівнює D_H, і

$$p'_M = 1 - e^{-\alpha_M D_H} \cdot \frac{\alpha_H}{\alpha_M + \alpha_H} = 1 - e^{-\alpha_M D_H} \cdot p_H;$$

$$p'_H = e^{-\alpha_M D_H} \cdot \frac{\alpha_H}{\alpha_M + \alpha_H} = e^{-\alpha_M D_H} \cdot p_H.$$

Імовірність атаки для протоколу консенсусу Proof-of-Stake (у спрощеній моделі з одним слот-лідером та нульовим часом затримки) [15].

Імовірність зміни історії блокчейну зловмисником, у якого доля обчислювальних потужностей становить p_M, після z блоків підтвердження дорівнює

$$P_z = 2I_{p_M}(z+1, z+1),$$

де $I_x(a,a)$ – регулярна неповна бета функція.

Імовірність атаки для протоколу консенсусу Proof-of-Proof (PoW-PoW) [16].

Імовірність зміни історії блокчейну зловмисником, у якого доля обчислювальних потужностей становить p_M, після (n,k) блоків підтвердження у відповідних блокчейнах не перевищує

$$\left(\sum_{l=1}^{n-1}\left(\sum_{i=0}^{k-1} C_{n-l+i-1}^{i}\, {p_P}^i {p_I}^{n-l}\right) \times \frac{e^{-\alpha_I k D_P} \cdot (\alpha_I k D_P)^{l-1}}{(l-1)!} + \frac{e^{-\alpha_I k D_P} \cdot (\alpha_I k D_P)^{n-1}}{(n-1)!}\right) \times e^{-\alpha_I k D_P} \cdot \sum_{l=0}^{n-1} \frac{(\alpha_I k D_P)^l}{l!} + \left(1 - e^{-\alpha_I k D_P} \cdot \sum_{l=0}^{n-1} \frac{(\alpha_I k D_P)^l}{l!}\right),$$

де α_I – доля зловмисника у блокчейні 2-го рівня, а D_P – час синхронізації мережі у блокчейні 1-го рівня.

Імовірність атаки для протоколу консенсусу Proof-of-Proof (PoS-PoS) [5].

Якщо після виходу блоку $B_P(B_I)$ у блокчейні 1-го рівня (SPB-блокчейні) вийшло не менше за k блоків у блокчейні 2-го рівня (SIB-блокчейні), то імовірність успіху атаки на підміну запису у блоку B_I не перевищує величину

$$P(\gamma, k) = \gamma + e^{-\frac{\Delta_I}{\Delta_P} \cdot k} \cdot \sum_{l=0}^{z_\gamma - 1} \left(\frac{1}{l!} \cdot \left(\frac{\Delta_I}{\Delta_P} \cdot k\right)^l\right),$$

де Δ_P та Δ_I довжини таймслотів у відповідних блокчейнах, $\gamma \in (0,1)$ – межа імовірності підміни блоку у SPB-блокчейні, z_γ відповідна кількість блоків підтвердження.

Зазначимо, що для інших протоколів консенсусу аналогічні результати ще не отримані, і ми сподіваємось отримати їх частково у нашій подальшій роботі.

За наведеними результатами можна зробити такі висновки: використання смарт-контрактів, в тому числі і їх юридичне обґрунтування, повинно враховувати той факт, що смарт-контракт можна вважати «стабільним» (тобто таким, що не зникне з історії блокчейну) лише після певної кількості блоків підтвердження, що залежить від протоколу консенсусу та від параметрів мережі. Зазначимо, що такий висновок стосується лише децентралізованих блокчейнів, а у централізованому блокчейні така загроза практично відсутня. Тому варто подумати про створення централізованого блокчейну на рівня держави, який буде обслуговуватись довіреними особами та гарантувати стійкість інших, «дочірніх» блокчейнів, кожен з яких може запускатись певною юридичною особою (нотаріусом, магазином, державним підприємством, тощо) для обслуговування своїх специфічних смарт-контрактів.

Правовий статус смарт-контрактів у різних країнах та досвід їх використання

З юридичної точки зору, смарт-контракт представляє собою цифрову ітерацію цивільно-правової угоди, виражену у вигляді програмного коду, що автоматично виконується в розподіленій мережі. За принципом, схожим на традиційний цивільно-правовий договір, смарт-контракт складається зі сторін, предмету договору та ключових умов. Законна діяльність смарт-контракту визначається його здатністю відповідати основним вимогам цивільно-правового договору, включаючи наміри сторін в установленні юридичних відносин, ясність умов контракту та забезпечення можливості примусового виконання. Важливими ризиками, пов'язаними з смарт-контрактами, є невизначеність щодо їх юридичного статусу, а також криптовалюти як засобу оплати, складність вираження умови контракту у математичному алгоритмі, потенційна помилка у програмному коді та загроза хакерських атак. Щоб вирішити ці проблеми та забезпечити адекватне юридичне регулювання смарт-контрактів в Україні, необхідно внести зміни до законодавства, які визнають криптовалюту як засіб платежу та встановлюють смарт-контракти як форму цивільно-правової угоди. Необхідно забезпечити,

щоб сторони договору були чітко ідентифіковані і підтвердили свою згоду з умовами договору, використовуючи цифровий електронний підпис. Крім того, смарт-контракт повинен супроводжуватися обов'язковим текстовим додатком, в якому будуть викладені ключові положення договору, які матимуть юридичну силу в разі судової суперечки. Окрім цього, важливим кроком є перехід державних реєстрів на технологію блокчейн з закріпленням юридичної можливості автоматичного внесення змін до реєстрів при виконанні смарт-контракту.

Аналізуючи юридичний статус смарт-контрактів у світі, можна зробити висновок, що найбільш розвинене правове регулювання смарт-контрактів у США. «Хоч розумні контракти можуть звучати по-новому, сама їхня концепція корініться у базовому договірному праві. Зазвичай суд виносить рішення щодо договірних суперечкам і забезпечує дотримання умов, але часто зустрічається й інший метод, особливо для міжнародних транзакцій За допомогою розумних контрактів комп'ютерна програма забезпечує виконання контракту, вбудованого в код» (Звіт Комітету з економіки Сенату США, 2018). У 2018 році кілька штатів прийняли перші закони, що визнають смарт-контракти (закон штату Арізона). Їх приклад унаслідували й інші. Так, за законом штату Теннессі «визнається право використовувати технологію блокчейн та смарт-контрактів під час проведення електронних транзакцій». Закон захищає права власності на інформацію, «захищену технологією блокчейн». За законом «жоден контракт, що стосується транзакції, не може бути позбавлений юридичної сили, дійсності чи позовної сили лише тому, що цей контракт містить умову смарт-контракту».

У США правовий статус смарт-контрактів було визнано в кількох штатах, включаючи Арізону та Теннессі. Ці штати прийняли законодавство, яке підтверджує юридичну силу розумних контрактів відповідно до законодавства штату.

На федеральному рівні Закон про електронні підписи в глобальній і національній торгівлі (E-SIGN Act) і Уніфікований закон про електронні транзакції (UETA) потенційно можуть тлумачитися як такі, що застосовуються до смарт-контрактів. Обидва акти надають юридичне визнання електронним підписам і записам, які можуть включати код, що лежить в основі смарт-контракту.

Однак у США все ще існує багато невирішених юридичних проблем, пов'язаних із смарт-контрактами, зокрема питання щодо юрисдикції, захисту споживачів і того, як традиційні принципи договірного права, такі як пропозиція, прийняття та розгляд, застосовуються в контексті смарт-контрактів.

Загалом законодавство штатів про смарт-контракти можна поділити на три групи.

Перша. Створення та діяльність робочих груп або комітетів для вивчення проблем, пов'язаних з смарт-контрактами, регулюється в таких штатах як Каліфорнія, Нью-Джерсі, Нью-Йорк, Техас і кілька інших. Ці документи передбачають створення груп або комітетів, які вивчають смарт-контракти та супутні питання. Завданням цих комітетів є підготовка звітів, які подаються законодавчим органам штату. Потім на основі цих звітів законодавчі органи приймають рішення про прийняття відповідного законодавства.

Друга. Закріплення основних концепцій смарт-контрактів. В кількох штатах, зокрема Арканзасі, Меріленді, Неваді, Нью-Йорку, Оклахомі, Південній Дакоті, Техасі та Юті, було прийнято законодавство, яке створює основу для використання смарт-контрактів у сфері торгівлі. Це відбувається шляхом додавання відповідних визначень до комерційного (торгового) кодексу кожного штату і визнання правомірності смарт-контрактів за певних умов.

Третя. Всебічне та детальне закріплення у законодавстві смарт-договорів. Тут штат Іллінойс видається лідером на глобальній арені. У 2020 році тут було прийнято Закон про технологію блокчейну. Цей закон визнає дійсність смарт-контрактів та записів на основі блокчейну в торгових відносинах. Важливо зазначити, що закон не дозволяє відмовляти смарт-контракту визнання юридичної дійсності або виключати його як доказ у суді тільки з причини використання ланцюжка комп'ютерних блоків для створення, зберігання або перевірки контракту. Крім того, закон обмежує можливість використання блокчейну в ситуаціях, коли інший федеральний або місцевий нормативний акт вимагає, щоб запис був відображений або переданий певним чином, або коли використання блокчейну може заважати іншій стороні зберігати або отримувати інформацію з блокчейну.

Окрім цього, варто зазначити, що поява смарт-контрактів породила нову хвилю юридичних і регуляторних міркувань як у Сполучених Штатах, так і у Великій Британії. Оскільки ці юрисдикції борються з правовим статусом смарт-контрактів, стає важливим вивчити схожість і відмінності в їхніх підходах. У цій статті ми досліджуємо правовий ландшафт, що оточує смарт-контракти в США та Великобританії, проливаючи світло на ключові нормативні рамки, судові рішення та нові тенденції. Ця стаття спрямована на те, щоб зрозуміти відмінні перспективи цих юрисдикцій, надати уявлення про зміну правового статусу смарт-контрактів і передбачити потенційні наслідки для компаній і фізичних осіб в інших юрисдикціях.

Проблеми правового регулювання смарт-контрактів та потенційні шляхи вирішення деяких з них

Зазначимо, що розуміння юридичних проблем при використанні смарт-контрактів виникло не зразу. Так, на початку так званої «інтернет-революції» виникла доктрина "код – це закон", згідно з якою функціонування кіберпростору визначається комп'ютерним кодом і не потребує юридичного регулювання . Дана доктрина не витримує критики з огляду на ряд причин. По-перше, існують проблеми трансформації юридичних правил у технічні. Прості однозначні угоди можна без проблем викласти в математичному алгоритмі, але це важко реалізувати для складних угод, що потребують інтерпретації, оцінювання і прийняття рішень під час їх виконання. По-друге, реалізація прав і виконання обов'язків, установлених у віртуальному світі, потребує юридичного забезпечення у реальному світі. По-третє, доктрина "код – це закон" передбачає, що сторони контракту приймають наслідки виконання програмного коду як істину в останній інстанції без можливості його зупинення, перегляду і виправлення у випадку необхідності. Це є найбільшим обмеженням застосування смарт-контрактів на сучасному етапі розвитку їх технології. Адже програмний код може не відповідати реальному волевиявленню сторін, містити помилки і навіть бути об'єктом злочинної хакерської атаки з негативними наслідками, як у наведеному вище прикладі з проектом "DAO". Згідно з висновками, викладеними в аналітичному дослідженні Центрального банку Європейського Союзу, юридична сила смарт-контракту залежить від того, чи має він базові властивості цивільно-правової угоди, основними з яких є докази того, що сторони мали наміри встановити цивільно-правові відносини, зрозумілість умов контракту і можливість зовнішнього примусу до виконання умов контракту. Наміри встановити цивільно-правові відносини є ключовою умовою існування будь-якого договору. Можливість визначити наявність намірів встановити цивільно-правові відносини у випадку зі смарт-контрактом значною мірою залежить від типу розподіленої мережі, у якій функціонує код смарт-контракту, – обмеженої чи необмеженої. В обмеженій мережі функціонує обмежена кількість учасників, які відповідають певним умовам, проходять ідентифікацію і перевірку та можуть довіряти один одному. У цьому випадку учасники смарт-контракту, погоджуючись на ідентифікацію, підтверджують свої наміри встановити цивільні права й обов'язки. У випадку необмеженої децентралізованої мережі кількість контрагентів не обмежена, вони не зобов'язані відповідати жодним критеріям і не проходять ідентифікацію, учасники довіряють не один одному, а самій технології функціонування смарт-контракту. У цьому випадку довести наявність намірів вступити у цивільні правовідносини з невідомими контрагентами може бути дуже складно. Ще однією проблемою ідентифікації намірів встановити цивільні правовідносини є автоматичне виконання контракту відповідно до програмного коду без можливості контрагентами оцінювати ситуацію і приймати певні рішення. У традиційних договорах активна участь сторін у визначенні своєї поведінки і ходу виконання контракту є підтвердженням їх волевиявлення. У випадку смарт-контракту сторони позбавлені можливості такого волевиявлення, хоча для розв'язання цієї проблеми при розробці смарт-контракту можна передбачити певні моменти, коли виконання контракту за наявності спільної згоди сторін буде зупинятися для можливості введення нових параметрів подальшого виконання.

Наступною важливою умовою юридичного існування смарт-контракту, згідно з висновками Центрального банку Європейського Союзу, є чіткість і зрозумілість його умов. Лише той смарт-контракт має юридичну силу, у програмному коді якого реалізовано умови, яких дійсно хотіли досягти сторони. Це легко досяжно в простих контрактах, таких як купівля-продаж товарів. Але в складних контрактах, де сторони навмисно включають неоднозначні умови, що потребують оціночних суджень і прийняття рішень під час виконання, надають їм певну гнучкість і можливість різних інтерпретацій, математично зафіксувати всі ці умови стає вкрай проблематичним завданням. Іншою проблемою виконання чіткості умов смартконтракту є його зовнішнє інформаційне забезпечення. Програма, що виконує смарт-контракт, самостійно шукає інформацію у заданих наперед джерелах і використовує її для визначення алгоритму виконання контракту. Верифікація і підтвердження коректності використаної програмою інформації покладається на програму, і у випадку автоматичного використання помилкової інформації смарт-контракт може призвести до наслідків, які не відповідають волі сторін угоди.

Можна виділити наступні етапи утворення смарт-договору:

1) оформлення домовленості в електронній формі, яка візуально може бути сприйнята сторонами (протокол);

2) компіляція у програмний код;

3) опублікування у блокчейні Ethereum або іншій системі.

У юридичній літературі висловлюються різні точки зору щодо визначення смартдоговору.

Смарт-договір – програмний код. Розумний контракт – це договір, який існує в формі програмного коду, що імплементовано на

платформі Blockchain, який забезпечує автономність і самовиконання умов такого договору у разі настання заздалегідь визначених у ньому обставин. Дане розуміння не відповідає нормам ЦК України. Договір є домовленістю двох або більше сторін, спрямованою на встановлення, зміну або припинення цивільних прав та обов'язків (ч. 1 ст. 626 ЦК України). Правочин може вчинятися усно або в письмовій (електронній) формі (ст. 205 ЦК України). Відповідно до п. 3 ст. 3 ЗУ «Про електронну комерцію», електронною формою представлення інформації вважається документування інформації, що дає змогу її відтворювати у візуальній формі, придатній для сприйняття людиною. Відповідно, за допомогою програмного коду сторони не можуть сприйняти умови, стосовно яких вони досягли згоди;

Смарт-договір може функціонувати як автоматизована система, спрямована на виконання договору, який був укладений в іншій формі. Сьогодні багато банків, телекомунікаційних операторів та інших підприємств активно використовують автоматизовані системи для виконання своїх договірних зобов'язань. Важливо розуміти, що ця позиція не передбачає розгляд смарт-договору як самостійної договірної конструкції; він слугує лише засобом автоматичного виконання раніше укладених угод. Наше бачення полягає в тому, що смарт-договір є правочином, який може бути спрямованим не лише на виконання існуючих договірних зобов'язань, але й на їх створення;

Смарт-договори – інноваційна форма контрактів, укладення, виконання та припинення яких відбувається з використанням мережевих комп'ютерних програмних та/або програмно-апаратних засобів, що мають взаємозв'язок з фізичними або цифровими об'єктами, за участю або без участі людини, що вимагає проведення системних і комплексних правових досліджень в рамках цивільного, фінансового та інформаційного права. Дана позиція передбачає потребу у виокремленні ще одного виду правочину, крім вже передбачених усної та письмової, – електронного. Електронною формою смарт-договору слід розуміти електронний документ, в якому сторони закріплюються істотні умови договору і який може бути візуально сприйнятий. Особливістю смарт-договору є те, що він набирає чинності з моменту компіляції програмного коду і розміщення в Ethereum або іншій системі. Зважаючи на те, що сторони договору не мають можливості перевірити коректність трансформації умов договору в програмний код, необхідним є збереження електронної форми договору, яка є основою для формування коду. Якщо через технічні неполадки відбувається спотворення інформації, смарт-контракт може бути визнаний недійсним. Електронна форма смарт-договору може слугувати доказом у судових справах у випадку виникнення конфліктів. Також не виключається можливість залучення фахівців для аналізу вмісту програмного коду.

Особливості укладення, виконання, зміни смарт-договорів, які дозволяють визначати смарт-договір як окремий вид договорів

Умови смарт-договору повинні бути конкретними (чіткими, однозначними), здійсненними (об'єктивними), правомірними. Умови смарт-договору можуть бути сформульовані за формулою «якщо..., тоді...». Не всі умови договору можуть бути автоматизовані. Не можуть використовуватися оціночні терміни, наприклад «прийнятний термін виконання». Смарт-договір існує в межах платформи блокчейну, тому можливі, як правило, трансакції в межах даної платформи (здійснення трансакцій криптовалюти). Умови смарт-контракту можуть передбачати забезпечувальні зобов'язання, зокрема, коли згідно зі смарт-договором блокується певна сума криптовалюти, перерахунок якої кредитору здійснюється лише після вчинення певних дій боржником. Це одна з основних переваг смарт-договору. Проблемним є отримання і підтвердження інформації зовні, тому умови про форс-мажорні обставини також не зазначаються в смарт-договорі. Тому можливими є два варіанти оформлення смарт-договорів: електронна форма повністю трансформована в програмний код або лише частина електронного договору відображена в програмному коді. Оскільки узгоджені сторонами умови домовленості трансформуються надалі в програмний код, необхідним є напрацювання програмістами разом з юристами шаблонів, протоколів щодо окремих видів договорів.

Виконання зобов'язань, передбачених смарт-договором, здійснюється в автоматичному порядку цифрових трансмісій у певній послідовності у разі настання певних обставин. Виконання смарт-договору здійснюється без участі сторін, на відміну від звичайного договору, коли сторони зобов'язані вчиняти певні дії.

Засобом розрахунків за смартконтрактами є криптовалюта. Одним із стримуючих факторів поширення смарт-договорів є відсутність правового регулювання статусу криптовалюти в Україні. Слід зазначити, що смарт-договір може забезпечити виконання лише зобов'язань у віртуальному світі, наприклад щодо проведення платежів криптовалютою. Виконання робіт та надання послуг здійснюється в реальному світі і не може бути опосередковано смарт-договором. Слід зазначити, що перед тим, як смарт-договір буде виконаний, потрібно оплатити трансакційну комісію.

З метою розширення сфери застосування смарт-договорів доцільно було б забезпечити доступ до інформації з Державного реєстру речових прав на нерухоме майно, Державного реєстру юридичних осіб, фізичних осіб – підприємців, громадських формувань та інших реєстрів.

Відсутня можливість змінити договір. Якщо смарт-договір почав виконуватися, його неможливо припинити або змінити. Це зумовлено технологією блокчейн. З одного боку, це є плюсом, адже сторона не може ухилитися від виконання угоди, а з іншої – мінусом, оскільки помилка або відсутність певної умови може унеможливити виконання домовленостей, які б відповідали б волевиявленням сторін. В останньому випадку можливе лише укладення нового смарт-договору.

З метою розв'язання проблем правового регулювання смарт-контрактів в Україні є необхідність внести зміни в законодавство, які дозволять урегулювати правовий статус криптовалют і визнати смарт-контракти формою цивільно-правового договору. Необхідно передбачити обов'язкову ідентифікацію сторін договору, підтвердження їх згоди з умовами договору шляхом накладення цифрового електронного підпису. Смарт-конракт має доповнюватися обов'язковим текстовим додатком з викладенням істотних умов договору, які матимуть юридичне значення при судовому розгляді. Потрібне переведення державних реєстрів на технологію блокчейн з юридичним закріпленням можливості автоматичного внесення змін до реєстрів при виконанні смартконтракту. Це дозволить не тільки знизити витрати й підвищити ефективність функціонування реєстрів, а й суттєво знизить рівень зловживань із реєстрами, які часто трапляються на практиці.

Для того, щоб досягти правового регулювання смарт-контрактів, сторони цивільних правовідносин, які уклали таку угоду, повинні мати можливість юридичного захисту порушених, невизнаних чи спірних прав. І тут виникає ряд проблем, без вирішення яких неможливо повністю гармонізувати смарт-контракти з правовим полем.

Найбільш суттєвими проблемами можна вважати такі:

- відсутність законодавчих документів та актів, що регулюють взаємовідносини сторін, що взаємодіють у смарт-контракті;
- складність аналізу коду, який відповідає смарт-контракту;
- складність аналізу вразливостей смарт-контракту та блокчейну, в якому він функціонує;
- невизначеність ступені відповідальності як за порушення контракту, так і за некоректне створення смарт-контракту та за навмисне чи ненавмисне нехтування можливими вразливостями, що виникають при його використанні, та атаками на смарт-контракт чи на відповідний блокчейн.

Нижче наведемо кілька можливих шляхів вирішення та їх порівняльний аналіз.

Повне дублювання цифрового смарт-контракту відповідним паперовим документом. Якщо ставити питання про повне юридичне ототожнення цифрових угод зі звичайними «паперовими», то тоді у відповідність кожній цифровій угоді потрібно поставити паперову, яка буде повним аналогом цифрової. Тут найбільша проблема полягає у тому, що особи, які за своїми обов'язками повинні захищати права сторін, що укладають угоду в цифровому вигляді, взагалі кажучи, не мають достатньої кваліфікації за програмним кодом (який і є смарт-контрактом) визначити, яка ж насправді угода укладена. Тобто для вирішення питання таким методом необхідна участь особи, яка є, в певному сенсі, «посередником» між програмістом, який пише відповідний смарт-контракт, і юристом або нотаріусом, який буде створювати, підписувати, засвідчувати відповідну паперову копію.

Переваги: повне юридичне врегулювання укладання контракту.

Недоліки:

1) в цьому випадку втрачається сама сутність смарт-контракту, яка полягає в тому, що для його укладання не потрібен посередник;

2) більш того, крім посередника-юриста потрібен ще посередник-«перекладач» з програмного коду на мову, що використовується для юридичних документів.

Створення юридично завірених шаблонів. Для найбільш поширених та актуальних смарт-контрактів створюється їх юридичний аналог, як у попередньому варіанті. Але цей аналог не один для кожного окремого контракту, а один для кожного типу контрактів. Тобто при укладання такого «типового» контракту сторони, грубо кажучи, беруть готовий шаблон з відкритим кодом, який має щось на зразок сертифікату, завіреного відповідальною особою, в якому детально описані умови контракту, і вписують у нього свої відповідні дані, які відповідають цьому конкретному контракту. Наприклад, якщо мова йде про контракт на p2p-закупівлю «зеленої» енергетики між домогосподарствами, то сторони узгоджують лише ціну 1 кВт електроенергії, кількість поставленої енергії та, можливо, якісь додаткові дані, такі як проміжок часу, протягом якого ця електроенергія має бути поставлена, частини, на які розбивається оплата, номера банківських карточок, з яких списуються кошти для оплати, відкриті ключі підписантів контракту, тощо.

Переваги: повне юридичне врегулювання укладання контракту та достатньо виконати юридичне дублювання типового контракту один

раз, а потім необмежену кількість разів використовувати.

Недоліки:

1) потрібні посередники (але лише один раз на кожен тип контракту);

2) якщо контракт не є типовим, навіть якщо зовсім трохи відрізняється від типового, то для його укладання теж потрібно один раз пройти описану вище процедуру юридичної верифікації, ну а потім вже можна використовувати цей новий тип контрактів. Більш того, крім посередника-юриста потрібен ще посередник-«перекладач» з програмного коду на мову, що використовується для юридичних документів.

Як бачимо, перший спосіб є більш універсальним, але при цьому втрачаються основні переваги смарт-контрактів (відсутність посередників); другий спосіб має дещо обмежене застосування, але дозволяє (після отримання сертифікату на смарт-контракт) укладати контракти без посередників.

Висновки

Деякі блокчейн-ентузіасти розглядають смарт-контракти як рішення, яке незабаром замінить та автоматизує більшу частину наших комерційних та бюрократичних систем. Хоча це можлива реальність, вона, певно, далека від того, щоб стати нормою. Смарт-контракти, безперечно, цікава технологія. Але будучи розподіленими, детермінованими, прозорими і певною мірою незмінними, вони можуть бути менш привабливими у деяких ситуаціях. Фактично, вся критика спирається на те що, що смарт-контракти не є правильним рішенням для безлічі реальних проблем. Насправді, деяким організаціям краще використовувати звичайні серверні альтернативи. Також слід зазначити, що важливим є питання стабілізації блоку(ів), в якому виконується смарт-контракт, і без забезпечення його стабілізації неможливо вирішити питання про юридичну значущість смарт-контрактів.

У порівнянні зі смарт-контрактами, централізовані сервери простіші і дешевші в обслуговуванні, і вони, як правило, мають вищу ефективність з точки зору швидкості та міжмережевої взаємодії (функціональної сумісності).

Доцільно доповнити ЦК України статтею 636-1 «Смарт-договір» з таким змістом.

Смарт-договір - це угода, яка укладається у формі електронного документа та набуває чинності після компіляції програмного коду та його розміщення в Ethereum або іншій системі з використанням технології блокчейну.

Ідентифікація сторін у смарт-договорі вважається еквівалентною використанню цифрових підписів в системі.

Для укладення смарт-договору застосовуються загальноприйняті протоколи, і умови повинні бути конкретними, виконуваними та законними.

Якщо програмний код реалізує умови, що не відповідають змісту електронного документа, смарт-договір може бути визнаний недійсним.

Виконання зобов'язань, передбачених смарт-договором, автоматично відбувається шляхом виконання цифрових передач у визначеній послідовності при настанні зазначених в договорі обставин.

Криптовалюта є законним засобом оплати при виконанні смарт-договорів.

Список літератури

1. Задобрюк Д. Що таке смарт-контракти і які принципи їх роботи. URL: https://lexinform.com.ua/yuridychna-praktyka/shho-take-smart-kontrakty-i-yaki-pryntsypy-yih-roboty/
2. Hilliard J. SIDECHAINS. URL: https://ethereum.org/en/developers/docs/scaling/sidechains/
3. Agrawal A. Product Types of Smart Contracts Market: Bitcoin, Sidechains, NXT, Ethereum. URL: https://customerthink.com/product-types-of-smart-contracts-market-bitcoin-sidechains-nxt-ethereum/
4. Кондратенко М.С. Використання технології блокчейну для побудови ієрархічної структури на множині державних реєстрів з метою захисту від підробки інформації. Електронне моделювання, 3(23), 2023. С. 43–56. URL: https://www.emodel.org.ua/images/em/45-3/45-3-4.pdf
5. Кондратенко М.С. Визначення кількості блоків підтвердження у блокчейні, в якому розміщено реєстр другого рівня у випадку, коли в обох блокчейнах використовується протокол консенсусу POS. XLI Науково-технічна конференція молодих вчених та спеціалістів інституту проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України. Збірник матеріалів конференції, 2023. С.188–190. URL: https://ipme.kiev.ua/wp-content/uploads/2023/05/Матеріали-конференції-2023.pdf6.
6. Що таке смарт-контакт? URL: https://bankchart.com.ua/finansoviy_gid/investitsiyi/statti/scho_take_smart_kontrakt
7. Що таке хешування? URL: https://academy.binance.com/uk/articles/what-is-hashing
8. Що таке смарт-контакти? URL:https://academy.binance.com/uk/articles/what-are-smart-contracts
9. Що таке Ethereum? URL: https://ethereum.org/uk/what-is-ethereum/
10. Solidity. A statically-typed curly-braces programming language designed for developing smart contracts that run on Ethereum, URL:https://docs.soliditylang.org/en/v0.8.21/

11. Зелений тариф. URL: https://uk.wikipedia.org/wiki/%D0%97%D0%B5%D0%BB%D0%B5%D0%BD%D0%B8%D0%B9_%D1%82%D0%B0%D1%80%D0%B8%D1%84
12. Kovalchuk L., Kaidalov D., Nastenko A., Rodinko M., Oliynykov R. Probability of double spend attack for network with non-zero synchronization time. 21th Central European Conference on Cryptology CECC, 2021. P. 52-54.
13. Nakamoto S. Bitcoin: A Peer-to-Peer Electronic Cash System. URL: https://bitcoin.org/bitcoin.pdf
14. Fitzi M., Gazi P., Kiayias A., Russell A. "Parallel chains: Improving throughput and latency of blockchain protocols via parallel composition", Cryptology ePrint Archive, Report 2018/1119, 2018. URL: https://eprint.iacr.org/2018/1119
15. Karpinski M., Kovalchuk L., Kochan R., Rodinko M., Wieclaw L. Blockchain technologies: Probability of double-spend attack on a proof-of-stake consensus. Sensors, 2021, 21(19). 6408. URL: https://www.scopus.com/record/display.uri?eid=2-s2.0-85115652254&origin=resultslist&sort=plf-f
16. Kovalchuk L., Kostanda V., Marukhnenko O., Pozhylenkov O. Achieving Security in Proof-of-Proof Protocol with Non-Zero Synchronization Time. Mathematics, 2022. 10(14). 2422 p.

References

1. Zadobiuk, D. "What are smart contracts and what are the principles of their operation?" ["Shcho take smart-kontrakty i yaki pryntsypy yikh roboty?"], available at: https://lexinform.com.ua/yuridychna-praktyka/shho-take-smart-kontrakty-i-yaki-pryntsypy-yih-roboty/
2. Hilliard, J. "SIDECHAINS", available at: https://ethereum.org/en/developers/docs/scaling/sidechains/
3. Agrawal, A. "Product Types of Smart Contracts Market: Bitcoin, Sidechains, NXT, Ethereum", available at: https://customerthink.com/product-types-of-smart-contracts-market-bitcoin-sidechains-nxt-ethereum/
4. Kondratenko, M. (2023), "Using blockchain technology to build a hierarchical structure on multiple state registers in order to protect against information forgery" ["Vykorystannia tekhnolohii blokcheinu dlia pobudovy iierarkhichnoi struktury na mnozhyni derzhavnykh reiestriv z metoiu zakhystu vid pidrobky informatsii"], *Electronic modeling – international scientific-theoretical journal* №3, 23, Kyiv, pp. 43-56.
5. Kondratenko, M. (2023), "Determination of the number of confirmation blocks of the blockchain in which the registry of the second level is placed in the case when both blockchains use the POS-consensus protocol" ["Vyznachennia kilkosti blokiv pidtverdzhennia ublokcheini, v yakomu rozmishcheno reiestr druhoho rivnia u vypadku, koly v obokh blokcheinakh vykorystovuietsia protokol konsensusu POS"], *XLI scientific and technical conference of young scientists and specialists of the G.E. Pukhov Institute for Modelling in Energy Engineering, Materials of the conference,* pp. 188-190.
6. "What is a smart contract?" ["Shcho take smart-kontrakt?"], available at: https://bankchart.com.ua/finansoviy_gid/investitsiyi/statti/scho_take_smart_kontrakt
7. "What is hashing?" ["Shcho take heshyvannya?"], available at: https://academy.binance.com/uk/articles/what-is-hashing
8. "What are smart-contracts?" ["Shcho take smart-contrakty?], available at: https://academy.binance.com/uk/articles/what-are-smart-contracts
9. "What is Ethereum?" ["Shcho take Ethereum?"], available at: https://ethereum.org/uk/what-is-ethereum/
10. "Solidity. A statically-typed curly-braces programming language designed for developing smart contracts that run on Ethereum", available at: https://docs.soliditylang.org/en/v0.8.21/
11. "Green tariff" ["Zelenyi taryf"], available at: https://uk.wikipedia.org/wiki/%D0%97%D0%B5%D0%BB%D0%B5%D0%BD%D0%B8%D0%B9_%D1%82%D0%B0%D1%80%D0%B8%D1%84
12. Kovalchuk, L., Kaidalov, D., Nastenko, A., Rodinko, M., Oliynykov, R. (2021), "Probability of double spend attack for network with non-zero synchronization time", *21th Central European Conference on Cryptology CECC*, pp. 52-54.
13. Nakamoto, S. "Bitcoin: A Peer-to-Peer Electronic Cash System", available at: https://bitcoin.org/bitcoin.pdf
14. Fitzi, M., Gazi, P., Kiayias, A., Russell, A. (2018), "Parallel chains: Improving throughput and latency of blockchain protocols via parallel composition", *Cryptology ePrint Archive, Report 2018/1119, 2018*, available at: https://eprint.iacr.org/2018/1119
15. Karpinski, M., Kovalchuk, L., Kochan, R., Rodinko, M., Wieclaw, L. (2021),"Blockchain technologies: Probability of double-spend attack on a proof-of-stake consensus", *Sensors*, 21(19), 6408, available at: https://www.scopus.com/record/display.uri?eid=2-s2.0-85115652254&origin=resultslist&sort=plf-f
16. Kovalchuk, L., Kostanda, V., Marukhnenko, O., Pozhylenkov, O. (2022), "Achieving Security in Proof-of-Proof Protocol with Non-Zero Synchronization Time", *Mathematics*, 10(14), 2422

Надійшла до редакції 12.02.2023

M. KONDRATENKO
G.E. Pukhov Institute for Modelling in Energy Engineering, National Academy of Sciences of Ukraine, Kyiv, Ukraine
nikolay.ns95@gmail.com

COORDINATION OF MATHEMATICAL AND LEGAL ASPECTS WHEN USING SMART CONTRACTS

The article examines the issue of the legal status of smart contracts in Ukraine, problems and obstacles that arise when smart contracts are harmonized with the legislation of Ukraine. Considering the reliability and transparency of blockchain technology, which is the basis of smart contracts, their implementation at the legislative level can significantly improve the process of interaction between the parties in conditions of complete mistrust. Blockchain smart contracts allows to create protocols that do not require trust between the parties. This means that two parties can make commitments through the blockchain without knowing or trusting each other. They can be sure that if the conditions are not met, the contract will not be fulfilled. In addition, the use of smart contracts can eliminate the need for intermediaries, which significantly reduces operational costs The work provides basic information about smart contracts, their characteristics and features of application, compares smart contracts with traditional contracts taking into account the advantages and disadvantages of each of them, analyzes the legal experience of countries in which smart contracts are actively used, and proposes solutions problems that could arise

Key words: blockchain, smart contract, smart contract, smart contract, cryptocurrency, Bitcoin, Ethereum, legal status of smart contracts, legal regulation of smart contracts

УДК 004.49

[1,2] Д.І. Вініченко, магістр,
[2]АТ «МОТОР СІЧ» м. Запоріжжя, Україна
[3]М.Ю. Тягунова, канд. техн. наук, доцент,
[4]Т.В. Голуб Т.В., канд. техн. наук, доцент
[1,3,4] Національний університет «Запорізька політехніка» м. Запоріжжя, Україна
vinichenko92dima@gmail.com
mary.tyagunova@gmail.com
golub.tv6@gmail.com

Вибір оптимальної CAD-системи для розробки і проєктування ендопротезів

Метод ендопротезування суглобів є одним із найкращих варіантів оперативного лікування опорно-рухового апарату у пацієнтів. Головна мета цієї операції полягає у заміні патологічно змінених структур суглоба на штучний ендопротез, що має анатомічну форму здорового суглоба та забезпечує можливість виконання повноцінного обсягу рухів. Дана робота спрямована на вибір оптимальної CAD-системи для розробки і проєктування ендопротезів. У результаті аналізу систем для наочності порівняння виявлених особливостей, розроблено порівняльну таблиці. У ній чітко видно, яка саме система є найбільш ефективною саме для проєктування ендопротезів.

Ключові слова: проєктування, CAD-система, ендопротезування, якість, 3D-моделювання

DOI: 10.31474/1996-1588-2023-2-37-67-71

Вступ

Операція ендопротезування є одним з кращих методів лікування захворювань опорно-рухового апарату. Головною метою процедури є заміна ураженого суглоба на штучний ендопротез, який має анатомічну форму здорового суглоба та забезпечує можливість повноцінного руху. У багатьох випадках це остання надія для пацієнта на відновлення функцій опорно-рухового апарату та продовження життя (наприклад, при онкологічних ураженнях) [1]. Для якісного створення ендопротезу першочерговим є розробка його 3D-моделі. Швидкість та якість такого моделювання безпосередньо пов'язана з CAD-системою, на якій це буде відбуватися. Тому у даній роботі саме розглядається аналіз CAD-систем з метою виявлення найбільш вдалої саме для завдання ендопротезування.

Аналіз останіх досліджень та публікацій

За світовою статистикою, річна потреба в ендопротезуванні суглобів складає від 500 до 1000 на кожен мільйон населення. У розвинених країнах Європи ця цифра ще вища. Проте в Україні проводять у 10 разів менше операцій, ніж необхідно [1-2]. Це пов'язано з відсутністю якісних вітчизняних ендопротезів. Важливість ендопротезування суглобів в Україні набуває особливого значення в контексті бойових дій та отриманих поранень воїнів та мирних громадян.

CAD-системи є досить поширеними і використовуються для створення різноманітних деталей та складних механізмів, включаючи ендопротези. Вони надають інженерам-конструкторам зручні інструменти для створення, модифікації, аналізу та оптимізації деталей та конструкцій, що зменшує кількість етапів розробки та поліпшує якість виробів.

CAD-системи дозволяють виробляти точні 3D-моделі ендопротезів, з яких можна виготовляти прототипи та зразки методами 3D-друку або фрезерування. Крім того, ці системи дозволяють проводити віртуальне тестування ендопротезів, що зменшує кількість потрібних фізичних тестів та поліпшує їхню ефективність.

Враховуючи широкий спектр функцій, які надають CAD -системи, їх використання може бути досить складним для новачків. Однак, з достатньою практикою та підготовкою, інженери-конструктори можуть ефективно використовувати ці інструменти для розробки та вдосконалення ендопротезів.

Метою даної роботи є порівняльний аналіз CAD-систем, що дозволить полегшити фахівцям з галузі комп'ютерного протезування зробити вибір оптимальної системи автоматизованого проєктування для розробки і проєктування ендопротезів.

Аналіз CAD-систем

Проаналізуємо найбільш поширені CAD-системи та їх особливі риси.

Fusion 360 – це хмарна система автоматизованого проектування, розроблена компанією Autodesk [3]. Система може виконувати широкий спектр завдань, таких як 2D-та 3D-моделювання, робота з збірками, обчислення та оптимізація. Основні особливості включають сучасний інтерфейс користувача, покращені інструменти збірки, автономний режим, доступну цінову політику, можливості симуляції, 3D друк та вбудовану підтримку комп'ютерної підтримки виробництва (CAM).

Таким чином, Fusion 360 наступні переваги:

- є можливість працювати над проектами з будь-якого місця з доступом до інтернету;
- має великий функціонал: багато інструментів для вирішення різних задач, таких як моделювання в 2D та 3D, складання, обчислення та оптимізація;
- програма має вбудовані інструменти оптимізації для зменшення часу на вирішення задач;
- дозволяє проводити імітацію різних процесів, що відбуваються в системі;
- дозволяє автоматизувати процеси створення деталей;
- простий і зручний для використання модерний інтерфейс;
- наявність різноманітних тарифних планів з різними можливостями, в залежності від потреб користувача.

Недоліками Fusion 360 при цьому є:

- залежність від Інтернету, оскільки може бути проблемою для тих, хто не має постійного доступу до Інтернету.
- відсутність можливості використання програми безкоштовно: Fusion 360 пропонує пробний період, але після його закінчення, користувач повинен купити підписку на програму.
- багато функцій програми можуть бути складними для новачків.

Не підходить для деяких галузей – Fusion 360 не є ідеальним рішенням для всіх галузей та видів дизайну, зокрема для дизайну електронних пристроїв та програмного забезпечення.

SolidEdge - система моделювання твердих та поверхневих тіл, розроблена компанією Siemens PLM Software [4]. Серед її переваг можна відзначити параметричне, синхронне та комбіноване середовище моделювання, інструменти для креслення згідно з європейськими стандартами, можливість розробляти литі деталі та обладнання для їхнього виробництва, 3D друк, кінематичний аналіз механізмів та бібліотеку стандартних виробів.

Переваги SolidEdge включають:

- параметричне, синхронне та комбіноване середовище моделювання, яке дозволяє швидко та зручно створювати моделі будь-якої складності;
- інструменти для створення креслень відповідно до європейських стандартів;
- можливість розробляти відливки та обладнання для їх виготовлення;
- 3D-друк;
- кінематичний аналіз механізмів;
- бібліотека стандартних продуктів.

Недоліки SolidEdge:

- в порівнянні з іншими CAD-системами він може бути менш зручним у використанні;
- він не має такої популярності, як, наприклад, AutoCAD чи SolidWorks, що може змінювати динаміку ринку робочих місць;
- він може бути менш підходящим для розробки деяких типів проектів, порівняно з іншими CAD-системами.

AutoCAD, розроблений та постачається компанією Autodesk, є популярною системою CAD з найбільшим попитом у світі [5]. Він дозволяє проектувати в середовищах 2D та 3D, створювати та форматувати проектні документи та інше. AutoCAD не має конкретної спрямованості на певну галузь дизайну, тому використовується для розробки будівництва, механічної інженерії, науково-дослідних проектів та іншого. Основні особливості включають стандартну систему CAD, гнучкий інтерфейс та інструменти, можливість створювати програми з використанням вбудованих мов (AutoLISP тощо) та API, а також має велику кількість додатків від сторонніх розробників.

SOLIDWORKS – це популярний пакет програмного забезпечення для комп'ютерної проектування (CAD), розроблений компанією Dassault Systèmes[6]. Він використовується для проектування, моделювання та підготовки продуктів до виробництва. SOLIDWORKS пропонує широкий спектр функцій та інструментів для 2D та 3D моделювання, креслення, моделювання, рендерингу, анімації та багатьох інших. Це програмне забезпечення широко використовується у галузях, таких як машинобудування, авіаційна та автомобільна промисловість, виробництво споживчих товарів та промисловий дизайн. Переваги SOLIDWORKS:

- досягнення високої продуктивності та ефективності проектування завдяки розширеній функціональності, такі як велика бібліотека стандартних елементів, шаблони проектів, автоматизація процесів і багато іншого;
- широкий спектр інструментів для 2D та 3D моделювання та редагування, а також інструментів для аналізу та імітації різних умов;
- зручний інтерфейс та можливості налаштування під конкретні потреби користувачів;
- легка інтеграція з іншими системами проектування та програмним забезпеченням для управління даними проектів;

- підтримка різних форматів файлів для імпорту та експорту.

Недоліки SOLIDWORKS:

- висока вартість ліцензій та обладнання для користувачів;

- потребує значних ресурсів комп'ютера, що може вимагати оновлення апаратної частини;

- обмежені можливості для користувачів веб-версії SOLIDWORKS;

- потребує певного рівня знань та навичок для використання всіх функцій та можливостей системи;

- можливість виникнення проблем з сумісністю при імпорті та експорті файлів з іншими системами проектування.

NX є потужним інтегрованим пакетом програмного забезпечення для машинобудування, розробки продуктів і виробництва, розроблений компанією Siemens PLM Software [4]. Ось кілька ключових переваг NX:

- має широкий набір інструментів для 2D і 3D моделювання, включаючи розробку поверхонь, пластичних деталей і інструментів для генерації моделей складних форм;

- має потужні інструменти для аналізу та моделювання, що дозволяє інженерам виконувати технічний аналіз і оптимізацію проектів на ранніх етапах розробки;

- забезпечує зручне керування даними, включаючи зберігання, організацію і обмін файлами з іншими системами CAD;

- дозволяє кільком інженерам працювати над проектами одночасно, зберігаючи всі зміни в реальному часі і спільно використовуючи всі ресурси;

- має інструменти для моделювання систем керування, електронних і механічних систем з метою підвищення ефективності і точності розробки;

- включає потужний CAM-модуль, що дозволяє створювати і оптимізовувати траєкторії різання, забезпечує зручний інтерфейс для програмування верстатів і станків з ЧПУ;

- має потужну архітектуру, що дозволяє ефективно використовувати ресурси комп'ютера, зменшуючи час, необхідний для виконання завдань.

CATIA (Computer-Aided Three-dimensional Interactive Application) - це потужний комп'ютерний програмний пакет для проектування, розробки та виробництва продуктів в різних галузях промисловості. CATIA створена французькою компанією Dassault Systemes і використовується в галузях, таких як авіація, автомобілебудування, виробництво промислового устаткування та інших [7].

Деякі з основних особливостей CATIA:

- пропонує повний набір інструментів для моделювання в 3D, що дозволяє користувачам створювати детальні моделі високої якості;

- надає інтегроване робоче середовище, яке дозволяє користувачам працювати з моделями з більшою ефективністю і точністю;

- має широкий набір інструментів для різних задач, таких як моделювання, креслення, монтаж, аналіз, візуалізація та ін.;

- має можливості співпраці, що дозволяють користувачам спільно працювати над проектом, обмінюватися даними та виконувати інші завдання в режимі реального часу;

- має велику групу розробників та користувачів, які надають міцну підтримку і допомогу вирішенню проблем.

Autodesk Inventor є потужним програмним забезпеченням для комп'ютерного проектування (CAD), яке використовується для проектування, моделювання та розробки продуктів у різних галузях, таких як машинобудування, аерокосмічна промисловість, автомобільна промисловість.

Основні особливості Autodesk Inventor:

- надає комплексні можливості для 3D-моделювання, малювання, рендерингу та анімації;

- інструменти внутрішнього програмування, або API, дозволяють користувачам розширювати можливості програми та адаптувати її до своїх потреб;

- інтуїтивний інтерфейс та інструменти навігації допомагають користувачам швидко навчитися використовувати програму;

- дозволяє автоматизувати рутинні завдання та процеси, що дозволяє економити час та зусилля.

- підтримує широкий спектр файлових форматів, що дозволяє легко обмінюватися даними з іншими програмами та користувачами;

- моделювання деталей дозволяють виконувати різноманітні завдання, такі як розробка механізмів та інженерних систем;

- надає засоби для розробки складних збірок, з можливістю створення змінних та залежностей між деталями;

- інструменти Autodesk Inventor дозволяють розробляти та аналізувати конструкції з метою визначення оптимального варіанту.

Creo (раніше відомий як Pro/Engineer) є популярним програмним забезпеченням для проектування та розробки продуктів, що дозволяє інженерам створювати складні деталі та збірки з високою точністю та ефективністю [8]. Деякі з особливостей Creo включають наступне:

- має можливість створювати моделі в реальному часі, що дозволяє інженерам відразу бачити результати своєї роботи;

- дозволяє інженерам працювати з декількома дисциплінами, такими як механіка,

електрика та електроніка, щоб створювати інноваційні продукти;

- має широкі можливості моделювання, включаючи 3D-моделювання, поверхневе моделювання;

- інтегрується з іншими програмами, такими як Mathcad, Windchill і Arbortext, що дозволяє інженерам працювати з даними з різних джерел;

- має можливість автоматизувати процеси проектування і зменшити час на підготовку до виробництва;

- має інтуїтивний і легкий у використанні інтерфейс, що дозволяє інженерам швидко навчитися програмі та виконувати завдання ефективно.

У версії Creo 7.0 реалізовані революційні інновації в областях генеративного дизайну, проєктування багатомодульних систем, адитивного виробництва і симуляції. Для аналізу всіх описаних систем складемо порівняльну таблицю 1 .

Таблиця 1 – Порівняльна таблиця сучасних CAD–систем

	Легкість опанування та використання	Пряма інтеграція з верстатами ЧПК	3d-моделювання	Підтримка ЕСКД	Наявність баз даних стандартних	Ціна ліцензії на рік
Fusion 360 CLOUD Commercial	+	+	+	-	-	21752 грн [9]
SolidEdge SHINING 3D EDITION SOFTWARE	+	-	+	+	+	65800 грн [10]
AutoCAD 2022	+	-	-	-	-	74 412 грн [9]
SOLIDWORKS Professional	+	+	+	+	+	139308 грн [9]
NX	-	+	+	+	+	Ціну необхідно уточнювати у виробника
CATIA	-	+	+	-	-	-
INVENTOR Professional 2023	-	-	+	-	-	91869,80 грн[9]
Creo	-	+	+	-	-	Ціну необхідно уточнювати у виробника

Проаналізувавши результати таблиці бачимо, що система SOLIDWORKS Professional повністю відповідає всім необхідним критеріям, які потрібні для розробки та проектування ендопротезів.

Висновки

Таким чином, проаналізувавши ряд систем автоматизованого проєктування виявлено, що найбільш оптимальною для проєктування та розробки ендопротезів є система SOLIDWORKS, бо має продуманий інтерфейс користувача та є достатньо простою та зручною для опанування. Вона орієнтована як на конструкторську, так і на технологічну підготовку виробництва. Система має розпізнавання та параметризацію імпортованої геометрії, також в ній відсутні обмеження по кількості компонентів складних креслень. SOLIDWORKS має можливість випробувань на міцність спроектованих моделей в наближених реальних умовах та проведення аналізу кінематики та процесу динаміки. Professional є прийнятною та повністю відповідає якості функціонального наповнення системи. SOLIDWORKS також пропонує рішення 3DEXPERIENCE WORKS, що дозволяє об'єднати фахівців, ідеї, дані і рішення в інтерактивному середовищі для спільної роботи, забезпечуючи зв'язок між віртуальним середовищем і реальним світом.

Список літератури

1. Алгоритм розробки показань та протипоказань до ендопротезування кульшового суглоба / Лоскутов А. Е., Олейник А. Е., Головаха М. Л. Вісник ортопедії, травматології та протезування. 2004. № 3. С. 8-12.

2. Двостороннє тотальне ендопротезування кульшових суглобів / Лоскутов О.Є., Синєгубов Д.А., Головаха М. Л., Олійник О.Є., К.: Дніпропетровська державна медична академія МОЗ України, 2005. 3 с.
3. Офіційний сайт Fusion-360. URL: https://www.autodesk.com/products/fusion-360/overview?term=1 YEAR&tab=subscription
4. Офіційний сайт Siemens Digital Industries Software. URL: https://www.plm.automatio.ens.com.n.siem
5. Офіційний сайт Autodesk. URL: https://www.autodesk.com/
6. Офіційний сайт SOLIDWORKS. URL: https://www.solidworks.com
7. Офіційний сайт Dassault Systèmes. URL: https://www.3ds.com
8. Офіційний сайт PTC. URL: https://www.ptc.com
9. Постачальник ліцензій та програмного забезпечення. URL: https://www.softkey.ua/
10. Постачальник ліцензій та програмного забезпечення. URL: https://3ddevice.com.ua/product/ %D0%BF% D0%BE-solid-edge-shining-3d-edition-software/

References

1. Loskutov, A. E., Oleinyk, A. E., Holovakha, M. L. (2004), "Algorithm for the development of indications and contraindications to endoprosthetic replacement of the culchus subglobe " [Alhorytm rozrobky pokazan′ ta protypokazan′ do endoprotezuvannya kul′shovoho suhloba. Visnyk ortopediyi, travmatolohiyi ta protezuvannya.] *Journal of Orthopedics, Traumatology and Prosthetics*, № 3., pp. 8−12.
2. Loskutov, O.Y., Syniehubov, D.A., Holovakha, M. L., Oliinyk, O.Y. (2005), "Bilateral total hip arthroplasty " [Dvostoronnye total′ne endoprotezuvannya kul′shovykh suhlobiv. K.: Dnipropetrovs′ka derzhavna medychna akademiya MOZU], *Dnipropetrovs'k State Medical Academy of the Ministry of Health of Ukraine*, 31p.
3. Official website Fusion-360, available at: https://www.autodesk.com/products/fusion-360/overview?term=1 YEAR&tab=subscription
4. Official website Siemens Digital Industries Software, available at: https://www.plm.automatioens.com. n.siem
5. Official website autodesk, available at: https://www.autodesk.som/
6. Official website SOLIDWORKS, available at: https://www.solidworks.com
7. Official website Dassault Systèmes, available at: https://www.3ds.com
8. Official website PTC, available at: https://www.ptc.com
9. License and software provider [Postachal′nyk litsenziy ta prohramnoho zabezpechenya], available at: https://www.softkey.ua/
10. License and software provider [Postachal′nyk litsenziy ta prohramnoho zabezpechenya], available at: https://3ddevice.com.ua/product/%D0%BF%D0%BE-solid-edge-shining-3d-edition-software

Надійшла до редакції 28.06.2023

D. VINICHENKO, M. TIAHUNOVA, T. HOLUB
Національний університет «Запорізька політехніка», АТ «МОТОР СІЧ» м. Запоріжжя, Україна
vinichenko92dima@gmail.com, mary.tyagunova@gmail.com, golub.tv6@gmail.com

SELECTION OF THE OPTIMAL CAD SYSTEM FOR THE DEVELOPMENT AND DESUNG OF ENDOPROSTHESES

The method of joint replacement is one of the best options for surgical treatment of the musculoskeletal system in patients. The main purpose of this operation is to replace pathologically altered joint structures with artificial endoprosthesis, which has an anatomical form of a healthy joint and provides the possibility of performing a full volume of movements. In many cases, replacement of the affected joint is the last hope for the patient to restore the full functioning of the musculoskeletal system and in some cases, for the extension of life as a whole (for example, with cancer joints). This work is aimed at choosing the optimal CAD-system for the development and projection of endoprostheses. The following are the following CAD-systems: Fusion 360, Solidedge, AutoCad, SOLIDWORKS, NX, Inventor, Creo. As a result of the analysis of systems, comparative tables have been developed to compare the identified features. It clearly which system is the most effective for the projection of endoprostheses. The most optimal for the promotion and development of endoprostheses is the SOLIDWORKS system, because it has a thoughtful user interface and is simple enough and easy to master. It is focused on both design and technological preparation of production. The system has the recognition and parameterization of imported geometry and no restrictions on the number of components of complex drawings. SOLIDWORKS has the ability to test for the strength of designated models in approximate real test conditions and has the ability to analyze kinematics and the dynamics process.

Keywords: projection, CAD system, endoprosthetics, quality, 3D modeling

УДК 519.65:514.182.7:004.92

О. В. Фролов, канд. техн. наук, доц.
Харківський національний економічний університет імені Семена Кузнеця, м. Харків, Україна
frolgx@gmail.com

Моделювання зон розсіювання похибок положення стовбуру бурових свердловин та їх відображення

Розглядається проблема визначення точності положення свердловини в просторі. Актуальність дослідження зумовлена поширеністю використання просторової інформації, що базується на даних геофізичних вимірювань свердловин, для вирішення інженерних завдань. Метою роботи є розробка конструктивної та алгоритмічної моделей визначення зон розсіювання похибок інклінометричних вимірювань, що утворюються навколо бурових свердловин та визначають їхнє положення в просторі надр. Для моделювання був прийнятий метод середнього кута, який відтворює ділянки стовбура свердловини між точками вимірів ланками ламаної лінії, що визначається середніми значеннями вимірювань кутів напрямів. Отримані вирази, що визначають параметри форми та положення зон розсіювання середньо квадратичних похибок у формі еліпсоїдів для просторової моделі та еліпсів – для плоскої моделі. Також була розглянута спрощена модель у формі сфер та кіл середньо квадратичних похибок, для яких визначаються лише параметри радіусів. Далі на цій основі отримана модель лінійної інтерполяції значень параметрів зон розсіювання між точками вимірів. Були визначені параметри моделі лінійної інтерполяції, що представляють зони розсіювання в просторі у вигляді конусів, дотичних до сфер розсіювання в точках інклінометричних досліджень. Розроблений алгоритм для комп'ютерного моделювання зон розсіювання на планах, який здійснює побудову зон в формі кіл проекцій та дотичних до них прямих ліній. Також отримані конструктивний та аналітичний вираз для розрахунку радіуса кола розсіювання в будь - якій точці траєкторії стовбура. Наведені приклади моделювання зон розсіювання похибок навколо бурових свердловин на основі даних реальних даних.

Ключові слова: стовбур свердловини, інклінометричні дослідження, геоінформаційна система, інтерполяція, ламана

DOI: 10.31474/1996-1588-2023-2-37-72-79

Вступ

Основною метою проектування, моніторингу та контролю траєкторії свердловини є визначення просторового розташування траєкторії свердловини. Інформація про просторове розташування свердловин є дуже важливою з точки зору її використання в геоінформаційних системах, оскільки вона потрібна на всіх етапах освоєння родовищ корисних копалин.

На етапі розвідки родовищ просторова інформація по свердловинам використовується для створення уяви про геологічну будову товщі порід (розташування в надрах продуктивних пластів, покладів та порід, структурних особливостей родовища) та запаси корисних копали.

На етапах будівництва та експлуатації родовищ вимоги до точності та оперативності отримання просторових даних значно підвищуються:

- сучасні методи геологічного моделювання, складення схем розробки родовищ потребують достовірної оцінки початкових та залишкових запасів, локалізації геологічних структур, які ускладнюють ведення робіт;

- бурові свердловини, що перетинають затоплені виробки або водоносні горизонти, і свердловини неякісного затампонування належать до небезпечних зон, при цьому ці виробки можливо віднести до виробок з недостовірним контуром, оскільки положення точок перетину та кінцевих точок свердловин визначається, як правило, з досить суттєвими похибками;

- видобуток нафти та газу потребує направленого буріння, контроль за яким використовує просторову інформація для цілей орієнтування бурового обладнання, недопущення перетину нової свердловини з вже існуючими, попередження бурильників про потенціальні проблеми, визначення місць аварій, викидів та загорянь.

Однак визначення траєкторії свердловини неможливо зробити абсолютно точним через похибки вимірювань і розрахунків. Точність позиціонування траєкторії свердловини можна підвищити за допомогою корекції помилок, але повністю усунути помилку неможливо, тому траєкторія свердловини має невизначеність. При розробці нафтових і газових родовищ із щільною структурою свердловин і тонкими шарами нафти, а також при бурінні морських платформ і

розвантажувальних свердловин, особливо важливо кількісно охарактеризувати невизначеність траєкторії свердловини, щоб зменшити ризик буріння. та покращити ефект від розробки нафтогазових родовищ [1].

Таким чином, проблема моделювання зони розсіювання положення стовбура свердловини є актуальною науковою проблемою.

Метою роботи є розробка алгоритму побудови зон розсіювання похибок положення навколо бурових свердловин на планах і картах щодо подальшого використання в геоінформаційних системах.

Інклінометричні дослідження бурових свердловин та їх похибки

Інклінометричні дослідження проводять в окремих точках траєкторії стовбуру свердловини за допомогою інклінометрів – приборів, що вимірюють параметри траєкторії за допомогою магнітних (або гіроскопічних) та гравітаційних датчиків [1]. Ці вимірювання можуть проводитись як в процесі буріння свердловини, так і після його закінчення.

Параметрами траєкторії стовбура свердловини, що вимірюються при інклінометричних дослідженнях є величини азимуту α та зенітного кута θ в точках траєкторії стовбура, а також відстань Δl між точками вимірювань.

Із свердловиною пов'язують систему прямокутних координат, що має початок в усті свердловини та наступні напрями осей:

z – донизу за напрямом прямовисних ліній;

x – за північним напрямом географічного або магнітного меридіану;

y - за східним напрямом відповідної паралелі або перпендикулярно до осей x та z.

Задача побудови траєкторії стовбура свердловини полягає в тому [2], щоб за відомим масивом інклінометричних вимірів визначити координати відповідних точок в системі координат, що пов'язана з устям свердловини, тобто визначити глибину та горизонтальні зміщення по осям x та y.

Проекції (dx, dy, dz) елементарного інтервалу (dl) стовбуру свердловини:

$$dx = dl\cos\alpha\sin\theta,$$
$$dy = dl\sin\alpha\sin\theta, \quad (1)$$
$$dz = dl\cos\theta.$$

Інтегруючи ці вирази по довжині інтервалу можливо отримати елементарні переміщення, але, оскільки значення функцій $\alpha = \alpha(l)$ та $\theta = \theta(l)$ відомі тільки в точках вимірювань, відомі методи обчислення координат траєкторії стовбуру свердловини використовують встановлені закономірності зміни зенітного кута та азимуту (їх згладжені залежності), або вважають відомою геометрію самого інтервалу [3].

Одним із простих та ефективних методів обчислення координат точок траєкторії свердловини є метод середнього кута [4]. Розрахункові формули координат з урахуванням припущень для цього методу мають вигляд:

$$x = \sum_{i=2}^{n} l_i \cdot \sin\frac{\theta_{i-1}+\theta_i}{2}\cos\frac{\alpha_{i-1}+\alpha_i}{2};$$
$$y = \sum_{i=2}^{n} l_i \cdot \sin\frac{\theta_{i-1}+\theta_i}{2}\sin\frac{\alpha_{i-1}+\alpha_i}{2}; \quad (2)$$
$$z = \sum_{i=2}^{n} l_i \cdot \cos\frac{\theta_{i-1}+\theta_i}{2},$$

де l_i - довжина інтервалу між точками *i-1* та *i*;

θ_{i-1}, θ_i – зенітні кути відповідно в верхній та нижній точках інтервалу вимірювань;

α_{i-1}, α_i – азимути в цих точках.

Детальне вивчення похибок інклінометричних досліджень можна знайти в роботі [5], де зазначені похибки були умовно поділені на три групи:

- інструментальні;
- технологічні;
- похибки обробки результатів вимірювань.

Перші дві групи складають похибки прибору як засобу вимірювань та похибки пов'язані з технологією проведення вимірювань. Ці похибки враховуються даними про точність вимірювань зенітних кутів та азимутів конкретного інклінометра та поправками до виміряних кутів. При цьому в діапазоні малих зенітних кутів (0^o - 50^o) похибки азимутів, як правило, корелюють зі значеннями зенітних кутів [4].

Третя група похибок характеризується обраним методом розрахунку координат точок свердловини [4, 5]. Різна точність методів обумовлюється різним характером «чутливості» методів до похибок виміряних величин.

Методичні похибки, властиві різним методам розрахунку координат, мають досить значний розкид від методу до методу і залежать, в першу чергу, від обраного методу розрахунку координат, кроку вимірювань і параметрів викривлення осі свердловини. Гранична (максимальна) і середньоквадратична похибки визначення координати, що характеризують розсіювання координат точок осі свердловини (так звані кола або еліпси розсіювання), визначаються за формулами [6]:

$$\delta_k = \sum_{i}^{n}\sum_{j}^{s}\left|\frac{\partial F_{ik}}{\partial x_j}\right|\delta_{x_j}, \quad (3)$$

$$m_k = \sqrt{\sum_{i}^{n}\sum_{j}^{s}\left(\frac{\partial F_{ik}}{\partial x_j}m_{x_j}\right)^2}. \quad (4)$$

де δ_k, m_k - гранична та середньоквадратична

похибки відповідно;

F_k - функція, за якої обчислюється k-та координата на і-му інтервалі;

x_j - аргументи, що вимірюються (довжина, азимут та зенітний кут).

Моделі розсіювання похибок у точках вимірювань

Розглянемо найбільш повну модель розподілу похибок у просторі у вигляді еліпсоїда, яка враховує нерівномірність розсіювання за напрямами [3, 5].

Для визначення параметрів еліпсоїда похибок необхідно скласти коваріаційну матрицю похибок координат точки траєкторії:

$$K_{xyz} = \begin{pmatrix} m_x^2 & K_{xy} & K_{xz} \\ K_{yx} & m_y^2 & K_{yz} \\ K_{zx} & K_{zy} & m_z^2 \end{pmatrix}, \quad (5)$$

де K_{xyz} – коваріаційна матриця, яка є симетричною відносно головної діагоналі, тобто, наприклад, $K_{xy} = K_{yx}$;

m_x, m_y, m_z – середньоквадратичні похибки (СКП) функцій, що визначають координати точки;

K_{xy}, K_{xz}, K_{yz} – кореляційні моменти відповідних оцінюваних функцій.

Вирази для СКП координат точки траєкторії отримують за відомою формулою похибки функції виміряних величин [4], що виражається формулою (4).

Складемо вирази квадрату СКП методу середнього кута – (2).

Будемо мати:

$$m_x^2 = \sum_{i=2}^{n} \sin^2 \frac{\theta_{i-1}+\theta_i}{2} \cos^2 \frac{\alpha_{i-1}+\alpha_i}{2} m_l^2 + \frac{1}{2}\Big(\sum_{i=2}^{n} l_i^2 \Big(\cos^2 \frac{\theta_{i-1}+\theta_i}{2} \cos^2 \frac{\alpha_{i-1}+\alpha_i}{2} m_\theta^2 + \sin^2 \frac{\theta_{i-1}+\theta_i}{2} \sin^2 \frac{\alpha_{i-1}+\alpha_i}{2} m_\alpha^2\Big)\Big);$$

$$m_y^2 = \sum_{i=2}^{n} \sin^2 \frac{\theta_{i-1}+\theta_i}{2} \sin^2 \frac{\alpha_{i-1}+\alpha_i}{2} m_l^2 + \frac{1}{2}\Big(\sum_{i=2}^{n} l_i^2 \Big(\cos^2 \frac{\theta_{i-1}+\theta_i}{2} \sin^2 \frac{\alpha_{i-1}+\alpha_i}{2} m_\theta^2 + \sin^2 \frac{\theta_{i-1}+\theta_i}{2} \cos^2 \frac{\alpha_{i-1}+\alpha_i}{2} m_\alpha^2\Big)\Big); \quad (6)$$

$$m_z^2 = \sum_{i=2}^{n} \Big(\cos^2 \frac{\theta_i+\theta_{i+1}}{2} m_l^2 + \frac{1}{2} l_i^2 \sin^2 \frac{\theta_i+\theta_{i+1}}{2} m_\theta^2\Big).$$

Вирази для кореляційних моментів отримаємо на основі наступної формули [6]

$$K_{ls} = \sum_j^k \frac{\partial f_l}{\partial x_j} \frac{\partial f_s}{\partial x_j} m_{x_j}^2, \quad (7)$$

де f_l, f_s – вирази функцій, між якими визначається кореляційний момент.

Отримаємо:

$$K_{xy} = \sum_{i=2}^{n} \sin^2 \frac{\theta_{i-1}+\theta_i}{2} \sin \frac{\alpha_{i-1}+\alpha_i}{2} \cos \frac{\alpha_{i-1}+\alpha_i}{2} m_l^2 + \frac{1}{2}\sum_{i=2}^{n} l_i^2 \sin \frac{\alpha_{i-1}+\alpha_i}{2} \cos \frac{\alpha_{i-1}+\alpha_i}{2} \Big(\cos^2 \frac{\theta_{i-1}+\theta_i}{2} m_\theta^2 - \sin^2 \frac{\theta_{i-1}+\theta_i}{2} m_\alpha^2\Big);$$

$$K_{xz} = \sum_{i=2}^{n} \sin \frac{\theta_{i-1}+\theta_i}{2} \cos \frac{\theta_{i-1}+\theta_i}{2} \cos \frac{\alpha_{i-1}+\alpha_i}{2} \cdot \cdot\Big(m_l^2 - \frac{1}{2} l_i^2 m_\theta^2\Big); \quad (8)$$

$$K_{yz} = \sum_{i=2}^{n} \sin \frac{\theta_{i-1}+\theta_i}{2} \cos \frac{\theta_{i-1}+\theta_i}{2} \sin \frac{\alpha_{i-1}+\alpha_i}{2} \cdot \cdot\Big(m_l^2 - \frac{1}{2} l_i^2 m_\theta^2\Big).$$

Головні радіуси еліпсоїда похибок знайдемо, як власні значення матриці похибок, із характеристичного рівняння третього порядку [7]:

$$\lambda^3 - I_1\lambda^2 + I_2\lambda - I_3 = 0, \quad (9)$$

де: I_1, I_2, I_3 - інваріанти рівняння (9), що у випадку матриці (5) мають вигляд:

$$I_1 = m_x^2 + m_y^2 + m_z^2;$$

$$I_2 = \begin{vmatrix} m_x^2 & K_{xy} \\ K_{yx} & m_y^2 \end{vmatrix} + \begin{vmatrix} m_x^2 & K_{xz} \\ K_{zx} & m_z^2 \end{vmatrix} + \begin{vmatrix} m_y^2 & K_{yz} \\ K_{zy} & m_z^2 \end{vmatrix}; \quad (10)$$

$$I_3 = \begin{vmatrix} m_x^2 & K_{xy} & K_{xz} \\ K_{yx} & m_y^2 & K_{yz} \\ K_{zx} & K_{zy} & m_z^2 \end{vmatrix}.$$

Розв'язавши рівняння (9) відносно λ, отримаємо три корені λ_1, λ_2 та λ_3, з яких можна знайти головні радіуси

$$a = \sqrt{\lambda_1}, b = \sqrt{\lambda_2}, c = \sqrt{\lambda_3}; \quad (11)$$

де для більшого кореню використано позначення λ_1, середнього - λ_2, а для меншого - λ_3.

Напрями головних осей еліпсоїда отримаємо із системи рівнянь[7]

$$\begin{cases} (m_x^2 - \lambda)\cos\alpha + K_{xy}\cos\beta + K_{xz}\cos\gamma = 0; \\ K_{yx}\cos\alpha + (m_y^2 - \lambda)\cos\beta + K_{yz}\cos\gamma = 0; \\ K_{zx}\cos\alpha + K_{zy}\cos\beta + (m_z^2 - \lambda)\cos\gamma = 0, \end{cases} \quad (12)$$

де λ – корінь характеристичного рівняння (9), якому відповідає напрямок, що утворює з координатними осями x, y, z відповідно кути α, β, γ.

Підставляючи до системи (12) значення коренів λ_1, λ_2, та λ_3 отримаємо напрями головних діаметрів (α_1, β_1, γ_1), (α_2, β_2, γ_2) та (α_3, β_3, γ_3), які є взаємно перпендикулярними.

Щоб перейти до побудови еліпсоїда напишемо його параметричні рівняння відносно головних осей, за які прийняті осі x, y, z декартової просторової системи координат

$$x' = a\cos u\cos v, y' = b\sin u\cos v, z' = c\sin v. \quad (13)$$

Підставляючи праву частину цих формул

заміст x', y', z' до формул перетворення координат відносно локальної системи координат свердловини:

$$\begin{aligned} x &= x_i + l_1 x' + l_2 y' + l_3 z'; \\ y &= y_i + m_1 x' + m_2 y' + m_3 z'; \\ z &= z_i + n_1 x' + n_2 y' + n_3 z', \end{aligned} \qquad (14)$$

де: x_i, y_i, z_i - координати точки осі свердловини, в якій визначається еліпсоїд розсіювання;

$l_i = \cos\alpha_i, m_i = \cos\beta_i, n_i = \cos\gamma_i$ - коефіцієнти перетворення.

Щоб спросити модель, разом із еліпсоїдом похибок, що визначає просторовий розподіл, застосовують сферу похибок [4], радіус якої дорівнює

$$\mathrm{M}_{\text{пр}} = \sqrt{m_x^2 + m_y^2 + m_z^2}\,. \qquad (15)$$

За аналогією з просторовою моделлю, задля спрощення обчислень на практиці є розповсюдженими побудови на планах гірничих робіт кіл розсіювання, що замінюють еліпси радіуси яких визначаються за формулою (15) у випадку якщо потрібно врахувати похибки за глибиною, або за виразом:

$$\mathrm{M}_{\text{пл}} = \sqrt{m_x^2 + m_y^2}\,, \qquad (16)$$

якщо потрібно враховувати тільки похибки положення на плані.

Отримані залежності дозволяють будувати зони розсіювання в точках інклінометричних досліджень. З метою отримання оцінювання розмірів зон уздовж всієї свердловини, розробимо модель інтерполяції значень параметрів зон між точками вимірів.

Алгоритм побудови зон розсіювання похибок положення навколо бурових свердловин

В роботі [6] була запропонована модель зони розсіювання похибок у вигляді еліптичних циліндрів, що є дотичними до еліпсоїдів похибок у точках вимірювань. Зона розсіювання є поверхнею, що огинає сімейство таких циліндрів. Реалізація такої моделі потребує відносно складних розрахунків і на практиці може бути спрощена за рахунок заміни еліпсів та еліпсоїдів відповідними колами та сферами, отриманими на основі формул (16) та (15).

Отже, згідно з прийнятим методом розрахунку траєкторії стовбура свердловини, який передбачає лінійну інтерполяцію ділянок стовбуру між точками вимірів, буде прийнятним розповсюдження лінійної залежності між параметрами зон розсіювання. При цьому для подальшого використання бажано мати придатну для конструктивних побудов модель. Тому зупинимось на моделі представлення зони розсіювання в кожній точці стовбуру у вигляді сфери радіусу (15) – для просторового представлення, або відповідних кіл радіусів (15) або (16) – для представлення на плані:

$$r_i = \sqrt{m_{xi}^2 + m_{yi}^2 + m_{zi}^2}\,,$$

або

$$r_i = \sqrt{m_{xi}^2 + m_{yi}^2}\,. \qquad (17)$$

Лінійно інтерполюючи радіуси між обчисленими в кінцевих точках ланок просторової ламаної осі свердловини будемо мати конуси як огинаючи поверхні відповідних сімей сфер. Аналогом такої інтерполяції на площині буде побудова спільних зовнішніх дотичних до кіл розсіювання у точках вимірів. Перейдемо до розгляду аналітичної моделі. На основі виразів (17) отримаємо послідовності відповідних радіусів, а на основі виразів (2) знайдемо координати відповідних їм центрів сфер та кіл. Введемо позначення для радіусів r_i, для центрів – x_{ci}, y_{ci}, z_{ci} (i=0..k), де x_{c0}, y_{c0}, z_{c0} визначають координати устя свердловини, x_{ck}, y_{ck}, z_{ck} відповідають розрахунку координат в останній точці вимірювань.

Далі, для кожної пари сфер (кіл) визначимо параметри перетворення координат до нової системи з початком в поточній точці P_i.

Для просторової моделі вісь z' нової системи сумістимо з напрямом осі свердловини від поточної точки P_i в сторону точки P_{i+1}. Щоб отримати орт осі свердловини, скористаємось наступними виразами:

$$\begin{aligned} x_{i,\,i+1} &= x_i - x_{i+1}, \\ y_{i,\,i+1} &= y_i - y_{i+1}, \ , \\ z_{i,\,i+1} &= z_i - z_{i+1} \end{aligned} \qquad (19)$$

Напрямні косинуси цієї оси матимуть вигляд:

$$l_{3i} = \frac{x_{i,i+1}}{\left|\mathbf{l}_{i,\,i+1}\right|}, m_{3i} = \frac{y_{i,i+1}}{\left|\mathbf{l}_{i,\,i+1}\right|}, n_{3i} = \frac{z_{i,i+1}}{\left|\mathbf{l}_{i,\,i+1}\right|}, \qquad (20)$$

де $\left|\mathbf{l}_{i,\,i=1}\right| = \sqrt{x_{i,i+1}^2 + y_{i,i+1}^2 + z_{i,i+1}^2}$.

В умовах моделі ми вільні в обранні напрямів осей x', y' в площині, що перпендикулярного до обраного напряму та проходить через точку P_i. Задля визначеності приймемо, наприклад, за напрямок осі x' напрям відрізка, що з'єднує проекцію устя свердловини на цю площину з точкою P_i (у випадку, коли ця проекція є невизначеною вісь свердловини збігається за напрямом з віссю z, а, отже, перетворення координат буде паралельним перенесенням). Координати цього відрізка

$$x_{0,i} = x_i, y_{0,i} = y_i, z_{0,i} = -\frac{x_i l_i + y_i m_i}{n_i}. \qquad (21)$$

Нормуючи вектор цього напряму, отримаємо напрямні косинуси осі x':

$$l_{1,i} = \frac{x_{0,i}}{\left|\mathbf{p}_{0,i}\right|}, m_{1,i} = \frac{y_{0,i}}{\left|\mathbf{p}_{0,i}\right|}, n_{1,i} = \frac{z_{0,i}}{\left|\mathbf{p}_{0,i}\right|}. \qquad (22)$$

Напрямок орта осі y' можна знайти за звичайними формулами векторного добутку вже знайдених ортів:

$$\begin{aligned} l_{2,i} &= m_{3,i}n_{1,i} - m_{1,i}n_{3,i}, \\ m_{2,i} &= n_{3,i}l_{1,i} - n_{1,i}l_{3,i}, \\ n_{2,i} &= l_{3,i}m_{1,i} - l_{1,i}m_{3,i} \end{aligned} \quad . \qquad (23)$$

Конус, дотичний до сфер розсіювання похибок в точках P_i та P_{i+1}, має своєю віссю вже визначений напрямок поточної ділянки осі свердловини. Його параметричні рівняння в новій системі можна записати у вигляді:

$$\begin{aligned} x' &= \left(\frac{r_i}{\cos\alpha} + u\sin\alpha\right)\cos v, \\ y' &= \left(\frac{r_i}{\cos\alpha} + u\sin\alpha\right)\sin v, \\ z' &= u\cos\alpha \end{aligned} \qquad (24)$$

де α – кут нахилу твірної дотичного конусу.

Цей кут, його косинус та синус можливо отримати з виразів:

$$\cos\alpha_i = \frac{\sqrt{\left|\mathbf{l}_{i,i+1}\right|^2 - \left(r_{i+1} - r_i\right)^2}}{\left|\mathbf{l}_{i,i+1}\right|}, \sin\alpha_i = \frac{r_{i+1} - r_i}{\left|\mathbf{l}_{i,i+1}\right|}, \qquad (25)$$

$$\alpha_i = \operatorname{arctg}\frac{r_{i+1} - r_i}{\sqrt{\left|\mathbf{l}_{i,i+1}\right|^2 - \left(r_{i+1} - r_i\right)^2}}$$

Межі зміни параметру v є стандартними – [0, 2π], параметр u повинен змінюватись в межах кіл дотику конуса зі сферами, цей інтервал може бути отриманий з виразів:

$$\left[-r_i\sin\alpha_i, \left|\mathbf{l}_{i,i+1}\right| - r_i\sin\alpha_i\right]. \qquad (26)$$

Перейдемо до розгляду побудов на плані. В цьому випадку напрямок осі x' нової системи будемо визначати горизонтальною проекцією $\bar{\mathbf{l}}_i$ поточної ланки ламаної осі свердловини. Цей напрямок може бути отриманий за аналогією з просторовою схемою:

$$\cos\varphi_i = \frac{x_{i,i+1}}{\left|\bar{\mathbf{l}}_{i,i+1}\right|}, \sin\varphi_i = \frac{y_{i,i+1}}{\left|\bar{\mathbf{l}}_{i,i+1}\right|}, \qquad (27)$$

де $\left|\bar{\mathbf{l}}_{i,i=1}\right| = \sqrt{x_{i,i+1}^2 + y_{i,i+1}^2}$ – довжина горизонтальної проекції.

Тоді кут між дотичними до кіл розсіювання та новою віссю x' може бути визначений як

$$\alpha_i = \operatorname{arctg}\frac{r_{i+1} - r_i}{\sqrt{\left|\bar{\mathbf{l}}_{i,i+1}\right|^2 - \left(r_{i+1} - r_i\right)^2}}. \qquad (28)$$

Рівняння спільної дотичної кіл r_i, r_{i+1}

$$y' = y'_{1i} + \frac{y'_{2i} - y'_{1i}}{x'_{2i} - x'_{1i}}(x' - x_{1i}) \qquad (29)$$

де $x_{1i} = -r_i\sin\alpha_i$, $y_{1i} = \pm\, r_i\cos\alpha_i$,

$$x_{2i} = \left|\bar{\mathbf{l}}_{i,i+1}\right| - r_{i+1}\sin\alpha_{i+1},\ y_{2i} = \pm\, r_{i+1}\cos\alpha_{i+1}. \qquad (30)$$

Підставивши рівняння (29) замість y' при обчислених параметрах до формул перетворення координат на площині, отримаємо рівняння прямої:

$$\begin{aligned} x &= x_i + x'\cos\varphi_i - y'\sin\varphi_i, \\ y &= x_i + x'\sin\varphi_i + y'\cos\varphi_i. \end{aligned} \quad . \qquad (31)$$

Щоб побудувати два відрізка прямих ліній дотичних до кіл розсіювання, потрібно у вирази (29) та (31) при різних значеннях подвійного знаку в (30) підставити значення x' межах зміни від x_{1i} до x_{2i}.

Отже, на основі отриманих залежностей можна сформулювати наступний алгоритм побудови зони розсіювання похибок навколо свердловини на плані:

- побудувати кола радіусів r_j в точках $Pj\ (j = 0, \ldots, k)$;
- для кожної і-ї точці інклінометричних досліджень від 0 до k-1:

1) обчислити довжину проекції $\left|\bar{\mathbf{l}}_{i,i+1}\right|$;

2) перевірити значення $\left|\bar{\mathbf{l}}_{i,i+1}\right|$ на рівність нулеві, якщо $\left|\bar{\mathbf{l}}_{i,i+1}\right|$=0 перейти до наступної точці;

3) при $\left|\bar{\mathbf{l}}_{i,i+1}\right|$>0 розрахувати значення cos φ_i та sin φ_i за формулами (27) та визначити кут α_i за формулою (25);

4) розрахувати параметри x_{1i}, y_{1i}, x_{2i}, y_{2i} за виразами (30);

5) побудувати два відрізка прямих ліній за виразами (29) та (31) при різних значеннях подвійного знаку в (30) та межах зміни x' від x_{1i} до x_{2i}.

На рис. 1 зображено результати моделювання зон розсіювання побудовані на основі даних реальних інклінометричних досліджень свердловин Н5552 та Н 3956 шахти «Капітальна» ДП «Мирноградвугілля», що були побудовані за запропонованим алгоритмом. В інженерній практиці може виникнути потреба побудови зони розсіювання похибки на певному горизонті (рівні глибини свердловини). Виконавши графічні побудови за розробленим алгоритмом, розв'язок цієї задачі не визиває ускладнень.

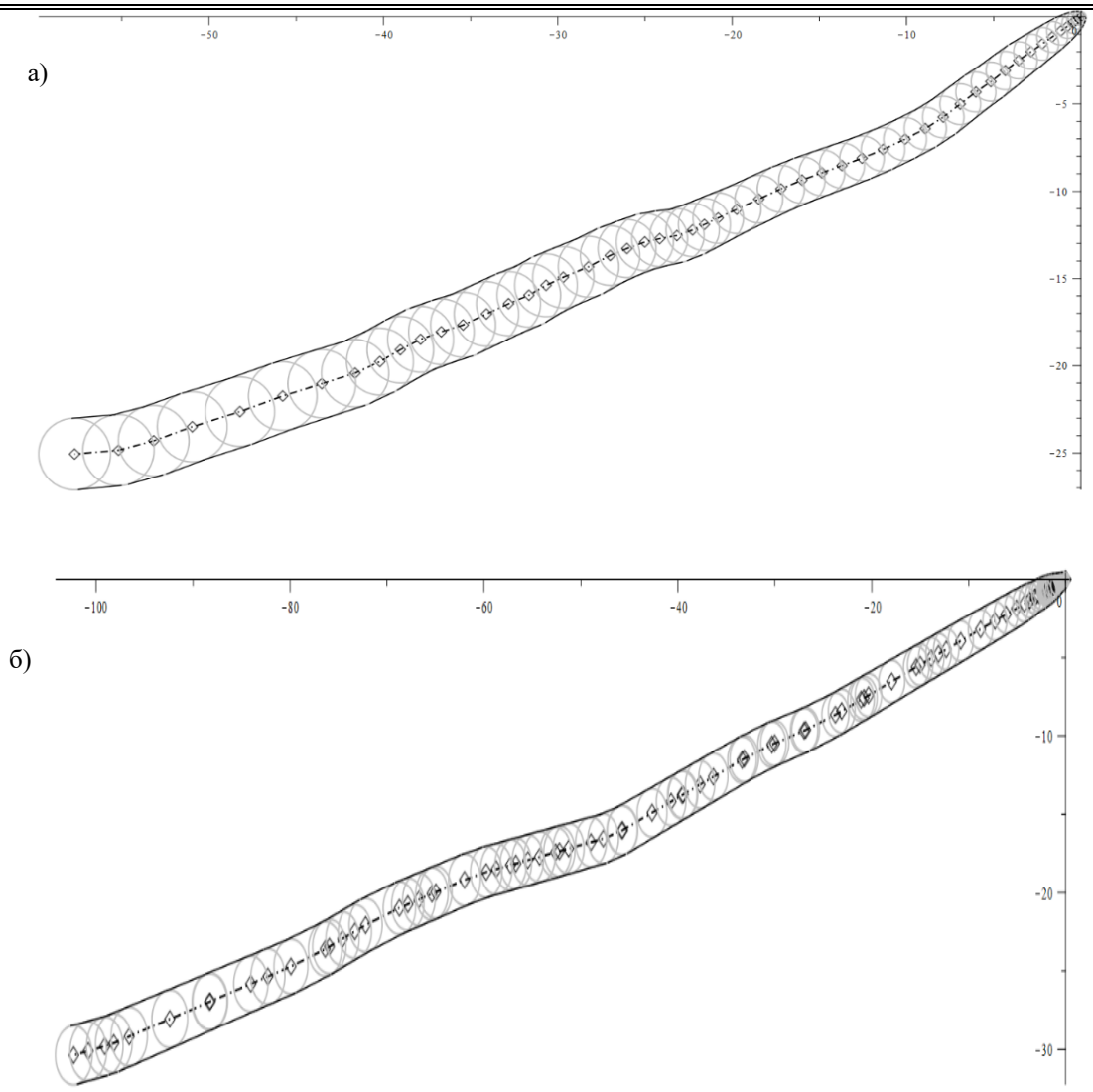


Рисунок 1 – Результати моделювання зон розсіювання побудовані на основі даних реальних інклінометричних досліджень свердловин Н5552 (а) та Н 3956 (б) шахти «Капітальна» ДП «Мирноградвугілля»

Можна запропонувати, як конструктивний, так аналітичний варіант розв'язку. За конструктивним варіантом спочатку методом графічної інтерполяції знаходиться точка на осі стовбуру свердловини – P_{-126} (рис. 2), що має задану глибину.

Далі, з цієї точки будується перпендикуляр до будь-якої з двох відповідних дотичних на плані і на перетині з дотичною отримується точка ($\overline{P}_{-126}$), відстань від точки стовбура до якої визначатиме радіус (r_{-126}) шуканого кола розсіювання. Аналітичний розв'язок для радіусу залежить від відстані x'_j (між відомою i -ю точкою на плані та потрібною точкою осі) дає вираз:

$$r_j = r_i + x'_j \sin\alpha_i . \qquad (32)$$

Обговорення результатів та висновки

В роботі було розглянуто залежності, що дозволяють розраховувати параметри еліптичних та кругових зон розсіювання похибок положення точок стовбуру свердловин на плані та в просторі. На цій основі було запропоновано спрощений алгоритм для побудови зон у вигляді кіл (сфер) та дотичних до них (дотичних конусів). Проведено комп'ютерне моделювання за розробленим алгоритмом побудови зон розсіювання. Отримана в роботі математична модель побудови зон розсіювання похибок навколо бурових свердловин є простою та ефективною з точки зору її реалізації у геоінформаційних системах [8], таких як, наприклад, K-MINE [9]. Алгоритм, що реалізує зазначену модель, визначається лінійною складністю – $O(n)$.

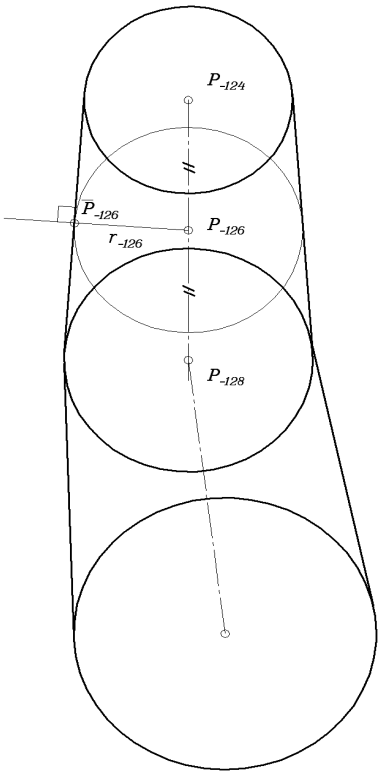


Рисунок 2 – Побутова кола розсіювання на заданій глибині

Список літератури

1. Well Trajectory Measurement Error Based on Gyro Survey / Diao B. et. al.; 5th International Conference on Green Energy and Sustainable Development.E3S Web Conf., 2022. vol. 358. 5 p. DOI: https://doi.org/10.1051/e3sconf/202235801008/
2. Liu X. Quantitative recognition method for borehole trajectory models. Petrol. Explor. Develop., 2018. Vol. 45, No. 1. P. 154–158. DOI: https://doi.org/10.1016/S1876-3804(18)30015-6
3. Близнюк В.О., Фролов В.О. Моделювання траєкторій скривлених свердловин методом гелікоїдальних дуг. Наукові праці Донецького національного технічного університету. Серія: Інформатика, кібернетика та обчислювальна техніка, Покровськ, 2017. №2, No 25. С. 12–21.
4. Халимендик Ю.М., Фролов В.О. Врахування похибок інклінометричних досліджень при визначенні меж небезпечних зон за проривами води у бурових свердловин. Вісті Донецького гірничого інституту, Покровськ, 2018. №1(42). С. 7–15. DOI: https://doi.org/10.31474/1999-981x-2018-1-7-15/
5. Diao B., Gao B., Jiang C. Contribution Rate of MWD Survey Error Sources on Well Trajectory Measurement Error. 2019 International Conference on Oil & Gas Engineering and Geological Sciences 28–29 September 2019, Dalian, China. IOP Conference Series: Earth and Environmental Science, 2019. Vol. 384. 8 p. DOI: https://doi.org/10.1088/1755-1315/384/1/012226
6. Liu X. Borehole trajectory uncertainty and its characterization. Petrol. Explor. Develop., 2019. Vol.46(2). P. 407–412. DOI: https://doi.org/10.1016/S1876-3804(19)60021-2.
7. Каленюк П., Рибицька О., Івасик Г. Лінійна алгебра та аналітична геометрія: навч. посіб. / перекл. Яромира Войтовича; М-во освіти і науки України, Нац. ун-т «Львів. політехніка». Львів: Вид-во Львів. політехніки, 2019. 160 с.
8. Xiao N. GIS Algorithms: Theory and Applications for Geographic Information Science & Technology. SAGE Publications Ltd, 2016. 336 p.
9. K-MINE: Комплексні рішення для гірничодобувної промисловості. URL: https://k-mine.com/ua/ (Дата звернення 02.01.2024)

References

1. Diao, B., et al. (2022), "Well Trajectory Measurement Error Based on Gyro Survey", *5th International Conference on Green Energy and Sustainable Development*, 358, 5 p.

2. Liu, X. (2018), "Quantitative recognition method for borehole trajectory models", *Petrol. Explor. Develop.*, 45(1), pp. 154–158.
3. Bliznuk, V., Frolov, O. (2017), "Modeling of well tractories by the method of helikoidal arcs" [Modeliuvannia traiektorii skryvlenykh sverdlovyn metodom helikoidalnykh duh], Scientific papers of Donetsk National Technical University. Series: "Informatics, Cybernetics and Computer Science", 2(24), pp. 12-21.
4. Khalymendyk, Y., Frolov, O. (2018), "Accounting of inclinimeter survey errors and determination of borders of dangerous areas around drilling wells" [Vrakhuvannia pokhybok inklinometrychnykh doslidzhen PRY vyznachenni mezh nebezpechnykh zon za proryvamy vody u burovykh sverdlovyn], *Journal of Donetsk Mining Institute*, 1(48), pp. 7-15.
5. Diao, B., Gao, B., Jiang, C. (2019), "Contribution Rate of MWD Survey Error Sources on Well Trajectory Measurement Error", *2019 International Conference on Oil & Gas Engineering and Geological Sciences 28–29 September 2019, Dalian, China*, 384, 8 p.
6. Liu, X. (2019), "Borehole trajectory uncertainty and its characterization" *Petrol. Explor. Develop.,* 46(2), pp. 407–412.
7. Kaleniuk, P., Rybytska, O., Ivasyk, H. (2019), Linear Algebra and Analytic Geometry. [Liniina alhebra ta analitychna heometriia: navch. posib. / perekl. Yaromyra Voitovycha; M-vo osvity i nauky Ukrainy, Nats. un-t «Lviv. politekhnika»] Basic Course, Lviv Polytechnik, 160 p.
8. Xiao, N. (2016), GIS Algorithms: Theory and Applications for Geographic Information Science & Technology, SAGE Publications Ltd, 336 p.
9. K-MINE: End-to-end Mining Industry Software & Mining Consulting [Kompleksni rishennia dlia hirnychodobuvnoi promyslovosti], available at: https://k-mine.com/

Надійшла до редакції 13.12.2023

O. FROLOV
Simon Kuznets Kharkiv National University of Economics, Kharkiv, Ukraine
frolgx@gmail.com

DISPERSION ZONES OF BOREHOLE POSITION ERRORS AND THEIR DISPLAY

The problem of determining the accuracy of the well position in space is considered. The relevance of the study is due to the prevalence of the use of spatial information, based on geophysical measurements of wells, to solve engineering problems. The aim of the work is to develop constructive and analytical models for defining the distribution zones of inclinometric measurement errors that form around boreholes and determine their position in the underground space. The methods of calculation of wellbore trajectory points and methods of calculation of errors of their position were analyzed in the work. For modeling, the method of the mean angle was adopted, which reproduces the sections of the wellbore between the measurement points by the segments of the polyline, which is determined by the average values of the angle measurements of the directions. The expressions that define the parameters of the shape and position of the distribution zones of the mean square errors have the form of ellipsoids for the spatial model and ellipses for the plane model are obtained. A simplified model in the form of spheres and circles of mean square errors, for which only the parameters of radii are determined, was also considered. The obtained expressions were used in the model of linear interpolation of the parameters values of the zones between the measurement points. The parameters of the linear interpolation model representing the distribution zones in space as the form of cones tangent to the spheres at the points of inclinometric researches were determined. An algorithm for computer modeling on plans has been developed, which constructs zones in the form of circles of projections and straight lines tangent to them. A constructive and analytical expression for calculating the radius of the circle at any point of the trajectory is also obtained. Examples of modeling of error distribution zones around boreholes based on real data are given.

Keywords: wellbore, inclinometric observations, geoinformation system, interpolation, polylin

НАУКОВЕ ВИДАННЯ

Наукові праці Донецького національного технічного університету

Серія: “Інформатика, кібернетика та обчислювальна техніка”

№2(37), 2023

(українською, англійською мовами)

Редактор, коректор *О.В. Рябчук*

Формат 60×84 1/8
Ум. друк. арк. 10,3
Тираж 150 прим. Замовлення №3

Видавець та виготовлювач: Державний вищий навчальний заклад «Донецький національний технічний університет», пл. Шибанкова, 2, м. Покровськ, 85300, Україна

Свідоцтво про державну реєстрацію суб’єкта видавничої справи: ДК 4911 від 09.06.2015.