УДК 519.71

В. ХИЛЕНКО
Національний університет біоресурсів і природокористування України, Київ, Україна; Словацький технічний університет, Братіслава, Словаччина, e-mail: *vkhilenko@ukr.net*.

Б. АХМЕТОВ
Казахський національний педагогічний університет імені Абая, Алмати, Казахстан.

Р. БЕРДИБАЄВ
Алматинський університет енергетики та телекомунікацій, Алмати, Казахстан.

В. ЛАХНО
Національний університет біоресурсів і природокористування України, Київ, Україна.

Ю. ХАРЧЕНКО
Національний університет біоресурсів і природокористування України, Київ, Україна.

ВЕН-ЛІАНГ ХВАНГ
Інститут інформаційних наук, Academia Sinica, Тайбей.

В. ХИЛЕНКО мол.
Словацький технічний університет, Братислава, Словаччина.

ПІДВИЩЕННЯ ШВИДКОДІЇ БАНКІВСЬКИХ СИСТЕМ КІБЕРБЕЗПЕКИ НА ОСНОВІ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ТА АЛГОРИТМІВ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ПРОГНОЗУВАННЯ КІБЕРАТАК. Ч. 1

Анотація. Розглянуто підвищення швидкодії та якості роботи систем кіберзахисту банківських установ в умовах постквантумної ери. Запропоновано математичний апарат для систем прогнозування кібератак та алгоритм визначення моменту включення режиму підвищеної захищеності. Враховано можливість організації кібератак за допомогою нейромереж та алгоритмів штучного інтелекту. Наведено приклад формування та аналізу кластера підозрілих операцій із використанням мови Julia.

Ключові слова: кіберзахист банківських установ, загрози постквантумної ери, система прогнозування кібератак, запобігання кібератакам, кластеризація.

Природний розвиток та вдосконалення апаратних засобів, насамперед розвиток квантових комп'ютерів та програмного забезпечення, а також поява нових технологій — технологій штучного інтелекту (ШІ), підвищують загрози кібератак на банківські установи, а статистика свідчить про збільшення втрат банківських установ, спричинених кібератаками. Тому зростає важливість удосконалення та підвищення якості систем кіберзахисту банків та фінансових установ. Це зумовлює потребу раннього виявлення можливих напрямків атак та обходу контрольних (захисних) алгоритмів системи кібербезпеки. Упровадження таких аналітичних підсистем (підблоків) «раннього запобігання», що прогнозують у реальному режимі часу можливу підготовку кібератак до систем кібербезпеки банківських установ, потрібне для зменшення часу реагування на початок кібератак, раннього припинення розвитку негативних процесів (операцій) і мінімізації заподіяної шкоди.

Вважатимемо, що загальна база даних (БД) та інформаційні потоки банківської установи відповідають принаймні двом параметрам, які дають змогу стверджувати про їхню належність Big Data [1, 2]: показнику обсягу даних та показнику швидкості їхньої модифікації. З огляду на це однією з розглядуваних проблем ефективності систем кібербезпеки є оперативне виявлення даних, що вказують на спроби дестабілізації системи кібербезпеки та можливої підготовки до кібератаки. Аналіз таких даних з урахуванням прогнозів стійкого функціонування банківської установи в нормальному режимі потрібен для оперативного

реагування системи кібербезпеки на початок кібератаки, щоб за допомогою відповідних захисних механізмів унеможливити розвиток негативних процесів.

Вважатимемо, що початкову БД було кластеризовано з використанням алгоритмів (методів) кластеризації, орієнтованих на роботу з Big Data (використовують як ієрархічні, так й ітераційні методи) [1, 2]. У результаті сформовано кластер C, що поєднує інформацію стосовно неуспішних та нестандартних (підозрілих) запитів. Оскільки загальні обсяги статистичної інформації з банківських операцій потрапляють під визначення Big Data, системи передбачення та прогнозування кібератак (СППКА) насамперед використовуватимуть інформацію цього кластера C, розв'язуючи такі основні задачі:

— прогнозування розмірів кластерів та порівняння прогнозних даних з фактичними;

— аналіз та кластеризація адрес, з яких надходять нестандартні (підозрілі) запити;

— прогнозування за кластером нестандартних (підозрілих) запитів, з яких адрес (браузерів) чекати на подальші надходження звернень, які потенційно будуть включені до кластера C.

Розв'язання зазначених задач та мінімізація часу виявлення потенційно небезпечних адресатів спрямовані на скорочення часу реагування системи кібербезпеки. Перший етап — це час виявлення небезпечних ситуацій, а також підвищення надійності виявлення небезпечних запитів. Метою подальшого аналізу кластеризованих даних є прогнозування типу очікуваної атаки хакерів і вибір відповідних методів підвищення рівня кіберзахисту (так, для певного типу загроз — це блокування адрес, для іншого типу — скорочення часу зміни шифру). Однак оскільки в загальному випадку невідомим є рівень технічних засобів, які може використовувати зловмисник для зламування системи кіберзахисту [3], в умовах постквантумної ери вкрай важливим є встановлення спеціалізованого кіберзахисту квантових комп'ютерів — захисту постквантумного рівня [4].

У цілому алгоритм фукціонування СППКА полягає у виконанні таких основних етапів.

1. Аналіз даних та формування кластерів у стандартному режимі функціонування.

2. Формування та розрахунок математичної моделі динаміки кластерів.

3. Визначення допустимого діапазону варіативності кластера підозрілих звернень та запитів.

4. Формування підкластерів односпрямованих (однотипних) запитів.

5. Відстеження змін та аналіз у реальному режимі часу параметрів кластера та підкластерів.

6. Ухвалення рішень про корегування параметрів системи кіберзахисту.

Кількість елементів у кластері (розмір кластера) C, сформованому на першому етапі алгоритму СППКА, є випадковою величиною α, що належить зазвичай деякому обмеженому діапазону скінченної множини Z. Тобто вважаємо $\alpha \in Z$ і $\alpha < C_{\max}$, де $C_{\max}$ — деякий екстремум, значення якого задає користувач відповідно до емпіричних чи інших знань. Якщо значення α перевищує екстремум заданого діапазону $C_{\max}$, то потрібен додатковий аналіз та дослідження причин, що це зумовили. При цьому значення $C_{\max}$ корегуватиметься зі зміною часу відповідно до статистичного аналізу даних та змін характеристик роботи банківської установи (збільшення або зменшення кількості клієнтів, їхньої активності тощо).

Для певності будемо вважати, що виокремлення відповідних груп даних у кластер C може виконуватися із застосуванням, наприклад, методу k-се-

редніх [5, 6], а як метрику використовують Евклідову відстань [7]:

$$d(x_i, x_j) = \sqrt{\sum_{n=1}^{m} (x_{in} - x_{jn})^2},$$

де $x_i = x_i(x_{i1}, \ldots, x_{in})$, $x_j = x_j(x_{j1}, \ldots, x_{jn})$.

Загалом зміна розмірів кластерів зумовлена різними факторами і визначається нелінійною залежністю

$$\dot{x}_i = f(x, \alpha, t), \; x_i(t_0) = x_i^0, \tag{1}$$

де x — елементи деякого нормованого простору E, $f(x)$ — вектор функцій, заданих на множині G, природа якої визначається насамперед структурою моделі ($G \in E$), α — вектор випадкових параметрів, t_0 — початковий момент інтегрування. Оскільки в нормальному режимі функціонування банківської установи усереднені характеристики змінюються досить повільно, а також враховуючи математичні труднощі розв'язання нелінійної оберненої задачі доцільно перейти від системи (1) до кусково-лінеаризованої моделі. У цілому формування системи (1) може бути залежним від розв'язання некоректних обернених задач. Як на цьому етапі, так і на наступних етапах аналізу моделювання та прийняття рішень це зумовлює використання глибоких нейронних мереж [8] та пов'язаних з цим напрямком задач, що потребують окремого розгляду.

Використовуючи статистичні залежності та застосовуючи методи регресійного аналізу [9], перейдемо від системи (1) до моделі динаміки зміни розміру кластерів у вигляді системи лінійних звичайних диференціальних рівнянь:

$$\dot{v} = Av + B(\alpha, t),$$

де $A = [a_{ij}]$, $B = [b_i(\alpha, t)]$ $(i, j = \overline{1, n})$ — матричні коефіцієнти, $v = (v_1, \ldots, v_n)^t$ — вектор шуканих змінних, $t \in [t', t'']$, де t', t'' — початкова та кінцева точки інтервалу лінеаризації.

Відповідно до технологічного режиму роботи банківської установи і під впливом різних випадкових факторів кількість запитів і операцій є послідовністю випадкових величин, динаміка яких залежить від змінних, що швидко змінюються. Усереднені значення, що моделюються на тривалому інтервалі часу, є повільно змінюваними функціями. Динаміка цих функцій є непрямою характеристикою процесів розвитку чи стагнації банківської установи і дає змогу зіставляти прогнози та реальні тенденції (характеристики, показники) динаміки об'єкта, а зміна складових, що швидко змінюються, відображає оперативний стан операційної активності учасників процесів.

Зазвичай динамічна модель варіювання розміру кластерів банківської установи належить класу систем з однорідним примежовим шаром. Але для збереження загальності визначимо формули перетворення змінних для випадку ступінчастого примежового шару [10], формуючи на кожному кроці відповідну апроксимувальну систему. Застосовуючи алгоритми та обчислювальні схеми методу зниження порядку [10], послідовно вилучатимемо на кожному кроці перетворення вихідної моделі відповідну «швидку» групу змінних, яка визначається аналогічно випадку однорідного примежового шару. Дотримуючись методу зниження порядку, запишемо змінні, які підлягають виродженню на r-му кроці:

$$v_i^r(t) = y_i(t) - \sum_{j=k_{r-1}+1}^{k_r} a_{ij}^r v^r{}_{(j+q_r-1)}, \tag{2}$$

де k_{r-1} — кількість змінних, вилучених під час послідовного виконання на $(r-1)$-му кроці зниження порядку початкової системи рівнянь, а величина $k_r - k_{r-1} + 1$ визначає кількість вилучених змінних на *r*-му кроці перетворення. Величини k_r і k_{r-1} обчислюються у разі послідовного зниження порядку початкової системи відповідно до алгоритму визначення квазіприєднаної системи рівнянь [9]. Елементи матриці A^r вигляду

$$a_{ij}^r = a_{ij}^r(a_{ij}^{r-1}) \tag{3}$$

визначаються під час розв'язання вироджених $(k_{r-1}+1), \ldots, k_r$ рівнянь системи на $(r-1)$-му кроці відносно V_i^{r-1} .

Систему рівнянь відносно $y(t)$ отримаємо у разі підстановки виразів (2) в $(k_{r-1}+1), \ldots, k_r$ рівняння системи на $(r-1)$-му кроці. Тоді початкова система рівнянь на *r*-му кроці набуде вигляду

$$\frac{dy}{dt} = D^r y + c^r v^r + G^r(t), \qquad \frac{dv^{r1}}{dt} = A^{r1} v^{r1} + B^{r1}(t), \tag{4}$$

де

$$D^r = [d_{ij}^r],\ c^r = [c_{im}^r],\ A^{r1} = [a_{lm}^{r1}],\ B^{r1} = [b_l^{r1}],\ d_{ij} = d_{ij}(a^1, a^2),$$

$$a_{lm}^r = a_{lm}^r(a_{gs}^{r-1})\ \ (i, j = \overline{k_{r-1}+1, k_2};\ l, m = \overline{k_2+1, n};\ g, s = \overline{k_{r-1}+1, n}).$$

Отже, наявність різношвидкісних процесів у системі (1) дає змогу, використовуючи метод зниження порядку [9], записати її у вигляді квазітихоновської [11] системи (4), де підвектори y і v^{r1} описують складові розв'язку системи (1), що змінюються з суттєво різною швидкістю. Процес розрахунку динаміки «повільних» складових v^{r1} дає змогу оцінювати поточний рівень кібератак чи некоректних запитів та прогнозувати природні зміни цих показників. Відхилення від прогнозних даних може свідчити про початок кібератаки або підготовки до кібератаки. Для прийняття системою кібербезпеки банківської установи технологічних рішень з корегування слід визначити діапазон допустимого розміру кластера *C*, який не потребує миттєвих корегувань режимів її роботи, а також швидкість зміни розміру кластера, що відповідає стандартному режиму роботи банківської установи. Поєднання двох складових: потужності множини *C* та швидкості збільшення розміру кластера розглядатимемо як параметр для оцінювання поточного рівня кібератак (некоректних запитів) і відповідно прийняття управлінських рішень.

Як приклад реалізації першого етапу роботи СППА розглянемо формування фрагмента кластера підозрілих запитів та звернень. Вибірка кластера формувалась з використанням мови Julia [12].

На першому кроці проведемо аналіз даних журналу доступу до вебсервера відповідно до такої інформації (далі ключових даних): **ip** — ip-адреса, з якої зроблено запит; **met** — http-метод запиту; **ret** — код результату запиту; **br** — агент, яким виконано запит.

Після завантаження файлів з допомогою відповідної програми формуємо таблицю ключових даних, фрагмент якої наведено нижче:

log =

	ip	ts	met	ret	
1	"86.179.233.90"	2022-03-04T00:10:21	"GET / HTTP/1.1"	200	"Moz
2	"106.75.223.50"	2022-03-04T00:40:02	"GET / HTTP/1.0"	200	"-"
3	"109.237.103.118"	2022-03-04T01:23:51	"GET /.env HTTP/1.1"	400	"Moz
4	"109.237.103.118"	2022-03-04T01:23:51	"GET /.env HTTP/1.1"	404	"Moz
5	"124.217.226.56"	2022-03-04T01:56:43	"GET /wp-login.php HTTP/1.1"	404	"Moz
6	"31.173.207.151"	2022-03-04T02:02:11	"GET / HTTP/1.1"	200	"Moz
7	"35.233.62.116"	2022-03-04T02:15:41	"GET / HTTP/1.1"	200	"pyt
8	"31.220.3.140"	2022-03-04T02:53:03	"GET / HTTP/1.1"	200	"Moz
9	"35.195.93.98"	2022-03-04T02:53:24	"GET / HTTP/1.1"	200	"pyt
10	"20.79.40.222"	2022-03-04T02:56:55	"GET / HTTP/1.1"	200	"Pyt
⋮ more					
16525	"114.119.131.131"	2022-03-13T17:34:24	"GET /robots.txt HTTP/1.1"	404	"Moz

Використовуючи початкову та сформовану базу даних, проведемо з подальшим розрахунком статистичних показників формування підкластерів для ключових даних за такими факторами:

- count — кількість записів в групі;
- part — відносна частка групи;
- gridx — ідентифікатор групи.

Наведемо фрагменти таблиць групування за окремими ключовими даними.

Фрагмент таблиці групування за ip-адресою:

	ip	count	part	gridx
1	"46.185.24.82"	5782	0.349894	137
2	"8.42.51.131"	4245	0.256884	295
3	"176.196.70.35"	1545	0.0934947	906
4	"198.27.82.185"	206	0.012466	819
5	"45.90.58.54"	202	0.0122239	357
6	"45.146.165.37"	200	0.0121029	11
7	"172.104.159.48"	133	0.00804841	849
8	"3.9.22.86"	132	0.0079879	147
9	"112.220.133.212"	132	0.0079879	262
10	"113.53.231.82"	132	0.0079879	266
⋮ more				
1003	"114.119.131.131"	1	6.05144e-5	1003

```
sort(D["ip"]["wdf"], [order(:count, rev=true)])
```

Фрагмент таблиці групування за http-методом запиту:

	met	count	part	gridx
1	"CONNECT www.msftncsi.com:443 HTTP/1.1	11453	0.693071	136
2	"GET / HTTP/1.1"	979	0.0592436	1
3	"GET /robots.txt HTTP/1.1"	189	0.0114372	19
4	"GET / HTTP/1.0"	150	0.00907716	2
5	"GET /favicon.ico HTTP/1.1"	130	0.00786687	21
6	"GET /.env HTTP/1.1"	87	0.00526475	3
7	"GET /sitemap.xml HTTP/1.1"	56	0.0033888	22
8	"POST / HTTP/1.1"	39	0.00236006	58
9	"GET /HNAP1/ HTTP/1.1"	39	0.00236006	66
10	"PRI * HTTP/2.0"	35	0.002118	6
⋮	more			
925	"\\x16\\x03\\x01\\x00\\xB2\\x01\\x00\\	1	6.05144e-5	925

```
sort(D["met"]["wdf"], [order(:count, rev=true)])
```

Фрагмент таблиці групування за агентом запиту:

	br	count	part	gridx
1	"-"	12376	0.748926	2
2	"Mozilla/5.0 (Windows NT 5.1; rv:9.0.1	996	0.0602723	5
3	"Mozilla/5.0 (Windows NT 10.0; Win64;	498	0.0301362	58
4	"Mozilla/5.0 (Windows NT 10.0; Win64;	184	0.0111346	9
5	"Go-http-client/1.1"	157	0.00950076	23
6	"Mozilla/5.0 (X11; Linux x86_64) Apple	135	0.00816944	3
7	"Mozilla/5.0 zgrab/0.x"	104	0.00629349	7
8	"Mozilla/5.0 (Windows NT 6.1) AppleWeb	102	0.00617247	25
9	"curl/7.54.0"	102	0.00617247	218
10	"Mozilla/5.0 (Windows NT 10.0; Win64;	95	0.00574887	51
⋮	more			
246	"Mozilla/5.0 (Windows NT 10.0; Win64;	1	6.05144e-5	246

```
sort(D["br"]["wdf"], [order(:count, rev=true)])
```

Проаналізуємо зв’язок між ключовими даними (факторами) з використанням формули [13]:

$$\mathbf{d} = \mathbf{p}_{ab} - \mathbf{p}_a * \mathbf{p}_b,$$

де $\mathbf{p}_{ab}$ — парна частота факторів a і b, $\mathbf{p}_a * \mathbf{p}_b$ — добуток індивідуальних частот факторів a і b; **d** — оцінка зв’язку, де додатне значення — зв’язок по-

зитивний, від'ємне значення — зв'язок негативний. На підставі аналізу сформуємо таблицю зв'язків між даними. Фрагмент таблиці зв'язків «код результату — агент запиту» наведено нижче:

	d	pab	papb	pa	pb	a	
1	0.17657	0.717337	0.540768	0.722057	0.748926	400	"-"
2	0.0468316	0.0579123	0.0110806	0.183843	0.0602723	404	"Mozilla/5.0 (Wi
3	0.0245958	0.0301362	0.00554031	0.183843	0.0301362	404	"Mozilla/5.0 (Wi
4	0.00563611	0.00738275	0.00174664	0.183843	0.00950076	404	"Go-http-client/
5	0.00503316	0.00708018	0.00204702	0.183843	0.0111346	404	"Mozilla/5.0 (Wi
6	0.004368	0.00586989	0.00150189	0.183843	0.00816944	404	"Mozilla/5.0 (X1
7	0.00328279	0.00441755	0.00113476	0.183843	0.00617247	404	"curl/7.54.0"
8	0.00321086	0.00405446	0.000843605	0.075764	0.0111346	200	"Mozilla/5.0 (Wi
9	0.00313951	0.00429652	0.00115701	0.183843	0.00629349	404	"Mozilla/5.0 zgr
10	0.00293706	0.00399395	0.00105689	0.183843	0.00574887	404	"Mozilla/5.0 (Wi
⋮ more							
403	-0.124916	0.0127685	0.137685	0.183843	0.748926	404	"-"

```
sort(get_mat_p(D, "ret", "br"), [order(:d, rev=true)])
```

Окрім статистичного аналізу додаткову інформацію для прогнозування і прийняття рішень надає аналіз процесів доступу до сервера. Стуктурна схема процесів доступу до сервера містить таке:

$$\text{ip-адреса} \rightarrow \text{агент} \rightarrow \text{метод} \rightarrow \text{код результату}$$

і дає змогу прогнозувати результати запиту та самого адресата стосовно цілей його запиту і можливих подальших звернень за першим фактором.

Відповідно до наведеного вище з ip-адреси 66.249.72.227 слід очікувати запити від агента "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" та методами "GET /search.php HTTP/1.1", "GET / HTTP/1.1", "GET /robots.txt HTTP/1.1", "GET /notifications.php HTTP/1.1", а з ip-адреси 103.203.57.10 можливе використання двох агентів "Mozilla/5.0 (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/45.0.2454.101 Safari/537.36" та "HTTP Banner Detection (https://security.ipip.net)". Отже, побудована структурна модель процесу дає можливість за допомогою вихідних даних прогнозувати появу ключових даних запитів, які можуть являти собою потенційну небезпеку.

СПИСОК ЛІТЕРАТУРИ

1. Ghavami P. Big Data management: Data governance principles for Big Data analytics. Berlin: De Gruyter, 2020. 174 p.
2. Balusamy B., Nandhini A.R., Kadry S., Gandomi A.H. Big Data: Concepts, technology, and architecture. Wiley, 2021, 368 p.
3. Fikar J. Chinese scientists claim to be able to break the RSA cipher with a new quantum computer algorithm. 6.1.2023. URL: https://www.root.cz/zpravicky/cinsti-vedci-tvrdi-ze-umi-prolomit-rsa-sifru-novym-algoritmem-pro-kvatovy-pocitac/.

4. Khilenko V.V. Formation of a new conception and a paradigm of constructing cybersecurity systems. *Cybernetics and Systems Analysis*. 2019. Vol. 55, N 3. P. 354–358. https://doi.org/10.1007/s10559-019-00141-8.

5. Coates A., Ng A.Y. Learning feature representations with *k*-means. In: Neural Networks: Tricks of the Trade. Montavon G., Orr G.B., Müller K.R. (Eds). *Lecture Notes in Computer Science.* Berlin; Heidelberg: Springer, 2012. Vol. 7700. P. 561–580. https://doi.org/10.1007/978-3-642-35289-8_30.

6. Celebi M.E., Kingravi H.A., Vela P.A. A comparative study of efficient initialization methods for the *k*-means clustering algorithm. *Expert Systems with Applications*. 2012. Vol. 40, N 1. P. 200–210. https://doi.org/10.1016/j.eswa.2012.07.021.

7. Deza M.M., Deza E. Encyclopedia of distances. 4th ed. Berlin; Heidelberg: Springer, 2016. 756 p. https://doi.org/10.1007/978-3-662-52844-0.

8. Hwang W.-L. , Tung S.-S. Analysis of function approximation and stability of general DNNs in directed acyclic graphs using un-rectifying analysis. arXiv:2206.05997v1 [cs.LG] 13 June 2022. https://doi.org/10.48550/arXiv.2206.05997.

9. Fahrmeir L., Kneib T., Lang S., Marx B. Regression models, methods and applications. Berlin; Heidelberg: Springer, 2022. 746 p.

10. Грищенко А.З., Хиленко В.В. Метод понижения порядка и исследование динамических систем. Киев: УМК ВО, 1988. 164 с.

11. Моисеев Н.Н. Математические задачи системного анализа. Москва: Наука, 1981. 488 с.

12. Шарингтон М. Осваиваем язык Julia. Москва: ДМК Пресс, 2017. 416 с.

13. Миркин Б.Г. Анализ качественных признаков и структур. Москва: Статистика, 1980. 319 с.

V. Khilenko, B. Akhmetov, R. Berdibayev, V. Lakhno, Yu. Harchenko, Wen-Liang Hwang, V. Khylenko, Jr.

INCREASING THE SPEED OF BANKING CYBER SECURITY SYSTEMS BASED ON INTELLIGENT DATA ANALYSIS AND ARTIFICIAL INTELLIGENCE ALGORITHMS FOR PREDICTING CYBER ATTACKS. P. 1

Abstract. An increase in the speed and quality of the cyber protection systems of banking institutions in the post-quantum era is considered. A mathematical apparatus for cyber attack prediction systems and an algorithm for choosing the moment of switching on the enhanced security mode are proposed. The possibility of organizing cyber attacks using neural networks and AI algorithms is taken into account. An example of the formation and analysis of a cluster of suspicious transactions using the Julia language is considered.

Keywords: cyber protection of banking institutions, threats of the post-quantum era, system of prediction and prevention of cyber attacks, clustering.

Надійшла до редакції 02.03.2023