КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 3 (23), 2024
ISSN 2663 - 4023
DOI 10.28925/2663-4023.2024.23.5670
УДК 004.056.53

Добришин Юрій Євгенович
кандидат технічних наук, доцент
Національна академія Служби безпеки України, Київ, Україна
ORCID 0000-0003-2473-9507
ydobryshyn@gmail.com

ВИКОРИСТАННЯ СТАТИСТИЧНИХ МЕТОДІВ ЩОДО ПРОГНОЗУВАННЯ ФІШИНГОВИХ АТАК

Анотація. У статті запропонована методика щодо прогнозування так званих фішингових атак, які є поширеною формою здійснення кіберзлочинів, кількість яких з кожним роком зростає, а також збільшується рівень їхнього шкідливого впливу на інформаційні системи об'єктів критичної інфраструктури. Для аналізу трендів та прогнозування фішингових атак використані статистичні дані, що опубліковані в наукових працях вітчизняних та зарубіжних дослідників, а також оприлюднені інтернет-виданнями провідних консалтингових компаній, що працюють у сфері інформаційної безпеки та кібербезпеки. У якості інструментів дослідження та прогнозування фішингових атак були обрані статистичні методи на основі застосування часових рядів як одного із популярних підходів, що використовується до прогнозування різних технологічних та економічних процесів. Це дозволило здійснити аналіз типів та шаблонів фішингових атак, за яким зловмисники вчиняють дії щодо порушення роботи програмного забезпечення інформаційно-комунікаційних систем та автоматизованих комплексів. На основі аналізу часових рядів, здійснено побудову моделі тренду щодо кількості виявлених фішингових атак за період 2020–2023 рр. та виконано розрахунок прогнозованої кількості фішингових атак за 16 кварталів 2020–2023 років, а також ймовірний прогноз появи зазначених атак за чотири квартал 2024 року. Для покращення прогнозу, розраховано коефіцієнт що враховує фактор сезонності та виконано кореляційно-регресійний аналіз впливу фішингових атак на загальну кількість атак, які були виявлені протягом 2020–2023 років. Виконані розрахунки, свідчать про те що розбіжності прогнозованих значень не значні, наведені результати дозволяють здійснювати вибір оптимальної стратегії щодо виявлення, прогнозування та усунення комп'ютерних атак, які відносяться до фішингу. На основі моделі часових рядів та отриманих розрахунків, зроблено висновок про те, що статистичні методи прогнозування дозволяють побудувати прогноз фішингових атак, надають у подальшому можливість розробляти та опрацьовувати методики протидії зазначеним атакам, планувати заходи щодо підвищення рівня захищеності інформаційних ресурсів.

Ключові слова: прогнозування; фішингова атака; статистичні методи; часові ряди; автоматизована інформаційно-комунікаційна система; тренд; кореляційно-регресійний аналіз.

ВСТУП

Постановка проблеми. Сучасне інформаційне середовище характеризується складними процесами та технологіями, що розробляються, аналізуються, постійно супроводжуються персоналом, який здійснює експлуатацію та захист автоматизованих систем та комплексів. Окрім вищезазначеного, життєвий цикл захищених систем включає виконання процедур з виявлення, попередження, блокування, нейтралізації та усунення наслідків реалізації кіберзагроз [1] – [3]. Результати досліджень кіберзагроз використовуються розробниками захищених інформаційних систем та службами захисту

інформації використовуються на етапах проєктування та експлуатації для формування та реалізації вимог щодо кібербезпеки обєктів критиної інфраструктури [2], [4], [5].

Останнім часом статистичні дані свідчать про те, що у період з 2022 року по 2023 рік збільшились масштаби кіберзлочинів [6]. Зокрема, спостерігається тенденція поширення фішингових атак, що побудовані на основі різних технологій для досягнення головної мети — введення в оману користувачів систем та отримання важливої інформації, застосовуючи електрону пошту, веб-сторінки, соціальні мережі, телефонні дзвінки, текстові повідомлення тощо для.

На думку фахівців, серед головних причин зростання кількості кіберзагроз є: застосування нових інформаційних технологій, які мають проблеми з безпекою своїх програмних модулів [2], вразливості протоколів, компонентів інтерфейсів [5], систем управління базами даних тощо. Спостерігається тенденція орієнтації злочинців на доступ до інформації великих корпорацій, активна діяльність як поодиноких кіберзлочинців, так і організованих злочинних угрупувань.

Враховуючи вищезазначене, актуальними постають питання щодо виявлення фішингових атак, а саме, їх прогнозування за допомогою математичних методів, що включають в себе аналіз даних, статистику, машинне навчання тощо.

Одним із напрямком вирішення проблеми прогнозування фішингових атак є застосування статистичних методів, які грають важливу роль у прогнозуванні зазначених атак, тому що дозволяють аналізувати відомості щодо кількості та тенденцій появи фішингових атак, за яким зловмисники порушують працездатність автоматизованих комп'ютерних систем та комплексів. Окрім цього статистичні методи дозволяють виявити середньостроковий прогноз фішингових атак з урахуванням факторів сезонності та тривалості за часом, що суттєво впливає на стан необхідного рівня безпеки інформаційних ресурсів та запобігання появи різного роду інцидентів щодо можуть становити реальну загрозу.

Аналіз останніх досліджень і публікацій. Питання щодо застосування статистичних методів щодо прогнозування комп'ютерних атак з урахуванням особливостей використання інформаційно-телекомунікаційних систем та комплексів, а також існуючого програмного забезпечення на теперішній час є актуальними.

Підходи щодо прогнозування та виявлення комп'ютерних атак розглядаються у багатьох наукових працях вітчизняних та закордонних фахівців [7] – [17]. Більшість робіт присвячена вирішенню загальних проблем прогнозування та виявлення комп'ютерних атак, але у багатьох з них висвічуються сучасні підходи до виявлення та прогнозування загроз, які пов'язані з процесами експлуатації та супроводження комп'ютерних мереж.

Незважаючи на значну кількість існуючих сучасних методів та математичних моделей щодо виявлення та прогнозування комп'ютерних атак та загроз, для практичного вирішення та доповнення окремих технологій прогнозування різних комп'ютерних атак, використовуються статистичні методи на підставі застосування часових рядів.

Так у науковій роботі [7] розглядаються питання виявлення та прогнозування загроз для комп'ютерної мережі шляхом застосування спеціалізованої автоматизованої системи, яка передбачає використання як сигнатурних методів так і широкого спектру статистичних методів на базі часових рядів з метою аналізу мережевого трафіку. У якості вхідних даних для прогнозування загроз безпеки у системі застосовуються кількісні та якісні параметри трафіку, які сформовані у вигляді певних часових рядів, на приклад, інтерфейс, кількість байт, кількість помилок тощо.

Таким чином автоматизована система аналізує часові ряди показників мережевого трафіка та надає прогноз щодо можливих загроз інформації під час експлуатації комп'ютерної мережі.

Інтерес представляє наукова робота [8], у якій розглядаються технологічні підходи щодо прогнозування атак, що використовують програмне забезпечення для доступу до об'єктів комп'ютерної мережі з використанням протоколу SSH. В роботі розглядається модель часових рядів першого порядку та здійснюється аналіз атак за певний період часу, а саме, з 02 листопада 2014 року по 08 травня 2016 року. За висновками авторів, застосування зазначеної моделі часових рядів першого порядку дозволяє з впевненістю та достовірністю здійснювати прогнозування майбутніх атак.

Група авторів [9], [11] пропонують здійснювати прогноз інтенсивності атак на основі отриманих даних щодо кількості атак за добу, тиждень. Для прогнозування застосовуються часові ряди, опис яких здійснюється за допомогою моделі прогнозування ARIMA.

Приведений варіант прогнозування атак, з використанням зазначеної моделі, дозволяє виявляти та прогнозувати появу найбільш шкідливих чотирьох типів атак: відмова в обслуговуванні (DoS), шкідливі електронні листи, шкідливі URL-адреси та атаки на служби Інтернет.

У статті [10] представлена методика прогнозування уразливостей за період з січня 1999 року по січень 2016 року за допомогою часових рядів. Для прогнозування атак, у роботі використовувалися моделі ARCH, GARCH та SARIMA. Дані для аналізу були взяті з Національної бази даних уразливостей (NVD) станом на 2016 року. Результати прогнозів були переважно корисні щодо управління ризиками уразливостей. Автори у своїй роботі досліджують використання моделі ARIMA для прогнозування майбутніх випадків кібератак та їх інтенсивності. Для побудови часових рядів автори використовують різні періоди вимірювання для кращого моделювання часових шаблонів, унікальних для кожного типу атаки. Отримані результати надають додаткові докази на підтримку висновків моделі ARIMA.

Важливість використання часових рядів щодо прогнозування атак зазначається у роботах [12], [13] автори вважають, що прогнозування рядів має широкий спектр різних сфер застосування: від аналізу фондового ринку та оцінювання росту економіки окремих регіонів до моделювання поширення епідемій і прогнозу погоди. Крім цього для аналізу та прогнозування часових рядів може бути застосований математичний апарат прихованих марківських моделей.

Зазначені підходи, також є найбільш фундаментальними щодо виявлення та запобігання атакам типу фішинг. Методи та моделі прогнозування атак типу фішинг розглядаються фахівцями з інформаційної безпеки у багатьох наукових працях.

Актуальність прогнозування та виявлення фишингових атак зазначена у роботі [14]. Автори розглядають фітинг. як серйозну загрозу для користувачів. У роботі запропоновано метод щодо протидії фішинговим атакам за рахунок застосування динамічної моделі у вигляді найпростішої штучної нейронної мережі. Запропонований метод дозволяє з великою ймовірністю здійснювати виявлення та прогнозування фішингових атак, використовує для штучної нейронної мережі статистичні методи прогнозування отриманих результатів.

У статті [15] розглядаються проблеми дослідження та прогнозування фішингових атак. На підставі проведення дослідження, з метою оцінки ймовірності фішингових атак запропонований метод дискретного вейвлету перетворення, а також показані особливості використання основних характеристик вейвлет-аналізу для дослідження часових рядів.

У наукових дослідженнях [16] проблема прогнозування фішингу розглядається на підставі аналізу теоретичних та практичних робіт, які здійснюються із застосуванням штучного інтелекту. В роботах надається алгоритм роботи штучного інтелекту для виявлення фішингу та аналізу постійно змінних моделей фішингу.

Ряд авторів [17] вважають що на появу такої загрози, як фішинг суттєво впливають дії персоналу, який здійснює експлуатацію інформаційно-телекомунікаційних систем, тому для цього пропонується застосовувати соціоінженерний підхід, який дозволяє за допомогою статистичних методів виконати оцінку захищеності інформації в автоматизованих системах та комплексах, виявити та зпрогнозувати вразливості і загорози захисту інформації.

Таким чином проблеми виявлення та прогнозування комп'ютерних атак, особливо фішингових атак, залишаються актуальними на теперішній час.

Незважаючи на те, що існує достатньо методів та технологій виявлення та класифікації атак типу фішинг, їх застосування має обмежений успіх. Більшість методів виконують прогнозування в умовах обмеження та певних налаштувань програмного забезпечення автоматизованих систем та комплексів, але не спроможні виявити атаки, які раніше не траплялися.

Прогнозування фішингових атак має середньостроковій фактор, тому вимагає від фахівців з захисту інформації визначити області, у яких слід розробити довгострокове прогнозування, яке залежить від суб'єктивного досвіду експерта з безпеки інформації. Виявлення фішингових атак продовжує бути актуальним та потребує подальших наукових досліджень із застосуванням статистичних методів та моделей, які дозволяють на підставі аналізу часових рядів здійснити прогнозування появи зазначених атак, визначити їх властивості та ознаки загроз.

Мета дослідження. Зазначена стаття ставить за мету проведення дослідження щодо підвищення рівня захищеності ресурсів автоматизованих систем та комплексів шляхом прогнозування появи фішингових атак на підставі статистичних методів аналізу структури часових рядів. На основі моделі часових рядів пропонується побудувати прогноз фішингових атак з урахуванням фактору сезонності та здійснити кореляційно-регресійний аналіз, який дозволить забезпечити необхідний рівень безпеки інформаційних ресурсів автоматизованих систем та комплексів.

ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

Застосування статистичних методів виступає головним інструментом у прогнозуванні існуючих та нових фішингових атак. Тому головною задачею служб захисту інформації є постійний моніторинг та статистичний аналіз динаміки основних показників інформаційних ресурсів з метою прийняття правильних тактичних та стратегічних рішень.

Під інструментом мається на увазі не тільки використання математичних методів та моделей, а методологічний підхід щодо аналізу взаємозв'язків між властивостями кібератак, на підставі яких можливо прийняти правильні рішення щодо прогнозування та їх усунення.

На підставі даних, що отримані з [18] – [22] побудуємо модель тренду для кількості фішингових атак що спостерігалися за період 2020–2023 рр. Для цього кількість фішингових атак подаємо у вигляді часового ряду [23], який наведений у табл. 1.

ISSN 2663 - 4023

Таблиця 1

Кількість спостережень фішингових атак

Роки	Квартал	Кількість спостережень фішингових атак (одиниць)
2020	I	305506
	II	275760
	III	940609
	IV	1035542
2021	I	938259
	II	648706
	III	818575
	IV	933160
2022	I	1025968
	II	1165110
	III	1399671
	IV	1441661
2023	I	1624144
	II	1286208
	III	1672070
	IV	2340898

У часовому ряду, що наведений у таблиці, є дві змінні:

– період (квартал), за яким спостерігалися фішингові атаки;
– кількість виявлених фішингових атак (одиниць).

На підставі даних табл. 1 з використанням програмного забезпечення MS Office Excel побудований графік (рис. 1), який відображає динаміку появи фішингових атак, де по осі *X* періоди спостережень, по осі *Y* — кількість атак, що були виявлені. Для більш наочності, наведені лінія тренду та рівняння тренду на графіку, а також величина достовірності апроксимації *R*.

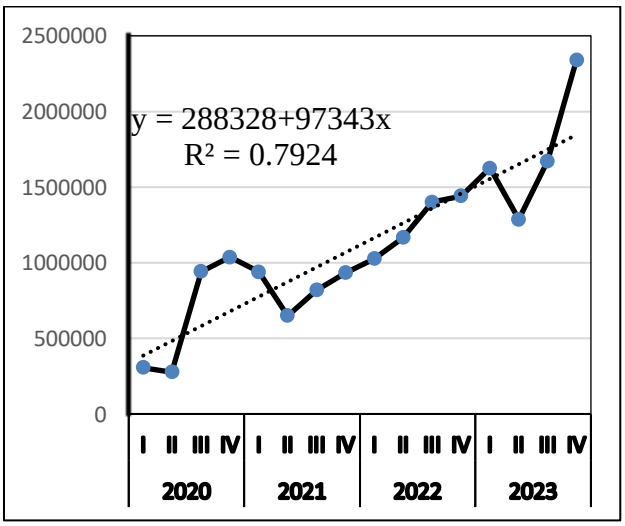

Рис. 1. Динаміка появи фішингових атак за період з 2020–2023 рр.

ISSN 2663 - 4023

Для проведення розрахунків, будемо застосовувати найбільш привабливий лінійний тренд, рівняння якого має наступний вигляд:

$$Y_x = a + b \cdot x, \quad (1)$$

де Y_x — кількість спостережених фішингових атак; x — період – квартал, на протязі якого були виявлені атаки; a — місце (точка) перетинання лінії тренду на графіку; b — величина, на яке збільшується наступне значення часового ряду.

На основі побудованого графіку рис. 1 визначені параметри a, b, що дає змогу рівняння тренду (1) подати у наступному вигляді:

$$Y_x = 288328 + 97343 \cdot x. \quad (2)$$

За допомогою значення тренду виконаємо розрахунок прогнозованої кількості фішингових атак за 16 кварталів 2020–2023 років, а також ймовірний прогноз появи зазначених атак на чотири квартали 2024 року. Результати розрахунків представимо у табл. 2.

Необхідно зазначити, що застосування моделі тренду з метою прогнозування атак може застосовуватися тільки до відносно короткого період їх виявлення. Використання моделі тренду для тривалого періоду може призвести до невірної оцінки (прогнозу).

Розрахунки, що були наведені можуть бути уточнені, у разі, коли під час аналізу атак фахівці з захисту інформації стикаються з циклічними коливаннями, які викликані сезонним характером появи таких атак.

Таблиця 2

Розрахунок (прогноз) кількості фішингових атак протягом 2020–2023 рр. та імовірний прогноз кількості таких атак на 2024 р.

a=288328	**b=97343**
X	**Y**
1	385671
2	483014
3	580357
4	677700
5	775043
6	872386
7	969729
8	1067072
9	1164415
10	1261758
11	1359101
12	1456444
13	1553787
14	1651130
15	1748473
16	1845816
17	1943159
18	2040502
19	2137845
20	2235188

З метою реалізації випереджаючих заходів щодо захисту інформації надзвичайно важливо вивчати і аналізувати тенденції сезонних коливань, що виникають, і розробляти прогноз на деяку перспективу, зокрема, на наступний рік.

Для аналізу сезонних коливань будемо використовувати спеціальні показники, сукупність яких утворює сезонну хвилю, безпосередньо показники будемо називати індексами сезонності.

Для розрахунків будемо використовувати отримані раніше значення тренду для кожного періоду (табл. 1) та величину відхилення фактичної кількості фішингових атак від значень, що отримані за допомогою тренду.

Величину відхилення визначимо за наступним рівнянням:

$$\delta = {V_p}/{Y_x}, \quad (3)$$

де δ — відхилення фактичних значень кількості атак; V_p — кількість спостережених фішингових атак; Y_x — значення тренду.

Таблиця 3

Величина відхилення фактичної кількості фішингових атак від значень, отриманих за допомогою тренду

Рік	Квартал	Кількість атак (одиниць) *Vp*	Значення тренду (*Yx*)	Відхилення δ
2020	I	305506	385671	0.792
	II	275760	483014	0.571
	III	940609	580357	1.621
	IV	1035542	677700	1.528
2021	I	938259	775043	1.211
	II	648706	872386	0.744
	III	818575	969729	0.844
	IV	933160	1067072	0.875
2022	I	1025968	1164415	0.881
	II	1165110	1261758	0.923
	III	1399671	1358101	1.031
	IV	1441661	1456444	0.990
2023	I	1624144	1553787	1.045
	II	1286208	1651130	0.779
	III	1672070	1748473	0.956
	IV	2340898	1845816	1.268

На підставі даних табл. 3 виконано розрахунок середнього відхилення кількості атак для кожного кварталу та загальний індекс сезонності I_{zs} за допомогою функції MS Office Excel (=СРЗНАЧ()). Далі для отримання коефіцієнтів сезонності для кожного кварталу K_s, виконане нормування середнього відхилення кількості фішингових атак для кожного кварталу на індекс сезонності:

$$K_s = {\delta_{sr}}/{I_{zs}}, \quad (4)$$

де K_s — коефіцієнт сезонності, визначений для кожного кварталу, виходячи з кількості атак; δ_{sr} — середнє відхилення кількості атак для кожного кварталу (2); I_{zs} — індекс сезонності, визначений за допомогою функції MS Office Excel (=СРЗНАЧ()).

Результати розрахунків наведені у табл. 4.

Таблиця 4

Узагальнені результати розрахунків коефіцієнтів сезонності для фішингових атак з урахуванням їх появи протягом 2020–2023 рр.

Рік	Квартал	Середнє відхилення об'ємів атак δsr	Коефіцієнти сезонності Ks
2020–2023	I	0.982	1.170
	II	0.754	1.048
	III	1.113	0.982
	IV	1.165	0.890

Дані з табл. 4 використані для побудови графіку коливань індексів сезонності спостережених фішингових атак для кожного кварталу (рис. 2).

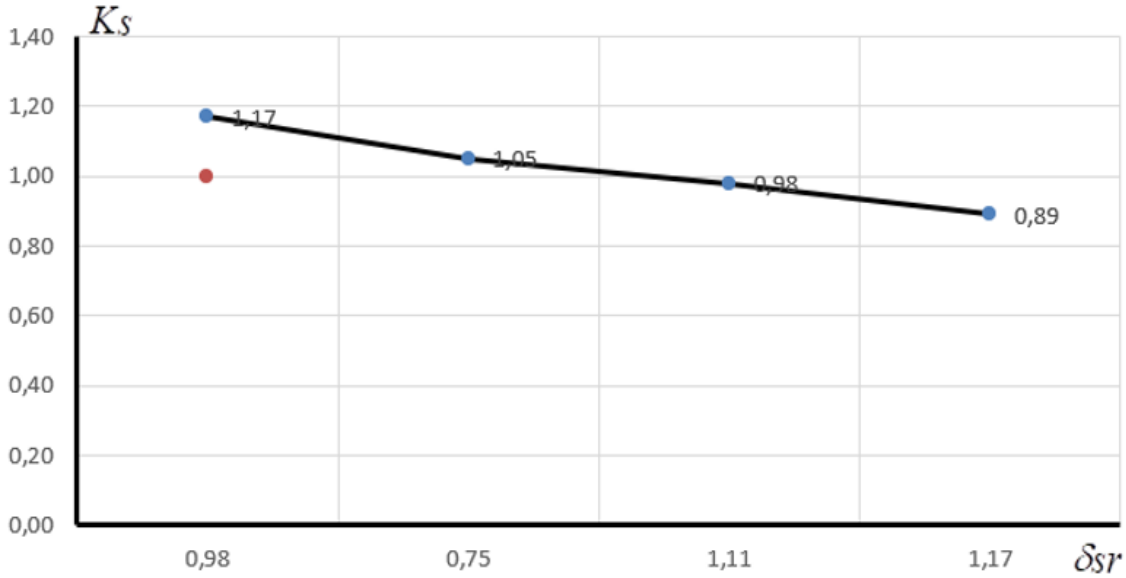

Рис. 2. Динаміка коливань індексів сезонності для фішингових атак, з урахуванням їх появи протягом 2020–2023 рр.

Наведені розрахунки не враховують впливу окремих факторів на кількість спостережених атак, а саме методи їхньої реалізації.

За даними [12] найчастіше фішингові атаки реалізуються з використанням електронної пошти, при цьому їх кількість складає приблизно 8.96% від загальної кількості атак. Серед інших технологій реалізації атак, крім тих що реалізуються за допомогою засобів електронної пошти, 3% у атак, які відбуваються через веб-сайти, і 1% атак що реалізуються з використанням телефонного зв'язку (голосове повідомлення або SMS).

Для визначення можливої залежності виконаємо кореляційно-регресійний аналіз впливу атак, реалізованих за допомогою засобів електронної пошти X, на загальну кількість атак Y, які були виявлені на протязі певних років.

Виходячи з інформації [12] розраховані значення відповідної кількості атак X, Y, результати розрахунків (табл. 5) далі використовуємо для обчислення значення коефіцієнта кореляції [18].

Таблиця 5

Відомості для розрахунку значення коефіцієнту кореляції

Рік	Розрахункова кількість фішингових атак, що реалізовані засобами електронної пошти (одиниць)	Загальна кількість спостережених атак (одиниць) [12]
	X	*Y*
2020	229145	$3 \cdot 10^6$
2021	299148	$6 \cdot 10^6$
2022	450904	$5 \cdot 10^6$
2023	620329	$7 \cdot 10^6$

Виходячи з даних табл 5 за допомогою програмного забезпечення MS Office Excel визначено значення коефіцієнта кореляції, яке дорівнює:

$$Kr = 0.765996 \approx 0.76$$

Значення коефіцієнту кореляції більше 0,76 свідчить про те, що існує достатньо висока ступень прямого лінійного взаємозв'язку між кількістю атак, з використанням засобів електронної пошти та загальним об'ємом спостережених атак протягом певного року.

Для визначення випадкового розподілу ймовірностей появи фішингових атак, проведено регресійний аналіз даних. За результатами зазначеного аналізу може бути наданий прогноз щодо подальшого виявлення атак такого типу, але результат не може бути передбаченим точно. Результати регресійного аналізу приведені у табл. 6.

Таблиця 6

Результати кореляційно — регресійного аналізу для фішингових атак

Регресійна статистика	
Лінійний коефіцієнт кореляції Kr	0,76
R^2	0,58
Нормований R^2	0,38
Стандартна помилка	134460688.65
Спостереження	4
Коефіцієнти	
Y-перетинання	223809630.39
Змінна *X1*	753.19

Наведені в табл. 6 коефіцієнти регресії показують, що в цьому випадку рівняння регресії має наступний вигляд:

$$Y = 223809630.39 + 753.19 \cdot X_1. \quad (5)$$

Лінійний коефіцієнт кореляції (*Kr*) свідчить про те, що між кількістю фішингових атак, що реалізовані за допомогою електронної пошти, та загальною кількістю атак існує зв'язок, а коефіцієнт детермінації (*R*-квадрат = 0,58) показує, що варіація між значенням *Y* зумовлюється варіацією *X* на 58%. Це пояснює, що параметри моделі на 58% залежать між собою. Чим вища така залежність, тим буде краще прогноз.

Взаємозв'язок між *X* та *Y* представимо у вигляді лінійного графіка, на який додаємо лінію тренду та зробимо позначення величини апроксимації (R^2).

ISSN 2663 - 4023

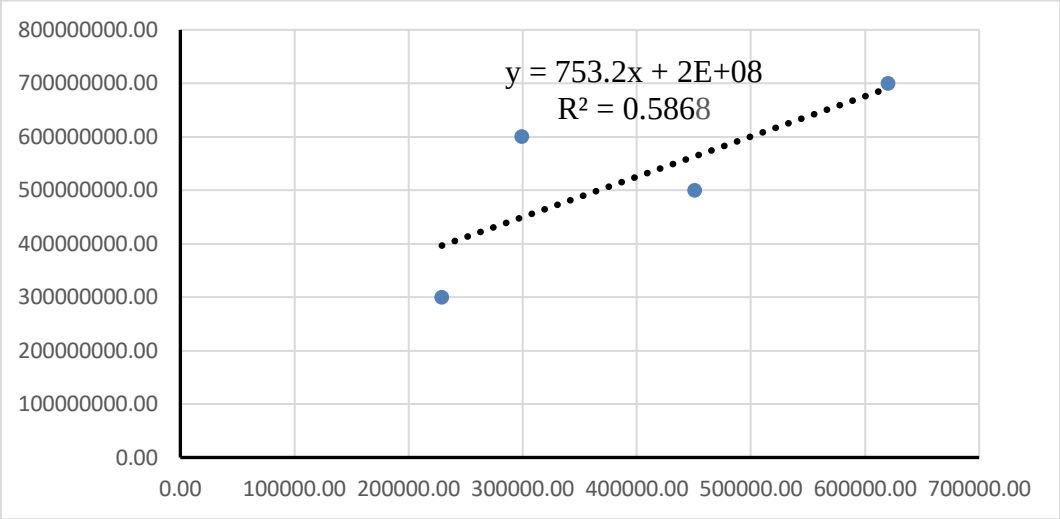

Рис. 3. Графік взаємозв'язку між показниками X та Y

Для покращення прогнозу, визначимо оцінку сумарного впливу різних видів фішингових атак на загальний об'єм виявлених атак за певними роками.

Комплексну взаємодію всіх факторів $(X_1, X_2, \dots, X_n)$ з результативним показником (Y) можна описати рівнянням лінійної багатофакторної регресії виду:

$$Y = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + \cdots + a_n \cdot x_n. \quad (6)$$

Відповідно до [12], [16] визначимо наступні види фішингових атак $(n = 3)$ для аналізу їх вплив на загальну кількість спостережених атак за певними роками:

- кількість фішингових атак, що реалізовані за допомогою електронної пошти (X_1);
- кількість фішингових атак, реалізованих через веб-сайт (X_2);
- кількість фішингових атак, що реалізовані з використанням телефону (X_3).

Зазначені показники визначим за період з 2020 по 2023 рік (табл. 6).

Таблиця 7

Дані для проведення кореляційно — регресійного аналізу впливу різних видів фішингових атак на загальні об'єми атак за роками

Рік	Загальна кількість атак (одиниць)	Кількість фішингових атак,		
		що реалізовані за допомогою e-mail (одиниць)	що реалізовані через веб-сайт (одиниць)	що реалізовані з телефоном (одиниць)
	Y	***X1***	***X2***	***X3***
2020	$3 \cdot 10^6$	229144	76723	25574
2021	$6 \cdot 10^6$	299148	100161	33387
2022	$5 \cdot 10^6$	450904	150972	50324
2023	$7 \cdot 10^6$	620329	207699	69233

На підставі розрахованих даних (табл. 7) за допомогою MS Office Excel отримуємо рівняння регресії та показники кореляційно — регресійного аналізу (табл. 7) в табл. 8:

$$Y = 4664545{,}8 + 921133{,}5 \cdot X_1 - 2751115{,}7 \cdot X_2 + 0 \cdot X_3, \quad (7)$$

Обчислені показники кореляційно — регресійного аналізу наведені в табл. 8:

Таблиця 8

Результати кореляційно — регресійного аналізу впливу різних типів фішингових атак на загальні об'єми атак за роками

Регресійна статистика	
Коефіцієнт кореляції R	0,90
R^2	0,81
Нормований R^2	-0,54
Стандартна помилка	1265378,39
Спостереження	4
Y-перетинання	4664545,8
Змінна *X1*	921133,5
Змінна *X*2	-2751115,7
Змінна *X*3	0

За результатами розрахунків виявлено, що R^2 — коефіцієнт детермінації на 81% показує суттєву залежність між параметрами, що приведені. Коефіцієнт 4664545,8 свідчить про те, яким буде *Y*, якщо змінні в будуть приймати значення 0. Коефіцієнти 921133,5, -2751115,7 показують вагомість змінної *X* на *Y*.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Таким чином на підставі публікацій, присвячених питанням виявлення та розповсюдження різного виду комп'ютерних атак, здійснені дослідні роботи щодо прогнозування атак, які відносяться до фішингових.

На основі аналізу часових рядів побудовані моделі тренду для кількості спостережених фішингових атак за період 2020–2023 рр та розрахунки прогнозу кількості таких атак за 16 кварталів 2020–2023 років, а також ймовірний прогноз появи зазначених атак за чотири квартал 2024 року.

За допомогою значення тренду розраховано коефіцієнт сезонності та виконаний кореляційно-регресійний аналіз впливу фішингових атак на загальну кількість атак, які протягом певних років. Проведені розрахунки свідчать про те, що розбіжності прогнозованих значень не значні.

Перспективи щодо подальшого дослідження проблеми прогнозування фішингових атак [6] представляються в застосуванні інших методів прогнозування та результатів їх аналізу, наприклад, таких як метод експертних оцінок, індуктивного прогнозування тощо.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Соколов, В., & Складанний, П. (2023). Методологія оцінки комплексних збитків від інцидентів інформаційної безпеки. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка», 1(21),* 99–120. https://doi.org/10.28925/2663-4023.2023.21.99120
2. Киричок, Р., Бжевска, З., Гулак, Г., Бессалов, А., & Астапеня, В. (2021). Правила реалізації експлойтів під час активного аналізу безпеки корпоративних мереж на основі нечіткої оцінки якості механізму валідації вразливості. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка», 2(14),* 148–157. https://doi.org/10.28925/2663-4023.2021.14.148157

3. Шевченко, С., Жданова, Ю., Складанний, П., & Бойко, С. (2022). Інсайдери та інсайдерська інформація: суть, загрози, діяльність та правова відповідальність. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка», 3(15),* 175–185. https://doi.org/10.28925/2663-4023.2022.15.175185
4. Романюк, О., Складанний, П., & Шевченко, С. (2022). Порівняльний аналіз рішень для забезпечення контролю та управління привілейованим доступом в IT-середовищі. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка», 4(16),* 98–112. https://doi.org/10.28925/2663-4023.2022.16.98112
5. Гулак, Г., Жданова, Ю., Складанний, П., Гулак, Є., & Корнієць, В. (2022) Уразливості шифрування коротких повідомлень в мобільних інформаційно-комунікаційних системах об'єктів критичної інфраструктури. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка», 1(17),* 145–158. https://doi.org/10.28925/2663-4023.2022.17.145158
6. *Війна в Україні: пульс кіберзахисту*. (2023). https://drive.google.com/drive/folders/1RjuBE_5Yznwnh1ELkppB94JCq3y17tTI7
7. Гребенник, А., Трунова, О., Казимир, В., & Мищенко, М. (2020). Виявлення та прогнозування рівня загроз для корпоративної комп'ютерної мережі. *Технічні науки та технології, 2(20),* 175–184.
8. Sokol, P., & Gajdos, A. (2017) Prediction of Attacks Against Honeynet Based on Time Series Modeling. *Applied Computational Intelligence and Mathematical Methods. CoMeSySo 2017, Advances in Intelligent Systems and Computing, 662.* https://doi.org/10.1007/978-3-319-67621-0_33
9. Werner, G., Yang, S., & McConky, K. (2017). Time series forecasting of cyber attack intensity. *Proceedings of the 12th Annual Conference on Cyber and Information Security Research, 18,* 1–3. https://doi.org/10.1145/3064814.3064831
10. Tang, M., Alazab, M., & Luo, Y. (2016). Exploiting vulnerability disclosures: statistical framework and case study. *Cybersecurity and Cyberforensics Conference*. https://doi.org/10.1109/CCC.2016.10
11. Hus´ak M., et al. (2021). Predictive methods in cyber defense: Current experience and research challenges. *Future Generation Computer Systems, 115,* 517–530.
12. Hyndman, J., & Athanasopoulos, G. (2021). *Forecasting: principles and practice.* OTexts.
13. Долгіх, А., & Байбуз, О. (2017). Огляд сучасних розробок прогнозування часових рядів з використанням прихованих марківських моделей. *Актуальні проблеми автоматизації та інформаційних технологій, 21,* 60–73.
14. Лахно, В., Малюков, В., Оган, А., Малюкова, І., Криворучко, О., Десятко, А., & Катерина В. (2023). Модель аналізу стратегій при динамічній взаємодії учасників фішингових атак. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка», 4(20),* 124–138.
15. Дейнеко, Ж., & Диана, Д. (2015). Дослідження динаміки фішинговий атак методом вейвлет-аналізу. *Інформаційні системи і технології ICT-2018,* 396–397.
16. Basit, A., et al. (2021). A comprehensive survey of AI-Enabled Phishing Attacks Detection Techniques. *Telecommun. Syst. 76,* 139–154.
17. Мохор В., Цуркан О., Герасимов Р., Крук О., Покровська, В. (2020). Модель аналізування уразливостей соціотехнічних систем до впливів соціальної інженерії. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка», 4(8),* 165–171.
18. *Кібератаки 2022–2023: огляд найбільших інцидентів, та що нас чекає у 2024 році*. (2023). https://www.h-x.technology/ua/blog-ua/cyber-threats-forecast-2024-ua
19. *Держспецзв'язку: Статистика кібератак за чотири місяці війни. (2023).* https://www.kmu.gov.ua/news/derzhspeczvyazku-statistika-kiberatak-za-chotiri-misyaci-vijni
20. *Актуальні кіберзагрози: I квартал 2023 року.* (2023). Positive Technologies. https://www.ptsecurity.com/ru-ru/research/analytics/cybersecurity-threatscape-2023-q1
21. *Інформаційне агентство Interfax Ukraine.* (2023). https://interfax.com.ua/news/telecom/943392.html
22. *Фішинг та цільовий фішинг: поради по захисту*. (2019). TechRepublic. https://www.imena.ua/blog/phishing-and-target-phishing

23. Cohen, J., et al. (2013). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.* LEA Publishers.

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 3 (23), 2024
ISSN 2663 - 4023

Yurii Dobryshyn
PhD, Associate Professor,
National Academy of the Security Service of Ukraine, Kyiv, Ukraine
ORCID 0000-0003-2473-9507
ydobryshyn@gmail.com

STATISTICAL METHODS FOR PREDICTING PHISHING ATTACKS

Abstract. The article proposes a methodology for predicting so-called phishing attacks, which are a common form of cybercrime, the number of which is growing every year, and the level of their harmful impact on the information systems of critical infrastructure objects is also increasing. To analyze trends and predict phishing attacks, we used statistical data published in scientific works of domestic and foreign researchers, as well as published by online publications of leading consulting companies working in the field of information security and cybersecurity. Statistical methods based on the use of time series, as one of the popular approaches used to predict various technological and economic processes, were chosen as tools for researching and predicting phishing attacks. This made it possible to analyze the types and patterns of phishing attacks that attackers use to disrupt the operation of software of information and communication systems and automated systems. Based on time series analysis, a trend model was built for the number of detected phishing attacks for the period 2020–2023. A calculation was made of the predicted number of phishing attacks for 16 quarters of 2020–2023, as well as the estimated forecast of the occurrence of these attacks for four quarters of 2024. To improve the forecast, a coefficient taking into account the seasonality factor was calculated and a correlation and regression analysis of the impact of phishing attacks on the total number of attacks detected during 2020–2023 was performed. Calculations have been performed, indicating that the discrepancies in the predicted values are not significant; the results presented allow us to select the optimal strategy for identifying, predicting and eliminating computer attacks related to phishing. Based on the time series model and the calculations obtained, it was concluded that statistical forecasting methods make it possible to build a forecast of phishing attacks, provide in the future the opportunity to develop and formulate methods for countering these attacks, and plan measures to increase the level of security of information resources.

Keywords: forecasting; phishing attack; statistical methods; time series; automated information and communication system; trend; correlation-regression analysis.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Sokolov, V., & Skladannyi, P. (2023). Methodology for Assessing Comprehensive Damages from an Information Security Incident. *Electronic Professional Scientific Edition "Cybersecurity: Education, Science, Technique", 1(21),* 99–120. https://doi.org/10.28925/2663-4023.2023.21.99120
2. Kyrychok, R., et al. (2021). Rules For The Implementation Of Exploits During An Active Analysis Of The Corporate Networks` Security Based On A Fuzzy Assessment Of The Quality Of The Vulnerability Validation Mechanism. Electronic Professional Scientific Edition «Cybersecurity: Education, Science, Technique», 2(14), 148–157. https://doi.org/10.28925/2663-4023.2021.14.148157
3. Shevchenko, S., et al. (2022). Insiders and Insider Information: Essence, Threats, Activities and Legal Responsibility. *Electronic Professional Scientific Edition "Cybersecurity: Education, Science, Technique", 3(15),* 175–185. https://doi.org/10.28925/2663-4023.2022.15.175185
4. Romaniuk, O., Skladannyi, P., & Shevchenko, S. (2022). Comparative Analysis of Solutions to Provide Control and Management of Privileged Access in the it Environment. *Electronic Professional Scientific Edition "Cybersecurity: Education, Science, Technique", 4(16),* 98–112. https://doi.org/10.28925/2663-4023.2022.16.98112
5. Hulak, H., et al. (2022). Vulnerabilities of Short Message Encryption in Mobile Information and Communication Systems of Critical Infrastructure Objects. *Electronic Professional Scientific Edition*

"Cybersecurity: Education, Science, Technique", *1(17)*, 145–158. https://doi.org/10.28925/2663-4023.2022.17.145158

6. *War in Ukraine: the pulse of cyber defense.* (2023). https://drive.google.com/drive/folders/1RjuBE_5Yznwnh1ELkppB94JCq3y17tTI7
7. Hrebennik, A., et al. (2020). Vyiavlennia ta Prognozuvannia Rivnia Zagroz Dlia Korporatyvnoi Kompiuternoi Merezhi. *Technichni Nauky ta Technologii, 2(20),* 175–184.
8. Sokol, P., & Gajdos, A. (2017) Prediction of Attacks Against Honeynet Based on Time Series Modeling. *Applied Computational Intelligence and Mathematical Methods. CoMeSySo 2017, Advances in Intelligent Systems and Computing, 662.* https://doi.org/10.1007/978-3-319-67621-0_33
9. Werner, G., Yang, S., & McConky, K. (2017). Time series forecasting of cyber attack intensity. *Proceedings of the 12th Annual Conference on Cyber and Information Security Research, 18,* 1–3. https://doi.org/10.1145/3064814.3064831
10. Tang, M., Alazab, M., & Luo, Y. (2016). Exploiting vulnerability disclosures: statistical framework and case study. *Cybersecurity and Cyberforensics Conference*. https://doi.org/10.1109/CCC.2016.10
11. Hus´ak, M., et al. (2021). Predictive Methods in Cyber Defense: Current Experience and Research Challenges. *Future Generation Computer Systems, 115,* 517–530.
12. Hyndman, J., & Athanasopoulos, G. (2021). *Forecasting: principles and practice*. OTexts.
13. Dolgikh, A., Baybuz, O. (2017). Overview of Modern Developments in Time Series Forecasting Using Hidden Markov Models. *Actual Problems of Automation and Information Technologies, 21,* 60–73.
14. Lakhno V., at al. (2023) Model of Strategy Analysis During the Dynamic Interaction of Phishing Attack Participants. *Electronic Professional Scientific Edition "Cybersecurity: Education, Science, Technique", 4(20),* 124–138. https://doi.org/10.28925/2663-4023.2023.20.124141
15. Deineko. Zh., & Diana, D. (2015). Study of the Dynamics of Phishing Attacks Using Wavelet Analysis. *Information Systems and Technologies IST-2018,* 396–397.
16. Basit, A., et al. (2021). A comprehensive survey of AI-enabled phishing attacks detection techniques. *Telecommun. Syst. 76,* 139–154.
17. Mokhor, V., Tsurkan, O., Herasymov, R., Kruk, O., & Pokrovska, V. (2020). A Model for Analyzing the Vulnerability of Sociotechnical Systems to the Influences of Social Engineering. *Electronic Professional Scientific Edition "Cybersecurity: Education, Science, Technique", 4(8),* 165–173. https://doi.org/10.28925/2663-4023.2020.8.165173
18. *Cyber attacks 2022-2023: an overview of the biggest incidents and what awaits us in 2024.* (2023). H-X. https://www.h-x.technology/ua/blog-ua/cyber-threats-forecast-2024-ua
19. *State Intelligence Service: Statistics of cyber attacks for four months of the war*. (2023). https://www.kmu.gov.ua/news/derzhspeczvyazku-statistika-kiberatak-za-chotiri-misyaci-vijni
20. *Actual cyber threats: I quarter of 2023.* (2023). Positive Technologies. https://www.ptsecurity.com/ru-ru/research/analytics/cybersecurity-threatscape-2023-q1
21. *Information agency Interfax Ukraine.* (2023). https://interfax.com.ua/news/telecom/943392.html
22. *Phishing and targeted phishing: protection tips.* (2019). TechRepublic. https://www.imena.ua/blog/phishing-and-target-phishing
23. Cohen, J., et al. (2013). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. LEA Publishers.