Міністерство освіти і науки України
Національний технічний університет
«Дніпровська політехніка»

Інститут електроенергетики
Факультет інформаційних технологій
Кафедра безпеки інформації та телекомунікацій

ПОЯСНЮВАЛЬНА ЗАПИСКА
кваліфікаційної роботи ступеня магістра

студента *Ангеловського Миколи Олексійовича*

академічної групи *125м-21-1*

спеціальності *125 Кібербезпека*

спеціалізації[1]

за освітньо-професійною програмою *Кібербезпека*

на тему *Детектування фішингових веб-сайтів за допомогою штучних нейронних мереж*

Керівники	Прізвище, ініціали	Оцінка за шкалою		Підпис
		рейтинговою	інституційною	
кваліфікаційної роботи	к.т.н., доц. Герасіна О.В.			
розділів:				
спеціальний	к.т.н., доц. Герасіна О.В.			
економічний				
Рецензент				
Нормоконтролер	ст. викл. Мєшков В.І.			

Дніпро
2022

ЗАТВЕРДЖЕНО:
завідувач кафедри
безпеки інформації та телекомунікацій
_____ д.т.н., проф. Корнієнко В.І.

«_____»_____20___року

ЗАВДАННЯ
на кваліфікаційну роботу
ступеня магістра

студенту *Ангеловському Миколі Олексійовичу* **академічної групи** *125м-21-1*
(прізвище ім'я по-батькові) (шифр)

спеціальності *125 Кібербезпека*

за освітньо-професійною програмою *Кібербезпека*

на тему *Детектування фішингових веб-сайтів за допомогою штучних нейронних мереж*

затверджену наказом ректора НТУ «Дніпровська політехніка» від _____ № ____

Розділ	Зміст	Термін виконання
Розділ 1	Аналіз принципів фішингових атак і існуючих методів захисту від них, основ побудови штучних нейронних мереж, постановка задачі класифікації.	03.09.2022 – 10.10.2022
Розділ 2	Дослідження використання нейронних мереж для виявлення фішингових веб-сайтів, розробка підходу до детектування фішингових URL-адрес із використанням нейромережевих класифікаторів і методу прямого випадкового пошуку та оцінка його ефективності.	11.10.2022 – 24.11.2022
Розділ 3	Розрахунки капітальних витрат, економічного ефекту та терміну окупності капітальних інвестицій застосування запропонованих рішень.	25.11.2022 – 04.12.2022

Завдання видано _____ Герасіна О.В.
(підпис керівника) (прізвище, ініціали)

Дата видачі: _____

Дата подання до екзаменаційної комісії: _____

Прийнято до виконання _____ Ангеловський М.О.
(підпис студента) (прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка: 92 с., 21 рис., 6 табл., 4 додатки, 52 джерела.

Об’єкт дослідження – фішингові та безпечні URL-адреси.

Предмет дослідження – підхід до детектування фішингових URL-адрес із використанням штучних нейронних мереж.

Мета кваліфікаційної роботи – дослідження та обґрунтування типу та параметрів нейромережевих класифікаторів для детектування фішингових URL-адрес.

Наукова новизна результатів полягає у використанні методу прямого випадкового пошуку для вибору типу та параметрів нейромережевого класифікатора для детектування фішингових веб-сайтів.

У першому розділі проаналізовано принципи фішингових атак і існуючих методів захисту від них, основи побудови штучних нейронних мереж, а також сформульовано задачу класифікації.

У спеціальній частині роботи досліджено використання нейронних мереж для виявлення фішингових веб-сайтів, запропоновано підхід до детектування фішингових URL-адрес із використанням нейромережевих класифікаторів і методу прямого випадкового пошуку та оцінено його ефективність. За наслідками досліджень зроблено висновки щодо рішення поставленої задачі.

У економічному розділі виконані розрахунки капітальних витрат, економічного ефекту та терміну окупності капітальних інвестицій застосування запропонованих рішень.

КОНФІДЕНЦІЙНІ ДАНІ, ШТУЧНІ НЕЙРОННІ МЕРЕЖІ, ФІШИНГОВІ АТАКИ, МЕТОД ПРЯМОГО ВИПАДКОВОГО ПОШУКУ, МЕТРИКИ ОЦІНКИ, КЛАСИФІКАЦІЯ, ЗАХИЩЕНЕ З'ЄДНАННЯ З ДІЙСНИМ СЕРТИФІКАТОМ, ІМІТАЦІЙНЕ МОДЕЛЮВАННЯ

ABSTRACT

Explanatory note: p. 92, fig. 21, tab. 6, 4 additions, 52 sources.

The object of the study is phishing and safe URLs.

The subject of the study is an approach to detecting phishing URLs using artificial neural networks.

The purpose of the qualification work is research and justification of the type and parameters of neural network classifiers for detecting phishing URLs.

The scientific novelty of the results is the use of the direct random search method to select the type and parameters of the neural network classifier for detecting phishing websites.

The first chapter analyzes the principles of phishing attacks and existing methods of protection against them, the basics of building artificial neural networks, and also formulates the problem of classification.

In a special part of the work, the use of neural networks to detect phishing websites was investigated, an approach to detecting phishing URLs using neural network classifiers and the direct random search method was proposed and its effectiveness was evaluated. Based on the results of the research, conclusions were made regarding the solution to the problem.

In the economic section, calculations of capital costs, economic effect and payback period of capital investments are performed using the proposed solutions.

CONFIDENTIAL DATA, ARTIFICIAL NEURAL NETWORKS, PHISHING ATTACKS, DIRECT RANDOM SEARCH METHOD, EVALUATION METRICS, CLASSIFICATION, SECURE CONNECTION WITH VALID CERTIFICATE, SIMULATION MODELING

СПИСОК УМОВНИХ СКОРОЧЕНЬ

ІТС – Інформаційно-телекомунікаційна система;

ІС – Інформаційна система;

НМ – Нейронна мережа;

ПЗ – Програмне забезпечення;

ШІ – Штучний інтелект;

DDoS attack – Distributed Denial of Service attack – Розподілена атака на відмову в обслуговуванні;

DoS attack – Denial of Service attack – Атака на відмову в обслуговуванні.

ЗМІСТ

с.

ВСТУП ... 8
1 СТАН ПИТАННЯ. ПОСТАНОВКА ЗАДАЧІ ... 11
1.1 Методи протидії використанню соціальної інженерії ... 11
1.1.1. Сутність соціоінженерного підходу ... 11
1.1.2. Форми маніпулювання персоналом ... 13
1.1.3. Методи соціальної інженерії ... 16
1.2 Фішинг ... 20
1.2.1 Поняття фішингу ... 20
1.2.2. Ретроспектива вірусної активності в Україні за 2021 рік ... 22
1.2.3. Ретроспектива вірусної активності в Україні за 2022 рік ... 23
1.3 Класифікація ... 32
1.3.1 Постановка задачі класифікації ... 32
1.3.2 Оцінка якості класифікації ... 34
1.4 Нейронні мережі ... 43
1.4.1 Перцептрони ... 45
1.4.2 Нейронні мережі Хопфілда ... 48
1.4.3 Нейронні мережі Хемінга ... 52
1.5 Висновок. Постановка задачі ... 54
2 СПЕЦІАЛЬНА ЧАСТИНА ... 58
2.1 Дослідження використання штучних нейронних мереж для виявлення фішингових веб-сайтів ... 58
2.2 Підхід до детектування фішингових URL-адрес із використанням нейромережевих класифікаторів та методу прямого випадкового пошуку ... 63
2.3 Оцінка ефективності підходу до детектування фішингових URL-адрес із використанням нейромережевих класифікаторів і методу прямого випадкового пошуку ... 67
2.4 Висновок ... 69

3 ЕКОНОМІЧНИЙ РОЗДІЛ 71
3.1 Розрахунок капітальних (фіксованих) витрат 71
3.2 Розрахунок поточних витрат 74
3.3 Оцінка можливого збитку 76
3.3.1 Загальний ефект від впровадження системи інформаційної безпеки 78
3.4 Визначення та аналіз показників економічної ефективності системи інформаційної безпеки 79
3.5 Висновок 80
ВИСНОВКИ 81
ПЕРЕЛІК ПОСИЛАНЬ 83
ДОДАТОК А. Відомість матеріалів кваліфікаційної роботи 89
ДОДАТОК Б. Перелік документів на оптичному носії 90
ДОДАТОК В. Відгук керівника економічного розділу 91
ДОДАТОК Г. Відгук керівника кваліфікаційної роботи 92

ВСТУП

З розвитком інтернет-технологій і розширенням інтернет-простору збільшується ймовірність зіткнення з шахрайством. Останнього часу популярність набув фішинг − різновид шахрайської діяльності, метою якої є отримання доступу до конфіденціальної інформації користувача (логінів, паролів).

Фішинг (від англ. «phishing» або «fishing» − риболовля) − це злочинне діяння в Інтернеті, яке відбувається, коли зловмисна веб-сторінка видає себе за законну веб-сторінку з метою отримання конфіденційної інформації від користувача [1-19].

З початку повномасштабного російського вторгнення українці все частіше стають цілями кіберзлочинців. Серед нещодавно виявлених загроз — шкідлива програма групи Armageddon, яка завантажується через відкриття вкладень в електронних листах нібито від Національної академії СБУ.

Крім того, в Інтернеті регулярно з'являються нові схеми онлайн-шахрайства, зокрема цього разу зловмисники поширювали повідомлення на тему «допомоги від Червоного Хреста» з фішинговими посиланнями у популярних месенджерах. Щоб викрасти дані для входу в онлайн-банкінг, шахраї маскувалися під представників українських банків.

Зазвичай кіберзлочинці маскуються під легітимні та надійні джерела, зокрема представників банків, технологічних компаній та роздрібних торговців. Ціль зловмисників — переконати жертв натиснути на посилання або відкрити вкладення в пошті. Після цього користувачі потрапляють на підроблений сайт, який просить ввести певну особисту інформацію, наприклад, облікові або фінансові дані, або приховано завантажує загрозу на пристрій.

Взагалі фішингові атаки часто є дуже масштабними подіями, які спрямовані на тисячі споживачів або більше, в надії, що певний відсоток з них захоче відповісти. Відносно великий відсоток одержувачів відповідають на електронні листи, оскільки вони видаються законними, і їх автентичність

неможливо легко перевірити. Оцінки рівня відповіді коливається від 1 до 20%, в залежності від атаки [12].

Оскільки між зловмисником та споживачем не існує особистого контакту, споживач має дуже мало інформації для роботи, щоб вирішити, чи є електронний лист або веб-сайт законним.

Таким чином, кількість та різноманітність фішингових сайтів дуже сильно впливають на безпеку в Інтернеті. Згідно зі звітами APWG (Anti-Phishing Work Group) останніми роками в Інтернеті з'являються все більше і більше шахрайських сторінок [13]. Разом із зростанням обсягу фішингу, зростає і кількість постраждалих від дій зловмисників. Слід також зазначити, що системи захисту від фішингових сайтів працюють з відносно невисокою швидкістю, тому користувачі найчастіше не попереджені про те, що сайт є підозрілим.

Деякі браузери використовують систему «блек-листу», яка передбачає пошук сайту, на який переходить користувач, серед переліку фішингових сайтів, який сформовано або розробником браузера, або однією з відомих асоціацій, які борються з фішингом в мережі Інтернет. Таку систему раніше використовували у таких відомих браузерах, як «Google Chrome», «Mozilla Firefox».

Наразі великі компанії користуються так званою системою «репортів». Зазначена система передбачає, що користувач, який потрапив на підозрілий сайт, відправляє скаргу на цей сайт і тоді команда спеціалістів перевіряє сторінку [14-16]. Якщо сайт виявився фішинговим, то він попадає до переліку підозрілих. Тепер, перед тим як зайти на такий сайт, браузер попередить користувача про те, що сайт є фішинговим. В комплексі з системою «репортів» використовують різні методи автоматичного виявлення фішингових сайтів, які базуються на інтелектуальних методах. Такі методи шукають фішингові сайти, а потім відправляють репорти, які будуть перевіряти спеціалісти.

Взагалі, задача детектування фішингових URL-адрес є задачею класифікації, яка може бути вирішена за допомогою методів систем штучного

інтелекту (ШІ): нейронних мереж (НМ) та систем з нечіткою логікою, які є універсальними ефективними апроксиматорами [20-32].

Таким чином, вдосконалення підходів до детектування фішингових URL-адрес із використанням методів систем штучного інтелекту наразі є актуальною задачею.

Метою роботи є дослідження та обґрунтування типу та параметрів нейромережевих класифікаторів для детектування фішингових URL-адрес.

Постановка задачі:

• проаналізувати принципи фішингових атак та існуючих методів захисту від них;

• сформулювати задачу класифікації та провести аналіз основи побудови штучних нейронних мереж;

• дослідити використання штучних нейронних мереж для виявлення фішингових веб-сайтів;

• запропонувати підхід до детектування фішингових URL-адрес із використанням нейромережевих класифікаторів та методу прямого випадкового пошуку;

• оцінити ефективність запропонованого підходу.

1 СТАН ПИТАННЯ. ПОСТАНОВКА ЗАДАЧІ

1.1 Методи протидії використанню соціальної інженерії

Інформаційні ресурси окремих організацій і фізичних осіб являють собою певну цінність, мають відповідне матеріальне вираження і вимагають захисту від різноманітних за своєю сутністю несприятливих впливів. Потенційно можливий несприятливий вплив тлумачиться загрозою. Тому захист інформації в інформаційно-телекомунікаційних системах (ІТС) полягає в створенні та підтриманні в дієздатному стані системи заходів для запобігання або ускладнення можливості реалізацій загроз, а також зменшення потенційних збитків.

Оскільки ІТС включає обчислювальну систему, фізичне середовище, персонал і оброблювану інформацію, то на оцінювання технічної захищеності інформації суттєво впливає врахування нетехнічного аспекту, зокрема, персоналу (наприклад, див. рис. 1.1, керівництва, адміністратора операційної системи, користувачів). Тому для цього в деяких джерелах пропонується соціоінженерний підхід [34-38].

1.1.1. Сутність соціоінженерного підходу

У рамках соціоінженерного підходу вразливості персоналу тлумачаться як його слабкості, потреби, манії (пристрасті), захоплення. Маніпулювання ними дозволяє отримати несанкціонований доступ до інформації без руйнування та перекручування головних для нього системоутворюючих якостей (цілісність, розвиток) [34].

Як наслідок, вищезазначене призводить до нової моделі поведінки персоналу, створення сприятливих умов реалізації загроз безпеці інформації і, як наслідок, зменшенню здатності системи захисту інформації протидіяти їх впливові (рис. 1.2).

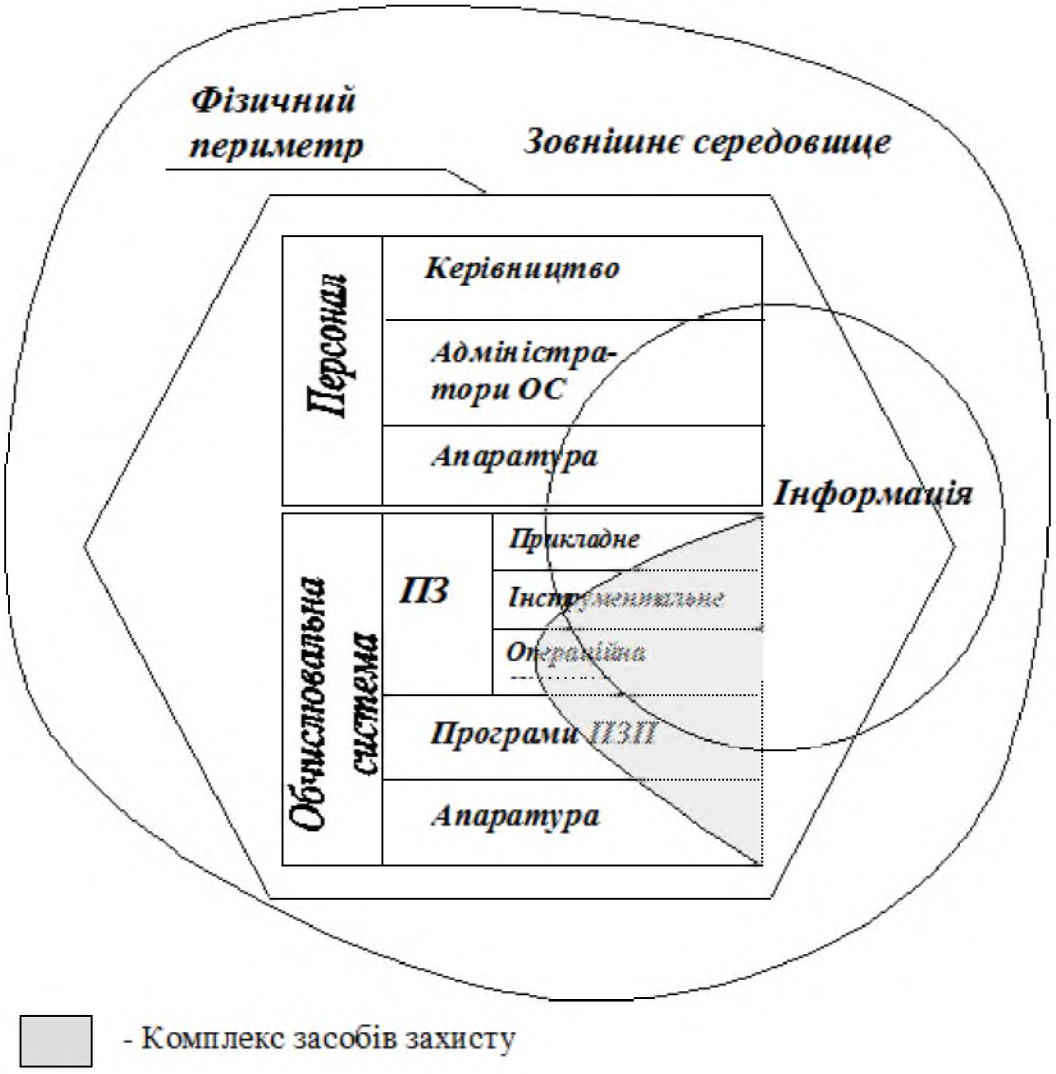


Рисунок 1.1 – Елементи ІТС

Це відображається в таких формах як, наприклад [34, 39-40], шахрайство, обман, афера, інтрига, містифікація, провокація. Використанню кожної з означених форм маніпулювання передує визначення її змісту шляхом ретельного планування, організування та контролювання

З огляду на рис. 1.2, використання соціоінженерного підходу до оцінювання захищеності інформації в ІТС передбачає цілеспрямований вплив на свідомість (підсвідомість) персоналу проти волі, але за його згодою. Такий вплив дозволяє управляти поведінкою керівництва, адміністратора, користувачів через слабкості, інтереси, потреби, схильності, переконання,

звички, психічний та емоційний стан. Тому маніпулювання цими уразливостями і виражається в таких формах як шахрайство, обман, афера, інтрига, містифікація, провокація. Разом з тим, використанню кожної з означених форм маніпулювання передує визначення їх сутності шляхом ретельних планування, організації та контролювання.

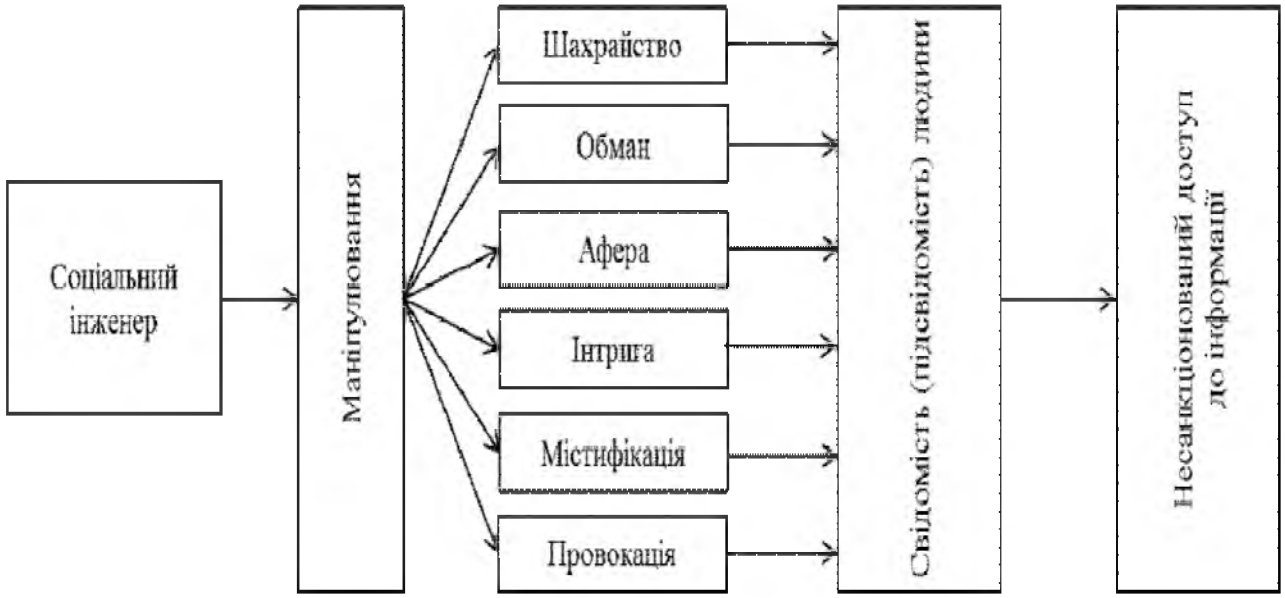


Рисунок 1.2 – Використання соціоінженерного підходу

У рамках соціоінженерного підходу використання атак соціальної інженерії орієнтоване на отримання «несанкціонованого» доступу до інформації при оцінюванні її захищеності шляхом «негативного» інформаційно-психологічного впливу на свідомість або підсвідомість персоналу (рис. 1.3 [41]).

1.1.2. Форми маніпулювання персоналом

Форми маніпулювання персоналом при оцінюванні захищеності інформації в інформаційно-комунікаційних системах змінюються залежно від різновиду атак соціальної інженерії, а саме [42-45]:

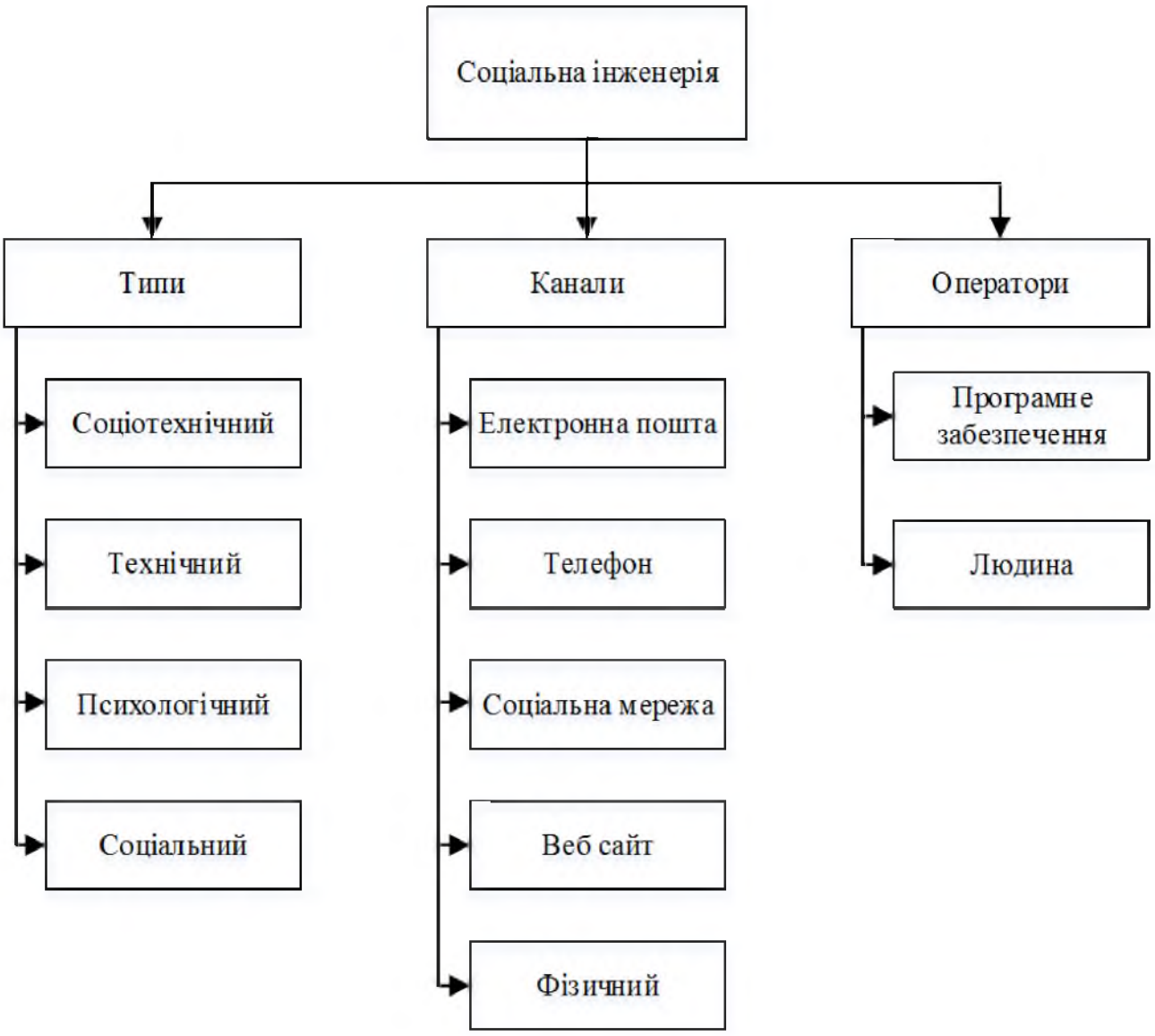


Рисунок 1.3 – Класифікування ознак реалізування атак соціальної інженерії

1. Фішінг (Phishing) – масове розсилання електронної пошти великій групі адресатів. Ознайомлення з електронними листами спонукає їх до, наприклад, відкриття вкладення до листа, переходу за посиланням на веб-сторінку. Його метою є виманювання у довірливого або неуважного персоналу комп'ютерної системи персональних даних.

2. Фармінг (Pharming) – перенаправлення користувачів на шахрайські сайти для отримання їх логіну та паролю. Це досягається завдяки

розповсюдженню електронної пошти серед користувачів, наприклад, соціальних мереж, онлайн-банкінгу, поштових веб-сервісів.

3. Прітекстінг (Pretexting) – отримання інформації або спонукання до вчинення певних дій обманом на основі заздалегідь складеного сценарію або створення фіктивної ситуації. Застосовується через телефон та потребує проведення попередніх досліджень для входження в довіру.

4. Смішінг (Smishing) – отримання інформації шляхом масового розсилання SMS повідомлень з посиланням на веб-ресурси або з реквізитами організацій (наприклад, фінансових). Внаслідок цього здійснюються відповідні дії, наприклад, дзвінок до банку для перевірки стану рахунку з зазначенням конфіденційних даних: номеру картки, терміну дії.

5. Вішінг (Vishing) – отримання інформації шляхом входження в довіру під час розмови через ір-телефон. При цьому в порушення конфіденційності здійснюється завдяки викладенню прохання у повідомленні зателефонувати на певний міський номер. Наприклад, вести номер карти, паролі, PIN-коди, коди доступу або іншу інформацію.

6. Спір фішінг (Spear Phishing) – надсилання листа електронної пошти конкретному адресату (наприклад, керівнику, адміністраторові, користувачеві), що спонукає його до обов'язкового перегляду та відповіді на отриманий лист.

7. Вейлінг (Whaling) – надсилання листа електронної пошти представнику керівництва організації, що спонукає його до обов'язкового перегляду та відповіді на отриманий лист.

Слід зазначити, що отримання несанкціонованого доступу до інформації за допомогою фішингу, фармінгу, смішінгу, вішінгу, спір фішінгу, вейлінгу здійснюється шляхом використання таких форм маніпулятивного впливу як шахрайство та обман (див. табл. 1.1). Тоді як основою для створення фіктивних ситуацій при прітекстінгу є афера, інтрига, містифікація та провокація.

Тому при оцінюванні захищеності інформації в ІТС за соціоінженерним підходом доцільно враховувати форми маніпулятивного впливу.

Таблиця 1.1 – Форми маніпулятивного впливу при соціоінженерних атаках

№ п/п	Різновид соціоінженерної атаки	Форми маніпулятивного впливу					
		Шахрайство	Обман	Афера	Інтрига	Містифікація	Провокація
1.	Фішінг	+	+	–	–	–	–
2.	Фармінг	+	+	–	–	–	–
3.	Прітекстінг	+	+	+	+	+	+
4.	Смішінг	+	+	–	–	–	–
5.	Вішінг	+	+	–	–	–	–
6.	Спір фішінг	+	+	–	–	–	–
7.	Вейлінг	+	+	–	–	–	–

1.1.3. Методи соціальної інженерії

Оцінювання захищеності інформації орієнтоване на отримання відомостей про ІТС – етап соціальної інженерії. Цей етап включено в аудит отримання і аналізування інформації зі зовнішнього середовища. Враховуючи високу ймовірність впливу людського фактору на захищеність інформації, на даному етапі вдало використовуються методи соціальної інженерії. Для успішного виконання запланованих дій соціальний інженер проводить роботу узгоджено з мережевим адміністратором. У його конструктивні дії входять: вивчення і аналізування змістовного боку комп'ютерної системи, пошук уразливостей, систематизування отриманих відомостей, розроблення схеми дій.

Використання методів соціальної інженерії для імітування дій порушника, що направлені на користувачів комп'ютерних систем організації, дозволяє оцінити рівень кваліфікації користувачів в області забезпечення безпеки інформації і ймовірність реалізування атак соціальної інженерії.

Вхідною інформацією для соціального інженера при спробі отримання інформації про ІТС може бути контактна інформація, що отримується з публічних джерел. Наприклад: прізвища, імена, посади користувачів. За отриманою вхідною інформацію вибираються, виокремлюються вірогідні уразливості в комп'ютерних системах, через які можливе реалізування загроз соціальної інженерії.

Основою використання методів соціальної інженерії є [34]:

• особливості, що керують людською свідомістю;

• аудиторія або поле діяльності;

• некомпетентність аудиторії у визначених термінах і предметних областях у сфері інформаційної безпеки;

• нестійкість психологічних властивостей особистості, що характеризуються поведінковими стереотипами; їх можна використовувати для маніпулювання через основні потреби, слабкості, бажання, ідеали.

Більшість соціальних інженерів діє за ідентичними або близькими шаблонами. Тому вивчення прийомів їх «роботи» дозволяє виокремити такі рівні взаємодії з об'єктом впливу як домінування, маніпулювання, суперництво, партнерство.

Всі методи соціальної інженерії можна поділити на дві групи [34, 42-45]:

1. Віддалена соціальна інженерія реалізується засобами сучасних телекомунікацій шляхом використання: телефону та глобальної мережі Інтернет.

Завдяки телефонії, соціальний інженер може залишатися анонімним і в той же час мати прямий зв'язок з об'єктом впливу. Останнє важливо тому, що безпосередній контакт не дає співрозмовнику часу обміркувати поведінку у вірогідних ситуаціях, зважати на всі за та проти. Вирішувати необхідно швидко, до того ж під тиском соціального інженера. Оскільки під час телефонної розмови відбувається обмін тільки звуковою інформацією, то велику роль у прийнятті рішень відіграє анотація і голос співрозмовника.

Дані характеристики підбираються у відповідності з моделлю поведінки соціального інженера для отримання інформації про об'єкт впливу, наприклад:

• начальник – людина, яка звикла віддавати команди, цінує свій час, досягає поставленої мети. Манера розмови жорстка, нетерпляча. Повна впевненість у собі і легка (або повна) зверхність до рядового персоналу. Своїм тоном показує, що проблема, з якою звернувся – дрібниця, яку необхідно вирішити якомога швидше. Ніяких прохань – тільки вимоги і вказівки. У відповідь на недовірливі або перевіряючі репліки – допустиме незадоволення і залякування співрозмовника;

• секретар – дівчина (здебільшого) з приємним голосом. Завдання – виконати конкретне доручення начальника, не відволікаючись на умовності. Вона володіє інформацією про начальника, його справи, у своїй мові користується достовірними або недостовірними фактами, які складно перевірити. Характер розмови – м'який, з легким фліртуванням (якщо співрозмовник – чоловік). Реакція на небажання співпрацювати – бурхливе розчарування, скарга, що скаже начальство;

• технічний співробітник – працівник організації, який характеризується поблажливим, але дружелюбним відношенням до клієнтів. Мета – усунути несправність. Супроводжується використанням специфічних термінів для відображення своєї компетентності. На відмову співпрацювати – реакція здивування, оскільки співпраця у першу чергу вигідна для клієнта. Жодних вмовлять – йому дається зрозуміти, що без його участі проблема тільки ускладнюється. Допустиме залякування тяжкими наслідками.

• користувач – працівник, що виконує свої обов'язки і наляканий виникненням неочікуваної проблеми. Чітко виражений мотив швидкого вирішення усіх проблем і повернення до своєї рутинної роботи. Відсутність уявлення про характер проблеми, зацікавленість тільки в її вирішенні. Характер спілкування – показати безнадійність свого положення і готовність віддатися у руки спеціалісту.

Найбільш розповсюдженими способами реалізації методів соціальної інженерії за допомогою глобальної мережі Інтернет є:

• проведення соціальної інженерії шляхом електронного листування;

• проведення соціальної інженерії через системи обміну повідомленнями (Skype, Viber тощо);

• соціальна інженерія на форумах, чатах, блогах.

У даних випадках вдале реалізування соціальної інженерії обумовлене правильністю розроблення сценарію спілкування.

2. Особистий контакт.

Найбільш складний і небезпечний метод соціальної інженерії. Крім перерахованих вимог до сценарію спілкування і моделі поведінки, соціальний інженер повинен приділяти увагу своїй зовнішності і манерам «живого» спілкування. Для правильного візуального сприйняття, необхідно правильно підібрати: колір одягу та взуття; манери та жести при спілкуванні; положення в просторі відносно співрозмовника.

Також при використанні методів соціальної інженерії необхідно характеризувати співрозмовника. За голосом або за зовнішністю, доцільно визначити яку його слабкість доцільно використовувати для досягнення поставленої мети.

До основних слабкостей людини, використання яких разом з правильно підібраною поведінкою і сценарієм розмови дозволяє досягнути очікуваного результату, належать, наприклад: довірливість, страх, жадібність, відкритість, зверхність, милосердя.

Основними причинами впливу на об’єкт соціальної інженерії є, наприклад: відчуття достоїнства, прагнення до успіху та матеріальна вигода. Використання прийомів прихованого і прямого маніпулювання персоналом дають можливість соціальному інженеру дізнаватися і у подальшому використовувати інформацію для оцінювання її захищеності в інформаційно-телекомунікаційній системі.

1.2 Фішинг

1.2.1 Поняття фішингу

Отже, фішинг – це форма атаки з використанням соціальної інженерії, в ході якої зловмисник, маскуючись під надійний суб'єкт, виманює конфіденційну інформацію жертв.

Концепція фішингу вперше була описана у 1987 році в документі з конференції під назвою «Безпека системи: перспективи хакера». В документі описувалась техніка зловмисників, яка полягає в імітації авторитетних організацій або сервісів.

Фішинг існує впродовж багатьох років, за цей час кіберзлочинці розробили широкий спектр методів інфікування жертв.

Найчастіше зловмисники, які займаються фішингом видають себе за банки чи інші фінансові установи, щоб змусити жертву заповнити фальшиву форму та отримати дані облікових записів.

У минулому для виманювання даних користувачів кіберзлочинці часто використовували неправильно написані або оманливі доменні імена. Сьогодні зловмисники використовують більш складні методи, завдяки чому фальшиві сторінки дуже схожі на свої легітимні аналоги.

Викрадені дані жертв, зазвичай, використовуються для викрадення коштів з банківських рахунків або продаються в Інтернеті. Подібні атаки здійснюються також через телефонні дзвінки та SMS-повідомлення.

Окремо слід виділити цілеспрямований фішинг. Кіберзлочинці, які використовують цей метод, зазвичай, заздалегідь детально досліджують свою ціль. Це значно ускладнює ідентифікацію вмісту як шкідливого.

Систематичні фішинг-атаки почалися в мережі America Online (AOL) у 1995 р. Щоб викрасти легітимні облікові дані, зловмисники зв'язувалися з жертвами через AOL Instant Messenger (AIM), видаючи себе за співробітників AOL, які перевіряють паролі користувачів. Термін «фішинг» з'явився в групі

новин Usenet, яка зосереджувалася на інструменті AOHell, який автоматизував цей метод, і так ім'я закріпилося. Після того, як AOL у 1997 році ввела контрзаходи, кіберзлочинці зрозуміли, що можуть використовувати таку ж техніку в інших галузях, зокрема й фінансових установах.

Одна з перших великих, хоча і невдалих, спроб була в 2001 році. Зловмисники, скориставшись хаосом від терористичних атак 9/11, розіслали потерпілим електронну розсилку нібито для перевірки посвідчення особи. Отримані дані використовувались для крадіжки банківських даних.

Вже у 2005 році за допомогою фішингу кіберзлочинці викрали у користувачів США понад 900 мільйонів доларів США.

Відповідно до дослідження глобального фішингу APWG, у 2016 році спостерігалося понад 250 тисяч унікальних фішингових атак, під час яких використовувалось рекордне число доменних імен, зареєстрованих зловмисниками, перевищуючи позначку в 95 тисяч. В останні роки кіберзлочинці намагалися зосередитися на банківських та фінансових послугах, користувачах електронного банкінгу, соціальних мереж, а також облікових даних електронної пошти.

Для виявлення кібератак наразі виділяють два підходи: детерміністичний та ймовірнісний.

У рамках першого зазвичай використовують сигнатури – унікальні послідовності байтів, що описують шкідливі об'єкти, які дозволяють однозначно ідентифікувати відомі кібератаки в автоматичному режимі.

Другий підхід здебільшого використовується для блокування невідомих загроз або загроз нульового дня при таргетованих атаках, коли ми заздалегідь не знаємо індикаторів компрометації. Цей підхід дозволяє виявляти нові кібератаки з певною ймовірністю, залишаючи останнє слово за користувачем системи або фахівцем з кібербезпеки. Саме ймовірнісний підхід і відкриває широке поле для використання систем штучного інтелекту – нейронних мереж, систем з нечіткою логікою тощо.

1.2.2. Ретроспектива вірусної активності в Україні за 2021 рік

Згідно огляду вірусної активності за 2021 рік, який представила антивірусна компанія Zillya!, проаналізувавши дані атак шкідливого програмного забезпечення (ПЗ) на користувачів в Україні, попередній багаторічний тренд спрямований на витіснення Adware з перших щаблів рейтингів найбільш вживаних видів шкідливого програмного забезпечення знаходить своє продовження [19].

За даним експертів з кібербезпеки антивірусної компанії Zillya! сукупна кількість зафіксованих випадків атак на ПК користувачів з використанням троянських програм різних видів та модифікацій склала 57% від загальної кількості зафіксованих випадків, що у порівнянні з 2020 роком свідчить про зростання на 12%.

Протилежний тренд на зниження активності шкідливого програмного забезпечення типу Adware, що примусово транслюють рекламу на ПК користувачів, за даними антивірусної компанії Zillya! Склав у загальній кількості виявлених атак на ПК користувачів приблизно 33%. Такі показники вказують на незначне падіння частоти використання Adware для зараження ПК користувачів в Україні на 2% у порівнянні з показником 35% у 2020 році.

Якщо раніше ми констатували, що низька популярність Adware серед кіберзлочинці, може свідчити про те, що тренди зміщуються в сторону використання інших видів шкідливого програмного забезпечення, зокрема троянських програм.

Кіберзлочинці не стоять на місці в питаннях створення нового шкідливого ПЗ та винайдення шпарин для використання у своїх злочинних цілях. Втім, експерти антивірусної компанії Zillya! відзначають, що головним каналом розповсюдження «вірусів» залишається електронна пошта, яка використовуються в СПАМ розсилках, заражених шкідливим ПЗ листів.

Сучасний етап технологічного розвитку та методів комунікації не дає можливості винайти щось кардинально нове для розповсюдження шкідливого

ПЗ. Саме тому, для інфікування пристроїв та мереж продовжують використовувати та вдосконалювати методи соціальної інженерії, психологічного впливу, направленого на несвідоме відкриття заражених документів у вкладеннях чи переходу за посиланнями в листах, що ведуть на заражені ресурси.

Окремо варто зазначити, що популярним каналом стала доставка шкідливого ПЗ на персональні ПК користувачів через уразливості, що виникають в результаті відмови користувачів встановлювати оновлення програмного забезпечення, а особливо операційних систем, що в першу чергу пов'язано з використанням піратських версій цих програм.

1.2.3. Ретроспектива вірусної активності за 2022 рік

Компанія ESET проаналізувала ситуацію в кіберпросторі від початку вторгнення Росії в Україну до сьогодні. Зокрема за цей період українські організації стали ціллю атак багатьох шкідливих програм. Переважно загрози націлювалися на знищення даних, ймовірно, з метою погіршення реагування на вторгнення Росії [46-48].

Крім цього, хакери спробували атакувати енергетичний сектор України за допомогою розгортання унікального програмного забезпечення на високовольтній електричній підстанції.

13 січня 2022 р. ряд українських державних установ, неурядових організацій та IT-компаній опинилися під прицілом шкідливої програми WhisperGate. За даними Microsoft, загроза виглядає як програма-вимагач та не має механізму відновлення після блокування, а її справжня ціль — зробити пристрої неробочими. Це може вказувати на зв'язок загрози із державними структурами агресора, а не з групами кіберзлочинців.

Наступного ранку веб-ресурси багатьох державних установ стали ціллю атак зловмисників, зокрема на цих сайтах з'явилися антиукраїнські зображення та повідомлення. Перед вторгненням державні та приватні організації

продовжували бути мішенню кіберзлочинців, включно з розподіленими атаками на відмову в обслуговуванні (DDoS attack – Distributed Denial of Service attack), які порушили роботу декількох важливих веб-сайтів в Україні. У той же час клієнти одного українського банку отримували SMS-повідомлення про фальшиві збої в роботі банкоматів банку.

Кібератаки в січні та на початку лютого 2022 р. були лише початком. Увечері 23 лютого, після ряду DDoS-атак на українські веб-сайти, спеціалісти ESET виявили нову загрозу HermeticWiper для знищення даних на сотнях машин у кількох організаціях в Україні. Тоді як часова мітка показує, що шкідливе програмне забезпечення було створено 28 грудня 2021 р. Тому можна припустити, що атака могла готуватися протягом деякого часу.

У деяких випадках одночасно з HermeticWiper використовувалась програма-вимагач HermeticRansom. Вперше про HermeticRansom стало відомо вранці 24 лютого. Програма-вимагач використовувалася як приманка, поки загроза HermeticWiper для знищення даних завдавала збитків (рис. 1.4).

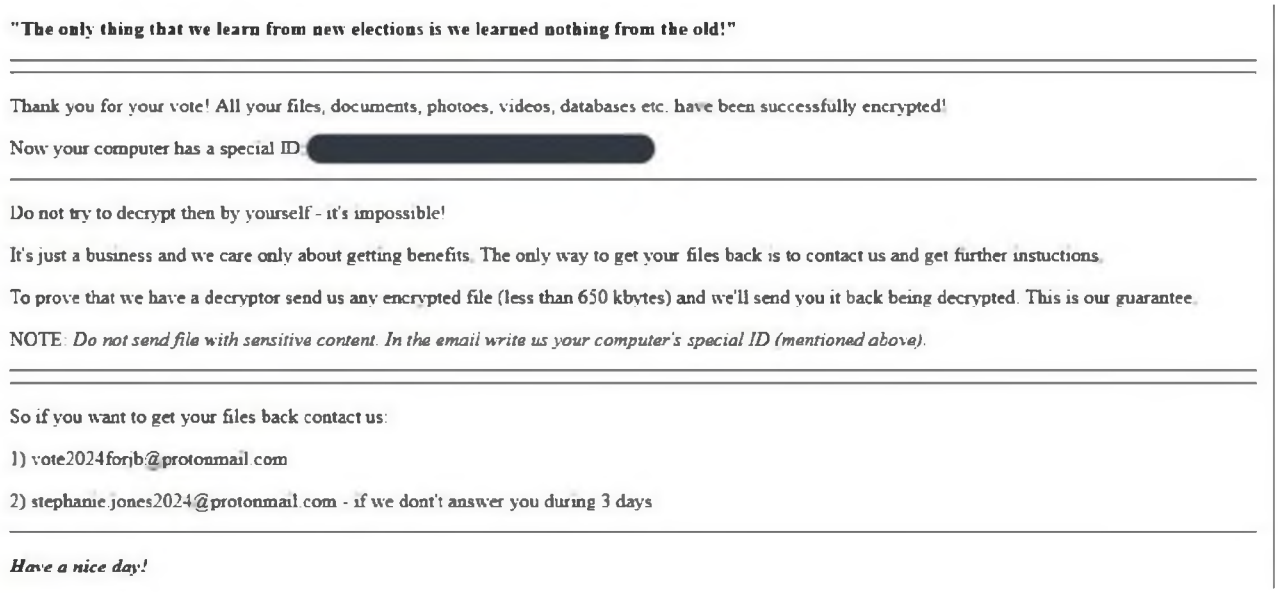


Рисунок 1.4 – Повідомлення про викуп у шкідливій програмі HermeticWiper

У день, коли відбулося військове вторгнення Росії в Україну, дослідники ESET помітили іншу шкідливу програму для знищення даних в українських системах – IssacWiper. Зокрема загроза була менш складною та не мала подібностей з HermeticWiper.

Дослідники ESET вважають, що різні атаки програм для знищення даних, включно з виявленою 14 березня CaddyWiper, були спрямовані на конкретні організації з метою погіршити їх реагування на вторгнення. Кіберзлочинці були націлені на фінансовий, медіа та державний сектори. Крім цього, спеціалісти ESET виявили зв'язок загроз HermeticWiper та CaddyWiper з групою Sandworm, яку США ідентифікували як частину російської військової розвідки.

Ця ж відома група кіберзлочинців також була причетна до спроби розгортання Industroyer2 на високовольтній електричній підстанції в Україні, викриття якого було вчасно здійснено завдяки співпраці між ESET та CERT-UA. Шкідливе програмне забезпечення є новою версією загрози Industroyer, яка використовувалася для атаки на українську електромережу ще в 2016 р.

Крім цього, відбувалися й інші види шкідливої діяльності, включно з DDoS-атаками, які спричинили збої в роботі медіа-ресурсів, неурядових організацій та телекомунікаційних провайдерів, а також державних установ (рис. 1.5).

Кіберзлочинці у всьому світі використовують тему війни в Україні у своїх схемах, зокрема створюючи фальшиві сайти та надсилаючи шкідливі спам-листи. Відразу після початку війни спеціалісти ESET виявили ряд підроблених ресурсів, на яких шахраї під виглядом фальшивих благодійних організацій виманювали пожертви для підтримки України.

Також, за даними ESET, було зафіксовано збільшення атак на криптовалютні платформи та поширення шкідливих програм, пов'язаних із криптовалютою. Такий ріст спричинила популярність цифрової валюти серед користувачів, яка стала поширеним способом оплати та надсилання коштів.

Відповідно до рейтингу кіберзагроз, перші місяці 2022 року показують руйнівну силу кібератак, які відбуваються паралельно з військовим

вторгненням. При цьому, кібератаки спрямовані не лише на державні установи, а й на компанії та звичайних користувачів в Україні.

Рисунок 1.5 – Атаки, виявлені дослідниками ESET до та після вторгнення Росії в Україну

Виходячи з вищевказаного, спеціалісти ESET та інших антивірусних компаній рекомендують дотримуватися основних правил онлайн-безпеки, зокрема уникати переходу за підозрілими посиланнями та завантаження вкладених файлів, перевіряти інформацію, перш ніж надсилати гроші, та використовувати надійні рішення для захисту пристроїв.

Тільки за січень-квітень 2022 р. загальна кількість виявлених зразків загроз зросла на 20% порівняно з останніми чотирма місяцями 2021 р. Зокрема збільшилась кількість шпигунських програм та загроз, які поширюються через електронну пошту (рис. 1.6).

Також у цей період значно вплинула на поширення загроз у світі війна в Україні. Зокрема, як вже було зазначено вище, ця тема активно використовувалася у спам-повідомленнях та на шкідливих сайтах. Крім цього, з

початком повномасштабного вторгнення Україна неодноразово ставала ціллю кіберзлочинців.

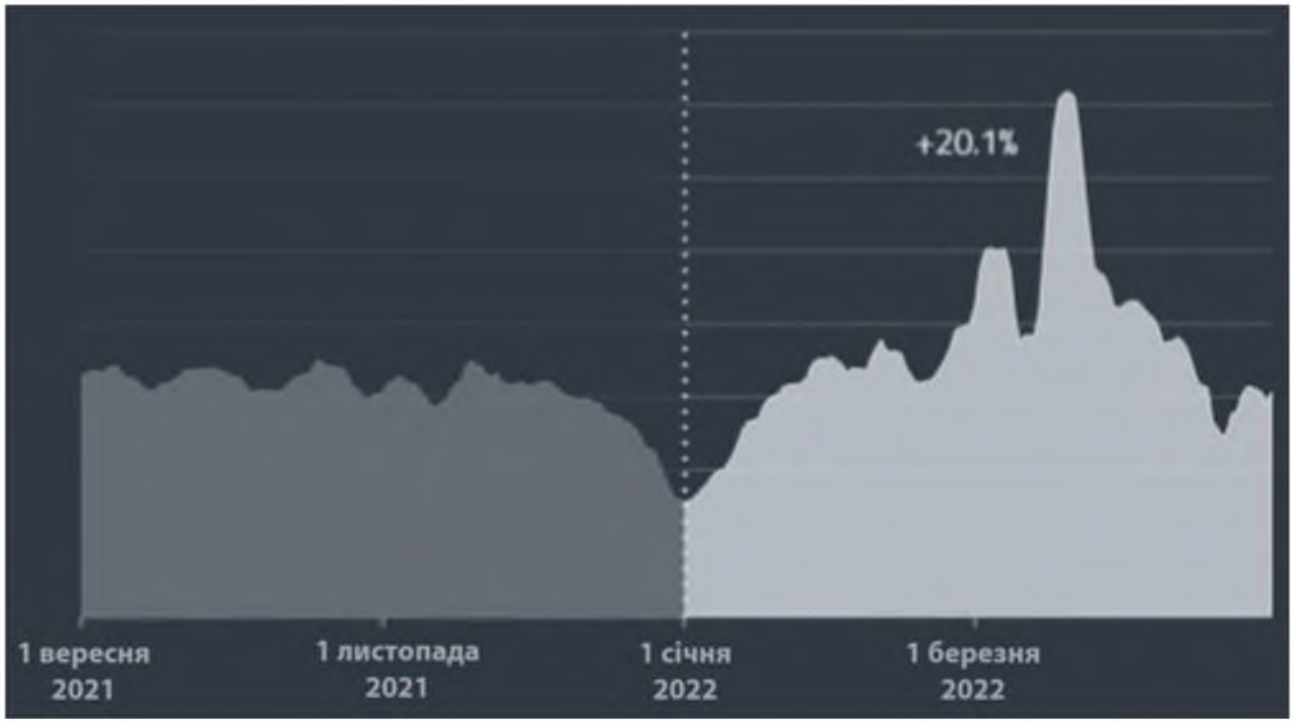


Рисунок 1.6 – Рівень виявлення загроз

За січень-квітень 2022 р. кількість загроз, які поширюються через електронні листи, різко збільшилася на 37%. Це найбільше зростання цього виду загроз із 2020 р. Така динаміка пов'язана з відновленням діяльності ботнета Emotet, який масово розсилав користувачам спам-повідомлення зі шкідливими вкладеннями. Крім звичайних тем електронних листів, таких як платежі, замовлення та доставки, з початку року росла кількість шкідливих повідомлень на тему подорожей. Також повернення Emotet вплинуло на зростання кількості завантажувачів на 121% у січні-квітні 2022 р.

Вперше за два роки безперервного росту кількість атак методом підбору пароля на протокол віддаленого робочого столу (RDP) знизилася на 41%. При цьому, 60% RDP-атак з початку 2022 р. відбувалися з території Росії.

Можливими причинами падіння цього виду шкідливої активності є зменшення кількості співробітників, які працюють віддалено, підвищення

обізнаності IT-відділів та поступове покращення безпеки корпоративного середовища, а також вторгнення Росії в Україну.

За перші 4 місяці 2022 р. кількість загроз для викрадення інформації зросла на 12%. При цьому найбільше зростання було у підкатегоріях шпигунського та шкідливого банківського програмного забезпечення. Зокрема кількість шпигунських програм у цей період зросла на близько 18%.

Попри незначний ріст кількості загроз для Android, на цій операційній системі найпоширенішими теж залишалися шпигунські програми. Активність таких програм, які можуть отримати доступ до різних функцій смартфона, зокрема здійснювати запис аудіо та відео, збільшилась на 170%. Зростання їх виявлення свідчить про пошук зловмисників способів заробітку на особистих або навіть корпоративних даних на пристроях Android.

Тоді як на операційній системі macOS майже половина всіх виявлених зразків становили потенційні небажані додатки.

Одразу після вторгнення Росії в Україну шахраї вирішили скористатися бажанням людей зі всього світу підтримати українців, зокрема зловмисники під цим приводом виманювали кошти у користувачів (рис. 1.7). Вже 24 лютого спеціалісти ESET виявили значне зростання спам-повідомлень та перші шахрайські домени, які використовують тему війни.

У середньому телеметрія ESET щоденно виявляла 4,8 мільйонів веб-загроз та 370 тисяч шкідливих URL-адрес у всьому світі. При цьому кількість заблокованих фішингових URL-адрес зросла майже на 30%. Найвищий рівень виявлення припав на 07 березня, збільшившись утричі за середній щоденний показник з початку цього року.

Крім цього, за даними ESET приблизно третина фішингових URL-адрес, виявлених у січні-квітні 2022 р., маскувалися під фінансові організації. Також як приманку зловмисники використовували фальшиві сторінки для входу у Facebook та WhatsApp.

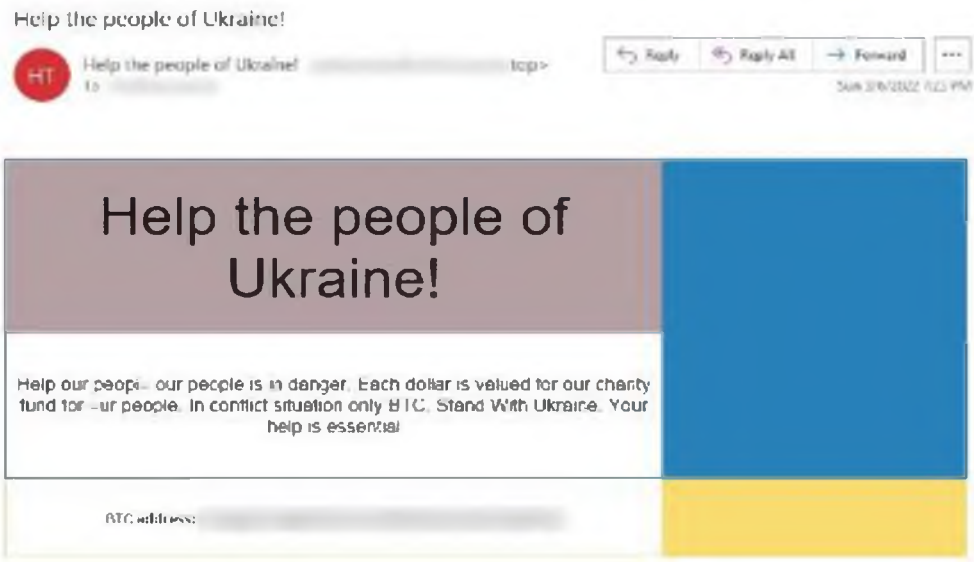


Рисунок 1.7 – Приклад того, як хакери активно використовували тему війни в Україні

За травень-серпень 2022 р. зросла кількість загроз для Android та число фішингових листів на тему доставки. Крім цього, Україна опинилася у п'ятірці країн, на які була націлена найбільша кількість атак програм-вимагачів.

На поширення загроз у цей період значно вплинуло стрімке падіння курсів криптовалют. Зокрема кіберзлочинці почали частіше викрадати криптовалюту замість її прихованого майнінгу, на що вказує зростання кількості фішингових листів на тему криптовалюти та загроз для її крадіжки.

Незважаючи на загальне зменшення активності програм-вимагачів, спостерігалося кілька піків їх активності. Один з них у червні 2022 р. був частково спричинений розповсюдженням в Україні версій шкідливої програми MSIL/Filecoder. Ця загроза є першою програмою-вимагачем з відкритим кодом. Також Україна опинилася на 5 місці серед кількості атак програм-вимагачів після США, Китаю, Ізраїлю та Росії (рис. 1.8).

Кількість фішингових листів на тему доставки зросла у 6 разів порівняно з попереднім періодом. Також зловмисники використовували тему фінансів та соцмереж для поширення фішингових загроз. За даними ESET, найчастіше хакери маскувались під компанію Facebook, зокрема використовуючи

Messenger для надсилання небезпечних посилань. Таким чином кіберзлочинці заманювали користувачів на фішингові сторінки, переконуючи їх ввести облікові дані свого акаунту.

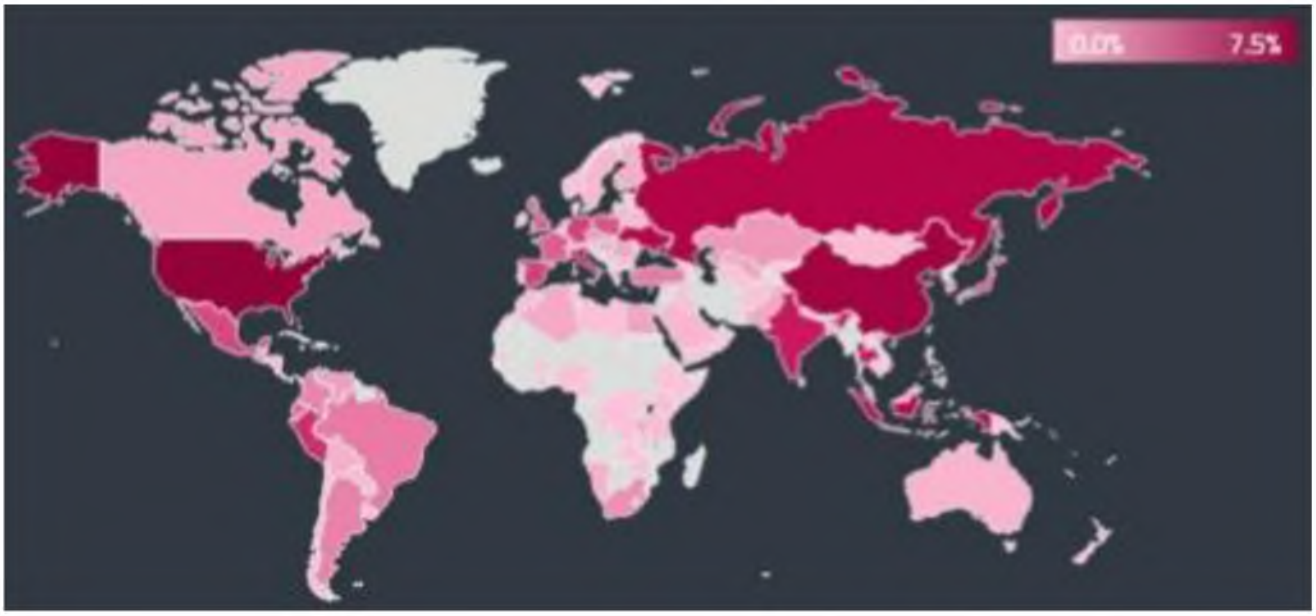


Рисунок 1.8 – Виявлення програм-вимагачів у світі у 2022 р.

Загалом кількість унікальних фішингових URL-адрес, які продукти ESET блокували щодня, у середньому досягала 38 000 за період травень-серпень 2022 р.

Серед загроз для викрадення даних збільшилась кількість банківських шкідливих програм майже на 10%. Найпоширенішою шкідливою програмою у цій категорії є Magecart, яка здатна викрадати дані банківських карток у онлайн-покупців. Тоді як серед загроз для викрадення інформації найбільш активним залишилося саме шпигунське ПЗ.

Крім цього, зловмисники все частіше вдаються до викрадення криптовалюти, а не до її незаконного майнінгу на пристроях жертв. Зокрема на 50% зросло число загроз для крадіжки цифрових коштів. Очікується, що і надалі їх активність буде рости разом з випадками прихованого майнінгу в браузері під час відвідування торент-сайтів та піратських ресурсів.

За даними телеметрії ESET, загальна кількість загроз для Android збільшилась на 9,5%. Найактивнішими на цій операційній системі були шпигунські програми та приховані додатки (рис. 1.9).

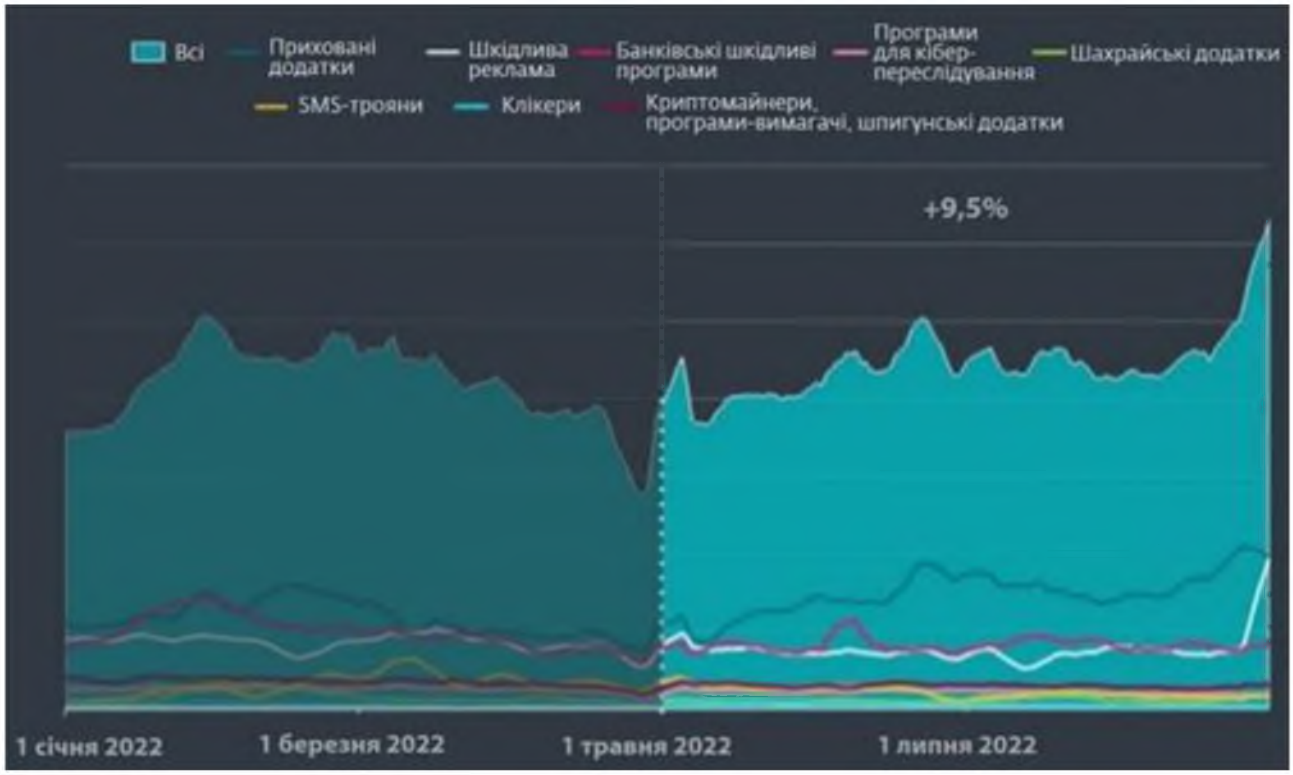


Рисунок 1.9 – Виявлення загроз для Android

Це зростання спричинено доступністю наборів шпигунських програм для Android на різноманітних онлайн-форумах. Таким чином навіть зловмисники-аматори можуть безкоштовно знайти трояни та розгорнути їх без певного рівня технічних навичок.

Незважаючи на зменшення активності загроз для macOS, атаки на користувачів цієї платформи продовжуються. Зокрема спеціалісти ESET виявили новий бекдор для macOS під назвою CloudMensis, який шпигує за користувачами. Загроза використовує виключно публічні служби хмарних сховищ для обміну даними зі своїми авторами.

Активність ботнета Mozi, який використовує застарілі уразливості, а в минулому році до складу якого входило близько 600 000 ботів, почала повільно

зменшуватися. Це спричинило появу нових ботнетів на базі Mirai, які стають дедалі потужнішими та поширенішими, експериментуючи з новими та відомими уразливостями.

1.3 Класифікація

1.3.1 Постановка задачі класифікації

Завдання класифікації полягає у розбитті об'єктів на декілька класів. Об'єкти в межах одного класу вважаються еквівалентними з погляду критерію розбиття.

Взагалі, у задачі класифікації та регресії потрібно визначити значення залежної змінної об'єкта на підставі значень інших змінних, що характеризують цей об'єкт.

Формально завдання класифікації та регресії можна описати наступним чином. Є безліч об'єктів:

$$I = \{i_1, i_2, ..., i_j, ..., i_n\}, \tag{1.1}$$

де i_j – досліджуваний об'єкт.

Кожен об'єкт характеризується набором змінних:

$$I_j = \{x_1, x_2, ..., x_h, ..., x_m, y\}, \tag{1.2}$$

де x_h – незалежні змінні, значення яких відомі і на підставі яких визначається значення залежної змінної y.

В Data Mining часто набір незалежних змінних позначають у вигляді вектора:

$$X = \{x_1, x_2, ..., x_h, ..., x_m\}. \tag{1.3}$$

Кожна змінна x_h може набувати значень з деякої множини:

$$C_h = \{c_{h1}, c_{h2}, ...\}. \tag{1.4}$$

Якщо значення змінної є елементи кінцевої множини, то кажуть, що вона має категоріальний тип.

Якщо безліч значень $C = \{c_1, c_2, ..., c_r, ..., c_k\}$ змінної *y* кінцева, то задача називається задачею класифікації. Якщо змінна *y* набуває значення на множині дійсних чисел R, то завдання називається завданням регресії.

Отже, класифікатор – це система, яка вводить (як правило) вектор дискретних і/або неперервних функцій і виводить одне дискретне значення класу [31-32].

Наприклад, фільтр спаму класифікує повідомлення електронної пошти на «спам» або «не спам», і його вхідними даними може бути вектором булевих значень $x=(x_1,...,x_j,...,x_d)$, де $x_j=1$, якщо j-е слово в словнику з'являється в електронному листі, а $x_j=0$ в іншому випадку. Учень уводить навчальний набір прикладів (x_i, y_i), де $x_i=(x_{i.1},...,x_{i,d})$ – це спостережуваний ввід, y_i – відповідний вихід і виводить класифікатор. Тест учня полягає в тому, чи дає цей класифікатор правильний вихід y_t для майбутніх прикладів x_t (наприклад, чи фільтр спаму правильно класифікує раніше невидимі електронні листи як спам чи не спам).

Для всіх класифікаторів найважливішими є 3 компоненти:

1. Представлення. Класифікатор повинен бути представлений за допомогою формальної мови, яку комп'ютер може обробляти. І, навпаки, вибір представлення для учня рівносильний вибору набору класифікаторів, яких він може навчитися. Цей набір називається гіпотезою простору учня. Якщо класифікатор не знаходиться в гіпотезі простору, то він не можу бути вивчений.

2. Оцінювання. Функція оцінювання (так звана цільова функція) потрібна для виділення «гарних» класифікаторів від «поганих».

3. Оптимізація. Також потрібен метод пошуку серед всіх класифікаторів такого, який буде класифікувати найбільш швидко й правильно. Вибір методу оптимізації є ключовим елементом ефективності учня, а також допомагає визначити вибраний класифікатор, якщо функція оцінки має більше ніж один оптимум. Для нових учнів найкраще почати використовувати загальноприйняті оптимізатори, які пізніше замінюються спеціально розробленими [32].

Класифікація може бути розбита на 2 частини:

• бінарна класифікація – групування результату в одну з двох груп (наприклад, фішинговий та безпечний веб-сайт);

• багатокласова класифікація – групування результату в одну з декількох (більше двох) груп.

Наразі для задач класифікації (до яких відноситься й детектування фішингових URL-адрес) все частіше використовують системи штучного інтелекту, зокрема нейронні мережі, які є універсальними ефективними апроксиматорами.

1.3.2 Оцінка якості класифікації

Для оцінки якості класифікації зазвичай використовують наступні показники (метрики):

• матриця помилок або неточностей (від англ. «Confusion Matrix»);

• акуратність (від англ. «Accuracy»);

• точність (від англ. «Precision»);

• повнота (від англ. «Recall»);

• F-міра (від англ. «F-score»);

• ROC-крива або крива робочих характеристик (від англ. «Receiver Operating Characteristics curve»);

• Precision-Recall (PR) крива.

Перед переходом до самих метрик необхідно ввести важливу концепцію для опису цих метрик у термінах помилок класифікації – confusion matrix (матрицю помилок або неточностей).

Припустимо, що у нас є два класи $y=\{0,1\}$ і алгоритм, який прогнозує (передбачає) належність кожного об'єкта одному з цих класів. Розглянемо приклад. Нехай система захисту мережі використовує систему класифікації для виявлення атаки: нормальна робота мережі чи аномальна (наявність атаки). При цьому у першому випадку система і далі нормально працює, а у другому – видається сигнал аномальної роботи. Таким чином, виявлення неадекватної

(аномальної) роботи мережі (y=1) можна розглядати як «сигнал тривоги», що повідомляє про можливі ризики.

Будь який реальний класифікатор робить помилки. При розгляді вищезазначеного прикладу таких помилок може бути дві:

• нормальна ситуація у мережі за даними трафіка розпізнається моделлю як аномальна (даний випадок можна трактувати як «помилкову тривогу»);

• аномальна ситуація розпізнається як нормальна і ніяких дій по захисту від атаки не відбувається (даний випадок можна розглядати як «пропуск цілі»).

Неважко помітити, що зазначені помилки нерівноцінні по зв'язаними з ними наслідками. У разі «помилкової тривоги» втрати складуть тільки марно потрачений час та ресурси на протидію неіснуючій загрозі. У разі ж «пропуску цілі» можна втратити набагато більше (інформацію, роботу мережі та інше, в залежності від виду атаки). Тому системі захисту важливіше не допустити «пропуску цілі», ніж «помилкової тривоги».

Оскільки з точки зору логіки задачі виявлення аномалій важливіше правильно розпізнати аномалію (атаку) з міткою y=1, ніж помилитися в розпізнаванні нормальної роботи мережі, відповідний результат класифікації зазвичай називають позитивним (аномалія чи атака виявлені вірно), а протилежний – негативним (аномалії чи атаки немає y=0).

Виходячи з вищевказаного можливі наступні чотири результати класифікації:

1. True Positive (TP) – наявність атаки класифікована як наявна атака, тобто позитивний клас розпізнано як позитивний. Спостереження, для яких це має місце називаються істинно-позитивними.

2. True Negative (TN) – нормальна робота мережі класифікована як нормальна робота без аномалій, тобто негативний клас розпізнано як негативний. Спостереження, яких це має місце, називаються істинно негативними.

3. False Positive (FP) – нормальна робота мережі класифікована як аномальна, тобто мала місце помилка, в результаті якої негативний клас був

розпізнаний як позитивний. Спостереження, для яких було отримано такий результат класифікації, називаються помилково-позитивними, а помилка класифікації називається помилкою I роду.

4. False Negative (FN) – атака чи аномальна робота мережі розпізнана як нормальна, тобто мала місце помилка, в результаті якої позитивний клас був розпізнаний як негативний. Спостереження, для яких було отримано такий результат класифікації, називаються помилково-негативними, а помилка класифікації називається помилкою II роду.

Таким чином, помилка I роду, або хибно-позитивний результат класифікації, має місце, коли негативне спостереження розпізнано моделлю як позитивне. Помилкою II роду, або хибно-негативних результатом класифікації, називають випадок, коли позитивне спостереження розпізнано як негативне.

Пояснимо вищенаведене за допомогою матриці помилок класифікації (табл. 1.2).

Таблиця 1.2 – Матриця помилок класифікації

	y=1	y=0
a(x)=1	Істинно-позитивний (True Positive – TP)	Помилково-позитивний (False Positive – FP)
a(x)=0	Помилково-негативний (False Negative – FN)	Істинно-негативний (True Negative – TN)

В табл. 1.1 введено такі позначення: a(x) – це відповідь алгоритму при конкретній ситуації; y – справжня мітка класу для цієї ситуації.

Отже, помилки класифікації бувають двох видів: False Negative (FN) і False Positive (FP). P означає що класифікатор визначає клас об'єкта як позитивний (N – негативний). T означає що клас передбачений вірно (відповідно F – невірно). Кожен рядок в матриці помилок представляє прогнозований клас, а кожен стовпець – фактичний клас.

Тобто у загальному випадку, матриця неточностей – це матриця розміром N на N, де N – кількість класів, яка представляє собою табличне представлення прогнозованих і фактичних значень для кожного можливого класу.

Матриця помилок, одна з наважливіших речей, на яку потрібно дивитися при оцінці моделі класифікації. Це матриця, яка візуалізує кількість фактичних екземплярів класу у порівнянні із прогнозованими екземплярами класу. Таке подання дозволяє швидко побачити кількість вірних і невірних прогнозів для кожної категорії.

На основі матриці помилок або неточностей будується ряд інших вищезазначених характеристик.

Акуратністю (від англ. «Accuracy») називається пропорція точних прогнозів по відношенню до загальної кількості прогнозів, тобто це ймовірність того, що клас буде передбачений правильно:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \quad (1.5)$$

Тобто, Accuracy – частка правильних відповідей алгоритму:

Хоча акуратність є швидким та інформативним індикатором продуктивності моделі, зазвичай не можна покладатися виключно на неї. Це пов'язано з тим, що вона приховує наявність зсуву в моделі, що є звичайним явищем, якщо набір даних незбалансований, тобто негативних моментів значно більше, ніж позитивних, або навпаки.

Отже, зазначену метрику не варто застосовувати в задачах з нерівними класами, в яких як варіант можна використовувати алгоритми семпліювання. Семпліювання (від англ. «Data sampling») – це метод коригування навчальної вибірки з метою балансування розподілу класів у початковому наборі даних.

Приклад. Нехай необхідно оцінити роботу спам-фільтра пошти. Є 100 не-спам листів, 90 з яких класифікатор визначив вірно (True Negative = 90, False Positive = 10), і 10 спам-листів, 5 з яких класифікатор також визначив вірно (True Positive = 5, False Negative = 5). Тоді Accuracy:

$$Accuracy = \frac{5 + 9_0}{5 + 90 + 10 + 5} = 86,4 \quad (1.6)$$

Але якщо просто прогнозувати, що всі листи не-спам, то буде отримано більш високу акуратність:

$$Accuracy = \frac{0 + 10_0}{0 + 100 + 0 + 1_0} = 90,9 \quad (1.7)$$

При цьому, модель абсолютно не володіє ніякою прогностичною силою, оскільки спочатку необхідно було визначати листи зі спамом. Подолати це допоможе перехід із загальної для всіх класів метрики до окремих показників якості класів.

Точністю (від англ. «Precision») називається частка правильних відповідей моделі в межах класу – це частка об'єктів, що дійсно належать даному класу, відносно всіх об'єктів які система віднесла до цього класу.

$$Precision = \frac{TP}{TP + FP} \quad (1.8)$$

Саме введення такої метрики, як Precision не дозволяє записувати всі об'єкти в один клас, оскільки в цьому випадку отримують зростання рівня False Positive.

Повнотою (від англ. «Recall») називається частка істинно позитивних класифікацій. Повнота показує, яку частку об'єктів, що реально належать до позитивного класу, було передбачено вірно. Або ж іншими словами: це частка варіантів, класифікованих як позитивні, які насправді виявилися позитивними.

$$Recall = \frac{TP}{TP + FN} \quad (1.9)$$

Отже, повнота демонструє здатність алгоритму виявляти даний клас взагалі.

Маючи матрицю помилок, дуже просто можна обчислити точність і повноту для кожного класу. Точність (Precision) дорівнює відношенню відповідного діагонального елемента матриці і суми усього рядка класу. Повнота (Recall) – відношенню діагонального елемента матриці і суми усього

стовпця класу. Оскільки класів може бути багато (не обов'язково два), то формально:

$$Precision_c = \frac{A_{c,c}}{\Sigma_{i=1}^{n} A_{c,i}} \tag{1.10}$$

$$Recall_c = \frac{A_{c,c}}{\Sigma_{i=1}^{n} A_{i,c}} \tag{1.11}$$

Тобто, результуюча точність класифікатора розраховується як середнє арифметичне його точності по всіх класах. Те ж саме з повнотою.

Precision і Recall не залежить від співвідношення класів (на відміну від Accuracy) і тому можуть бути застосовні в умовах незбалансованих вибірок. Часто в реальній практиці стоїть завдання знайти оптимальний (для замовника) баланс між цими двома метриками. Зрозуміло що чим вище точність і повнота, тим краще. Але у реальному житті максимальна точність і повнота недосяжні одночасно і доводиться шукати якийсь баланс. Тому, хотілося б мати якусь метрику яка об'єднувала б у собі інформацію про точність та повноту нашого алгоритму. У цьому випадку буде простіше приймати рішення про те, яку реалізацію запускати у виробництво (у кого більше той і крутіше). Саме такою метрикою і є F-міра.

F-міра (від англ. «F-score») є гармонійним середнім між точністю і повнотою. Вона прагне до нуля, якщо точність або повнота прагне до нуля.

$$F = \frac{2 \times precision \times recall}{precision + recall} \tag{1.12}$$

Формула (1.12) надає однакову вагу точності і повноти, тому F-міра буде падати однаково при зменшенні і точності і повноти.

Слід зауважити, що є можливість розрахувати F-міру надавши різну вагу точності і повноті, якщо свідомо віддається пріоритет одній з цих метрик при розробці алгоритму:

$$F_\beta = \frac{\left(1 + \beta^2\right) \times precision \times recall}{\left(\beta^2 \times precision\right) + recall} \tag{1.13}$$

де β приймає значення в діапазоні 0<β<1 якщо необхідно віддати пріоритет точності, а при β>1 пріоритет віддається повноті. При β=1 формула зводиться до попередньої і отримується збалансована F-міра (також її називають F_1) – рис 1.10-1.12.

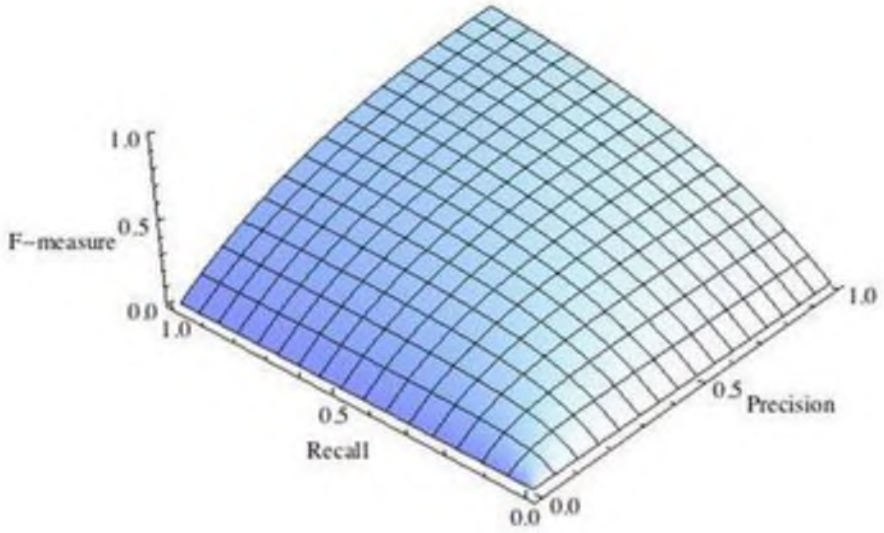


Рисунок 1.10 – Збалансована F-міра, β=1

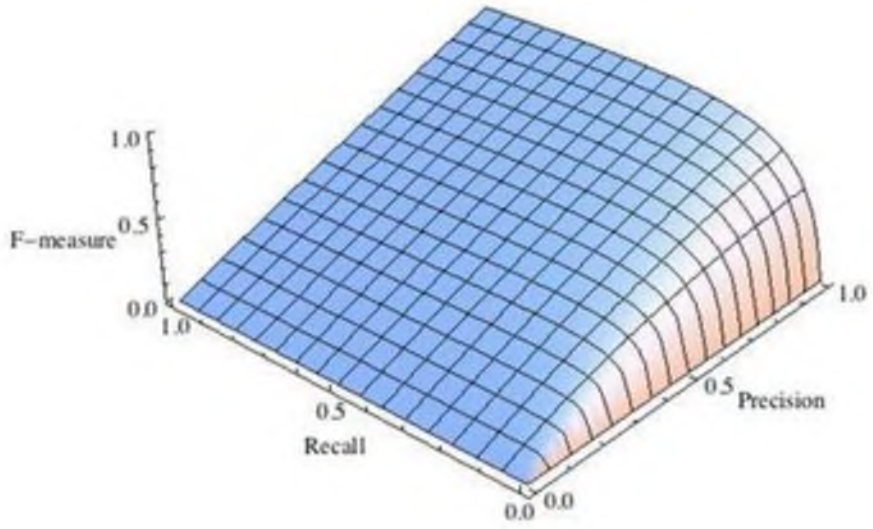


Рисунок 1.11 – F-міра з пріоритетом точності, $\beta^2=1/4$

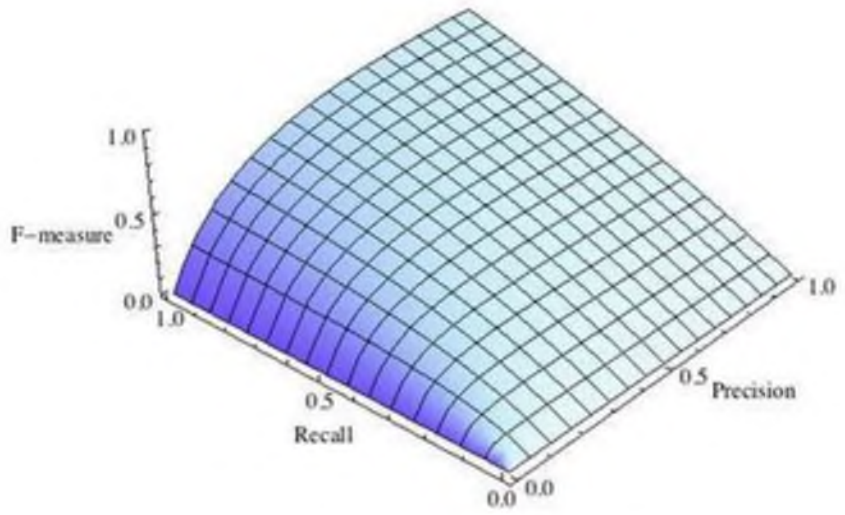


Рисунок 1.12 – F-міра с пріоритетом повноти, $\beta^2=2$

F-міра досягає максимуму при максимальній повноті і точності, і близька до нуля, якщо один з аргументів близький до нуля.

F-міра є хорошим кандидатом на формальну метрику оцінки якості класифікатора. Вона зводить до одного числа дві інші основоположні метрики: точність і повноту. Маючи F-міру набагато простіше відповісти на питання: «змінився алгоритм в кращу сторону чи ні?».

ROC-крива або крива робочих характеристик (від англ. «Receiver Operating Characteristics curve») використовується для аналізу поведінки класифікаторів при різних порогових значеннях. Дозволяє розглянути всі порогові значення для даного класифікатора.

Показує частку хибно позитивних прикладів (від англ. False Positive Rate, FPR) у порівнянні з часткою істинно позитивних прикладів (від англ. True Positive Rate, TPR) (рис. 1.13).

$$TPR = \frac{TP}{TP + FN} = Recall \tag{1.14}$$

$$FPR = \frac{FP}{FP + TN} \tag{1.15}$$

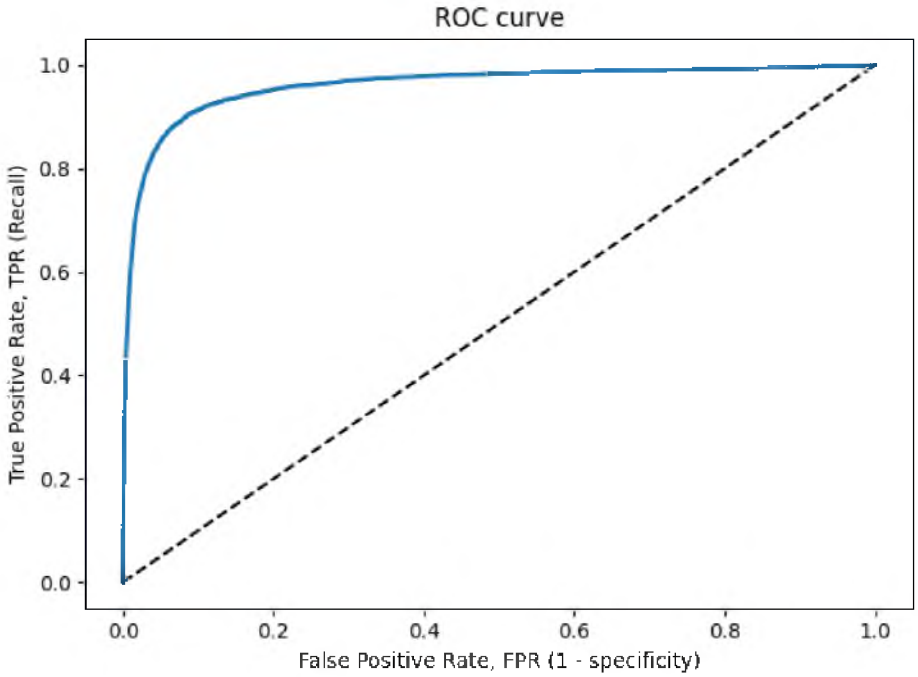


Рисунок 1.13 – ROC-крива

Частка FPR – це пропорція негативних зразків, які були некоректно класифіковані як позитивні.

$$FPR = 1 - TNR \tag{1.16}$$

де TNR – частка істинно негативних класифікацій (від англ. True Negative Rate), що представляє собою пропорцію негативних зразків, які були коректно класифіковані як негативні.

Частка TNR також називається специфічністю (від англ. «Specificity»). Отже, ROC-крива зображає чутливість (від англ. «Sensitivity»), тобто повноту, у порівнянні з різницею 1 – specificity.

Пряма лінія по діагоналі представляє ROC-криву чисто випадкового класифікатора. Хороший класифікатор тримається від зазначеної лінії настільки далеко, наскільки це можливо (прагнучи до лівого верхнього кута) (див. рис. 1.13).

Один із способів порівняння класифікаторів передбачає вимір площі під кривою (від англ. Area Under the Curve – AUC). Бездоганний класифікатор

матиме площу під ROC-кривою (ROC-AUC), що дорівнює 1, тоді як чисто випадковий класифікатор – площу 0.5.

Графік ROC (рис. 1.13) допомагає прийняти рішення про те, де встановити поріг класифікації, щоб максимізувати істинно позитивний рівень або мінімізувати псевдопозитивний показник, що у кінцевому підсумку є бізнес-рішенням.

Чутливість до співвідношення класів. Розглянемо задачу виділення математичних статей з безлічі наукових статей. Припустимо, що за все мається 1000100 статей, з яких лише 100 належать до математики. Якщо нам вдасться побудувати алгоритм a(x), що ідеально вирішує завдання, то його TPR буде дорівнює одиниці, а FPR – нулю. Розглянемо тепер поганий алгоритм, що дає позитивну відповідь на 95 математичних і 50000 нематематичних статтях. Такий алгоритм абсолютно даремний, але при цьому має TPR=0.95 і FPR=0.05, що вкрай близько до показників ідеального алгоритму.

Таким чином, якщо позитивний клас істотно менше за розміром, то AUC-ROC може давати неадекватну оцінку якості роботи алгоритму, оскільки вимірює частку невірно прийнятих об'єктів щодо загального числа негативних. Так, алгоритм b(x), що поміщає 100 релевантних документів на позиції з 50001-й по 50101-ю, матиме AUC-ROC 0.95.

Позбутися від зазначеної проблеми з незбалансованими класами можна, перейшовши від ROC-кривої до Precision-Recall (PR) PR-кривої. Вона визначається аналогічно до ROC-кривої, тільки по осях відкладаються не FPR і TPR, а повнота (по осі абсцис) і точність (по осі ординат) (рис. 1.14).

Слід зазначити, що критерієм якості сімейства алгоритмів виступає площа під PR-кривою (від англ. Area Under the Curve – AUC-PR).

1.4 Нейронні мережі

Нейронні мережі – це обчислювальні структури, які моделюють прості біологічні процеси, що зазвичай асоціюються з процесами людського мозку

[20-30]. НМ, що адаптуються і навчаються, представляють собою розпаралелені системи, здатні до навчання шляхом аналізу позитивних й негативних впливів. Елементним перетворювачем в даних мережах є штучний нейрон або просто нейрон, названий так за аналогією із біологічним прототипом.

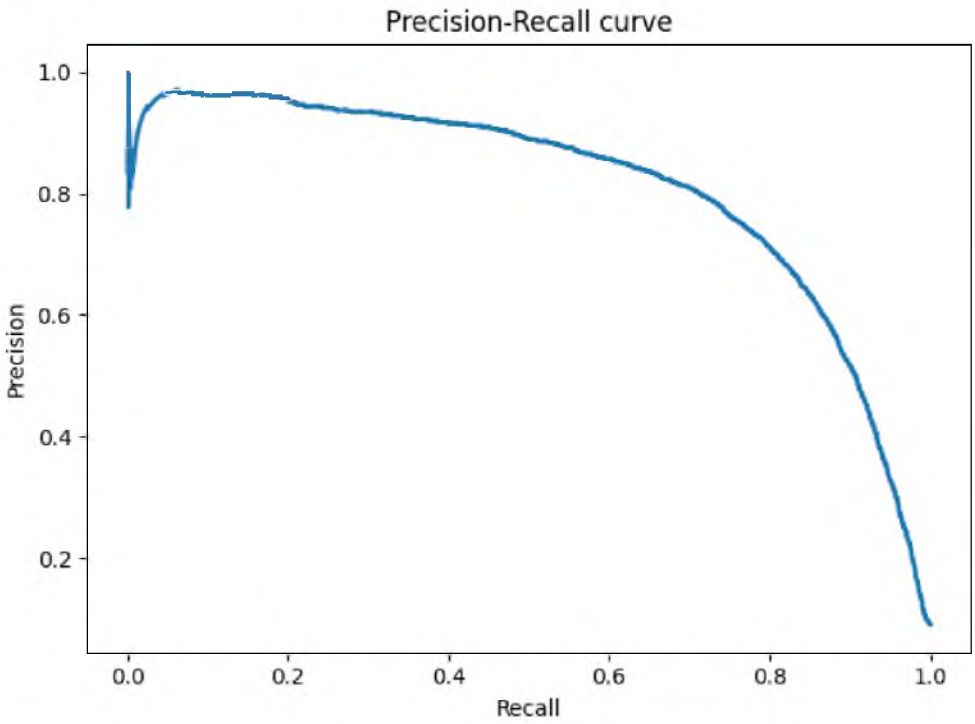


Рисунок 1.14 – Precision-Recall (PR) крива

НМ є універсальними структурами, що дозволяють реалізувати будь-який обчислювальний алгоритм. Ця ефективність НМ випливає з ряду так званих теорем про повноту.

Теорема про повноту. Будь-яка безперервна функція на замкнутій обмеженій множині може бути рівномірно наближена функціями, що обчислюються нейронними мережами, якщо функція активації нейрону двічі безперервно диференційована і безперервна.

У 1989 р. К. Фунахаші (Funahashi) показав, що нескінченно велика НМ з єдиним прихованим шаром здатна апроксимувати будь-яку безперервну функцію, та сформулював дане твердження у формі наступної теореми.

Теорема Фунахаші. Нехай $\phi(x)$ – непостійна, обмежена і монотонно зростаюча безперервна функція. Нехай, $U \subset \Re^n$ – обмежена множина і $f: U \to \Re$ – речова безперервна функція, що визначена на U. Тоді для довільного ε>0 існує ціле L і речові константи w_i, w_{ij}, такі, що апроксимація

$$\widetilde{f}(x_1, x_2, ..., x_n) = \sum_{i=1}^{L} w_i \phi(\sum_{j=1}^{n} w_{ij} x_j) \tag{1.17}$$

задовольняє нерівності

$$\left\| f - \widetilde{f} \right\|_\infty = \sup_{x \in U} \left| f(x) - \widetilde{f}(x) \right| \le \varepsilon . \tag{1.18}$$

Іншими словами, будь-яке безперервне відображення може бути апроксимоване в сенсі однорідної топології на U двошаровою НМ з активаційними функціями $\phi(x)$ для нейронів прихованого шару і лінійними активаційними функціями для нейронів вихідного шару.

Наразі запропоновано і вивчено велика кількість моделей нейроподібних елементів і нейромереж, але для вирішення задач класифікації зазвичай використовують перцептрони, НМ Хопфілда та НМ Хемінга.

1.4.1. Перцептрони

Ідею перцептрону (від латинського «perceptio» – сприйняття), запропонував американський нейрофізіолог Ф. Розенблатт. У 1960 році він представив перший нейрокомп'ютер – «Марк-1», який був здатний розпізнавати деякі букви англійського алфавіту.

У найбільш простому вигляді перцептрон (рис. 1.15) складається із сукупності чутливих (сенсорних) елементів (S-елементів), на які надходять вхідні сигнали. S-елементи випадковим чином пов'язані із сукупністю асоціативних A-елементів, вихід яких відрізняється від нуля тільки тоді, коли порушено досить велике число S-елементів, що впливають на один A-елемент. A-елементи з'єднані з реагуючими R-елементами, зв'язками, коефіцієнти посилення v яких змінюються в процесі навчання. Зважені комбінації виходів

R-елементів складають реакцію системи, яка вказує на належність об'єкта, який розпізнається, певному образу.

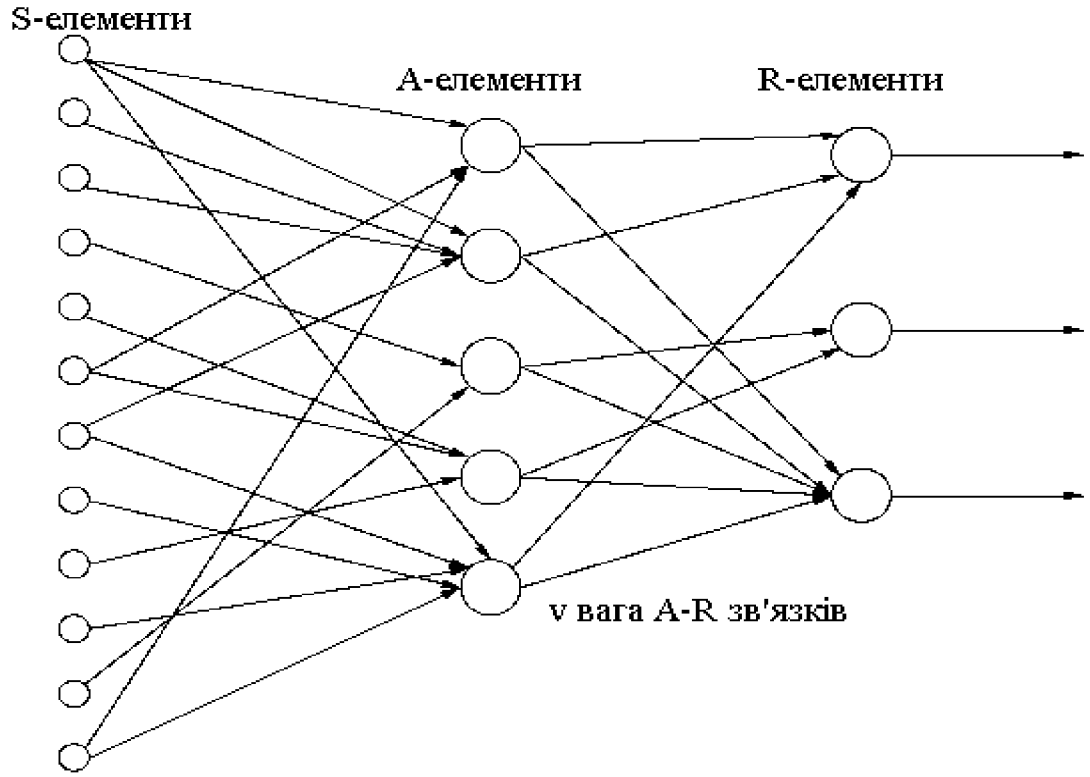


Рисунок 1.15 – Структурна схема перцептрону

Якщо розпізнаються тільки два образи, то перцептроном встановлюється тільки один R-елемент, який має дві реакції – позитивну і негативну. Якщо образів більше двох, то для кожного образу встановлюють свій R-елемент, вихід якого представляє лінійну комбінацію виходів A-елементів:

$$R_j = \Theta_j + \sum_{i=1}^{n} v_{ij} x_i \quad , \tag{1.19}$$

де R_j – реакція j-го R-елемента; x_i – реакція i-го A-елемента; v_{ij} – вага зв'язку від i-го A-елемента до j-го R елемента; j – поріг j-го R-елемента.

Аналогічно записується рівняння i-го A-елемента:

$$x_i = \Theta_i + \sum_{k=1}^{S} y_k \quad . \tag{1.20}$$

Сигнал y_k (1.20) може бути безперервним, але частіше за все він приймає тільки два значення: 0 або 1. Сигнали від S-елементів подаються на входи A-елементів з постійними вагами рівними 1, але кожен A-елемент пов'язаний тільки з групою випадково обраних S- елементів.

Припустимо, що потрібно навчити перцептрон розрізняти два образи V_1 і V_2. Будемо вважати, що в перцептроні існує два R-елемента, один з яких призначений образу V_1, а інший – образу V_2. Перцептрон буде навчений вірно, якщо вихід R_1 перевищує R_2, коли об'єкт, що розпізнається, належить образу V_1, і навпаки. Розподіл об'єктів на два образи можна провести і за допомогою тільки одного R-елемента. Тоді об'єкту образу V_1 повинна відповідати позитивна реакція R-елемента, а об'єктам образа V_2 – негативна.

Перцептрон навчається шляхом пред'явлення навчальної послідовності зображень об'єктів, що належать образам V_1 і V_2. У процесі навчання змінюються ваги v_i A-елементів. Зокрема, якщо застосовується система підкріплення з корекцією похибок, перш за все враховується вірність рішення, прийнятого перцептроном. Якщо рішення вірне, то ваги зв'язків всіх A-елементів, що спрацювали та ведуть до R-елементу, який видав правильне рішення, збільшуються, а ваги A-елементів, що не спрацювали, залишаються незмінними. Можна залишати незмінними ваги A-елементів, що спрацювали, але зменшувати ваги, що не спрацювали, тощо. Після процесу навчання перцептрон сам, без вчителя, починає класифікувати нові об'єкти.

Якщо перцептрон діє за описаною схемою і в ньому допускаються лише зв'язки, що йдуть від бінарних S-елементів до A-елементів та від A-елементів до єдиного R-елементу, то такий перцептрон прийнято називати елементарним α-перцептроном. Зазвичай класифікація задається вчителем. Перцептрон повинен виробити в процесі навчання класифікацію, задуману вчителем.

Про перцептрони було сформульовано і доведено кілька фундаментальних теорем, дві з яких, що визначають основні властивості перцептрона, наведені нижче.

Теорема 1. Клас елементарних α-перцептронів, для яких існує рішення для будь-якої задуманої класифікації, не є порожнім.

Ця теорема стверджує, що для будь-якої класифікації навчальної послідовності можна підібрати такий набір (з нескінченного набору) А-елементів, в якому буде здійснено задуманий поділ навчальної послідовності за допомогою лінійного вирішального правила.

Теорема 2. Якщо для деякої класифікації рішення існує, то в процесі навчання α-перцептрону із корекцією похибок, що починається з довільного початкового стану, це рішення буде досягнуто протягом кінцевого проміжку часу.

Сенс цієї теореми полягає у тому, що якщо щодо задуманої класифікації можна знайти набір А-елементів, в якому існує рішення, то в рамках цього набору воно буде досягнуто в кінцевий проміжок часу.

Зазвичай обговорюють властивості нескінченного перцептрона, тобто перцептрону з нескінченним числом А-елементів із усілякими зв'язками з S-елементами (повний набір А-елементів). В таких перцептронах рішення завжди існує, а раз воно існує, то воно і досяжне в α-перцептронах з корекцією похибок.

Дуже цікавою областю досліджень є також багатошарові перцептрони і перцептрони з перехресними зв'язками.

1.4.2. Нейронні мережі Хопфілда

Серед різних конфігурацій штучних НМ зустрічаються такі, при класифікації яких за принципом навчання, напевно, не підходять ні навчання з вчителем, ні навчання без вчителя. У таких мережах вагові коефіцієнти синапсів розраховуються тільки один раз перед початком функціонування мережі на основі інформації про оброблювані дані, і все навчання мережі зводиться саме до цього розрахунку. З одного боку, пред'явлення апріорної інформації можна розцінювати, як допомога вчителя, але з іншого – мережа

фактично просто запам'ятовує зразки до того, як на її вхід надходять реальні дані, і не може змінювати свою поведінку, тому говорити про зворотній зв'язок зі «вчителем» не доводиться.

Серед нейромереж з подібною логікою роботи найбільш відомі мережі Хопфілда і Хемінга, які зазвичай використовуються для організації асоціативної пам'яті.

Структурна схема мережі Хопфілда наведена на рис. 1.16. Вона складається з єдиного шару нейронів, число яких є одночасно числом входів і виходів мережі. Кожен нейрон пов'язаний синапсами з усіма іншими нейронами, а також має один вхідний синапс, через який здійснюється введення сигналу. Вихідні сигнали, як зазвичай, утворюються на аксонах.

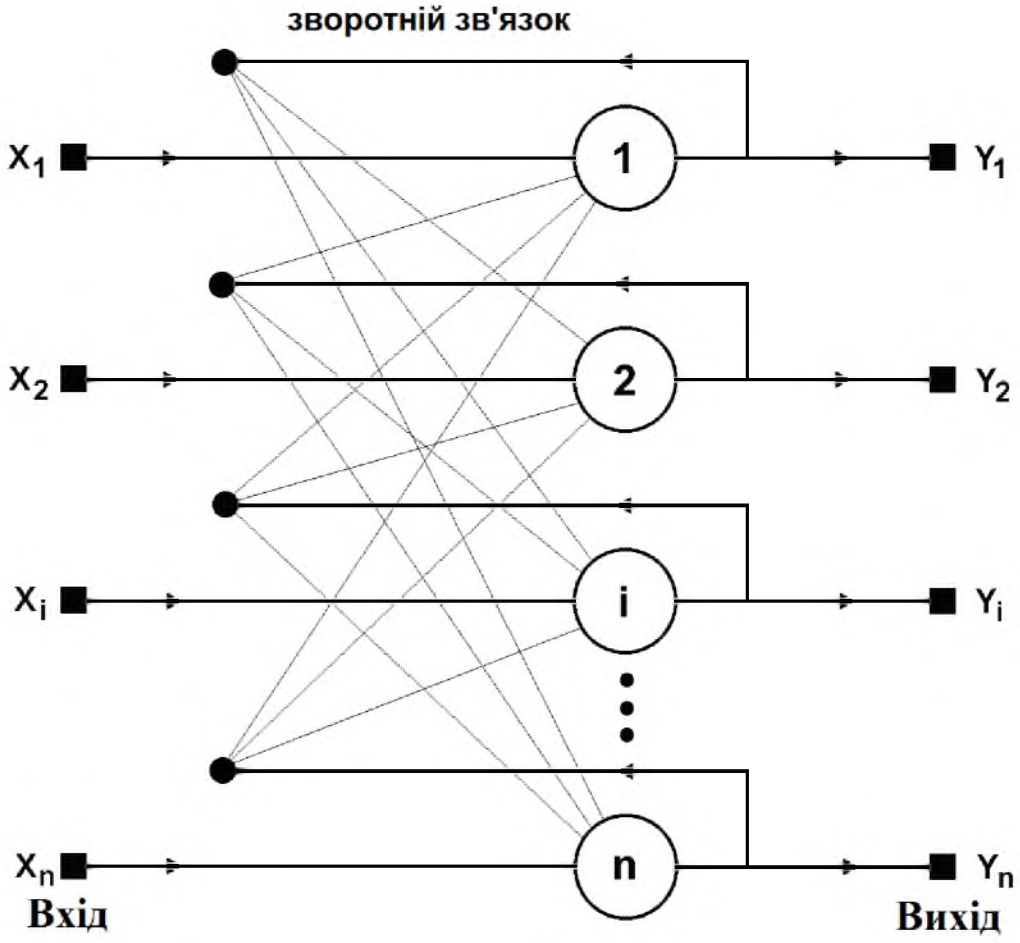


Рисунок 1.16 – Структурна схема НМ Хопфілда

Завдання, яке вирішується даною мережею в якості асоціативної пам'яті, зазвичай, формулюється у такий спосіб. Відомий деякий набір двійкових сигналів (зображень, звукових та інших даних, що описують якісь об'єкти або характеристики процесів), які вважаються зразковими. Мережа повинна вміти з довільного неідеального сигналу, поданого на її вхід, виділити («згадати» по частковій інформації) відповідний зразок (якщо такий є) або «дати висновок» про те, що вхідні дані не відповідають жодному із зразків. У загальному випадку, будь-який сигнал може бути описаний вектором $X=\{x_i: i=0...n-1\}$, n – число нейронів в мережі та розмірність вхідних і вихідних векторів. Кожен елемент x_i дорівнює або +1, або -1. Позначимо вектор, що описує зразок k, через X^k, а його компоненти, відповідно, – x_i^k, k=0 ... m-1, m – число зразків. Коли мережа розпізнає (або «згадає») будь-який зразок на основі пред'явлених їй даних, її виходи будуть містити саме його, тобто $Y=X^k$, де Y – вектор вихідних значень мережі: $Y=\{y_i: i=0, ... n-1\}$. В іншому випадку, вихідний вектор не співпаде ні з одним зразком.

Якщо, наприклад, сигнали являють собою зображення, то, відобразивши в графічному вигляді дані з виходу мережі, можна буде побачити картинку, яка повністю збігається з однією зі зразкових (у разі успіху) або ж «вільну імпровізацію» мережі (у разі невдачі).

На стадії ініціалізації мережі вагові коефіцієнти синапсів встановлюються таким чином:

$$w_{ij} = \begin{cases} \sum_{k=0}^{m-1} x_i^k x_j^k, i \neq j \\ 0, \quad i = j \end{cases}, \tag{1.21}$$

тут i і j – індекси, відповідно, предсинапсичного і постсинапсичного нейронів; x_i^k, x_j^k – i-ий і j-ий елементи вектору k-го зразка.

Алгоритм функціонування мережі наступний (p – номер ітерації):

1. На входи мережі подається невідомий сигнал. Фактично його введення здійснюється безпосередньою встановленням значень аксонів:

$$y_i(0)=x_i, \; i=0...n-1, \tag{1.22}$$

тому позначення на схемі мережі вхідних синапсів в явному вигляді носить чисто умовний характер. Нуль в дужках біля y_i означає нульову ітерацію в циклі роботи мережі.

2. Розраховується новий стан нейронів

$$s_j(p+1) = \sum_{i=0}^{n-1} w_{ij} y_i(p) \ , \text{j=0...n-1.} \tag{1.23}$$

і нові значення аксонів

$$y_j(p+1) = f\left[s_j(p+1)\right], \tag{1.24}$$

де f – активаційна функція у вигляді стрибка, що приведена на рис. 1.17,а.

3. Перевірка, чи змінилися вихідні значення аксонів за останню ітерацію. Якщо так – перехід до пункту 2, інакше (якщо виходи застабілізовано) – кінець алгоритму. При цьому вихідний вектор являє собою зразок, який найкращим чином поєднується з вхідними даними.

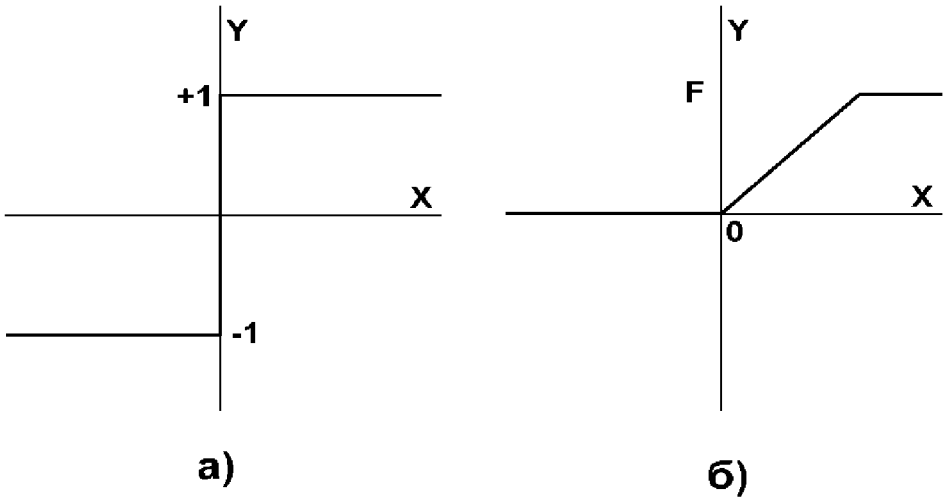


Рисунок 1.17 – Активаційні функції, які використовуються в НМ Хопфілда (а) та Хемінга (б)

Слід зазначити, що іноді мережа не може провести розпізнавання і видає на виході неіснуючий образ. Це пов'язано із проблемою обмеженості можливостей мережі.

Для НМ Хопфілда число образів, що запам'ятовуються не повинно перевищувати величини, приблизно рівної 0.15•n. Крім того, якщо два образи A і B дуже схожі, вони, можливо, будуть викликати у мережі Хопфілда перехресні асоціації, тобто пред'явлення на входи мережі вектора A призведе до появи на її виходах вектору B, і навпаки.

1.4.3. Нейронні мережі Хемінга

Коли немає необхідності, щоб НМ у явному вигляді видавала зразок, тобто досить, скажімо, отримувати номер зразка, асоціативну пам'ять успішно реалізує мережа Хемінга. У порівнянні з НМ Хопфілда, НМ Хемінга характеризується, меншими витратами на пам'ять і обсягом обчислень, що стає очевидним з її структури (рис. 1.18).

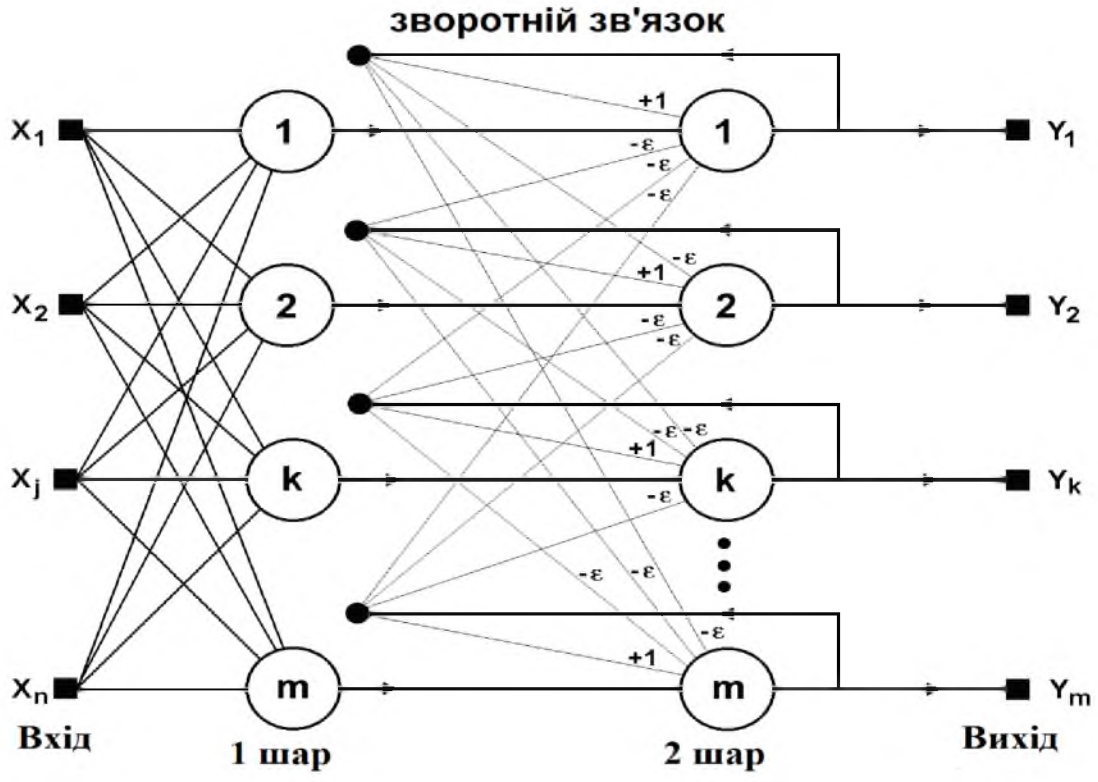


Рисунок 1.18 – Структурна схема НМ Хемінга

Мережа складається з двох шарів, які мають по m нейронів, де m – число зразків. Нейрони першого шару мають по n синапсів, з'єднаних зі входами мережі (утворюють фіктивний нульовий шар). Нейрони другого шару пов'язані між собою інгібіторними (негативними зворотними) синапсичними зв'язками. Єдиний синапс з позитивним зворотнім зв'язком для кожного нейрону з'єднаний з його ж аксоном.

Ідея роботи мережі полягає в знаходженні відстані Хемінга від тестованого образу до всіх зразків. Відстанню Хемінга називається число бітів, що відрізняються, в двох бінарних векторах. Мережа повинна вибрати зразок з мінімальною відстанню Хемінга до невідомого вхідного сигналу, в результаті чого буде активований тільки один вихід мережі, що відповідає цьому зразку.

На стадії ініціалізації ваговим коефіцієнтам першого шару і порогу активаційної функції привласнюються наступні значення:

$$w_{ik} = \frac{x_i^k}{2},\ i=0...n-1,\ k=0...m-1, \tag{1.25}$$

$$T_k = n/2,\ k=0...m-1, \tag{1.26}$$

де x_i^k – елемент і зразка k.

Вагові коефіцієнти гальмуючих синапсів у другому шарі беруть рівними деякій величині $0<\varepsilon<1/m$. Синапс нейрону, пов'язаний з його ж аксоном має вагу +1.

Алгоритм функціонування НМ Хемінга наступний:

1. На входи мережі подається невідомий вектор $X=\{x_i;\ i=0...n-1\}$, за яким розраховуються стани нейронів першого шару (верхній індекс у дужках вказує номер шару):

$$y_j^{(1)} = s_j^{(1)} = \sum_{i=0}^{n-1} w_{ij} x_i + T_j,\ j=0...m-1 \tag{1.27}$$

Після цього отриманими значеннями ініціалізуються значення аксонів другого шару:

$$y_j^{(2)} = y_j^{(1)},\ j=0...m-1. \tag{1.28}$$

2. Обчислити нові стани нейронів другого шару:

$$s_j^{(2)}(p+1) = y_j(p) - \varepsilon \sum_{k=0}^{m-1} y_k^{(2)}(p), k \neq j, j = 0...m-1, \qquad (1.29)$$

і значення їх аксонів:

$$y_j^{(2)}(p+1) = f\left[s_j^{(2)}(p+1)\right], j = 0...m-1. \qquad (1.30)$$

Активаційна функція f має вигляд порога (рис. 1.17,б), причому величина F повинна бути досить великою, щоб будь-які можливі значення аргументу не призводили до насичення.

3. Перевірити, чи змінилися виходи нейронів другого шару за останню ітерацію. Якщо так – перейди до кроку 2. Інакше – кінець алгоритму.

З оцінки алгоритму видно, що роль першого шару досить умовна: скориставшись значеннями його вагових коефіцієнтів один раз на кроці 1, мережа більше не звертається до нього, тому перший шар може бути взагалі виключений з НМ (замінений на матрицю вагових коефіцієнтів).

Отже, нейронні мережі Хопфілда і Хемінга дозволяють просто і ефективно вирішити задачу відтворення образів по неповній та спотвореній інформації.

Невисока ємність зазначених нейронних мереж (число образів, що запам'ятовуються) пояснюється тим, що вони не просто запам'ятовують образи, а дозволяють проводити їх узагальнення.

Так, наприклад, за допомогою НМ Хемінга можлива класифікація за критерієм максимальної правдоподібності. Разом з тим, легкість побудови програмних і апаратних моделей роблять ці НМ привабливими для багатьох застосувань.

1.5 Висновок. Постановка задачі

Захист інформації в ІТС орієнтований на збереження її властивостей конфіденціності, цілісності та доступності від різноманітних за своєю сутністю несприятливих впливів. Потенційно можливий несприятливий вплив тлумачиться загрозою. Для запобігання або ускладнення можливості реалізацій

загроз, зменшення потенційних збитків створюється та підтримується у дієздатному стані система заходів захисту інформації в комп'ютерних системах. Така система включає обчислювальну систему, фізичне середовище, персонал та інформацію. На збереження її властивостей в ІТС суттєво впливає врахування нетехнічного аспекту, зокрема, персоналу (наприклад, керівника, адміністратора, користувача). З огляду на це, для оцінювання технічної захищеності інформації пропонується соціоінженерний підхід. У рамках такого підходу вразливості персоналу тлумачяться як його слабкості, потреби, манії (пристрасті), захоплення. Маніпулювання ними дозволяє отримати несанкціонований доступ до інформації без руйнування та перекручування головних для нього системоутворюючих якостей (цілісність, розвиток). Як наслідок, це призводить до нової моделі поведінки персоналу, створення сприятливих умов реалізації загроз безпеці інформації і, як наслідок, зменшенню здатності системи захисту інформації протидіяти їх впливові. Це відображається в таких формах як шахрайство, обман, афера, інтрига, містифікація, провокація. Використанню кожної з означених форм маніпулювання передує визначення її змісту шляхом ретельного планування, організування та контролювання. Означені дії є основою методів соціальної інженерії. З одного боку, вони реалізуються засобами сучасних телекомунікацій. Тоді як з іншого, передбачається встановлення особистого контакту з персоналом. Таким чином, шляхом використання методів соціальної інженерії можливе виявлення, нейтралізування, запобігання появі уразливостей інформації в ІТС. Цим підвищується її захищеність з урахуванням нетехнічного аспекту.

Фішингові URL – це шкідливі веб-сайти, які маскують себе під безпечні для отримання конфіденційних даних, таких як номери кредитних карт, інформація щодо імені користувача та паролю для входу тощо. Фішинг використовує соціальну інженерію та технічний обман, щоб отримати приватну інформацію від веб-користувача. Наразі фішинг є однією з найбільших загроз в Інтернеті. Взагалі каналами фішингових атак є електронна пошта, текстові

повідомлення та посилання в соціальних мережах. Нерідко цільова сторінка фішингового веб-сайту також робить спробу проникнення в комп'ютер відвідувача та встановлення шкідливого програмного забезпечення.

Таким чином, фішингові атаки є значним ризиком як для фізичних осіб, так і для організацій, оскільки представляють собою загрозу розголошення або отримання конфіденційної особистої та корпоративної інформації. Отже, автоматичне виявлення та класифікація фішингових веб-сайтів є актуальною задачею.

Для виявлення кібератак наразі виділяють два підходи: детерміністичний та ймовірнісний. У рамках першого зазвичай використовують сигнатури – унікальні послідовності байтів, що описують шкідливі об'єкти, які дозволяють однозначно ідентифікувати відомі кібератаки в автоматичному режимі. Другий підхід здебільшого використовується для блокування невідомих загроз або загроз нульового дня при таргетованих атаках, коли ми заздалегідь не знаємо індикаторів компрометації. Цей підхід дозволяє виявляти нові кібератаки з певною ймовірністю, залишаючи останнє слово за користувачем системи або фахівцем з кібербезпеки. Саме ймовірнісний підхід і відкриває широке поле для використання систем ШІ.

Встановлено, що наразі для задач класифікації, до яких відноситься детектування фішингових URL-посилань все частіше використовують штучні нейронні мережі, які є універсальними ефективними апроксиматорами. Використання нейронних мереж потребує великої кількості даних для якісного навчання. Для вирішення задач класифікації (до якої відноситься й детектування фішингових веб-сайтів) зазвичай використовують наступні НМ: перцептрони, НМ Хопфілда та НМ Хемінга.

Отже, висновки, які отримані в цьому розділі, визначають подальші цілі і завдання, та підтверджують актуальність роботи.

Таким чином, для виконання мети кваліфікаційної роботи необхідно:

• дослідити використання штучних нейронних мереж для виявлення фішингових веб-сайтів;

• запропонувати підхід до детектування фішингових URL-адрес із використанням нейромережевих класифікаторів та методу прямого випадкового пошуку;

• оцінити ефективність запропонованого підходу.

2 СПЕЦІАЛЬНА ЧАСТИНА

2.1 Дослідження використання штучних нейронних мереж для виявлення фішингових веб-сайтів

Для детектування фішингових веб-сайтів використовувався набір даних із 100 фішингових і 100 безпечних URL-адрес, який включав в себе наступну інформацію:

- регістратор доменного імені;
- час життя домену, в роках;
- геолокація хостинг-сервера;
- наявність захищеного з'єднання з дійсним сертифікатом (https).

Для збору фішингових URL-адрес використовувалось джерело класифікованих фішингових посилань – Phishtank.org. Це база валідованих фішингових URL-адрес, куди завантажують нові зразки посилань як антивірусні компанії, так й індивідуальні дослідники. Для збору безпечних URL-адрес використовувалась платформа Similarweb.com, яка надає дані про онлайн-трафік, що дозволяє в реальному часі давати об'єктивне уявлення про веб-сайти та програми.

Дослідження штучних нейронних мереж для виявлення фішингових веб-сайтів проводилось за допомогою стандартних та розроблених програм в середовищі Matlab.

Зібраний набір даних з 200 фішингових і безпечних URL-адрес був розділений на навчальну вибірку, яка склала 70 % даних і тестову – 30 % даних.

Як класифікатори використовувались наступні НМ: перцептрон, НМ Хопфілда та НМ Хемінга. При цьому значення параметрів НМ класифікаторів були наступними: для перцептрону – це 30 нейронів прихованого шару, функція активації прихованого шару – порогова. Інші параметри перцептрону, а

також параметри НМ Хопфілда та НМ Хемінга встановлюються в середовищі Matlab за замовчуванням.

Дослідження штучних НМ для виявлення фішингових веб-сайтів проводилось за різних розмірів навчальної вибірки. Слід зазначити, що порівняння показників різних методів систем штучного інтелекту при різних розмірах навчальних вибірок – поширений спосіб їх дослідження для вирішення різних задач в галузі кібербезпеки [49].

У роботі розглядалося 5 різних навчальних вибірок, які мають розміри 60, 80, 100, 120 та 140 фішингових і безпечних URL-адрес, відповідно.

Для оцінки класифікації URL-адрес використовувались наступні метрики:

- TP (True Positive) – кількість безпечних посилань які було класифіковано як безпечні;

- TN (True Negative) – кількість фішингових посилань які було класифіковано як фішингові;

- FP (False Positive) – кількість безпечних посилань які було класифіковано як фішингові;

- FN (False Negative) – кількість фішингових посилань які було класифіковано як безпечні;

І похідні від них:

- True positive rate:

$$TPR = TP/(TP + FN); \tag{2.1}$$

- True negative rate:

$$TNR = TN/(TN + FP); \tag{2.2}$$

- False positive rate:

$$FPR = FP/(FP + TN); \tag{2.3}$$

- False negative rate:

$$FNR = FN/(FN + TP); \tag{2.4}$$

- Positive predictive value:

$$PPV = (TP + TN)/(TP + TN + FP + FN); \tag{2.5}$$

- Negative predictive value:

$$NPV = TN/(TN + FN); \quad (2.6)$$

- F-measure – гармонійне середнє між TPR і PPV:

$$F - measure = 2 * PPV * TPR/(PPV + TPR). \quad (2.7)$$

Значення метрик оцінки детектування фішингових URL-адрес для різних нейромережевих класифікаторів наведено в табл. 2.1-2.3.

Таблиця 2.1 – Значення метрик оцінки детектування URL-адрес при використанні перцептрону

Розмір навчальної вибірки	TP	TN	FP	FN	TPR	TNR	FPR	FNR	PPV	NPV	F-measure
60	22	25	5	8	0,73	0,83	0,17	0,27	0,78	0,76	0,754
80	23	27	3	7	0,76	0,9	0,1	0,23	0,83	0,79	0,793
100	23	28	2	7	0,76	0,93	0,07	0,23	0,85	0,8	0,802
120	24	30	0	6	0,8	1	0	0,2	0,9	0,83	0,847
140	24	30	0	6	0,8	1	0	0,2	0,9	0,83	0,847

Таблиця 2.2 – Значення метрик оцінки детектування URL-адрес при використанні НМ Хопфілда

Розмір навчальної вибірки	TP	TN	FP	FN	TPR	TNR	FPR	FNR	PPV	NPV	F-measure
60	19	23	7	11	0,63	0,77	0,23	0,37	0,7	0,68	0,663
80	19	24	6	11	0,63	0,8	0,2	0,37	0,72	0,69	0,672
100	20	24	6	10	0,67	0,8	0,2	0,33	0,73	0,7	0,699
120	20	25	5	10	0,67	0,83	0,17	0,33	0,75	0,71	0,708
140	20	25	5	10	0,67	0,83	0,17	0,33	0,75	0,71	0,708

Таблиця 2.3 – Значення метрик оцінки детектування URL-адрес при використанні НМ Хемінга

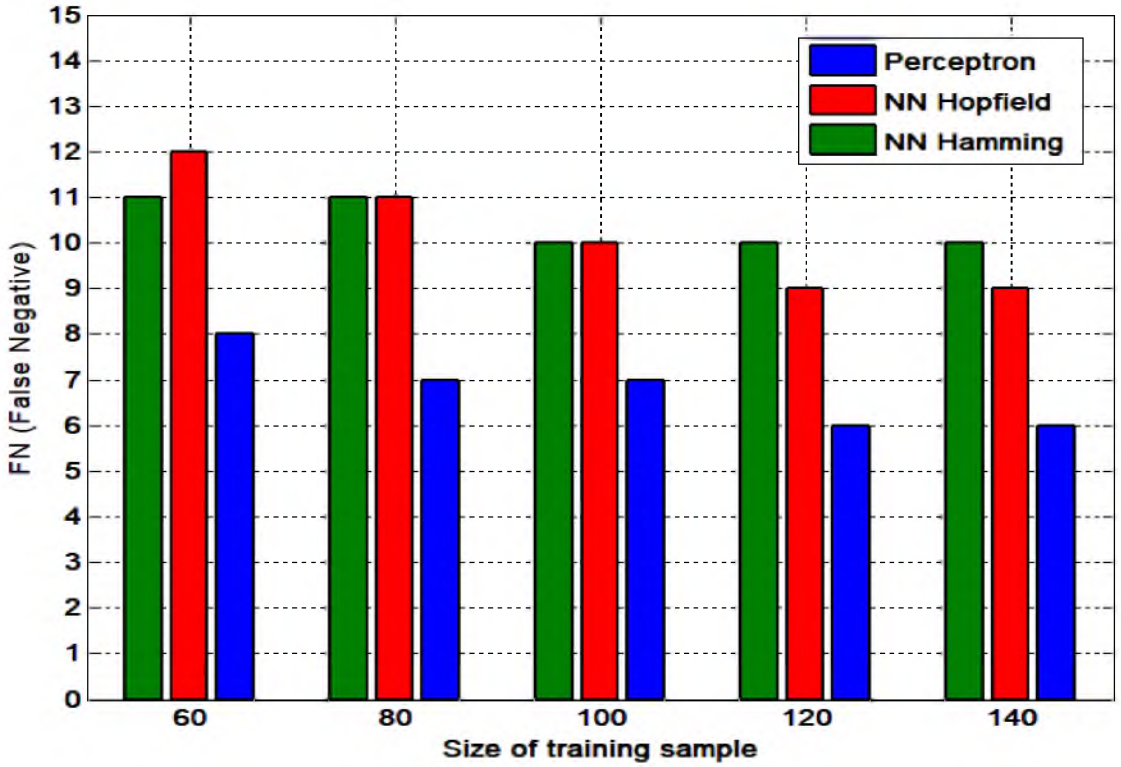
Розмір навчальної вибірки	TP	TN	FP	FN	TPR	TNR	FPR	FNR	PPV	NPV	F-measure
60	18	23	7	12	0,6	0,77	0,23	0,4	0,68	0,66	0,638
80	19	24	6	11	0,63	0,8	0,2	0,37	0,72	0,69	0,672
100	20	25	5	10	0,67	0,83	0,17	0,33	0,75	0,71	0,708
120	21	25	5	9	0,7	0,83	0,17	0,3	0,77	0,73	0,733
140	21	26	4	9	0,7	0,87	0,13	0,3	0,78	0,74	0,738

Оскільки у контексті завдання виявлення кібератак найбільш важливими є метрики FN та FNR, які дають кількісну оцінку невиявленим атакам, значення цих метрик для різних нейромережевих класифікаторів було додатково зображено на рис. 2.1.

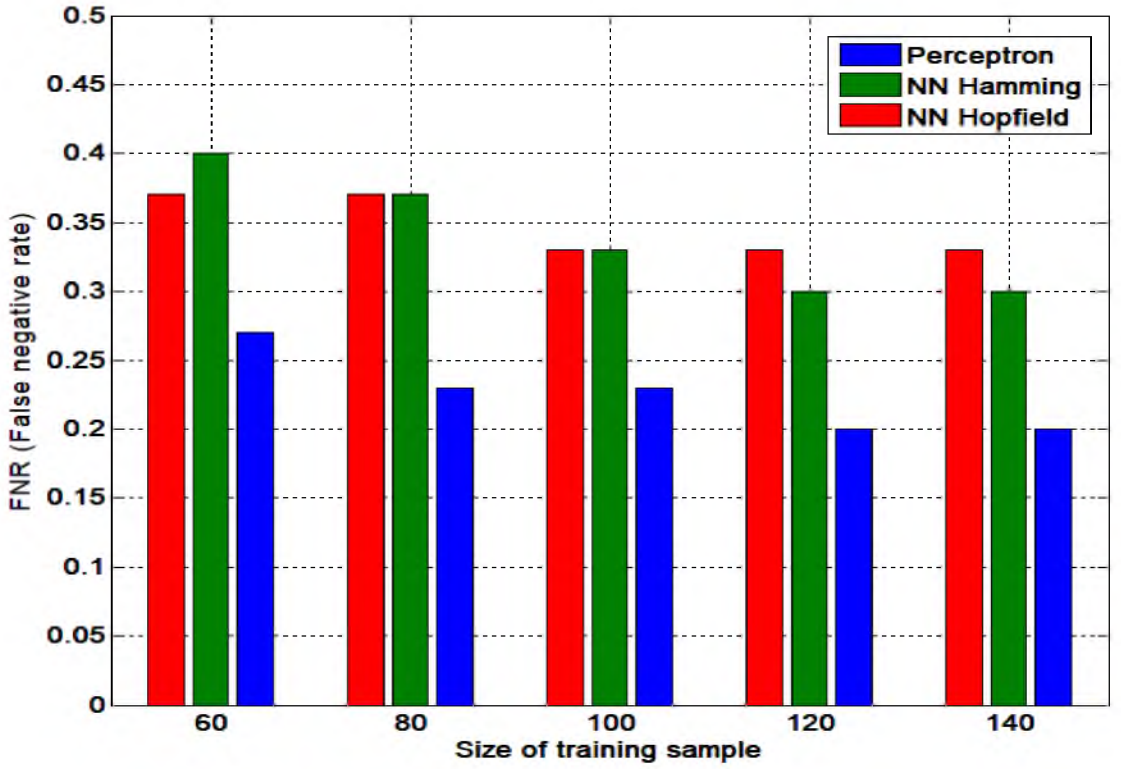
Було встановлено, що усі фішингові URL-адреси, які були класифіковані як безпечні (метрика FN) мали захищене з'єднання з дійсним сертифікатом (https-протокол).

Встановлено (див. табл. 2.1-2.3 та рис. 2.1), що найкращі значення показав класифікатор на основі перцептрону. Результати класифікаторів на основі НМ Хопфілда і Хемінга показали приблизно однакові результати.

Також встановлено, що для всіх нейромережевих класифікаторів результати детектування є кращими із збільшенням розміру навчальної вибірки від 60 до 120 URL-адрес. Результати ж детектування для розмірів вибірки в 120 і 140 URL-адрес виявились однаковими, тому немає необхідності у використанні навчальних вибірок більшого розміру, оскільки це збільшує час класифікації.



а



б

Рисунок 2.1 – Значення метрик FN (а) та FNR (б) оцінки детектування URL-адрес для різних нейромережевих класифікаторів

2.2 Підхід до детектування фішингових URL-адрес із використанням нейромережевих класифікаторів і методу прямого випадкового пошуку

Метод прямого випадкового пошуку (ПВП) дозволяє з певною ймовірністю знайти глобальний екстремум функцій кількох змінних [50]. Принцип роботи даного методу полягає у зміні початкового розподілу таким чином, щоб при випадковому викиданні точок зі зміненим розподілом зростала ймовірність попадання в ε-околицю екстремального значення цільової функції.

На першому етапі пошук відбувається на всій області визначення функції. На наступних кроках враховується інформація попередніх і відбувається обмеження околиці, в якій найбільш очікуваний вміст точки екстремуму, за рахунок зменшення дисперсії розподілу випадкових точок. Далі в цій околиці пошук йде з більшою інтенсивністю.

Нехай є цільова функція C(x[1:d]), де x[1:d] – вектор аргументів розмірності d. Вважатимемо, що компоненти x[k], де kє1:d, лежать в інтервалі [0; 1], оскільки інші обмеження можна врахувати під час побудови функції C(x), наприклад, з допомогою штрафних функцій. Розглянемо пошук мінімуму цільової функції.

Випадковий пошук складається з n_{stage} етапів, у свою чергу кожен етап ділиться на $m_{step}[i]$ кроків, де *i* – номер етапу. На кожному етапі задається певний закон, за яким випадково вибираються значення аргументів x[1:d]. Позначимо за $x_j[1:d]$ випадкові значення аргументів, вибрані на кроці *j*. Тоді $C^j=C(x_j[1:d])$ – значення цільової функції, обчислене в точці $x_j[1:d]$. Далі обчислюється мінімум із усіх значень C^j, отриманих на одному етапі, за формулою:

$$C^j_{\min} = C(x^*_j[1:d]) = \min\{C^j; C^{j-1}_{\min}\}, \tag{2.8}$$

де $x^*_j[1:d]$ – значення аргументу x[1:d], при якому цільова функція набуває найменшого значення за j кроків. Після виконання $m_{step}[i]$ кроків на і-му етапі змінюється закон вибору випадкових параметрів x[1:d] таким чином, щоб на

наступному етапі за нашими припущеннями збільшити ймовірність попадання в околиці глобального мінімуму цільової функції.

Передбачається, що на початку кожного етапу для параметрів x[k] виділяється інтервал I[k,i]≤[0;1] (i – номер етапу, kϵ1:d), в якому передбачається найімовірніше знаходження оптимального значення аргументу x[k]. Ширина інтервалу I[k,i] дорівнює 2q[i]. Вона однакова для будь-яких k і залежить від номера етапу i. Після кожного зменшення значення цільової функції перераховуються координати центру інтервалу I[k,i] за формулою:

$$x_j^0[k] = \max\{q[i]; \min\{x_j^*[k]; 1 - q[i]\}\}. \tag{2.9}$$

З метою спрощення процесу моделювання параметри x[k] вибираються з рівномірним розподілом всередині та зовні інтервалу I[k,i]. Їх щільності записуються відповідним чином:

$$H[i] = p[i]/s[i], \tag{2.10}$$

$$h[i] = (1 - p[i])/(1 - s[i]), \tag{2.11}$$

де p[i] – ймовірність того, що значення x[k]ϵI[k,i], а $s[i]=(2q[i])^d$ – d-вимірний обсяг звужуваної області.

На першому етапі пошук відбувається рівномірно по всьому проміжку [0;1] для кожного параметра x[k], тому що немає жодної початкової інформації про поведінку цільової функції. Тому вважатимемо, що

$$p[1]=1 \quad \text{та} \quad q[1]=1/2. \tag{2.12}$$

Пошук мінімуму ведеться як усередині інтервалу I[k,i], так і зовні його. На практиці в процесі пошуку звуження інтервалу I[k,i] відбувається до заданої величини ε.

Нехай i – номер етапу, тоді можна вважати, що

$$\lim_{i\to\infty} p[i] = 1 \quad \text{та} \quad \lim_{i\to\infty} q[i] = 0. \tag{2.13}$$

З припущень (2.12) та (2.13) випливає, що p[i] досягає свого мінімуму на деякому етапі з номером i_{min}, тобто $p_{min}= p[i_{min}]$.

Усередині інтервалу I[k,i] пошук проходить інтенсивніше, тому що там з більшою ймовірністю очікується оптимальне значення параметра x[k]. Тому справедливі співвідношення

$$1/2 \le p[i] \le 1 \quad \text{та} \quad h[i] \le 1 \le H[i]. \tag{2.14}$$

Як функцію p[i], будемо використовувати наступний варіант:

$$p[i] = \begin{cases} \frac{s[i](p_{\min}-1)}{s_{\min}} + 1 & \text{якщо } 0 \leqslant s[i] \leqslant s_{\min}, \\ \frac{s[i](1-p_{\min})}{1-s_{\min}} + \frac{p_{\min}-s_{\min}}{1-s_{\min}} & \text{якщо } s_{\min} \leqslant s[i] \leqslant 1. \end{cases} \tag{2.15}$$

Функція p[i] задовольняє всім необхідним умовам (2.12)-(2.14) і проста для моделювання.

За s_{min} прийматимемо d-вимірний обсяг перспективної області на етапі, якому відповідає мінімальне значення ймовірності p[i]. Тоді $s_{min}=(2q_{min})^d$, де $q_{min}=q[i_{min}]$ та i_{min} – номер етапу, на якому досягається мінімум ймовірності $p[i_{min}]=p_{min}$.

При даному поданні ймовірності p[i] H[i] та h[i] матимуть вигляд:

$$H[i] = \begin{cases} \frac{p_{\min}-1}{s_{\min}} + \frac{1}{s[i]} & \text{якщо } 0 \leqslant s[i] \leqslant s_{\min}, \\ \frac{1-p_{\min}}{1-s_{\min}} + \frac{p_{\min}-s_{\min}}{s[i](1-s_{\min})} & \text{якщо } s_{\min} \leqslant s[i] \leqslant 1. \end{cases} \tag{2.16}$$

$$h[i] = \begin{cases} \frac{(1-p_{\min})s[i]}{(1-s[i])s_{\min}} + \frac{1}{s[i]} & \text{якщо } 0 \leqslant s[i] \leqslant s_{\min}, \\ \frac{1-p_{\min}}{1-s_{\min}} & \text{якщо } s_{\min} \leqslant s[i] \leqslant 1. \end{cases} \tag{2.17}$$

Таким чином, запропонований підхід до детектування фішингових URL-адрес із використанням нейромережевих класифікаторів і методу прямого випадкового пошуку полягає у використанні глобальної оптимізації (методу ПВП) для вибір типу та параметрів нейромережевого класифікатора (на основі перцептрону, НМ Хопфілда та НМ Хемінга).

Узагальнена структура алгоритму детектування фішингових URL-адрес згідно запропонованого підходу із використанням нейромережевих класифікаторів і методу ПВП зображена на рис. 2.2.

Рисунок 2.2 – Узагальнена структура алгоритму детектування фішингових URL-адрес згідно запропонованого підходу із використанням нейромережевих класифікаторів і методу прямого випадкового пошуку

2.3 Оцінка ефективності підходу до детектування фішингових URL-адрес із використанням нейромережевих класифікаторів і методу прямого випадкового пошуку

Оцінка ефективності підходу виконувалась в середовищі Matlab за допомогою стандартних і розроблених програм.

При цьому використовувалась навчальна вибірка з 120 фішингових і безпечних URL-адрес (див. розділ 2.1).

Як критерій глобальної оптимізації використовувався критерій незміщенності (мінімуму зсуву), який не чутливий до рівня шуму у вхідних даних і при збільшенні завад їх мінімум не зміщується в область простіших моделей [51]:

$$C_{см} = \frac{\left\| \hat{Y}_A[m+n] - \hat{Y}_B[m+n] \right\|}{\left\| Y^*[m+n] \right\|}, \tag{2.18}$$

де $\hat{Y}_A[m+n]$ і $\hat{Y}_B[m+n]$ – виходи моделей, які навчені на вибірках A і B, відповідно.

Метод ПВП мав адаптивний крок пошуку і повний пошук навколо поточної ітерації. Кількість ітерацій для ПВП обмежувалося на рівні 100.

При оптимізації варіювалися наступні характеристики класифікаторів:

- архітектура НМ – перцептрон, НМ Хопфілда і НМ Хемінга;
- кількість нейронів в прихованому шарі (для перцептрону);
- функція активації прихованого шару (для перцептрону).

Результати глобальної оптимізації методом прямого випадкового пошуку для знаходження типу і параметрів нейромережевого класифікатора наведені на рис. 2.3.

В результаті моделювання (див. рис. 2.3) встановлено, що ПВП виходить в область оптимальних рішень після 15 ітерацій. Його швидкодія – 8 с на ітерацію.

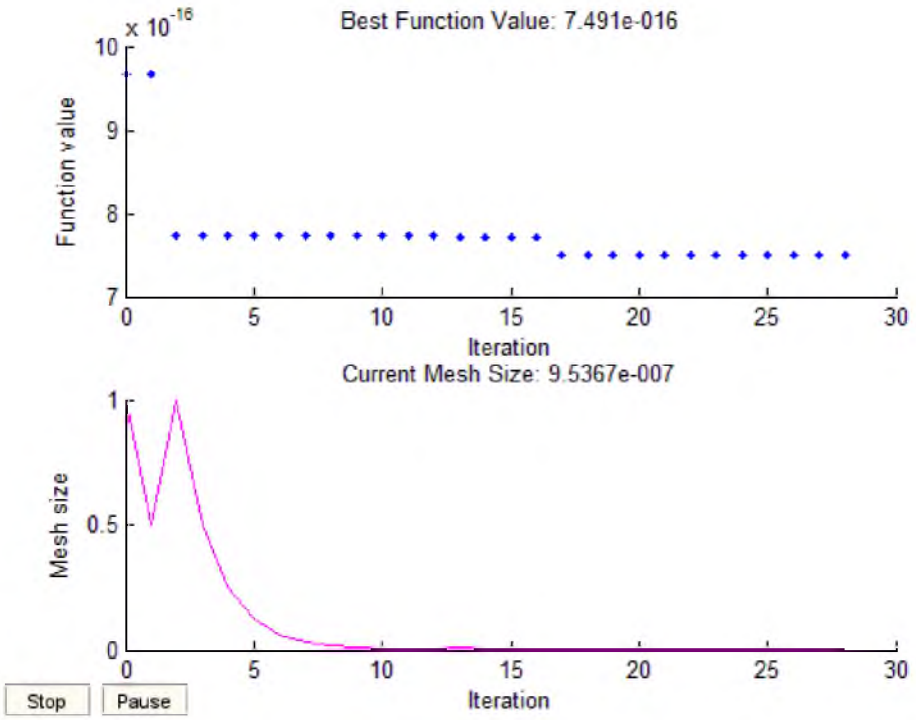


Рисунок 2.3 – Результат глобальної оптимізації для знаходження типу і параметрів нейромережевого класифікатора за допомогою ПВП

Встановлено, що мінімуму критерію зсуву (2.18) відповідає класифікатор на основі перцептрону із 43 нейронами у прихованому шарі, функцією активації прихованого шару – логістичною сигмоїдальною.

Значення метрик оцінки детектування фішингових URL-адрес для цього класифікатора наведено в табл. 2.4. Тут встановлено, що усі фішингові URL-адреси, які були класифіковані як безпечні (FN=4) мали захищене з'єднання з дійсним сертифікатом (https).

Таблиця 2.4 – Значення метрик оцінки детектування URL-адрес при використанні нейромережевого класифікатора, що обраний методом ПВП

TP	TN	FP	FN	TPR	TNR	FPR	FNR	PPV	NPV	F-measure
26	30	0	4	0,87	1	0	0,13	0,93	0,88	0,89

Отже, в розділі було запропоновано підхід до детектування фішингових URL-адрес із використанням нейромережевих класифікаторів та методу прямого випадкового пошуку та оцінено його ефективність. Запропоновані рішення можуть бути використані як частина комплексної системи розпізнавання фішингових сайтів в браузері

Встановлено, що усі фішингові URL-адреси, які помилково були класифіковані як безпечні, мали наявність захищеного з'єднання з дійсним сертифікатом. Таким чином, подальші дослідження мають бути спрямовані на дослідження додаткових інформативних атрибутів (ознак), які могли б дозволити більш якісно розділяти фішингові і безпечні URL-адреси.

2.4 Висновки

В розділі було досліджено використання штучних нейронних мереж (перцептрону, НМ Хопфілда та НМ Хемінга) для виявлення фішингових веб-сайтів в залежності від розмірів навчальної вибірки (60, 80, 100, 120 та 140 фішингових і безпечних URL-адрес).

Набір експериментальних даних включав в себе наступну інформацію:

- регістратор доменного імені;
- час життя домену, в роках;
- геолокація хостинг-сервера;
- наявність захищеного з'єднання з дійсним сертифікатом (https).

Дослідження проводилось за допомогою стандартних та розроблених програм в середовищі Matlab.

Встановлено (див. табл. 2.1-2.3 та рис. 2.1), що найкращі значення показав класифікатор на основі перцептрону. Результати класифікаторів на основі НМ Хопфілда і Хемінга показали приблизно однакові результати.

Також встановлено, що для всіх нейромережевих класифікаторів результати детектування є кращими із збільшенням розміру навчальної вибірки від 60 до 120 URL-адрес. Результати ж детектування для розмірів вибірки в 120

і 140 URL-адрес виявились однаковими, тому немає необхідності у використанні навчальних вибірок більшого розміру, оскільки це збільшує час класифікації.

Запропоновано підхід до детектування фішингових URL-адрес із використанням нейромережевих класифікаторів і методу прямого випадкового пошуку, який полягає у використанні глобальної оптимізації (методу ПВП) для вибір типу та параметрів нейромережевого класифікатора (на основі перцептрону, НМ Хопфілда та НМ Хемінга). Запропоновані рішення можуть бути використані як частина комплексної системи розпізнавання фішингових сайтів в браузері.

Як критерій глобальної оптимізації використовувався критерій незміщенності (мінімуму зсуву). Встановлено, що його мінімуму відповідає класифікатор на основі перцептрону із 43 нейронами у прихованому шарі, функцією активації прихованого шару – логістичною сигмоїдальною.

Встановлено, що усі фішингові URL-адреси, які помилково були класифіковані як безпечні, мали наявність захищеного з'єднання з дійсним сертифікатом. Таким чином, подальші дослідження мають бути спрямовані на дослідження додаткових інформативних атрибутів (ознак), які могли б дозволити більш якісно розділяти фішингові і безпечні URL-адреси.

3 ЕКОНОМІЧНИЙ РОЗДІЛ

Запропоновано підхід до детектування фішингових веб-сайтів із використанням нейромережевих класифікаторів та методу прямого випадкового пошуку, який може бути використаний як частина комплексної системи розпізнавання фішингових сайтів в браузері.

Метою даного розділу є обґрунтування економічної доцільності детектування фішингових веб-сайтів із використанням нейромережевих класифікаторів та методу прямого випадкового пошуку.

Для досягнення цієї мети необхідно визначити:

– величини капітальних витрат на розробку запропонованого підходу та експлуатаційних витрат на його реалізацію;

– економічний ефект від впровадження запропонованого підходу;

– показники економічної ефективності, зокрема коефіцієнт повернення інвестицій та період окупності.

3.1 Розрахунок капітальних (фіксованих) витрат

Фіксованими витрати називаються тому, що робляться, як правило, один раз, на початкових етапах створення інформаційної системи (ІС). До фіксованих належать наступні витрати: вартість розробки і впровадження проекту; залучення зовнішніх консультантів; первинні закупівлі основного ПЗ; первинні закупівлі додаткового ПЗ; первинні закупівлі апаратного забезпечення тощо.

Витрати на детектування фішингових веб-сайтів із використанням нейромережевих класифікаторів та методу прямого випадкового пошуку визначаються, виходячи з трудомісткості кожної операції, та належать до капітальних витрат.

Визначення трудомісткості розробки підходу до детектування фішингових веб-сайтів із використанням нейромережевих класифікаторів та методу прямого випадкового пошуку

Трудомісткість розробки системи захисту інформації на підприємстві визначається тривалістю кожної робочої операції, до яких належать наступні:

– тривалість складання технічного завдання для детектування фішингових веб-сайтів із використанням нейромережевих класифікаторів та методу прямого випадкового пошуку, $t_{тз}$=16 годин;

– тривалість аналізу можливих загроз безпеки інформації, $t_{аз}$=40 годин;

– тривалість імітаційного моделювання для дослідження використання штучних нейронних мереж (перцептрону, НМ Хопфілда та НМ Хемінга) для виявлення фішингових веб-сайтів в залежності від розмірів навчальної вибірки, $t_{нд}$=39 годин;

– тривалість вивчення технічного завдання, літературних джерел за темою тощо, $t_{вз}$=32 години;

– тривалість розробки та оцінка ефективності підходу до детектування фішингових веб-сайтів із використанням нейромережевих класифікаторів та методу прямого випадкового пошуку, $t_{озб}$=86 годин;

td – тривалість документального оформлення запропонованого підходу до детектування фішингових веб-сайтів із використанням нейромережевих класифікаторів та методу прямого випадкового пошуку, $t_д$=15 годин.

Отже,

$$t = t_{тз}+ t_{аз}+ t_{нд}+ t_{вз}+ t_{озб}+ t_д=16+40+39+32+86+15 = 228 \text{ годин.}$$

Розрахунок витрат на розробку підходу до детектування фішингових веб-сайтів із використанням нейромережевих класифікаторів та методу прямого випадкового пошуку

Витрати на розробку системи захисту інформації на підприємстві $К_{рп}$ складаються з витрат на заробітну плату спеціаліста з інформаційної безпеки $З_{зп}$

і вартості витрат машинного часу, що необхідний для розробки запропонованого підходу $З_{мч}$.

$$К_{рп} = З_{зп} + З_{мч}\ .$$

$$К_{рп} = З_{зп} + З_{мч} = 47424 + 1162{,}8 = 48586{,}8 \text{ грн.}$$

$$З_{зп} = t\ З_{пр} = 228*208 = 47424 \text{ грн.}$$

де t – загальна тривалість розробки політики безпеки, годин;

$Зіб$ – средньогодинна заробітна плата спеціаліста з інформаційної безпеки з нарахуваннями, грн/годину.

Вартість машинного часу для розробки політики безпеки інформації на ПК визначається за формулою:

$$З_{мч} = t*С_{мч} = 228*5{,}1 = 1162{,}8 \text{ грн.}$$

де $t_д$ – трудомісткість підготовки документації на ПК, годин;

$С_{мч}$ – вартість 1 години машинного часу ПК, грн./година.

Вартість 1 години машинного часу ПК визначається за формулою:

$$С_{мч} = 0{,}7 \cdot 3 \cdot 1{,}68 + \frac{5200 \cdot 0{,}4}{1920} + \frac{3100 \cdot 0{,}3}{1920} = 5{,}1 \text{ грн.}$$

Оцінка ефективності підходу до детектування фішингових веб-сайтів із використанням нейромережевих класифікаторів та методу прямого випадкового пошуку виконувалась за допомогою стандартних та розроблених програм в середовищі Matlab. При цьому використовувалась безкоштовна навчальна версія пакета прикладних програм Matlab&Simulink, тому додаткові капітальні витрати не виникають.

Витрати на налагодження системи інформаційної безпеки заплановані величиною 2000 грн.

Таким чином, капітальні (фіксовані) витрати на розробку засобів підвищення рівня інформаційної безпеки складуть:

$$К = К_{рп} + К_{зпз} + К_{пз} + К_{аз} + К_{навч} + К_{н} = 48586{,}8 + 2000 = 50586{,}8 \text{ грн.}$$

де $К_{рп}$ – вартість розробки заходів із забезпечення інформаційної безпеки та залучення для цього зовнішніх консультантів, тис. грн;

$К_{зпз}$ – вартість закупівель ліцензійного основного й додаткового ПЗ, тис. грн;

$К_{пз}$ – вартість створення основного й додаткового програмного забезпечення, тис. грн;

$К_{аз}$ – вартість закупівлі апаратного забезпечення та допоміжних матеріалів, тис. грн;

$К_{навч}$ – витрати на навчання технічних фахівців і обслуговуючого персоналу,

$К_{н}$ – витрати на встановлення обладнання та налагодження системи інформаційної безпеки.

3.2 Розрахунок поточних витрат

Річні поточні витрати на функціонування системи інформаційної безпеки складають:

$$С = С_{в} + С_{к} + С_{ак}, \text{ грн.}$$

де $С_{в}$ - вартість відновлення й модернізації системи ($С_{в} = 0$);

$С_{к}$ - витрати на керування системою в цілому;

$С_{ак}$ - витрати, викликані активністю користувачів системи інформаційної безпеки ($С_{ак} = 0$ грн.).

Витрати на відновлення й модернізації системи інформаційної безпеки заплановані в розмірі 4000 грн.

Витрати на керування системою інформаційної безпеки ($С_{к}$) складають:

$$С_{к} = С_{н} + С_{а} + С_{з} + С_{ел} + С_{о} + С_{тос}, \text{ грн.}$$

Витрати на навчання адміністративного персоналу та кінцевих споживачів складуть 8000 грн.

Річний фонд заробітної плати інженерно-технічного персоналу, що обслуговує систему інформаційної безпеки ($С_{з}$), складає:

$$С_{з} = З_{осн} + З_{дод}, \text{ грн.}$$

Основна заробітна плата визначається, виходячи з місячного посадового окладу, а додаткова заробітна плата – в розмірі 8-10% від основної заробітної плати.

Основна заробітна плата одного спеціаліста з інформаційної безпеки на місяць складає 17000 грн. Додаткова заробітна плата – 8% від основної заробітної плати. Виконання роботи щодо впровадження системи захисту інформації на підприємстві потребує залучення спеціаліста інформаційної безпеки на 0,2 ставки.

Отже,

$$С_з = (17000*12 + 17000*12*0,08)*0,2= 44064 \text{ грн.}$$

Ставка ЄСВ для всіх категорій платників з 01.01.2021 р. складає 22%.

$$С_{єв} = 44064*0,22 = 9694,08 \text{ грн.}$$

Вартість електроенергії, що споживається апаратурою системою інформаційної безпеки протягом року ($С_{ел}$), визначається за формулою:

$$С_{ел} = Р \cdot F_р \cdot Ц_е, \text{ грн.},$$

де Р – встановлена потужність апаратури інформаційної безпеки, (Р=1,2 кВт);

$F_р$ – річний фонд робочого часу системи інформаційної безпеки ($F_р$ = 1920 год.);

$Ц_е$ – тариф на електроенергію, ($Ц_е$ = 1,68 грн./кВт за годину).

Вартість електроенергії, що споживається апаратурою системою інформаційної безпеки протягом року:

$$С_{ел} = 1,2*1920*1,68 = 3870,72 \text{ грн.}$$

Витрати на технічне й організаційне адміністрування та сервіс системи інформаційної безпеки визначаються у відсотках від вартості капітальних витрат – 3%

$$С_{тос} = 50586,8*0,03=1517,6 \text{ грн.}$$

Витрати на керування системою інформаційної безпеки ($С_к$) визначаються:

$$С_к = 8000+44064+9694,08+3870,72+1517,6 = 67146,4 \text{ грн.}$$

Таким чином, річні поточні витрати на функціонування системи інформаційної безпеки складають:

$$С = 67146,4 \text{ грн.}$$

3.3 Оцінка можливого збитку від атаки на вузол або сегмент мережі

Для розрахунку вартості такого збитку можна застосувати наступну спрощену модель оцінки для умовного підприємства.

Необхідні *вихідні дані* для розрахунку:

$t_п$ – час простою вузла або сегмента корпоративної мережі внаслідок атаки, 1 години;

$t_в$ – час відновлення після атаки персоналом, що обслуговує корпоративну мережу, 2 години;

$t_{ви}$ – час повторного введення загубленої інформації співробітниками атакованого вузла або сегмента корпоративної мережі, 4 годин;

$З_о$ – заробітна плата обслуговуючого персоналу (адміністраторів та ін.), 18000 грн./міс.;

$З_с$ – заробітна плата співробітників атакованого вузла або сегмента корпоративної мережі, 16000 грн./міс.;

$Ч_о$ – чисельність обслуговуючого персоналу (адміністраторів та ін.), 1 особи;

$Ч_с$ – чисельність співробітників атакованого вузла або сегмента корпоративної мережі, 5 осіб.;

О – обсяг прибутку атакованого вузла або сегмента корпоративної мережі, 300 тис. грн. у рік;

$П_{зч}$ – вартість заміни встаткування або запасних частин, 2000 грн;

I – число атакованих сегментів корпоративної мережі, 1;

N – середнє число атак на рік, 24.

Упущена вигода від простою атакованого сегмента корпоративної мережі становить:

$$U = П_п + П_в + V,$$

де $П_п$ – оплачувані втрати робочого часу та простої співробітників атакованого вузла або сегмента корпоративної мережі, грн;

$П_в$ – вартість відновлення працездатності вузла або сегмента корпоративної мережі (переустановлення системи, зміна конфігурації та ін.), грн;

V – втрати від зниження обсягу продажів за час простою атакованого вузла або сегмента корпоративної мережі, грн.

Втрати від зниження продуктивності співробітників атакованого вузла або сегмента корпоративної мережі являють собою втрати їхньої заробітної плати (оплата непродуктивної праці) за час простою внаслідок атаки:

$$П_п = \frac{\sum З_с}{F} t_п = \frac{16000 * 5}{176} * 1 = 454{,}55 \text{ грн},$$

де F – місячний фонд робочого часу (при 40-а годинному робочому тижні становить 176 ч).

Витрати на відновлення працездатності вузла або сегмента корпоративної мережі включають кілька складових:

$$П_в = П_{ви} + П_{пв} + П_{зч},$$

де $П_{ви}$ – витрати на повторне уведення інформації, грн.;

$П_{пв}$ – витрати на відновлення вузла або сегмента корпоративної мережі, грн;

$П_{зч}$ – вартість заміни устаткування або запасних частин, грн.

Витрати на повторне введення інформації $П_{ви}$ розраховуються виходячи з розміру заробітної плати співробітників атакованого вузла або сегмента корпоративної мережі $З_с$, які зайняті повторним введенням втраченої інформації, з урахуванням необхідного для цього часу $t_{ви}$.

Отже:

$$П_{ви} = \frac{\sum З_с}{F} t_{ви} = \frac{16000 * 5}{176} * 4 = 1818{,}18 \text{ грн.}$$

Витрати на відновлення сегмента корпоративної мережі $П_{пв}$ визначаються часом відновлення після атаки $t_в$ і розміром середньогодинної заробітної плати обслуговуючого персоналу (адміністраторів):

$$П_{пв} = \frac{\sum З_о}{F} t_в = \frac{18000 * 1}{176} * 2 = 204{,}55 \text{ грн.}$$

Тоді витрати на відновлення працездатності вузла або сегмента корпоративної мережі складуть:

$$П_в = 1818{,}18+204{,}55+2000 = 4022{,}73 \text{ грн.}$$

Втрати від зниження очікуваного обсягу прибутків за час простою атакованого вузла або сегмента корпоративної мережі визначаються виходячи із средньогодинного обсягу прибутку і сумарного часу простою сегмента корпоративної мережі:

$$V = \frac{O}{Fг} \cdot (t_П + t_В + t_{ВИ})$$

$$V = \frac{300000}{2080} \cdot (1 + 2 + 4) = 1009{,}62 \text{ грн.}$$

де $F_г$ – річний фонд часу роботи філії (52 робочих тижні, 5-ти денний робочий тиждень, 8-ми годинний робочий день) становить близько 2080 год.

U = 454,55+4022,73+1009,62=5486,9 грн.

Таким чином, загальний збиток від атаки на сегмент корпоративної мережі організації складе:

$$B = \sum_i \sum_n U = \sum_1 \sum_{24} 5486{,}9 = 131685{,}6 \text{ грн.}$$

3.3.1 Загальний ефект від впровадження системи інформаційної безпеки

Загальний ефект від впровадження системи інформаційної безпеки визначається з урахуванням ризиків порушення інформаційної:

$$E = B \cdot R - C \text{ грн.},$$

де B – загальний збиток від атаки у разі перехоплення інформації, тис. грн.;

R – вірогідність успішної реалізації атаки на сегмент мережі, частки одиниці (60%);

С – щорічні витрати на експлуатацію системи інформаційної безпеки.

Загальний ефект від впровадження системи інформаційної безпеки визначається з урахуванням ризиків порушення інформаційної безпеки:

E = 131685,6*0,6 – 67146,4 = 11864,96 грн.

3.4 Визначення та аналіз показників економічної ефективності системи інформаційної безпеки

За методикою сукупної вартості володіння (ТСО) визначають такі показники економічної ефективності системи інформаційної безпеки як коефіцієнт повернення інвестицій (ROSI) та термін окупності капітальних інвестицій (T_o).

Коефіцієнт повернення інвестицій ROSI показує, скільки гривень додаткового прибутку приносить одна гривня капітальних інвестицій на впровадження системи інформаційної безпеки:

$$ROSI = \frac{E}{K}, \text{ частки одиниці,}$$

де E – загальний ефект від впровадження системи інформаційної безпеки грн.;

К – капітальні інвестиції за варіантами, що забезпечили цей ефект, грн.

Коефіцієнт повернення інвестицій ROSI:

ROSI=11864,96 /50586,8 =0,23, частки одиниці,

Проект визнається економічно доцільним, якщо розрахункове значення коефіцієнта повернення інвестицій перевищує величину річної депозитної ставки з урахуванням інфляції:

$$ROSI > (N_{деп} - N_{інф})/100),$$

де $N_{деп}$ – річна депозитна ставка, (6%);

$N_{інф}$ – річний рівень інфляції, (5%).

Розрахункове значення коефіцієнта повернення інвестицій:

0,23 > (6 – 5)/100 = 0,23 > 0,01.

Термін окупності капітальних інвестицій T_o показує, за скільки років капітальні інвестиції окупляться за рахунок загального ефекту від впровадження системи інформаційної безпеки:

$$T_o = \frac{К}{Е} = \frac{1}{ROSI}$$

T_0= 1 / 0,23 = 4,35 років.

3.5 Висновок

Розробка підходу до детектування фішингових веб-сайтів із використанням нейромережевих класифікаторів та методу прямого випадкового пошуку є економічно доцільною відповідно до отриманих значень показників економічної ефективності, зокрема: коефіцієнт повернення інвестицій ROSI складає 0,23 грн./грн. (тобто на 1 гривню капітальних витрат припадає 0,23 грн. економічного ефекту). При цьому величина економічного ефекту складає 11864,96 грн. Отримане значення коефіцієнта ROSI перевищує дохідність альтернативного вкладення коштів. Термін окупності при цьому складатиме 4,35 років. Капітальні витрати визначено обсягом 50586,8 грн.

ВИСНОВКИ

Основні результати кваліфікаційної роботи полягають у наступному:

1. Встановлено, що наразі фішинг є однією з найбільших загроз в Інтернеті. З початку повномасштабного російського вторгнення українці все частіше стають цілями кіберзлочинців. Фішингові атаки часто є дуже масштабними подіями, які спрямовані на тисячі споживачів або більше, в надії, що певний відсоток з них захоче відповісти, все це дуже сильно впливає на безпеку в Інтернеті.

2. Встановлено, що наразі великі компанії користуються так званою системою «репортів», яка передбачає, що користувач, який потрапив на підозрілий сайт, відправляє скаргу на нього і тоді команда спеціалістів перевіряє сторінку. Якщо сайт виявився фішинговим, то він попадає до переліку підозрілих. Тепер, перед тим як зайти на такий сайт, браузер попередить користувача про те, що сайт є фішинговим. В комплексі з системою «репортів» використовують різні методи автоматичного виявлення фішингових сайтів, які базуються на інтелектуальних методах. Такі методи шукають фішингові сайти, а потім відправляють репорти, які будуть перевіряти спеціалісти.

3. Встановлено, що наразі для задач класифікації, до яких відноситься й детектування фішингових URL-посилань все частіше використовують штучні нейронні мережі, які є універсальними ефективними апроксиматорами.

4. В результаті дослідження використання штучних нейронних мереж (перцептрону, НМ Хопфілда та НМ Хемінга) для виявлення фішингових веб-сайтів в залежності від розмірів навчальної вибірки встановлено, що найкращі значення показав класифікатор на основі перцептрону. Результати класифікаторів на основі НМ Хопфілда і Хемінга показали приблизно однакові результати.

Також встановлено, що для всіх нейромережевих класифікаторів результати детектування є кращими із збільшенням розміру навчальної вибірки

від 60 до 120 URL-адрес. Результати ж детектування для розмірів вибірки в 120 і 140 URL-адрес виявились однаковими, тому немає необхідності у використанні навчальних вибірок більшого розміру, оскільки це збільшує час класифікації.

5. Запропоновано підхід до детектування фішингових URL-адрес із використанням нейромережевих класифікаторів і методу прямого випадкового пошуку, який полягає у використанні глобальної оптимізації (методу ПВП) для вибір типу та параметрів нейромережевого класифікатора (на основі перцептрону, НМ Хопфілда та НМ Хемінга). Запропоновані рішення можуть бути використані як частина комплексної системи розпізнавання фішингових сайтів в браузері.

6. В результаті оцінки ефективності запропонованого підходу встановлено, що мінімуму критерію глобальної оптимізації відповідає класифікатор на основі перцептрону із 43 нейронами у прихованому шарі, функцією активації прихованого шару – логістичною сигмоїдальною.

Встановлено, що усі фішингові URL-адреси, які помилково були класифіковані як безпечні, мали наявність захищеного з'єднання з дійсним сертифікатом. Таким чином, подальші дослідження мають бути спрямовані на дослідження додаткових інформативних атрибутів (ознак), які могли б дозволити більш якісно розділяти фішингові і безпечні URL-адреси.

ПЕРЕЛІК ПОСИЛАНЬ

1. Інформаційна та кібербезпека: соціотехнічний аспект: підручник / [В.Л. Бурячок, В.Б. Толубко, В.О. Хорошко, С.В. Толюпа]; за заг. ред. д-ра техн. наук, професора В.Б. Толубка. – К.: ДУТ, 2015. – 288 с.

2. Бурячок В.Л. До питання організації та проведення розвідки у кібернетичному просторі / В.Л. Бурячок, Г.М. Гулак, В.О. Хорошко // Наука і оборона. – 2011. – № 2. – С. 19-23.

3. Бурячок В.Л. Поняття кібервійни та розвідки інформаційно-телекомунікаційних систем у контексті захисту держави від стороннього кібернетичного впливу / В.Л. Бурячок, О.А. Ільяшов, Г.М. Гулак // Збірник матеріалів круглого столу «Актуальні питання підготовки фахівців із розслідування кіберзлочинів», 25.11.2011. – К.: Наук.-вид. відділ НА СБ України, 2011. – С. 27-32.

4. Шахрайство за допомогою фармінгу: перенаправлення на фальшиві сайти [Електронний ресурс]. – Режим доступу: http://www.microsoft.com/rus/athome/security/privacy/pharming.mspx.

5. Бабок В.П. Інформаційна безпека та сучасні мережені технології: англ.-укр.-рос. словник термінів / В. П. Бабок, В. Г. Корченко. – К.: НАУ, 2003. – 670 с.

6. Mitnik Kevin U. The Art of Deception / Kevin U. Mitnik, William L. Simon, Steve Wozniak. – Wiley, 2002. – 304 с.

7. Корченко А.Г. Несанкционированный доступ к компьютерным системам и методы защиты: учеб. пособие / А.Г. Корченко. – К.: КМУГА, 1998. – 116 с.

8. Cialdini Robert B. The Science of Persuasion / Robert B. Cialdini // II Scientific American Magazine. – 2001. – № 2. – P.76-81.

9. Корченко О.Г. Класифікація методів соціального інжинірингу / О.Г. Корченко, Є.В. Паціра, Д.А. Пуха // Захист інформації. – К.: НАУ. – 2007. – № 4. – С. 37-45.

10. How to Protect Insiders from Social Engineering Threats [Електронний ресурс]. – Режим доступу: http://www.microsoft.com/downloads/details.aspx?FamilyID=05033e55-aa96-4d49-8f57- c47664107938&DisplayLang=en.

11. Jeeva S.C. Intelligent phishing URL detection using association rule mining / S.C. Jeeva, E.B. Rajsingh // Human-centric Computing and Information Sciences. – Num. 6 (1). – 2016.

12. Tally G., R. Thomas and Tom Van Vleck. Anti-Phishing: Best Practices for Institutions and Consumers. [Електронний ресурс]. – Режим доступу: https: //www.semanticscholar.org/paper /Anti-Phishing%3A-Best-Practices-for-Institutions-and-Tally Thomas/3e5ae0fb6cba 7c975bb2ca2da50b659e98493441.

13. APWG. Phishing activity trends reports. [Електронний ресурс]. – Режим доступу: https://apwg.org/trendsreports/.

14. Google Support. Як захиститися від фішингових атак і повідомляти про них. [Електронний ресурс]. – Режим доступу: https://support.google.com/websearch/answer/106318?hl=uk.

15. Microsoft Support. Захист від фішингу. [Електронний ресурс]. – Режим доступу: https://support.microsoft.com/uk-ua/windows/захист-від-фішингу-0c7ea947-ba98-3bd9-7184-430e1f860a44.

16. ESET. Фішинг. [Електронний ресурс]. – Режим доступу: https://www.eset.com/ua/support/information/entsiklopediya-ugroz/fishing/.

17. Звіт компанії NSS Labs про тестування захисту від фішингу за другий квартал 2020 р. [Електронний ресурс]. – Режим доступу: https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RWAIQX.

18. Фішинг: що це таке і як себе убезпечити? [Електронний ресурс]. – Режим доступу: https://zillya.ua/index.php?q=fishing-shcho-tse-take-i-yak-sebe-ubezpechiti.

19. 2021 рік: Ретроспектива вірусної активності. [Електронний ресурс]. – Режим доступу: https://zillya.ua/index.php?q=2021-rik-retrospektiva-virusno-aktivnosti.

20. Корнієнко В.І. Інтелектуальне моделювання нелінійних динамічних процесів в керуванні, кібербезпеці, телекомунікаціях: підручник / В.І. Корнієнко, О.Ю. Гусєв, О.В. Герасіна. – Міністерство освіти і науки України, Національний технічний університет «Дніпровська політехніка». – Дніпро, НТУ «ДП», 2020. – 531 с.

21. Глибовець М. М. Штучний інтелект : підручник для студ. вищих навч.закладів / М. М. Глибовець, О.В. Олецький. – К. : КМ Академія, 2002. – 369 с.

22. Ланде Д.В., Субач І.Ю., Бояринова Ю.Є. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки: навчальний посібник. — К.: ІСЗЗІ КПІ ім. Ігоря Сікорського», 2018. — 297 с.

23. Зайченко, Ю.П. Нечіткі моделі і методи в інтелектуальних системах. - К: Слово, 2008. - 344 с.

24. Гулак Г.М. Методологія захисту інформації. Аспекти кібербезпеки: підручник. – К.: Видавництво НА СБ України, 2020. – 256 с.

25. Nelles O. Nonlinear System Identification: From Classical Approaches to Neural and Fuzzy Models / O. Nelles. – Berlin: Springer, 2001. – 785 pp.

26. Cheat Sheets for AI, Neural Networks, Machine Learning, Deep Learning & Big Data / Stefan Kojouharov [Електронний ресурс]. – Режим доступу: https://becominghuman.ai/cheat-sheets-for-ai-neural-networks-machine-learning-deep-learning-big-data-678c51b4b463.

27. Глибовець М. М. Штучний інтелект : підручник для студ. вищих навч.закладів / М. М. Глибовець, О.В. Олецький. – К. : КМ Академія, 2002. – 369 с.

28. Іванченко Г. Ф. Системи штучного інтелекту : навч. посібник /Г. Ф. Іванченко. – К., 2011. – 382 с.

29. Субботін С. О. Нейронні мережі : теорія та практика: навч. посіб./ С. О. Субботін. – Житомир : Вид. О. О. Євенок, 2020. – 184 с.

30. Ротштейн А. П. Интеллектуальные технологии идентификации: нечеткие множества, генетические алгоритмы, нейронные сети / А.П. Ротштейн. – Винница: «УНІВЕРСУМ-Вінниця», 1999. – 320 с.

31. Rao C. Handbook of Statistics: Machine Learning: Theory and Applications, // C. Rao, V. Govindaraju. – Oxford: North Holland & IFIP, 2013. – 552 с.

32. Кравченко С.М. Методи класифікації машинного навчання з використанням бібліотеки Scikit-Learn / С.М. Кравченко, Є.О. Гришкун, О.В. Власенко // Вчені записки ТНУ імені В.І. Вернадського. Серія: технічні науки. – 2020. – № 31 (70). – С. 121-125.

33. Даник Ю.Г. Основи кібербезпеки та кібероборони: підручник / Ю.Г Даник, П.П. Воробієнко, В.М. Чернега. – О.: ОНАЗ ім. О.С. Попова, 2018. – 228 с.

34. Оцінювання захищеності інформації в комп'ютерних системах за соціоінженерним підходом / В. Мохор, О. Цуркан, В. Цуркан, Р. Герасимов // Selected Papers of the XVII International Scientific and Practical Conference "Information Technologies and Security". – Kyiv, 2017. – P. 1-6.

35. Мохор В. В. Наставления по кибербезопасности (ISO/IEC 27032:2013) / В. В. Мохор, А. М. Богданов, А. С. Килевой. – К. : ООО "Три-К". – 2013. – 129 с.

36. Information technology. Security techniques. Information security management systems. Requirements : ISO/IEC 27001:2013. – Second edition 2013-10-01. – Geneva, 2013. – P. 23.

37. Загальні положення щодо захисту інформації в комп'ютерних системах від несанкціонованого доступу : НД ТЗІ 1.1-002-99. – Чинний від 1999-04-28. – К. : ДСТСЗІ СБ України, 1999. – 15 с.

38. Про захист інформації в інформаційно-телекомунікаційних системах : Закон України від 05.07.1994 № 80/94-ВР // Верховна Рада України. [Електронний ресурс]. – Режим доступу: http://zakon2.rada.gov.ua/laws/show/80/94-%D0%B2%D1%80.

39. Остроухов В. В. Інформаційно-психологічна безпека особи: соціально-правові аспекти / В. В. Остроухов // Інформаційна безпека людини, суспільства, держави. – 2010. – № 1(3). – С. 38-41.

40. Жарков Я. М. Інформаційно-психологічне протиборство (еволюція та сучасність): Монографія / Я. М. Жарков, В. М. Петрик, М. М. Присяжнюк та ін. – К.: ПАТ "Віпол", 2013. – 248 с.

41. Krombholz K. Advanced social engineering attacks / K. Krombholz, H. Hobel, M. Huber, E. Weippl // Journal of information security and applications. – 2014. – P. 1-10. – http://dx.doi.org/10.1016/j.jisa.2014.09.005.

42. Winterfeld S. The Basics of Cyber Warfare: Understanding the Fundamentals of Cyber Warfare in Theory and Practice / S. Winterfeld, J. Andress. – Waltham : Elsevier, 2013. – 152 p.

43. Mouton F. Necessity for ethics in social engineering research / Francois Mouton, Mercia M. Malan c, Kai K. Kimppa d, H.S. Venter // ScienceDirect. – 2015. – http://dx.doi.org/10.1016/j.cose.2015.09.001.

44. Mouton F. Social engineering attack examples, templates and scenarios / F. Mouton, L. Leenen, H. Venter // Computers & Security. – 2016. – http://dx.doi.org/doi: 10.1016/j.cose.2016.03.004.

45. Junger M. Priming and warnings are not effective to prevent social engineering attacks / M. Junger, L. Montoya, F.- J. Overink // Computers in Human Behavior. – 2017. – http://dx.doi.org/10.1016/j.chb.2016.09.012.

46. Понад 100 днів війни: як Україна протистояла атакам на кіберфронті. [Електронний ресурс]. – Режим доступу: https://eset.ua/ua/news/view/978/boleye-100-dney-voyny-kak-ukraina-protivostoyala-atakam-na-kiberfronte.

47. Рейтинг Інтернет-загроз: вплив війни в Україні та найактивніші шкідливі програми. [Електронний ресурс]. – Режим доступу: https://eset.ua/ua/news/view/977/reyting-internet-ugroz-kak-izmenilas-aktivnost-khakerov-vo-vremya-voyny.

48. Рейтинг Інтернет-загроз: Україна у п’ятірці цілей програм-вимагачів. [Електронний ресурс]. – Режим доступу: https://eset.ua/ua/news/view/990/reyting-internet-ugroz-ukraina-v-pyaterke-celey-programm-vymogateley.

49. Network Traffic Classification Using Correlation Information / Jan Zhang, Yang Xiang, Yu Wang, Wanlei Zhou, Yong Xiang, Yong Guan // IEEE Transactions on parallel and distributed systems. – 2013. – Vol. 24, issue 1. – P. 104-117.

50. On the Local Convergence of Pattern Search / Elizabeth D. Dolan, Robert Michael Lewis, Virginia Torczon // SIAM Journal on Optimization. – 2003. – Vol. 14, Iss. 2. – P. 234-251.

51. Ivakhnenko A.G. Inductive learning algorithms for complex systems modeling / A.G. Ivakhnenko, H.R. Madala – London, Tokyo: CRC Press, 1994. – 384 p.

52. Методичні рекомендації до виконання кваліфікаційних робіт бакалаврів спеціальності 125 Кібербезпека/ Упоряд.: О.В. Герасіна, Д.С. Тимофєєв, О.В. Кручінін, Ю.А. Мілінчук – Дніпро: НТУ «ДП», 2020. – 47 с.

ДОДАТОК А. Відомість матеріалів кваліфікаційної роботи

№	**Формат**	**Найменування**	**Кількість листів**	**Примітки**
Документація				
1	А4	Реферат	2	
2	А4	Список умовних скорочень	1	
3	А4	Зміст	2	
4	А4	Вступ	3	
5	А4	Стан питання. Постановка задачі	47	
6	А4	Спеціальна частина	13	
7	А4	Економічний розділ	8	
8	А4	Висновки	2	
9	А4	Перелік посилань	6	
10	А4	Додаток А	1	
11	А4	Додаток Б	1	
12	А4	Додаток В	1	
13	А4	Додаток Г	1	

ДОДАТОК Б. Перелік документів на оптичному носії

1 Презентація Ангеловський.ppt

2 Диплом Ангеловський.doc

ДОДАТОК В. Відгук керівника економічного розділу

Керівник розділу _____ _____

(підпис) (прізвище, ініціали)

ДОДАТОК Г. Відгук керівника кваліфікаційної роботи

В І Д Г У К

на кваліфікаційну роботу студента групи 125м-21-1 Ангеловського М.О. на тему: «Детектування фішингових веб-сайтів за допомогою штучних нейронних мереж»

Пояснювальна записка складається зі вступу, трьох розділів і висновків, розташованих на 92 сторінках.

Мета роботи є актуальною, оскільки вона спрямована на дослідження нейромережевих класифікаторів для детектування фішингових URL-адрес.

При виконанні роботи автор продемонстрував відмінний рівень теоретичних знань і практичних навичок. На основі аналізу принципів фішингових атак і існуючих методів захисту від них, основ побудови штучних нейронних мереж, а також постановки задачі класифікації в ній сформульовані задачі, вирішенню яких присвячений спеціальний розділ. У ньому було досліджено використання нейронних мереж для виявлення фішингових веб-сайтів, запропоновано підхід до детектування фішингових веб-сайтів із використанням нейромережевих класифікаторів і методу прямого випадкового пошуку та оцінено його ефективність.

Практична цінність роботи полягає у тому, що запропоновані рішення можуть бути використані як частина комплексної системи розпізнавання фішингових сайтів в браузері або як система попередження про загрозу фішингу в інших програмах, які працюють з сайтами.

Рівень запозичень у кваліфікаційній роботі відповідає вимогам «Положення про систему виявлення та запобігання плагіату».

В цілому робота задовольняє усім вимогам, а її автор Ангеловський М.О. заслуговує на оцінку « » та присвоєння кваліфікації «Магістр з кібербезпеки» за спеціальністю 125 Кібербезпека.

Керівник роботи,
к.т.н., доцент **О.В. Герасіна**