УДК 004.056.5:37.018.43:355(477) DOI: 10.31673/2412-4338.2022.023345

Шапран О.О.
Державний університет телекомунікацій, м. Київ

МОДЕЛІ ПІДВИЩЕННЯ ЗАХИЩЕНОСТІ ПЕРСОНАЛЬНИХ ДАНИХ КОРИСТУВАЧІВ СИСТЕМИ ДИСТАНЦІЙНОГО НАВЧАННЯ ЗБРОЙНИХ СИЛ УКРАЇНИ

Анотація: *Обґрунтовано проблему удосконалення існуючих та розробки нових моделей підвищення захищеності персональних даних користувачів системи дистанційного навчання ЗС України на основі штучного інтелекту, як часткову проблему загальної проблеми кібербезпеки даної системи. Доведено те, що для забезпечення високого рівня захищеності персональних даних користувачів системи дистанційного навчання в сучасних умовах активно використовують прогресивні організаційні, апаратні та програмні рішення. Надано аналіз закордонного та вітчизняного досвіду розробки та впровадження систем захисту персональних даних користувачів системи дистанційного навчання та зроблено висновок про можливість значного підвищення їх ефективності за рахунок розвитку математичного та програмного забезпечення. Обґрунтовано те, що найбільш актуальним в цьому напрямку є використання моделей та методів штучного інтелекту, а саме, нечіткої логіки та гібридних мереж. Представлено матеріали дослідження щодо розробки методики підвищення захищеності персональних даних користувачів системи дистанційного навчання ЗС України, яка забезпечує ефективне реагування на потік загроз та базується на основі впровадження моделей та методів нечіткої логіки та гібридних мереж; описана модель визначення стану системи захисту персональних даних та метод прогнозування стану системи захисту персональних даних. Представлено результати комп'ютерного моделювання.*

Ключові слова: *захист, модель, персональні данні, дистанційне навчання, нечітка логіка, гібридна мережа.*

Shapran O.O.
State University of Telecommunications, Kyiv

MODELS FOR ENHANCED PROTECTION OF PERSONAL DATA OF USERS OF THE DISTANCE LEARNING SYSTEM OF THE ARMED FORCES OF UKRAINE

Abstract: *The problem of improving existing and developing new models and methods of increasing the security of personal data of users of the distance learning system of the Armed Forces of Ukraine based on artificial intelligence is substantiated. It has been proven that to ensure a high level of security of personal data of users of distance learning systems in modern conditions, progressive organizational, hardware and software solutions are actively used. An analysis of foreign and domestic experience in the development and implementation of personal data protection systems for users of the distance learning system is provided, and a conclusion is made about the possibility of significantly increasing their efficiency due to the development of mathematical and software. It is justified that the most relevant in this direction is the use of models and methods of artificial intelligence, namely, fuzzy logic and hybrid networks. Research materials are presented on the development of a methodology for improving the security of personal data of users of the distance learning system of the Armed Forces of Ukraine, which provides an effective response to the flow of threats and is based on the implementation of models and methods of fuzzy logic and hybrid networks; the model for determining the state of the personal data protection system and the method of forecasting the state of the personal data protection system are described. The results of computer modeling are presented.*

Keywords: *protection, model, personal data, distance learning, fuzzy logic, hybrid network.*

1. Вступ

Тривіальним є висновок про те, що в сучасних умовах технології дистанційного навчання мають розвиватися швидкими темпами. На думку експертів очікується значне збільшення обсягу використання даних технологій у різних сферах суспільства. Це стосується і дистанційного навчання в ЗС України. Особлива увага при використанні таких систем приділяється заходам кібербезпеки, а саме, забезпеченню захищеності персональних даних користувачів систем дистанційного навчання. Даний напрям розвитку відповідає вимогам законодавства Європейського Союзу щодо захисту персональних даних усіх осіб у межах Європейського Союзу та Європейської економічної зони. Це підкреслює актуальність наукових досліджень за темою.

2. Постановка проблеми

Відомо те, що для забезпечення високого рівня захищеності персональних даних користувачів системи дистанційного навчання в сучасних умовах активно використовують прогресивні організаційні, апаратні так програмні рішення. Аналіз закордонного та вітчизняного досвіду розробки та впровадження як систем кібербезпеки в загалі, так і систем захисту персональних даних користувачів системи дистанційного навчання свідчить про можливість значного підвищення їх ефективності за рахунок розвитку математичного та програмного забезпечення [1-9]. Дослідження показало те, що найбільш актуальним в цьому напрямку є використання моделей та методів штучного інтелекту, а саме, нечіткої логіки та гібридних мереж. Незважаючи на стрімкий розвиток теорії штучного інтелекту взагалі, актуальним є завдання удосконалення існуючого та розробки нового математичного та програмного забезпечення системи захисту персональних даних користувачів системи дистанційного навчання на основі нечіткого виведення.

Отже, при вирішенні загальної проблеми кібербезпеки системи дистанційного навчання закладів ЗС України, актуальною є часткова проблема удосконалення існуючих та розробки нових моделей та методів підвищення захищеності персональних даних користувачів системи дистанційного навчання ЗС України на основі штучного інтелекту, вирішенню цього завдання і присвячена дана стаття.

3. Мета і задачі дослідження

Мета дослідження – підвищення ефективності системи управління способами, засобами та процедурами захисту персональних даних користувачів системи дистанційного навчання ЗС України на основі моделей штучного інтелекту.

Задачі дослідження – розробити методику підвищення захищеності персональних даних користувачів системи дистанційного навчання ЗС України, яка забезпечує ефективне реагування на потік загроз та базується на основі впровадження моделей та методів нечіткої логіки та гібридних мереж; удосконалити модель визначення стану системи захисту персональних даних; розробити метод прогнозування стану системи захисту персональних даних.

4. Аналіз останніх досліджень і публікацій

В сучасних наукових публікаціях є інформація про використання штучного інтелекту в кібератаках. Безумовно ця тенденція вплине на майбутнє кіберзлочинності. Дослідження показало те, що у кібербезпеці використовується кілька різних методів та моделей штучного інтелекту, а саме, нечітка логіка, генетичні алгоритми, нейронні мережі та машинне навчання. Найпоширеніші зі стратегій, які використовують штучний інтелект вирішують завдання ідентифікації та моніторингу зловмисних дій, виявлення кіберзагроз і захисту мереж організації. Наприклад, аналітик зловмисного програмного забезпечення може використовувати алгоритми машинного навчання, щоб навчити систему захисту, яка

побудована на принципах штучного інтелекту, виявляти шкідливі файли або підозрілі комп'ютери. Система штучного інтелекту також може відстежувати поведінку окремої людини або групи, а саме, виявляти зміни в соціальних мережах або аналізувати трафік співробітників, щоб визначити тих, хто може планувати кіберзлочини. Під час впровадження штучного інтелекту в кібербезпеку ключовими питаннями є: як керувати даними, та як структурувати дані, щоб зробити їх доступними для програм, які можуть включати контроль людини.

Разом з цією тенденцією кібербезпека також все більше використовує когнітивні технології. Когнітивні технології на основі штучного інтелекту є невід'ємною частиною цілісного підходу до кібербезпеки, в якому людський фактор керує процесом і відіграє ключову роль. Загалом кіберзахист – це простір, що постійно змінюється, де характер загроз безпеці змінюється з кожним новим кроком розвитку. Аналітики вважають, що професіонали з кібербезпеки, які можуть застосувати успішні когнітивні технології, матимуть більший успіх у захисті від кібератак. Галузь також охопила певні тенденції, які формувались роками, наприклад, роль технології блокчейн, як засобу кіберзахисту та збільшення потреби в штучному інтелекті для кібербезпеки. На думку фахівців прогнозується, що в результаті цього ефективність IT-безпеки зросте.

Варто підкреслити те, що раніше фахівці з кібербезпеки зосереджувалися на моніторингу загроз і захисті від них. Тепер вони більше стурбовані оцінкою та зменшенням ризиків, що дозволяє їм уникнути експлойтів (фрагментів програмного коду або послідовності команд, що використовують вразливості в програмному забезпеченні), які можуть завдати шкоди. Аналіз сучасного стану дозволяє зробити висновок про те, що перспективні стратегії кібербезпеки зосереджені на зменшенні ризиків і оцінці ймовірності, а не на моніторингу загроз.

Ці зміни створюють абсолютно нові тенденції, які пов'язані з штучним інтелектом. Загальна класифікація методів штучного інтелекту, які використовуються для кібербезпеки, включає експертні системи та інтелектуальні агенти. Штучний інтелект можна використовувати для виявлення та припинення кібератак шляхом імітації людського інтелекту. Він може виявляти моделі поведінки, щоб визначити потенційні сигнали загрози, які вказують на потенційну атаку. Машинне навчання можна використовувати в кібербезпеці для виявлення та запобігання цілеспрямованим атакам на промислові системи управління.

Розглянемо основні завдання кібербезпеки, які вирішують за допомогою штучного інтелекту.

Ідентифікація зловмисного програмного забезпечення: штучний інтелект може ідентифікувати шкідливі файли до того, як вони досягнуть кінцевого користувача, і, таким чином, може забезпечити значні переваги безпеки. Для виявлення зловмисного програмного забезпечення використовувалося багато різних підходів штучного інтелекту та машинного навчання [2]. Наприклад, за допомогою технології SourceFinder використовується машинне навчання та інтелектуальний аналіз даних для пошуку сховищ вихідного коду зловмисного програмного забезпечення [3]. У [4] представлене спосіб, який застосовує машинне навчання для пошуку рядків у файлах, які можуть вказувати на наявність зловмисного програмного забезпечення або зловмисного коду. Інший підхід полягає у виявлені шаблонів у двійкових виконуваних файлах і визначенні їх шкідливості [5]. У роботі [6] опубліковано метод блокування зловмисного програмного забезпечення, сутність якого у використанні візуальних двійкових шаблонів у коді, і типу самоорганізованої мережі, яка адаптується з часом. Нейронні мережі, які виконують автоматичне виявлення зловмисного програмного забезпечення, не захищені від цих типів атак, в [7] пропонується спосіб для вирішення цієї проблеми.

Техніка обфускації коду є проблемою для методів на основі сигнатур, які використовуються вдосконаленим зловмисним програмним забезпеченням для ухилення від інструментів захисту від зловмисного програмного забезпечення. Щоб вирішити цю проблему, автори роботи [8] обговорюють підхід, який вони використали для покращення точності виявлення невідомого вдосконаленого зловмисного програмного забезпечення, і

запропонували новий метод, який використовує критерій Фішера для вибору функцій і п’ять класифікаторів для виявлення невідомого шкідливого програмного забезпечення.

Штучний інтелект доцільно використовувати для виявлення спаму шляхом аналізу змісту повідомлення та пошуку шаблонів спаму. Це робиться за допомогою алгоритмів машинного навчання, навчених на великій кількості прикладів спаму та інших повідомлень. У процесі виявлення спаму штучний інтелект часто замінює роль людини та зможе виявляти, що повідомлення є спамом, не потребуючи втручання людини.

Метою системи виявлення фішингу є автоматичне виявлення електронних листів, які містять фішингові посилання, і звітування про них. Наприклад, у [9] пропонують використовували нейронну мережу для виявлення фішингових веб-сайтів за допомогою алгоритму Монте-Карло та підходу до мінімізації ризиків.

Варто зазначити те, що використання штучного інтелекту для систем виявлення вторгнень є новою галуззю, що розвивається. Вони призначені для виявлення зловмисної поведінки та припинення її до того, як вона спричинить будь-яку шкоду. Він може бути реалізований як окрема система або як додатковий модуль до іншого програмного забезпечення безпеки, наприклад, антивірусних програм. Система виявлення вторгнень зазвичай налаштована за допомогою набору правил, які визначають, що є атакою. Прикладом такого типу реакції може бути система, яка вже встановлена на місці, відстежує зміни в мережевому трафіку навколо свого периметра, а потім ініціює звіт про ці зміни в центральний центр моніторингу, а також зберігає дані в аналітичній базі даних. Детальний огляд робіт з виявлення вторгнень за останні кілька десятиліть наведено в [9]; перерахувавши багато робіт, автори дійшли висновку, що методи гібридного машинного навчання широко використовуються.

Отже, моделі та методи штучного інтелекту можуть виявляти та зупиняти кіберзагрози в реальному часі з обмеженими ресурсами. Варто підкреслити те, що технології штучного інтелекту не є високоефективними для всіх форм безпеки. Але підходи на основі штучного інтелекту стають все більш поширеними та економічно ефективними в більшості аспектів кібербезпеки. Дослідження показують, що штучний інтелект позитивно вплинув на кібербезпеку та ризики. Впровадження штучного інтелекту та машинного навчання виведе сферу кібербезпеки на новий рівень.

5. Результати дослідження

5.1. Методика підвищення захищеності персональних даних користувачів системи дистанційного навчання

Концепція підвищення захищеності персональних даних, яка реалізована в розробленій методиці, базується на наступних положеннях:

1. Використання алгоритмів прийняття рішення, які побудовані на детермінованих або стохастичних моделях, не є доцільним, так як недостатньо або повністю відсутні дані для формалізації аналітичних виразів. Отже, доцільним є використання апарату математичної логіки, а саме, алгебри логіки, який дозволяє формально визначати систему логічних операції над висловлюваннями та в бінарному вигляді представити оцінку істинності цих висловлювань. Для захисту персональних даних саме ці висловлювання описують систему управління способами, засобами та процедурами кібербезпеки.

2. Відомо те, що математичні моделі алгебри логіки мають певні обмеження щодо їх застосування, а саме, бінарна оцінка істинності або хибності висловлювання – це те саме обмеження. Ідея нечітких множин бінарну оцінку 0 або 1 заміняє оцінкою на число з інтервалу [0,1].

3. Класичні моделі та методи числення предикатів першого порядку, правила виведення числення висловлювань та продукційних систем виведення є основою для так званого нечіткого логічного виведення - адекватної та ефективної моделі прийняття рішення при підвищенні захищеності персональних даних.

4. Адаптивна нечітка нейронна мережа (гібридна мережа) для прогнозування стану системи захисту персональних даних є адекватною та ефективною моделлю. Основна ідея гібридних мереж – використовувати існуючу вибірку даних для визначення параметрів функцій приналежності, що найкраще відповідають деякій системі нечіткого виведення. При цьому для визначення параметрів функцій приналежності використовуються алгоритми навчання нейронних мереж.

Методика підвищення захищеності персональних даних користувачів системи дистанційного навчання ЗС України, яка запропонована автором статті особисто забезпечує ефективне реагування на потік загроз та базується на моделях та методах нечіткої логіки і нейронних мереж. Методика поєднує моделі та методи, які спрямовані на комплексне вирішення наступних завдань: 1) ідентифікація загрози та визначення стану системи захисту персональних даних, а саме, система забезпечує захист в повному обсязі або не забезпечує захист та вимагає застосування способів, засобів та процедур, які раніше не були використані; 2) формування відповідної протидії на загрозу з арсеналу існуючих методів, способів, засобів та процедур або у випадку їх низької ефективності формування стратегії удосконалення та розробки нових; 3) прогнозування стану системи захисту в майбутньому з метою попередження шкодливого впливу загроз. Практична реалізація методики дозволяє забезпечити постійний захист персональних даних користувачів системи дистанційного навчання ЗС України.

5.2. Модель оцінки стану системи захисту персональних даних

Процедура оцінки стану системи захисту персональних даних користувачів системи дистанційного навчання ЗС України є важливим етапом методики підвищення захищеності персональних даних. Дослідження існуючого науково-методичного апарату розроблення моделей даного класу дозволило зробити висновок щодо доцільності використання нечітких моделей виведення нових знань, а саме, алгоритму Мамдані [10]. Даний алгоритм є прикладом так званих систем нечіткого виведення, які в свою чергу займають вагоме місце в нечіткій логіці та моделях штучного інтелекту.

Постановка задачі: за допомогою системи нечіткого виведення алгоритму Мамдані, яка реалізована в пакеті Matlab побудувати модель оцінки рівня захищеності для подальшого формування заходів щодо підвищення захищеності персональних даних користувачів системи дистанційного навчання ЗС України. В моделі забезпечити метод дефазифікації з мінімальною нев'язкою, раціональну кількість правил, оптимальну вагомість правил та оптимальні параметри функцій приналежності.

Етапи моделі.

1. Обґрунтування вхідних та вихідних лінгвістичних змінних.

Опишемо та обґрунтуємо 1 вихідну та 3 вхідні лінгвістичні змінні, а саме, ефективність захищеності (висока, середня, низька) залежить від: 1) ступеня навчання співробітників заходам кібербезпеки (висока, середня, низька); 2) часу, який минув після останніх заходів щодо підвищення захищеності персональних даних (мало, багато, дуже багато); 3) оцінки співробітниками існуючого рівня захищеності (високий, середній, низький).

Вхідні лінгвістичні змінні та відповідні терми:

β_1– ступінь навчання співробітників заходам кібербезпеки.

T_1={«висока», «середня», «низька»}, в моделі позначено, як EmpTraining={H, M, L}.

Визначається в діапазоні [0, 100]. Значення показника тим більше, чим складніше завдання.

β_2 – час, який минув після останніх заходів щодо підвищення захищеності персональних даних.

T_2={«дуже багато», «багато», «мало»}, в моделі позначено, як SecurityUpdateFrequency={H, M, L}. Визначається в діапазоні [0, 100]. Значення показника тим більше, чим більше час.

β_3– інтегральна оцінка співробітниками існуючого рівня захищеності.

T_3={«низька», «середня», «висока»}, в моделі позначено, як InformationAvailability={L, M, H}. Визначається в діапазоні [0, 100]. Значення показника тим більше, чим більше оцінка.

Вихідна лінгвістична змінна та її терми:

β_4– ефективність захищеності.

T_4={«мала», «не дуже мала», «середня», «не дуже висока», «висока»}, в моделі позначено, як ProtectionLevel={L, LA, M, HA, H}. Визначається в діапазоні [0, 1].

2. Обрання методу дефазифікації з мінімальною нев'язкою.

Для дослідження залежності нев'язки від методу дефазифікації використана технологія порівняння експериментальних значень з значеннями, які отримані на основі моделі алгоритму Мамдані.

Аналітичний вираз залежності ефективності системи захисту від 3-х чинників має вигляд

$$\boldsymbol{y = x_1^2 x_2^2 x_3}\ ,\ \ \boldsymbol{x_1 \in [0,1]; x_2 \in [0,1]; x_3 \in [0,1]}. \quad (1)$$

Базу правил створено шляхом візуального аналізу результатів моделювання. Розроблено програма, яка дозволяє розраховувати нев'язку для різних методів дефазифікації. Результати моделювання представлені в таблиці 1.

На відміну від виразу (1) обрана наступна аналітична залежність

$$\boldsymbol{y = x_1 x_2 x_3}\ ,\ \ \boldsymbol{x_1 \in [0,1]; x_2 \in [0,1]; x_3 \in [0,1]}. \quad (2)$$

Результати моделювання представлені в таблиці 1.

Таблиця 1

Залежність нев'язки від методу дефазифікації

Методи дефазифікації	Середня абсолютна нев'язка	
	$\boldsymbol{y = x_1 x_2 x_3}$	$\boldsymbol{y = x_1^2 x_2^2 x_3}$
centroid	0,3707	0,3701
bisector	0,3702	0,3621
mom	0,3709	0,3429
lom	0,4178	0,1809
som	0,2281	0,2062

Отже, з метою забезпечення мінімізації середньої абсолютної нев'язки в моделі доцільно використовувати метод лівого модального значення («som») [10].

3. Визначення раціональної кількості правил.

Для розробленої моделі встановлено метод дефазифікації – метод лівого модального значення («som»), так як він є найкращим і гарантує найменшу абсолютну нев'язку. Варто нагадати про те, що максимальна кількість правил в нечіткій базі знань дорівнює

$$N_{max} = l_1 l_2 l_n\ , \quad (3)$$

де: l_i – кількість термів для оцінки і-ої вхідної змінної (i =1,n);

n – кількість вхідних змінних.

Також відомо те, що однією з переваг систем нечіткого логічного виведення є їх ефективна робота при раціональній кількості правил в базі знань, що менша N_{max}. Звичайно, збільшення кількості правил покращує якість роботи системи, але в реальних умовах кількість достовірної інформації для їх створення може бути відсутня. Крім того, актуальна відома властивість залежності «ефективність – вартість», так звана, насиченість, суть якої полягає в тому, що, починаючи з певного значення кількості правил, додавання нового правила фактично не покращує характеристики системи виведення.

На першому інтуїтивному формуванні 22 правил середня абсолютна нев'язка дорівнює 0,2281. Далі в базі знань була залишена мінімальна кількість правил – 3. Обчислена нев'язка – 0,3078. Отже, нев'язка зросла на 34,9%. Алгоритм дослідження базується на послідовному додаванні по 3 правила доти, поки кількість правил не стане максимальною – 27. Результати представлені в таблиці 2.

Таблиця 2

Залежність нев'язки від кількості правил

Кількість правил	Середня абсолютна нев'язка	Кількість правил	Середня абсолютна нев'язка
3	0,3320	18	0,1865
6	0,2762	21	0,1753
9	0,2358	24	0,1456
12	0,1818	27	0,1287
15	0,1889		

Абсолютна нев'язка для системи з 24 правил дорівнює 0,1199. Результати дослідження для такої системи в залежності від методу дефазіфікації представлені в таблиці 4.

Таблиця 4

Залежність нев'язки від методу дефазифікації для системи з 24 правил

Метод дефазифікації	Нев'язка	Метод дефазифікації	Нев'язка
centroid	0,2165	lom	0,1325
bisector	0,2126	som	0,1199
mom	0,2043		

Отже, метод лівого модального значення є найкращим методом денацифікації при раціональній кількості правил 24. Важливим висновком є те, що збільшення правил з 12 не дає значного приросту ефективності, тому у випадку відсутності достовірної інформації для формування правил логічного виведення дана кількість є досить раціональною.

4. Формування раціональної ваги правил.

Ваговий коефіцієнт продукційного правила бази знань є число в діапазоні [0,1], якому відповідає ранг даного правила в процесі формування рішення. Чим більше значення вагового коефіцієнта, тим важливіше відповідне правило, і тим якісніше воно описує реальний об'єкт. Вагові коефіцієнти правил впливають на результат нечіткого логічного

виведення. В зв'язку з цим, виникає задача підбору таких значень вагових коефіцієнтів правил, які забезпечують найкращу якість нечіткої бази знань. Якість нечіткої бази знань можна оцінити шляхом обчислення нев'язки між експериментальними даними та результатом нечіткого логічного висновку. Підбір вагових коефіцієнтів вирішується шляхом розв'язання відповідної задачі оптимізації. Цільовою функцією цієї задачі оптимізації є якість нечіткої бази знань, а керованими змінними є вагові коефіцієнти правил. Поставлена задача являє собою багатокритеріальну задачу оптимізації.

Цілком логічні наступні вимоги, умови та обмеження дослідження:

1) База знань повинна складатися з 12 або більше правил.

2) Критерієм якості нечіткої бази знань вважати середню квадратичну нев'язку між експериментальними даними та результатами нечіткого логічного виведення.

3) Експериментальні дані згенерувати на основі заданої аналітичної залежності (2).

Введена функція swrules, яка дозволяє програмно встановлювати вагові коефіцієнти правил. Дана функція повертає систему нечіткого висновку, яка відрізняється від початкової новими значеннями вагових коефіцієнтів. Вагові коефіцієнти задаються вектором w. Далі виконується розрахунок цільової функції. Алгоритм оптимізації використовує функцію constr бібліотеки Optimization Mathlab. Виконується оптимізація нечіткої бази знань шляхом зміни вагових коефіцієнтів правил.

Обрана система нечіткого логічного виведення з 23 правил. При зміні діапазону вихідної змінної від 0 до 1, середня квадратична нев'язка буде лише зменшуватися. Це не наглядно, тому доцільно використовувати значення від 0 до 10. Результат розрахунків: середня абсолютна нев'язка дорівнює 1,8092; середня квадратична нев'язка дорівнює 5,7062.

Приклад впливу ваги правила №23 на середню квадратичну нев'язку представлене в таблиці 5.

Таблиця 5

Середня квадратична нев'язка від ваги правила №23

Коефіцієнт	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Середня квадратич. нев'язка	5,5134	5,4867	5,4810	5,5198	5,5201	5,5171	5,4699	5,4677	5,4732	5,4706

Оптимальна вага правила №23 дорівнює 0,8, коли нев'язка дорівнює 5,4677. Далі виконується розрахунок нев'язки для системи з різними коефіцієнтами правил. Вагові коефіцієнти для всіх 23 правил до та після оптимізації представлені в таблиці 6.

Отже, частина моделі оцінки, а саме, алгоритми оптимізації значень вагових коефіцієнтів правил по критерію мінімальної середньої квадратичної нев'язки для конкретного випадку зменшує нев'язку з 5,7062 до 1,6709.

Таблиця 6

Вагові коефіцієнти правил

Правило	Коеф. без оптим.	Коеф. з оптим.
1,3,8,9,12-16,18,20	1	0,1
2,4,19	1	0,2
5,10,14	1	0,3
7,11	1	0,6
21-23	1	0,7

5. Формування оптимальних параметрів функцій приналежності.

Аналіз літератури з нечіткої логіки дозволяє зробити висновок про те, що характеристики функцій приналежності впливають на результат нечіткого логічного

виведення. Тому однією із часткових задач при розробці моделі оцінки стану системи захисту персональних даних є задача підбору таких форм цих функцій, які забезпечують найкращу якість результатів нечіткого виведення, тобто найменшу розбіжність між експериментальними даними та результатами моделювання. При параметричному заданні функцій приналежності сформульована часткова задача оптимізації запишеться таким чином: знайти такі параметри функцій приналежності, які забезпечують найкращу якість нечіткого логічного висновку, а саме, мінімальну нев'язку між експериментальними даними та результатами моделювання.

Спочатку генеруються експериментальні точки відповідно до виду функції, далі зчитується система нечіткого логічного виведення та обчислюються її значення; розраховується максимальна абсолютна нев'язка. В наступному циклі генеруються усі можливі значення термів та записуються оптимальні значення в змінні, які потім вносяться в систему нечіткого виведення. В якості критерію оптимальності використовується середня абсолютна нев'язка. За умови, що поточна нев'язка є нижчою ніж попередня, вона встановлюється як оптимальна. Нев'язка до оптимізації: 5,0371. Нев'язка після оптимізації: 0,4. Алгоритм для вихідної змінної аналогічний. Отже, результати рішення задачі це параметри функцій приналежності, які забезпечують найкращу якість нечіткого логічного висновку, а саме, мінімальну нев'язку між експериментальними даними та результатами моделювання. Оцінка стану системи захисту персональних даних користувачів системи дистанційного навчання навчального закладу ЗС України виконана на основі запропонованої та описаної в статті моделі, частина даних надана в таблиці 7. Данні результати дозволяють отримати кількісну оцінку ефективності системи, яка необхідна для формування стратегії подальшого удосконалення системи захисту. Отримані дані використані в подальшому як навчальні для нейронної мережі.

Таблиця 7

Результати оцінки стану системи захисту персональних даних

Дата дослідження	Ефективність системи	Дата дослідження	Ефективність системи
02.01.2023	0.9955	02.02.2023	0.9999
03.01.2023	0.9999	03.02.2023	0.9997
04.01.2023	0.9999	04.02.2023	0.9998
10.01.2023	0.9999	10.02.2023	0.9998
11.01.2023	0.9997	11.02.2023	0.9999
12.01.2023	0.9995	12.02.2023	0.9999
13.01.2023	0.9997	17.02.2023	0.9996
15.01.2023	0.9997	21.02.2023	0.9999
19.01.2023	0.9997	22.02.2023	0.9999
30.01.2023	0.9996	02.03.2023	0.9994
31.01.2023	0.9999	03.03.2023	0.9999
01.02.2023	0.9999	04.03.2023	0.9997

Отже, модель оцінки стану системи захисту персональних даних, яка запропонована базується на алгоритмі Мамдані (алгоритмі нечіткого логічного виведення). В моделі забезпечено оптимальний метод дефазифікації, раціональна вага та кількість правил, оптимальні параметри функцій приналежності. Застосування моделі дозволяє отримувати чітку кількісну оцінку ефективності системи, яка необхідна для прийняття рішення для подальшого удосконалення.

5.3. Метод прогнозування стану системи захисту персональних даних

В якості ефективного методу прогнозування стану системи захисту персональних даних обрана адаптивна мережева нечітка система виведення ANFIS, яка була запропонована Джи-Шинг Роджером Джангом (Jyh-Shing Roger Jang) в 1993 році [10].

Етапи методу прогнозування

1) Формування навчальних даних (Training Data).

В якості навчальних даних для моделі прогнозу використані дані екстерименту оцінки стану системи захисту персональних даних (таблиця 7). Дані з 28.02.2023р. по 04.03.2023р. використані для перевірки точності розробленої моделі прогнозу. В таблицях 8 та 9 представлено частину навчальних даних для створення гібридної мережі та перевірочні дані відповідно.

2) Завантаження навчальних даних.
3) Нечітка кластеризація.
4) Формування гібридної мережі (рис.1).
5) Навчання мережі гібридним методом.

Таблиця 8

Навчальні дані для моделі гібридної мережі

№ вхідної змінної				Вихідна змінна
1	2	3	4	
0.9954	0.9965	0.9999	0.9999	0.9998
0.9998	0.9954	0.9965	0.9999	0.9999
0.9999	0.9998	0.9954	0.9965	0.9999
0.9999	0.9999	0.9998	0.9954	0.9999
0.9999	0.9999	0.9999	0.9998	0.9997
0.9997	0.9999	0.9999	0.9999	0.9965
0.9965	0.9997	0.9999	0.9999	0.9951
0.9951	0.9965	0.9997	0.9999	0.9999
0.9999	0.9951	0.9965	0.9997	0.9999
0.9999	0.9999	0.9951	0.9965	0.9999
0.9999	0.9999	0.9999	0.9951	0.9998
0.9998	0.9999	0.9999	0.9999	0.9999
0.9999	0.9998	0.9999	0.9999	0.9999

Таблиця 9

Перевірочні дані

№ вхідної змінної				Вихідна змінна
1	1	1	1	
0.9997	0.9999	0.9999	0.9998	0.9998
0.9998	0.9997	0.9999	0.9999	0.9998
0.9998	0.9998	0.9997	0.9997	0.9999
0.9999	0.9998	0.9998	0.9997	0.9999
0.9999	0.9999	0.9998	0.9998	0.9997

Результати моделювання прогнозу ефективності системи свідчать про збіг даних з реальними значеннями, які були обрані для перевірки адекватності та точності методу (таблиця 10).

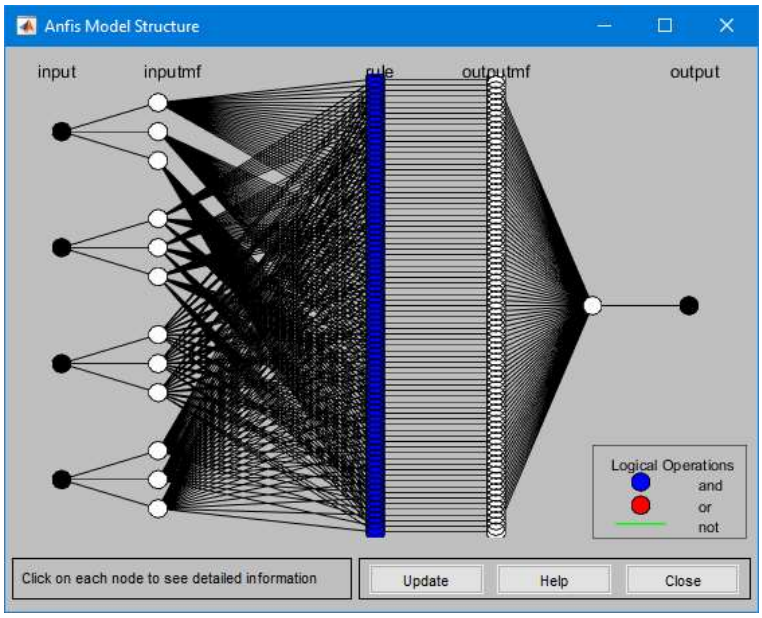

Рис. 1. Гібридна мережа

Результат функції evalfis:

```
>> out = evalfis([0.9997          0.9999 0.9999 0.9998;
0.9998    0.9999 0.9999 0.9998;
0.9998    0.9997 0.9999 0.9998;
0.9999    0.9998 0.9998 0.9997;
0.9999    0.9999 0.9998 0.9998;],1)
out =
   0.9997
   0.9998
   0.9999
   0.9999
   0.9996
```

Таблиця 10

Оцінка точності прогнозу

Експеримент	Прогноз	Похибка, %
0.9998	0.9997	0,01
0.9998	0.9998	0
0.9999	0.9999	0
0.9999	0.9999	0
0.9997	0.9996	0,01

Отримана середня похибка прогнозу – δ=0,004 %.

6. Висновки

1) В сучасних умовах існує тенденція постійного розвитку способів кіберзагроз та кібератак, а це в свою чергу вимагає збільшення уваги до нових розробок кібербезпеки, в тому числі для захисту персональних даних користувачів системи дистанційного навчання ЗС України. Забезпечення високого рівня ефективності таких розробок потребує сучасних, креативних підходів, з яких найбільш доцільним є той, що заснований на штучному інтелекті.

2) Дослідження визначило те, що застаріла технологія кібербезпеки зосереджувалася в основному на моніторингу загроз і захисті від них; сучасна та перспективна політика та стратегія кібербезпеки спрямована на оцінку та зменшенням ризиків, що забезпечує уникнення шкідливих фрагментів програмного коду або послідовності команд. Аналіз

публікацій за тематикою впровадження штучного інтелекту в методи, способи та процедури кібербезпеки визначає даний напрям як нову тенденцією. Дослідження показало те, що у кібербезпеці використовується кілька різних напрямів штучного інтелекту, а саме, м'які обчислення, нейронні мережі та машинні навчання. Найпоширеніші методики включають моделі, які використовують штучний інтелект для ідентифікації та моніторингу зловмисних дій, виявлення кіберзагроз і захисту мереж організації.

3) Методика підвищення захищеності персональних даних користувачів системи дистанційного навчання Збройних Сил України, яка запропонована автором статті, забезпечує ефективне реагування на потік загроз та базується на моделях і методах нечіткої логіки та нейронних мереж. Методика інтегрує моделі та методи, які спрямовані на комплексне вирішення низки завдань: 1) ідентифікація загрози та визначення стану системи захисту персональних даних; 2) формування відповідної протидії на загрозу з арсеналу існуючих методів, способів, засобів та процедур або команди на удосконалення та розробки нових; 3) прогнозування стану системи захисту в майбутньому з метою попередження шкідливого впливу загроз. Практична реалізація методики дозволяє забезпечити постійний захист персональних даних користувачів системи дистанційного навчання Збройних Сил України.

4) Модель оцінки стану системи захисту персональних даних, яка пропонується автором статті, на відміну від існуючих базується на нечіткому логічному виведенні (алгоритм Мамдані). В моделі забезпечено оптимальний метод дефазифікації, раціональна вага та кількість правил, оптимальні параметри функцій приналежності. Застосування моделі дозволяє отримувати чітку кількісну оцінку ефективності системи, яка необхідна для прийняття рішення щодо подальшого удосконалення.

5) Дослідження розробленого методу прогнозування стану системи захисту персональних даних на основі нечіткої моделі гібридної мережі дозволяє зробити висновок про високу ступінь його адекватності реальним вихідними даними, та рекомендувати цей метод до практичного використання. Також варто визначите те, що нечіткі моделі адаптивних систем нейронечіткого висновку можуть вважатися конструктивним інструментом для інших розробок в кібербезпеці.

Список використаної літератури

1. Savchenko V.,Matsko O.,Vorobiov O., Kizyak Y.Kriuchkova L.Tikhonov Y., Kotenko A. Network traffic forecasting based on the canonical expansion of a random process. Eastern European Journal of Enterprise Technologies. 2018. Vol 3, № 2 (93). pp. 33-41.

2. M. J. Hossain Faruk et al. Malware Detection and Prevention using Artificial Intelligence Techniques," 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 2021, pp. 5369-5377, doi: 10.1109/BigData52589.2021.9671434.

3. Rokon, M.O.F., Islam, R., Darki, A., Papalexakis, E., Faloutsos, M.: Sourcefinder: Finding malware source-code from publicly available repositories in github. No 10, 2020, pp. 149-163.

4. Shrestha, P., Maharjan, S., Ramirez-de-la Rosa, G., Sprague, A., Solorio, T., Warner, G.: Using string information for malware family identification. No 11, 2014, pp. 686–697. https://doi.org/10.1007/978-3-319-12027-0 55

5. Schultz, M., Eskin, E., Zadok, F., Stolfo, S.: Data mining methods for detection of new malicious executables. No 02, 2001, pp. 38–49.

6. I. Baptista, S. Shiaeles and N. Kolokotronis, "A Novel Malware Detection System Based on Machine Learning and Binary Visualization," 2019 IEEE International Conference on Communications Workshops (ICC Workshops), Shanghai, China, 2019, pp. 1-6, doi: 10.1109/ICCW.2019.8757060.

7. Bose, S., Barao, T., Liu, X.: Explaining ai for malware detection: Analysis of mechanisms of malconv. In: 2020 International Joint Conference on Neural Networks (IJCNN). pp. 1-8.

8. Sharma, S., Challa, R., Sahay, S.: Detection of advanced malware by machine learning techniques. No 03, 2019.

9. Abhilash Chakraborty, Anupam Biswas, Ajoy Kumar Khan. Artificial Intelligence for Cybersecurity: Threats, Attacks and Mitigation. 2022. https://doi.org/10.48550/arXiv.2209.13454

10. Jang J.-S. R. ANFIS: Adaptive-Network-based Fuzzy Inference System. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 1993. Vol. 23. no. 3. pp. 665−685.

11. Bohdan Zhurakovskyi, Serhii Toliupa, Volodymir Druzhynin, Andrii Bondarchuk, Mykhailo Stepanov . Calculation of Quality Indicators of the Future Multiservice Network. In: Future Intent-Based Networking: On the QoS Robust and Energy Efficient Heterogeneous Software Defined Networks. Springer International Publishing, 2022. p. 197-209.

References

1. Savchenko V.,Matsko O.,Vorobiov O., Kizyak Y.Kriuchkova L.Tikhonov Y., Kotenko A. Network traffic forecasting based on the canonical expansion of a random process. Eastern European Journal of Enterprise Technologies. 2018. Vol 3, № 2 (93). pp. 33-41.

2. M. J. Hossain Faruk et al. Malware Detection and Prevention using Artificial Intelligence Techniques," 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 2021, pp. 5369-5377, doi: 10.1109/BigData52589.2021.9671434.

3. Rokon, M.O.F., Islam, R., Darki, A., Papalexakis, E., Faloutsos, M.: Sourcefinder: Finding malware source-code from publicly available repositories in github. No 10, 2020, pp. 149-163.

4. Shrestha, P., Maharjan, S., Ramirez-de-la Rosa, G., Sprague, A., Solorio, T., Warner, G.: Using string information for malware family identification. No 11, 2014, pp. 686–697. https://doi.org/10.1007/978-3-319-12027-0 55

5. Schultz, M., Eskin, E., Zadok, F., Stolfo, S.: Data mining methods for detection of new malicious executables. No 02, 2001, pp. 38–49.

6. I. Baptista, S. Shiaeles and N. Kolokotronis, "A Novel Malware Detection System Based on Machine Learning and Binary Visualization," 2019 IEEE International Conference on Communications Workshops (ICC Workshops), Shanghai, China, 2019, pp. 1-6, doi: 10.1109/ICCW.2019.8757060.

7. Bose, S., Barao, T., Liu, X.: Explaining ai for malware detection: Analysis of mechanisms of malconv. In: 2020 International Joint Conference on Neural Networks (IJCNN). pp. 1-8.

8. Sharma, S., Challa, R., Sahay, S.: Detection of advanced malware by machine learning techniques. No 03, 2019.

9. Abhilash Chakraborty, Anupam Biswas, Ajoy Kumar Khan. Artificial Intelligence for Cybersecurity: Threats, Attacks and Mitigation. 2022. https://doi.org/10.48550/arXiv.2209.13454

10. Jang J.-S. R. ANFIS: Adaptive-Network-based Fuzzy Inference System. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 1993. Vol. 23. no. 3. pp. 665−685.

11. Bohdan Zhurakovskyi, Serhii Toliupa, Volodymir Druzhynin, Andrii Bondarchuk, Mykhailo Stepanov . Calculation of Quality Indicators of the Future Multiservice Network. In: Future Intent-Based Networking: On the QoS Robust and Energy Efficient Heterogeneous Software Defined Networks. Springer International Publishing, 2022. p. 197-209.