МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ АВІАЦІЙНИЙ УНІВЕРСИТЕТ

ПОГОРЕЛОВ Володимир Володимирович

УДК 004.056.5:004.8(043.3)

НЕЙРОМЕРЕЖЕВІ МОДЕЛІ ТА МЕТОДИ РОЗПІЗНАВАННЯ КОМП'ЮТЕРНИХ ВІРУСІВ

Спеціальність 05.13.21 – системи захисту інформації

Автореферат
дисертації на здобуття наукового ступеня
кандидата технічних наук

Київ – 2020

Дисертацією є рукопис.

Робота виконана на кафедрі безпеки інформаційних технологій Національного авіаційного університету Міністерства освіти і науки України

Науковий керівник: доктор технічних наук, професор
Терейковський Ігор Анатолійович,
професор кафедри безпеки інформаційних технологій Національного авіаційного університету.

Офіційні опоненти: доктор технічних наук, доцент
Опірський Іван Романович,
професор кафедри захисту інформації Національного університету «Львівська політехніка».

кандидат технічних наук
Фесенко Андрій Олексійович,
асистент кафедри кібербезпеки та захисту інформації Київського національного університету ім. Т. Шевченка.

Захист відбудеться «26» листопада 2020 р. о 15.00 на засіданні спеціалізованої вченої ради Д 26.062.17 при Національному авіаційному університеті за адресою: 03058, м. Київ. просп. Любомира Гузара, 1, ауд. 11-111.

З дисертацією можна ознайомитися в науково-технічній бібліотеці Національного авіаційного університету за адресою: 03058, м. Київ. просп. Любомира Гузара, 1.

Автореферат розісланий «27» жовтня 2020 р.

Учений секретар
спеціалізованої вченої ради,
доктор технічних наук, доцент С. О. Гнатюк

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. В теперішній час системи антивірусного захисту (САЗ) є одним з основних засобів захисту інформації більшості комп'ютерних систем і мереж. Не зважаючи на те, що такі системи використовуються вже не одне десятиліття і їх розробкою та створенням методологічної бази займаються висококваліфіковані фахівці, практичний досвід і результати багатьох науково-практичних досліджень вказують на наявність в сучасних антивірусах розпізнавання суттєвих недоліків. Основним з яких є недостатня точність розпізнавання всієї номенклатури комп'ютерних вірусів, що підтверджується відомими випадками успішних вірусних кібератак на вітчизняні та закордонні комп'ютерні системи і мережі. Однак впровадження відомих засобів розпізнавання комп'ютерних вірусів в вітчизняні системи захисту інформації викликає необхідність їх складної адаптації до очікуваних умов використання. Також недоліками відомих засобів розпізнавання є висока вартість і відсутність докладної науково-технічної документації.

Важливим напрямком підвищення точності розпізнавання є «інтелектуалізація» методів розпізнавання за рахунок використання теорії штучних нейронних мереж (НМ). Перспективність вказаного напрямку підтверджується окремими вдалими застосуваннями НМ в засобах розпізнавання комп'ютерних вірусів (антивірус з відкритим програмним кодом ClamAV, стартап Deep Instinct) та великою кількістю відповідних теоретико-практичних робіт.

Разом з тим, недостатня точність розпізнавання та недостатня адаптованість до умов експлуатації, закритість використаних рішень, значно обмежують сферу їх застосування. При цьому постійний прогрес в області теорії НМ вказує на можливість значного вдосконалення апробованих засобів розпізнавання.

В такій постановці проблеми є актуальною науково-прикладна задача розробки ефективних нейромережевих моделей та методів розпізнавання комп'ютерних вірусів, адаптованих до умов вітчизняних систем антивірусного захисту (САЗ).

Дослідження вітчизняних та зарубіжних вчених, зокрема И. Бенджио (Yoshua Bengio), Бодянського Є. В., Я. Лекуна (Yann LeCun), Різника О.М., Руденка О.Г., Д. Хінтона (Geoffrey Hinton), З. Хохрайтера (Sepp Hochreiter) вказують на те, що перспективним шляхом підвищення ефекивності антивірусного захисту є застосування апарату НМ для розпізнавання комп'ютерних вірусів. Це пояснюється тим, що задача розпізнавання комп'ютерних вірусів є однією із основних при розробці САЗ, що підтверджується ефективністю використання НМ для вирішення подібних задач оцінки параметрів безпеки інформаційних систем у відомих засобах захисту інформації (AVZ, продукція компаній Cisco, Symantec) та є доведеною адаптивністю нейромережевих засобів (НМЗ) розпізнавання до умов застосування в САЗ.

Методологічну і теоретичну основу для ефективного впровадження НМЗ розпізнавання комп'ютерних вірусів складають теоретичні розробки та досвід створення систем захисту інформації таких науковців як, Б. Ахметова,

Є. Бодянського, Д. Деннінга, О. Додонова, В. Лахна, А. Лукацького, О. Корченка, О. Петрова, О. Резніка, О. Руденка, С. Форестера, І. Терейковського, В. Харченка, В. Хорошка.

Таким чином, задача розробки ефективних нейромережевих моделей та методів розпізнавання комп'ютерних вірусів зумовила актуальність наукових досліджень і розробок, яким присвячена дисертаційна робота.

Зв'язок роботи з науковими програмами, планами, темами. Тематика дисертаційної роботи та одержані результати безпосередньо пов'язані з «Основними науковими напрямами та найважливішими проблемами фундаментальних досліджень у галузі природничих, технічних і гуманітарних наук НАН України на 2019-2023 роки», зі Стратегією кібербезпеки України від 15 березня 2016 року №96/2016. Результати роботи відображені у звітах держбюджетних науково-дослідних робіт Національного авіаційного університету «Квантово-криптографічні методи захисту критичної інформаційної структури держави» (д.р. № 0117U006770), «Системи мультирівневого розмежування доступу до інформаційних ресурсів» (№19/14.01.05) та «Дослідження ризиків інформаційної безпеки об'єктів критичної інфраструктури ГТС України та розробка методології поводження з ними» (д.р. № 0118U002371).

Мета і задачі дослідження. Метою дисертаційної роботи є підвищення ефективності протидії комп'ютерним вірусам за рахунок розробки і дослідження нових нейромережевих моделей, методів і засобів розпізнавання комп'ютерних вірусів, здатних оперативно пристосовуватись до умов використання і реагувати на виникнення нових видів вірусів.

Відповідно до поставленої мети визначено такі основні завдання дослідження:

– проведення аналізу можливостей нейромережевих засобів (НМЗ) розпізнавання комп'ютерних вірусів;

– розробка моделі формування параметрів та концептуальної моделі оцінювання глибоких нейронних мереж (ГНМ);

– розробка методу визначення архітектурних параметрів глибокої нейронної мережі та розвиток методу розпізнавання комп'ютерних вірусів;

– проведення експериментальних досліджень, спрямованих на верифікацію запропонованих рішень.

Об'єктом дослідження є процеси розпізнавання комп'ютерних вірусів.

Предметом дослідження є нейромережеві моделі та методи розпізнавання комп'ютерних вірусів.

Методи дослідження. Використано методи теорії захисту інформації, НМ, комп'ютерного моделювання, експертного і статистичного аналізу та оптимізації.

Наукова новизна отриманих результатів. Проведені у дисертаційній роботі дослідження дозволили розробити й науково обґрунтувати принципи, моделі та методи нейромережевого розпізнавання комп'ютерних вірусів. Зокрема отримані такі наукові результати.

Вперше розроблено:

– концептуальну модель оцінювання глибоких нейронних мереж, яка за

рахунок взаємопов'язаних принципів допустимості використання, визначення множини ефективних видів та оцінювання ефективності виду глибокої нейронної мережі дозволяє визначити множину сучасних нейромережевих моделей для побудови ефективних антивірусних засобів;

– модель формування параметрів навчальних прикладів глибокої нейронної мережі, яка за рахунок формального представлення закодованих значень викликів API-функцій, байт-послідовності N-грамів, опкодів, основних регістрів процесора, а також результатів статичного аналізу зразків шкідливих та безпечних програм, двомірної інтерпретації бінарного коду програми і параметрів графу залежностей значень та станів дозволяє будувати засоби нейромережевого аналізу обфускованого програмного коду;

– метод визначення архітектурних параметрів глибокої нейронної мережі, призначеної для розпізнавання вірусів, який за рахунок використання запропонованої концептуальної моделі оцінювання глибоких нейронних мереж та моделі формування параметрів навчальних прикладів, що використовуються для реалізації етапів визначення основних умов застосування, доцільності використання нейромережевої моделі та найбільш ефективної архітектури, а також формування параметрів навчальних прикладів та визначення параметрів архітектури найбільш ефективного виду глибокої нейронної мережі, дозволяє сформувати набір величин, які забезпечують пристосованість такої мережі до визначених умов застосування;

Отримав подальший розвиток:

– метод нейромережевого розпізнавання комп'ютерних вірусів, який, за рахунок визначення умов створення та застосування нейромережевих засобів, процесів формування портретів вірусів та безпечних програм, а також визначення архітектурних параметрів глибокої нейронної мережі та верифікації і оцінки ефективності нейромережевих засобів, забезпечує достатню похибку розпізнавання при різних умовах застосування з урахуванням обмежень щодо створення навчальної вибірки та обмежень щодо обчислювальних ресурсів системи антивірусного захисту.

Практичне значення одержаних результатів дисертаційного дослідження полягає у наступному:

– розроблене алгоритмічне та програмне забезпечення, що базується на створених нейромережевих методах та моделях, дозволило забезпечити достатню точність розпізнавання комп'ютерних вірусів та приблизно в 1,5 рази зменшити обчислювальні витрати, пов'язані з визначенням значень архітектурних параметрів ГНМ, що підтверджується актом впровадження в діяльність ТОВ «Сайфер ПРО» (акт впровадження від 17.08.2020);

– розроблені програми, що реалізують запропоновані моделі та методи, впроваджені в навчальний процес на кафедрі безпеки інформаційних технологій Національного авіаційного університету (акт впровадження від 25.02.2020).

– результати проведених розрахунків вказують на те, що ефективність розробленого НМЗ приблизно в 1,14 рази вища ніж у подібних відомих засобів. Таким чином, результати досліджень підтверджують можливість підвищення ефективності розпізнавання комп'ютерних вірусів за рахунок застосування розроблених нейромережевих моделей (НММ) та НМЗ, що підтверджується

актом впровадження в діяльність ТОВ «Сайфер ПРО» (акт впровадження від 17.08.2020)

Особистий внесок здобувача. Всі основні результати дисертаційної роботи отримані здобувачем самостійно. У роботах, опублікованих із співавторами, здобувачу належить: [1] – дослідження підходів до виявлення шкідливих програм за допомогою штучних НМ; [2] – розроблено нейромережеву модель призначену для розпізнавання комп'ютерних вірусів; [3] – розроблено модель деобфускації програмного коду для визначення вхідних параметрів нейромережевої моделі призначеної для розпізнавання комп'ютерних вірусів; [4, 5] – розроблено метод застосування ГНМ для розпізнавання комп'ютерних вірусів; [8] – розроблено нейромережеву модель для розпізнавання комп'ютерних вірусів; [9] – розроблено підхід до визначення ефективності НМЗ розпізнавання шкідливого програмного забезпечення; [10] – розроблено підхід до визначення вихідних сигналів нейромережевої моделі призначеної для розпізнавання шкідливого програмного забезпечення; [11] – розроблено метод розпізнавання шкідливого ПЗ; [14] – визначено перелік параметрів, що використовуються для розпізнавання комп'ютерних вірусів.

З робіт, що опубліковані у співавторстві, у дисертаційній роботі використовуються виключно результати, отримані особисто здобувачем.

Апробація результатів дисертації. Основні результати дисертації доповідались, обговорювались та отримали позитивні оцінки на наступних конференціях: Information Technologies, Management and Society, The 15th International Scientific Conference Information Technologies and Management (2017); Международной научно-практической конференции «Математи-ческие методы и информационные технологии макроэкономического анализа и экономической политики», посвященной празднованию 80-летнего юбилея академика НАН РК Абдыкаппара Ашимовича Ашимова (2017); Aviation in the XXI-st Century (2018); Advances in Computer Science for Engineering and Education. ICCSEEA (2018, 2019); The 10 th International Scientific Conference «ITSec» (2020); Стан та удосконалення безпеки інформаційно-телекомунікаційних систем (SITS' 2020); VI Міжнародна науково-практична конференція «Актуальні питання забезпечення кібербезпеки та захисту інформації» (2020).

Публікації. Основні положення дисертації опубліковано у 14 наукових працях, серед яких 7 наукових статей (4 статті у фахових наукових виданнях України, 3 – у міжнародних рецензованих виданнях, які входять до бази наукометричної бази даних SCOPUS), 1 закордонна колективна монографія, а також 6 матеріалів і тез доповідей на конференціях.

Структура та обсяг роботи. Дисертаційна робота складається зі вступу, чотирьох розділів, висновків та списку використаних джерел (106 найменувань) на 11 сторінках, 2 додатки на 11 сторінках. Загальний обсяг дисертації становить 166 сторінок, у тому числі 144 сторінки основного тексту, ілюстрацій – 38, таблиць – 9.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі обґрунтовано актуальність теми дисертації, визначено мету і задачі дослідження, розкрито наукову новизну та практичне значення отриманих результатів, наведені дані щодо їх апробації та впровадження.

У першому розділі охарактеризовано науково-прикладну задачу розробки засобів розпізнавання комп'ютерних вірусів в САЗ.

Проведено аналіз науково-практичних досліджень, присвячених вирішенню задачі розпізнавання комп'ютерних вірусів. Обґрунтована перспективність застосування в контурі розпізнавання САЗ НМЗ. При цьому показано можливість застосування НММ як в поведінкових аналізаторів, так і при використанні сигнатурного аналізу. Також для вітчизняних САЗ визначена множина очікуваних умов застосування означених НМЗ.

Разом з тим показано, що очікувані умови характеризують варіативність обмежень на термін розробки САЗ, обмеженість обчислювальних ресурсів, що можуть бути використані при побудові НММ, тип антивірусної системи, відсутність достатньо повних баз даних прикладів комп'ютерних вірусів, необхідних для проведення навчання НММ та забезпеченням можливості аналізу обфускованого програмного коду. Водночас, аналіз сучасних нейромережевих методів розпізнавання комп'ютерних вірусів показав їх недостатню адаптацію до означених умов застосування.

На основі результатів аналізу сучасних науково-практичних робіт присвячених нейромережевому розпізнаванню комп'ютерних вірусів та нейромережевому розпізнаванню інших типів кібератак визначена множина процедур, виконання яких впливає на ефективність застосування НМЗ для розпізнавання комп'ютерних вірусів.

Показано, що ефективність НМЗ розпізнавання комп'ютерних вірусів в значній мірі залежить від вибору виду та параметрів архітектури НММ. Визначено перспективність використання НММ на базі глибоких нейронних мереж (ГНМ), що на сьогодні представлені трьома базовими архітектурами: згорткові нейронні мережі (ЗНМ), ГНМ з преднавчанням, ГНМ без преднавчання. При цьому в доступній літературі не знайдено опису методу визначення архітектурних параметрів НМЗ розпізнавання на базі ГНМ.

Таким чином, на підставі проведеного аналізу показано, що важливим напрямком підвищення ефективності САЗ є впровадження нейромережевих моделей розпізнавання комп'ютерних вірусів, що базуються на сучасних рішеннях теорії штучних НМ. Для цього необхідно розвинути методологічну базу і розробити відповідні нейромережеві методи, адаптовані до очікуваних умов застосування.

Другий розділ присвячено розвитку методологічної бази нейромережевого розпізнавання комп'ютерних вірусів.

Розроблена концептуальна модель оцінювання ГНМ, яка дозволяє визначити множину сучасних нейромережевих моделей для побудови ефективних антивірусних засобів.

Обґрунтовано формулювання специфічних термінів: портрет комп'ютерного вірусу – множина параметрів, що використовуються для розпізнавання; портрет сигнатури – множина параметрів, що характеризують

сигнатуру комп'ютерного вірусу; портрет поведінки – множина параметрів, що характеризують події в операційній системі під час функціонування комп'ютерного вірусу.

До концептуальної моделі відносяться:

- *принцип допустимості використання виду ГНМ для розпізнавання комп'ютерних вірусів*, який полягає в тому, що серед множини доступних *i*-ий вид ГНМ ($\mathbf{DNN}_i$) входить до множини допустимих видів ($\mathbf{DNN_{avl}}$), якщо його основні характеристики ресурсоємність та час навчання ($Q(\mathbf{DNN}_i), \tau(\mathbf{DNN}_i)$) задовольняють вимогам щодо допустимого часу навчання (τ_{avl}) і допустимої ресурсоємності побудови (Q_{avl}) НМЗ:
if $\left(Q(\mathbf{DNN}_i) \le Q_{avl}\right) \wedge \left(\tau(\mathbf{DNN}_i) \le \tau_{avl}\right) \rightarrow \mathbf{DNN}_i \in \mathbf{DNN}_{avl}$;

- *принцип визначення множини ефективних видів ГНМ для розпізнавання комп'ютерних вірусів, який* полягає у тому, що серед множини доступних *i*-ий вид ГНМ ($\mathbf{DNN}_i$) входить до множини ефективних видів ($\mathbf{DNN_{eff}}$), якщо для нього значення функції ефективності ($V(\mathbf{DNN}_i)$) не менше допустимого значення: *if* $V(\mathbf{DNN}_i) \le V_d \rightarrow \mathbf{DNN}_i \in \mathbf{DNN}_{eff}$; при цьому $V(\mathbf{DNN}_i) = \sum_{k=1}^{K} \alpha_k H_k(\mathbf{DNN}_i)$, де $H_k(\mathbf{DNN}_i)$ – значення *k*-го критерію для ГНМ з *i*-ою архітектурою, $\alpha_k = |0..,1|$ – ваговий коефіцієнт *k*-го критерію ефективності, *K* – кількість критеріїв.

- *принцип оцінювання ефективності виду ГНМ*, призначеної для розпізнавання комп'ютерних вірусів, полягає в тому, що серед множини допустимих *i*-ий вид ГНМ ($\mathbf{DNN}_i$) є найбільш ефективним, якщо для нього функція ефективності (V_i) має максимальне значення: $\max_{V_i} = \{V_1, V_2, .. V_I\}$.

За допомогою концептуальної моделі визначено, що розвиток методологічної бази повинен забезпечити достатню точність розпізнавання з врахуванням обмежень щодо створення навчальної вибірки та обмежень щодо обчислювальних ресурсів системи антивірусного захисту. Для цього методологічна база доповнена рядом принципів та моделей.

Модель формування параметрів навчальних містить:

- *принцип представлення програмного забезпечення у вигляді графа залежностей значень та станів, який* полягає в тому, що поведінка програмного забезпечення представляється за допомогою спеціалізованого графа залежностей значень та станів $\mathbf{F =< T,S >}$, де $\mathbf{F}$ – множина, що описує дії програмного забезпечення, $\mathbf{T}$ – множина переходів, $\mathbf{S}$ – множина станів графу, що описує поведінку програмного забезпечення;

- *принцип оцінювання безпечності програмного забезпечення за допомогою графів, який* полягає в тому, що безпечність програмного забезпечення можливо оцінити за рахунок порівняння графа залежностей значень та станів піддослідної програми з відповідними графами безпечних програм та графами комп'ютерних вірусів: if $F_x = \mathbf{F}_s \rightarrow x \in \mathbf{S}$, if $\mathbf{F}_x = \mathbf{F}_v \rightarrow x \in \mathbf{V}$, де $\mathbf{F}_x$ – граф піддослідного програмного забезпечення, $\mathbf{F}_s$ – множина графів безпечних програм, $\mathbf{F}_v$ – множина графів комп'ютерних вірусів, x –піддослідне програмне забезпечення, $\mathbf{S}$ – множина безпечних програм, $\mathbf{V}$ – множина комп'ютерних вірусів.

Для визначення ефективних видів ГНМ формується низка правил, що базується на принципі допустимості застосування виду ГНМ та на принципі визначення множини ефективних видів ГНМ.

Визначення множини ефективних видів ГНМ представлено у вигляді $\mathbf{DNN_{ent}} \to \mathbf{DNN_{avl}} \to \mathbf{DNN_{eff}}$, де $\mathbf{DNN}_{\mathbf{ent}}$ – множина доступних видів ГНМ, $\mathbf{DNN}_{\mathbf{avl}}$ – множина допустимих видів ГНМ, $\mathbf{DNN}_{\mathbf{eff}}$ – множина ефективних видів ГНМ. При цьому $\mathbf{DNN_{ent}} = \{dnn_1, dnn_2, dnn_3, dnn_4\}$, де dnn_1 складають повнозв'язні ГНМ, при навчанні яких процедура преднавчання не передбачена; dnn_2 складають повнозв'язні ГНМ, при навчанні яких використовується процедура преднавчання; dnn_3 складають ГНМ типу ЗНМ з прямим поширенням сигналу; dnn_4 складають ГНМ типу рекурентних ЗНМ.

Визначено, що при оціночних розрахунках допустимість ГНМ типу dnn_1 може бути зумовлена виразом $0,4\bar{\lambda}N_x^2N_y^2(N_x+N_y)^2+0,02\vartheta_m N_x N_y \le t_{max}$, де N_x, N_y – кількість вхідних та вихідних параметрів ГНМ, ϑ_w – середній час, необхідний на створення одного навчального прикладу з очікуваним вихідним сигналом, $\bar{\lambda}$ – приведена тривалість однієї навчальної ітерації.

Допустимість ГНМ типу dnn_2 визначається виразом math, допустимість типу dnn_3 – виразом $0,4\bar{\lambda}N_x^2N_y^2(N_x^3+N_yN_x^2)+0,02\vartheta_m N_x N_y \le t_{max}$, а допустимість типу dnn_4 – виразом $0,8\bar{\lambda}N_x^2N_y^2(N_x^3+N_yN_x^2)+0,02\vartheta_m N_x N_y \le t_{max}$, де ϑ_{nm} – середній час, необхідний на створення одного навчального прикладу без очікуваного вихідного сигналу, t_{max} – максимально допустимий термін створення ГНМ.

Розробка наведених принципів та моделей забезпечила можливість створення ефективних нейромережевих методів розпізнавання комп'ютерних вірусів.

Третій розділ присвячено розробці нейромережевої моделі та методів розпізнавання комп'ютерних вірусів. *Розроблена модель формування параметрів навчальних прикладів глибокої нейронної мережі* передбачає можливість використання в якості вхідних параметрів закодованих значень:

– викликів API-функцій (Ψ_{api});

– байт-послідовності N-грамів (Ψ_{bs});

– опкодів, вилучених з дизасебльованих файлів (Ψ_{opc});

– результатів статистичного аналізу зразків шкідливих та безпечних програм (Ψ_{st});

– значень регістрів загального призначення EAX, EBX, EDX, EDI та регістрів для роботи зі стеком ESP, EBP (Ψ_{reg});

– параметрів графів викликів API-функції (Ψ_{gr});

– бінарного двохвимірного представлення програмного коду (Ψ_{bin}); параметрів PE-заголовку файлу (Ψ_{pe});

– параметрів графу залежностей значень та станів програмного забезпечення (Ψ_{dvs}).

Перетворення вхідної інформації моделі у набір вхідних параметрів ГНМ описується виразом $F_{fpe}(\mathbf{\Psi}) = \mathbf{X}$, де $\mathbf{X}$ –множина вхідних параметрів ГНМ, $\mathbf{\Psi} = \{\Psi_{api}, \Psi_{bs}, \Psi_{opc}, \Psi_{st}, \Psi_{reg}, \Psi_{gr}, \Psi_{bin}, \Psi_{pe}, \Psi_{dvs}\}$.

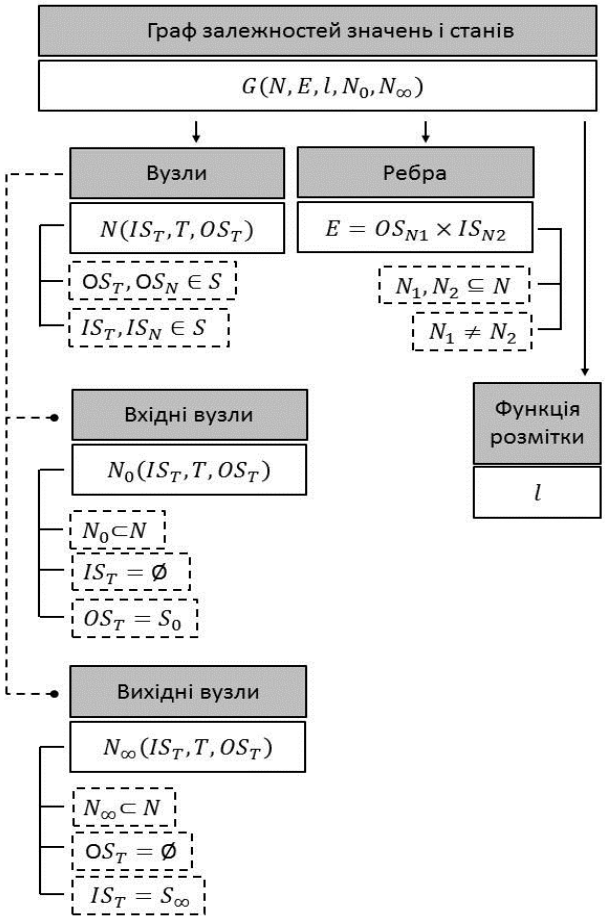

Рис. 1. Внутрішня структура графу залежностей значень і станів

На відміну від відомих моделей формування параметрів навчальних прикладів, в даній моделі передбачена можливість подання програмного забезпечення у вигляді графу залежностей значень та станів $\mathbf{G(N,E,l,N_0,N_\infty)}$, структура якого показана на рис. 1.

Основними складовими графу є множини вузлів $\mathbf{N(IS_T, T, OS_T)}$, що відповідають операціям з вхідною та вихідною інформацією вузлів; ребра $\mathbf{E = OS_{N1} \times IS_{N2}}$, що відповідають результатам операцій; функції розмітки l, що відповідають кожній операції переходу; вхідні вузли $N_0(\mathbf{IS_T, T, OS_T})$, що відповідають входам функцій; вихідні вузли $\mathbf{N_\infty(IS_T, T, OS_T)}$, що відповідають виходам функцій; T - вузли, що представляють операції, S - вузли, що представляють результати операції, S_0 - вузли, що представляють вхідні дані, IS_T - операції що відповідають входу вузла з множини T, OS_T - операції що відповідають виходу вузла з множини T, IS_N - операції що відповідають входу вузла з множини N, OS_N - операції що відповідають виходу вузла з множини N. В базовому варіанті $\mathbf{\Psi_{dvs} = \{N, E, l, N_0, N_\infty\}}$. Вказане передбачення забезпечує можливість нейромережевого розпізнавання обфускованого програмного коду, характерного для сучасних поліморфних вірусів.

Розроблено метод визначення архітектурних параметрів глибокої нейронної мережі. Метод базується на розроблених принципах підвищення ефективності НМЗ розпізнавання комп'ютерних вірусів, правилах визначення ефективних видів ГНМ та моделі формування параметрів навчальних прикладів ГНМ.

Перетворення інформації в цьому методі описується виразом: $\langle \mathbf{DNN}_{ent}, \mathbf{A}_{DNN}, \mathbf{R}, \mathbf{H}, \boldsymbol{\alpha}, \mathbf{V}, \Delta_d, \mathbf{M}_v, \mathbf{M}_p, \tau_{avl}, Q_{avl}, \varepsilon_{avl} \rangle \rightarrow \langle \mathbf{DNN}_{ea}, \mathbf{A}_{ea} \rangle$, де $\mathbf{H}$ – множина критеріїв ефективності виду ГНМ, $\mathbf{R}$ – множина значень критеріїв ефективності, $\boldsymbol{\alpha}$ – множина вагових коефіцієнтів критеріїв ефективності видів ГНМ, $\mathbf{DNN_{ent}} = \{dnn_1, dnn_2, dnn_3, dnn_4\}$ – множина доступних типів ГНМ, $\mathbf{A_{DNN}} = \{\mathbf{A}_{dnn_1}, \mathbf{A}_{dnn_2}, \mathbf{A}_{dnn_3}, \mathbf{A}_{dnn_4}\}$ – множина, що містить архітектурні параметри різних типів ГНМ, $\mathbf{A}_{dnn_1}, \mathbf{A}_{dnn_2}, \mathbf{A}_{dnn_3}, \mathbf{A}_{dnn_4}$ – множини архітектурних параметрів ГНМ без преднавчання, ГНМ з преднавчанням, ЗНМ та рекурентних ЗНМ, $\mathbf{DNN}_{ea}$ – множина найбільш ефективних та апробованих видів ГНМ, $\mathbf{A}_{ea}$ –

архітектурні параметри найбільш ефективних та апробованих видів ГНМ, Δ_d – допустиме значення функції ефективності виду ГНМ, $\mathbf{V}$ – множина видів комп'ютерних вірусів, що мають бути розпізнані, M_v – множина доступних портретів комп'ютерних вірусів, M_p – множина портретів безпечних програм, τ_{avl} – допустимий термін побудови ГНМ, Q_{avl} – допустима ресурсоємність побудови ГНМ, ε_{avl} – допустима похибка розпізнавання. Метод передбачає виконання 6 етапів (рис. 2).

Етап 1 – визначення основних умов застосування ГНМ.

На основі аналізу $\mathbf{V}$, M_v, M_p, τ_{avl}, Q_{avl}, ε_{avl} визначаються множини вхідних ($\mathbf{X}$) та вихідних параметрів ($\mathbf{Y}$) ГНМ, кількість вхідних (N_x) та вихідних параметрів (N_y), кількість видів комп'ютерних вірусів, що мають бути розпізнані (K_v), загальна кількість доступних прикладів комп'ютерних вірусів та безпечних програм (L), мінімальна (L_{min}) допустима кількість навчальних прикладів, допустима похибка навчання (ε_l), λ – очікувана тривалість однієї навчальної ітерації, ϑ_w – середній час, необхідний на створення одного навчального прикладу з очікуваним вихідним сигналом, ϑ_n – середній час, необхідний на створення одного навчального прикладу без очікуваного вихідного сигналу.

Етап 2 – визначення доцільності використання нейромережевої моделі типу ГНМ. Вхідними даними етапу є τ_{avl}, $\mathbf{DNN}_{ent}$, ϑ_w, λ, ϑ_n, L_{min}, N_x, N_y. Етап розділено на три кроки, кожен із яких співвідноситься з перевіркою допустимості компонентів множини $\mathbf{DNN}_{ent}$. Виходом етапу є множина допустимих видів ГНМ $\mathbf{DNN}_{avl}$.

Етап 3 – визначення найбільш ефективної архітектури ГНМ. Вхідні дані: $\mathbf{DNN}_{avl}$, $\boldsymbol{\alpha}$ та Δ_d. Етап розділено на три кроки. На першому та другому кроках для кожного типу допустимого виду ГНМ розраховується значення функції ефективності ($E(\mathbf{DNN}_i)$) та перевіряється допустимість її значення. На третьому кроці визначається множина найбільш ефективних видів архітектури ГНМ ($\mathbf{DNN}_{eff}$), що і є виходом етапу.

Етап 4 – формування параметрів навчальних прикладів. Етап орієнтовано на формування параметрів прикладів, що входять до складу навчальної та тестової вибірки ГНМ. Виконання етапу забезпечує можливість визначення параметрів архітектури найбільш ефективного виду ГНМ та можливість проведення експериментальних досліджень, спрямованих на апробацію розробленої ГНМ. На вхід етапу подаються M_v, M_p, K_v, L, $\mathbf{X}$, $\mathbf{Y}$, N_x, N_y, $\mathbf{DNN}_{eff}$. Виходом етапу є $\mathbf{S}=\{\mathbf{S_w},\mathbf{S_n}\}$ – множина навчальних даних ГНМ, де $\mathbf{S_w}=\left\{\mathbf{s}_w^{(V_k)},\mathbf{s}_n^{(P)}\right\}$ – множина, що містить приклади портретів програмного забезпечення з очікуваним вихідним сигналом, $\mathbf{S_n}=\left\{\mathbf{s}_n^{(V_k)},\mathbf{s}_n^{(P)}\right\}$ – множина, що містить приклади портретів програмного забезпечення без очікуваного вихідного сигналу, $\mathbf{s}_w^{(V_k)}$, $\mathbf{s}_n^{(V_k)}$ – навчальні приклади виду $\mathbf{S_w}$, $\mathbf{S_n}$ для комп'ютерних вірусів *k*-го виду, $\mathbf{s}_w^{(P)}$, $\mathbf{s}_n^{(P)}$ – навчальні приклади виду $\mathbf{S_w}$, $\mathbf{S_n}$ для безпечних програм.

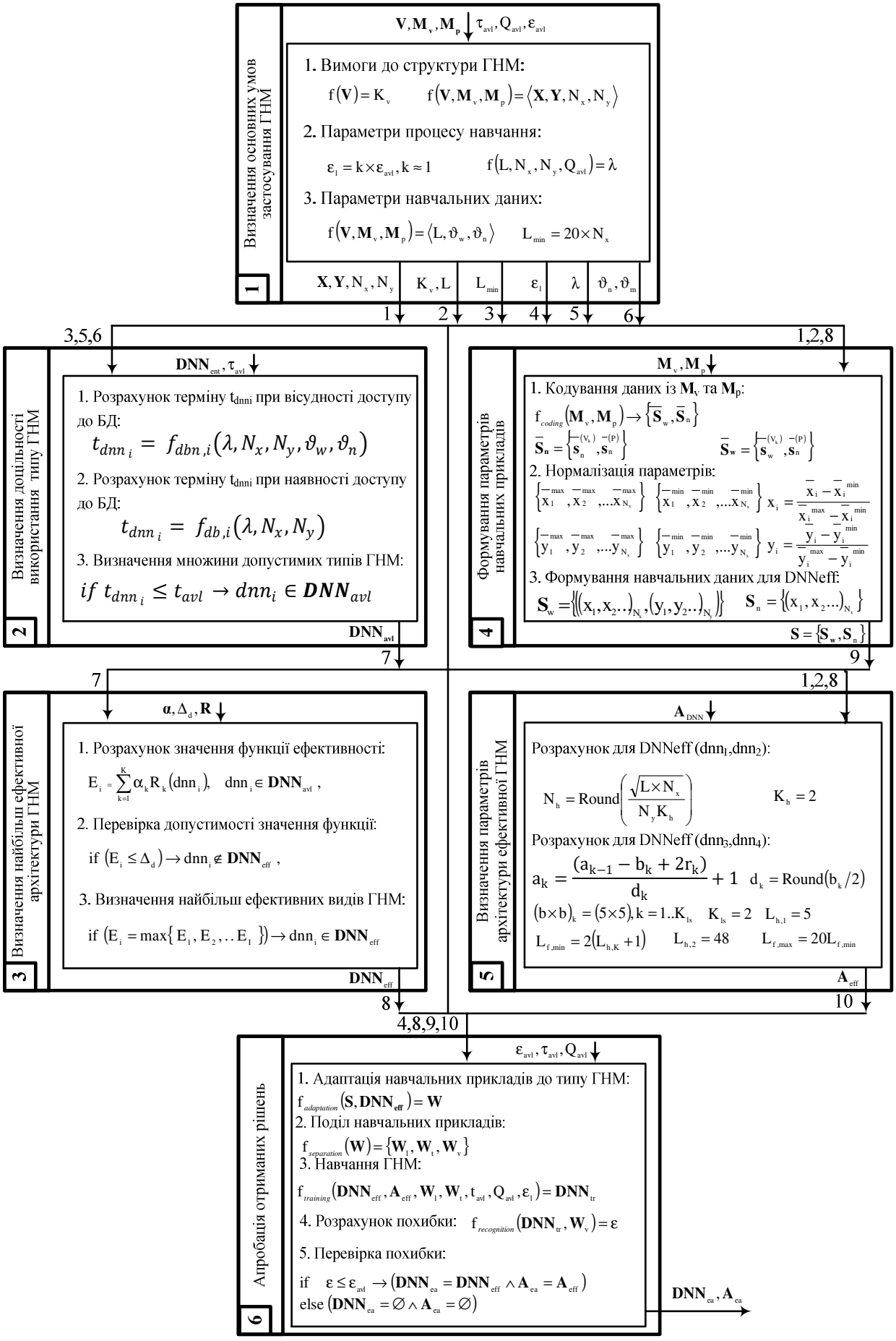

Рис. 2. Структурно-аналітична схема методу визначення архітектурних параметрів глибокої нейронної мережі

Етап 5 – визначення параметрів архітектури найбільш ефективного виду ГНМ. Вхідними даними етапу є $\mathbf{DNN}_{\mathbf{eff}}$, $\mathbf{A}_{\mathbf{DNN}}$, N_x, N_y, L, а виходом – $\mathbf{A}_{\text{eff}}$. При виконанні етапу використовуються вирази, що визначають значення архітектурних параметрів для тих видів ГНМ, що входять до складу $\mathbf{DNN}_{\mathbf{eff}}$.

Етап 6 – апробація отриманих рішень. Етап полягає у реалізації експериментальних досліджень, спрямованих на підтвердження достатньої точності розпізнавання розроблених ГНМ при їх застосуванні в очікуваних умовах. Входом етапу є ε_{avl}, τ_{avl}, Q_{avl}, $\mathbf{DNN}_{\text{eff}}$, $\mathbf{A}_{\text{eff}}$, $\mathbf{S}$, ε_1, а виходом $\mathbf{DNN}_{\text{ea}}, \mathbf{A}_{\text{ea}}$. Етап розділено на 5 кроків.

На першому кроці множина $\mathbf{S}$ адаптується до типу ГНМ: $f_{adaptation}(\mathbf{S}, \mathbf{DNN}_{\mathbf{eff}}) = \mathbf{W}$.

На другому кроці $\mathbf{W}$ розділяється на навчальну ($\mathbf{W}_{\text{l}}$), тестову ($\mathbf{W}_{\text{t}}$),та валідаційну вибірку ($\mathbf{W}_{\text{v}}$): $f_{separation}(\mathbf{W}) = \{\mathbf{W}_{\text{l}}, \mathbf{W}_{\text{t}}, \mathbf{W}_{\text{v}}\}$.

На третьому кроці з урахуванням τ_{avl} та Q_{avl} проводиться навчання та визначається похибка ГНМ на валідаційній вибірці: $f_{training}(\mathbf{DNN}_{\text{eff}}, \mathbf{A}_{\text{eff}}, \mathbf{W}_{\text{l}}, \mathbf{W}_{\text{t}}, \tau_{\text{avl}}, Q_{\text{avl}}) = \mathbf{DNN}_{\text{tr}}$, $f_{recognition}(\mathbf{DNN}_{\text{tr}}, \mathbf{W}_{\text{v}}) = \varepsilon$.

На четвертому та п'ятому кроці перевіряється допустимість похибки розпізнавання.

Якщо похибка перевищує допустиму, то $\mathbf{DNN_{ea}}=\varnothing$ та $\mathbf{A_{ea}}=\varnothing$.

Даний метод забезпечує можливість зменшення обсягу експериментальних досліджень, пов'язаних з визначенням архітектурних параметрів ГНМ, призначеної для використання в НМЗ розпізнавання комп'ютерних вірусів.

Подальший розвиток отримав метод нейромережевого розпізнавання комп'ютерних вірусів. Метод базується на розробленій моделі формування параметрів навчальних прикладів глибокої нейронної мережі та розробленому методі визначення архітектурних параметрів глибокої НМ.

Вхідними даними методу є $\mathbf{Q}_{\text{НМЗ}}$ – вимоги до НМЗ розпізнавання комп'ютерних вірусів, $\mathbf{X}_{\text{АПЗ}}$ – характеристики апаратно-програмного забезпечення НМЗ, $\mathbf{R}_{\text{НМЗ}}$ – ресурси на розробку НМЗ, $\mathbf{E}_{\text{НМЗ}}$ – експертні дані для побудови НМЗ, $\mathbf{DNN}_{\text{ent}}$ – доступні види НММ, $\mathbf{A}_{\text{DNN}}$ – параметри $\mathbf{DNN}_{\text{ent}}$, $\mathbf{V}$ – множина видів комп'ютерних вірусів, що мають бути розпізнані, **БД** – бази даних, в яких міститься інформація, що може бути використана для формування портретів комп'ютерних вірусів та портретів безпечних програм, $\mathbf{G}$ – процедури, що визначають ефективність НМЗ, $\mathbf{U}$ – визначена множина критеріїв ефективності, D_e – експертні дані для оцінки значущості критеріїв ефективності, що входять до складу $\mathbf{U}$.

Структурна схема методу нейромережевого розпізнавання комп'ютерних вірусів показана на рис. 3. Метод передбачає виконання 5 етапів.

Етап 1 – визначення умов створення та застосування НМЗ. На цьому етапі за допомогою процедури експертного оцінювання визначаються умови створення та застосування НМЗ розпізнавання комп'ютерних вірусів. Вхідними даними етапу є $\mathbf{Q}_{\text{НМЗ}}$, $\mathbf{X}_{\text{АПЗ}}$, $\mathbf{R}_{\text{НМЗ}}$, $\mathbf{E}_{\text{НМЗ}}$. Виходом етапу є параметри, що відповідають умовам: $\Theta_{\text{нв}}$ – формування навчальної вибірки, $\Theta_{\text{ГНМ}}$ – розробки ГНМ, $\Theta_{\text{АПЗ}}$ – використання АПЗ, $\Theta_{\text{чп}}$ – часу застосування НМЗ.

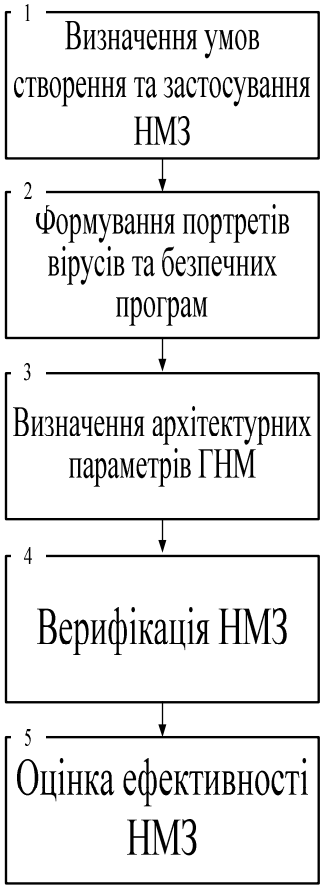

Рис. 3. Структурна схема методу нейромережевого розпізнавання вірусів

Етап 2 – формування портретів вірусів та безпечних програм. Виконання етапу полягає в аналізі доступних баз даних для формування множин портретів комп'ютерних вірусів та портретів безпечного програмного забезпечення. Входом даного етапу є множина доступних **БД**, а виходом кортеж $\langle \mathbf{M}_{v}, \mathbf{M}_{p} \rangle$.

Етап 3 – визначення архітектурних параметрів ГНМ. Даний етап співвідноситься із розробленим методом визначення архітектурних параметрів ГНМ. Етап розділено на 5 кроків, що відповідають 1-5 етапам вказаного методу. Виходом етапу є кортеж $\langle \mathbf{DNN}_{eff}, \mathbf{A}_{eff} \rangle$, а входом $\langle \mathbf{DNN}_{ent}, \mathbf{A}_{DNN}, \mathbf{R}, \mathbf{H}, \boldsymbol{\alpha}, \mathbf{V}, \Delta_{d}, \mathbf{M}_{v}, \mathbf{M}_{p}, \tau_{avl}, Q_{avl}, \varepsilon_{alw} \rangle$.

Етап 4 – верифікації НМЗ. Верифікація НМЗ проводиться з позицій допустимості похибки розпізнавання. Вхідні дані $\langle \mathbf{DNN}_{eff}, \mathbf{A}_{eff} \rangle$, $\Theta_{апз}$, $\Theta_{чп}$. Вихід $\mathbf{H}_{eff}$ – параметри верифікованих НМЗ.

Етап 5 – оцінка ефективності НМЗ. Оцінка реалізується з позицій забезпечення в побудованому НМЗ процедур, що забезпечують ефективне розпізнавання комп'ютерних вірусів в очікуваних умовах застосування. На вхід подається $\mathbf{H}_{eff}$, **G**, **U**, $\mathbf{D}_{e}$. Виходом етапу є $\mathbf{S}_{eff}$ – сигнал, що свідчить про достатню чи недостатню ефективність побудованого НМЗ.

Таким чином, отримав подальший розвиток метод нейромережевого розпізнавання комп'ютерних вірусів, який за рахунок визначення параметрів ефективної нейромережевої моделі на базі ГНМ забезпечує адаптацію створених на його основі НМЗ розпізнавання комп'ютерних вірусів до очікуваних умов застосування.

Четвертий розділ присвячено практичній реалізації та експериментальним дослідженням розроблених рішень. Розроблено методику проведення експерименту, обґрунтовано доцільність вибору бази експерименту, визначено мету та задачі експерименту, вхідні та вихідні параметри, гіпотезу і критерії дослідження, достатність експериментальних об'єктів та послідовність необхідних дій.

На першому етапі досліджень з урахуванням особливостей запропонованого нейромережевого методу розпізнавання комп'ютерних вірусів розроблена експериментальна нейромережева система (НМС), яка представлена структурно на рис. 5. Дана НМС дозволила перевірити отримані теоретичні результати. До елементів розробленої НМС належать такі її складові як БВП - блок визначення параметрів, БРНП - блок розробки навчальних прикладів, БРА - блок розробки архітектури ГНМ, БРКВ - блок розпізнавання комп'ютерних вірусів, ВДП - модуль визначення діагностичних параметрів, що використовуються для розпізнавання заданих типів комп'ютерних вірусів ($\boldsymbol{\Phi}_{k}$),

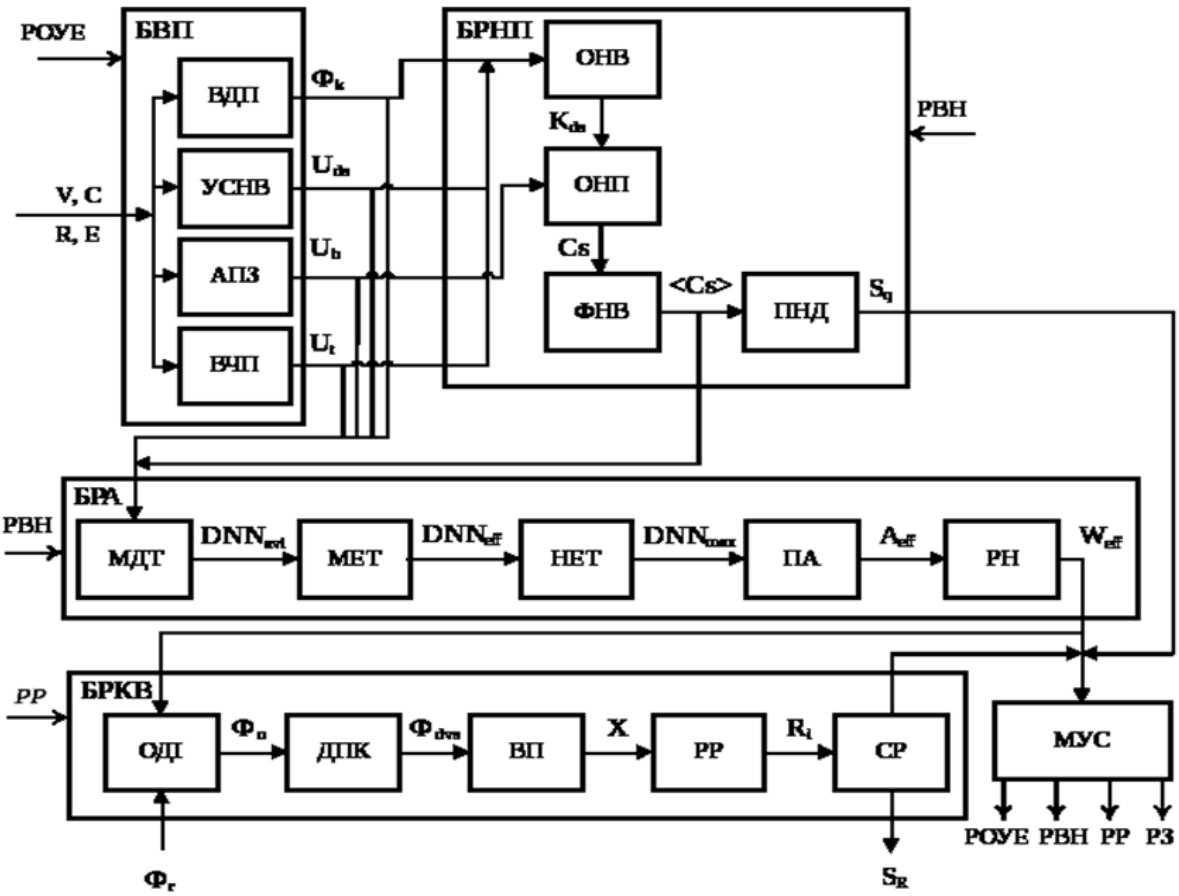

Рис. 5. Структура нейромережевої системи розпізнавання комп'ютерних вірусів

УСНВ - модуль визначення умов створення навчальної вибірки, що характеризує умови створення вибірки ГНМ ($\boldsymbol{U}_{ds}$), АПЗ - модуль опису апаратних засобів, що описують характеристики доступного апаратно-програмного забезпечення НСМ ($\boldsymbol{U}_{h}$), ВЧП - модуль визначення часових параметрів, що визначають допустимі часові показники побудови та створення НМС розпізнавання комп'ютерних вірусів ($\boldsymbol{U}_{t}$), ОНВ - модуль оцінки обсягу навчальної вибірки, ОНП - модуль обробки навчальних прикладів, ФНВ - модуль формування навчальної вибірки, що містить множини навчальних даних для кожного із допустимих типів ГНМ (<$\boldsymbol{Cs}$>), ПНД - модуль перевірки навчальних даних, МДТ - модуль формування множини допустимих типів ГНМ, МЕТ - модуль формування множини ефективних типів ГНМ, НЕТ - модуль визначення найбільш ефективного типу ГНМ, ПА - модуль визначення параметрів архітектури, РН - модуль реалізації навчання, призначений для навчання ГНМ з визначеною архітектурою, ОДІ - модуль отримання діагностичної інформації параметрів, ДПК - модуль деобфускації програмного коду, ВП - модуль визначення вхідних параметрів, РР - модуль реалізації розпізнавання, СР - модуль сигналізації, що видає сигнал про результати розпізнавання наперед визначених типів ПЗ ($\boldsymbol{S_R}$). До основних вхідних та вихідних параметрів модулів належать: множина зареєстрованих діагностичних параметрів ($\boldsymbol{\Phi_r}$), параметри вимог до НМС розпізнавання ($\boldsymbol{V}$), характеристики доступного програмно-апаратного забезпечення системи ($\boldsymbol{C}$), ресурси ($\boldsymbol{R}$), що виділяються на розробку системи та заданий термін її створення ($\boldsymbol{T}$), множина експертних даних ($\boldsymbol{E}$), що використовується для обробки $\boldsymbol{V}$, $\boldsymbol{R}$ та $\boldsymbol{C}$.

Передбачено, що розроблена НМС може функціонувати в таких режимах: окреслення умов експлуатації (РОУЕ); визначення налаштувань (РВН); розпізнавання комп'ютерних вірусів (РКВ); зупинки (РЗ). Переключення режимів функціонування реалізується за допомогою модулю управління системою (МУС).

Основною частиною НМС став створений за допомогою мови Python програмний комплекс для реалізації НММ. В процесі розробки комплексу використана загальнодоступна бібліотека TensorFlow (розробка компанії Google) призначена для моделювання ГНМ. Апаратне забезпечення експериментальної платформи базувалось на персональному комп'ютері (AMD FX-9800P (2.7 - 3.6 ГГц) / RAM 32 ГБ / HDD 1 ТБ / AMD Radeon RX 540, 4 ГБ), що функціонував під управлінням операційної системи Windows 10.

Задачею другого етапу досліджень було доведення ефективності розробленого методу визначення архітектурних параметрів ГНМ за умов: ГНМ використовується для розпізнавання Windows-орієнтованих комп'ютерних вірусів на основі аналізу використаних програмою потенційно небезпечних API-функцій операційної системи; на вхід ГНМ подається інформація, отримана в результаті сканування піддослідних файлів; допустимий термін створення ГНМ складає 1 місяць ($\tau_{avl} = 2{,}6 \times 10^6$ с); для навчання та тестування ГНМ використовується опублікована компанією Microsoft БД комп'ютерних вірусів BIG-2015.

Таблиця 1

Характеристика BIG-2015

Назва вірусу	Кількість прикладів
Ramnit	1541
Lollipop	2478
Kelihos_ver3	2942
Vundo	475
Simda	420
Tracur	751
Kelihos_ver1	398
Obfuscator.ACY	1228
Gatak	1013

В БД BIG-2015 представлено приклади сигнатур 9 комп'ютерних вірусів. Для формування вказаних сигнатур використано програмний комплекс Interactive DisAssembler, що дозволяє вилучити із бінарного файлу метадані, які стосуються інструкцій мови Assembler, вмісту регістрів та даних і функцій, імпортованих із DLL. Характеристики БД BIG-2015 частково наведено в табл. 1. При цьому застосування до дизасембльованого коду технології Flirt дозволяє визначити наявність в ньому потенційно небезпечних функцій управління розділами, управління файлами, роботи з реєстром, використання системної інформації, використання мережевих з'єднань, управління пам'яттю, використання сервісів, управління системою захисту об'єктів. В першому наближенні прийнято, що кількість таких функцій дорівнює 300. Таким чином, кількість вхідних параметрів ГНМ $N_x = 300$, а кількість вихідних параметрів $N_y = 10$.

Оскільки для формування навчальної вибірки передбачається використання БД, то $\vartheta_w = 0$, $\vartheta_n = 0$. Також визначено, що $\lambda = 0{,}01$ с. Підставивши ці дані в вирази, що представлені в другому розділі і визначають допустимість різних типів ГНМ розраховано час створення кожного із доступних типів ГНМ: $\tau_{cr}(dnn_1, dnn_4) \approx 3720$с, $\tau_{cr}(dnn_2) \approx 37200$с, $\tau_{cr}(dnn_3) \approx 186$с. Оскільки для всіх типів ГНМ цей час менший від допустимої, то доцільність

використання ГНМ можна вважати доведеною. На наступному етапі розробки ГНМ, за допомогою експертного методу парного порівняння, було проведене визначення значущості кожного із критеріїв ефективності. Отримані результати показані в табл. 2.

Таблиця 2

Вагові коефіцієнти критеріїв ефективності ГНМ

α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9
0,08	0,07	0,02	0,05	0,05	0,14	0,03	0,05	0,05
α_{10}	α_{11}	α_{12}	α_{13}	α_{14}	α_{15}	α_{16}	α_{17}	α_{18}
0,02	0,05	0,05	0,07	0,07	0,05	0,05	0,05	0,05

З урахуванням отриманих результатів функції ефективності апробованих видів ГНМ, дорівнюють $H_{dnn_1} = 0{,}72$, $H_{dnn_2} = 0{,}77$, $H_{dnn_3} = 0{,}6$, $H_{dnn_4} = 0{,}58$. Таким чином, найбільш ефективною архітектурою є ГНМ з можливістю навчання в процесі експлуатації.

Кількість схованих нейронних шарів обрано з позицій максимального спрощення структури ГНМ і дорівнює $K_h = 2$. Для розрахунку кількості нейронів у кожному із схованих шарів використано вираз $N_h = \text{Round}\left(\sqrt{L \times N_x}/N_y\right)$, де L – кількість навчальних прикладів. Відзначимо, що в БД кількість навчальних прикладів, що відповідають вірусам дорівнює 10868. Враховуючи, що в навчальній вибірці необхідне пропорційне представлення зразків комп'ютерних вірусів та безпечних програм визначено, що L=2*10868=21736. Таким чином, $N_h = 255$.

Навчання розробленої ГНМ проводилось на протязі 100 епох. Приблизно після 90 навчальних епох помилка навчання стабілізувалась на рівні 0.01. Після цього на вхід ГНМ із БД BIG-2015 були подані тестові приклади, що не використовувались при навчанні. Похибка розпізнавання для різних вірусів показана на рис. 6. Аналіз результатів з рис. 6 вказує на те, що найбільша похибка розпізнавання характерна для вірусів Simda, Tracur та Vundo. Це можна пояснити невеликою кількістю навчальних прикладів, що відповідають цим вірусам. При цьому середня похибка розпізнавання всіх видів вірусів дорівнює 0,036, що відповідає похибці сучасних антивірусних засобів. Зазначимо, що за рахунок використання запропонованого методу проектування архітектури ГНМ вдалось уникнути довготривалих чисельних експериментів, спрямованих на визначення доцільності її використання і на визначення її структурних параметрів, та приблизно в 1,5 рази зменшити обчислювальні витрати, пов'язані з визначенням вказаних архітектурних параметрів.

На заключному етапі досліджень проведено розрахунки, спрямовані на оцінювання ефективності запропонованого методу нейромережевого розпізнавання комп'ютерних вірусів. Для цього використаний загально-визнаний метод розрахунку ефективності НМЗ оцінювання параметрів безпеки інформаційних систем та запропоновані критерії ефективності, що характеризують особливості сучасних НМЗ розпізнавання комп'ютерних вірусів.

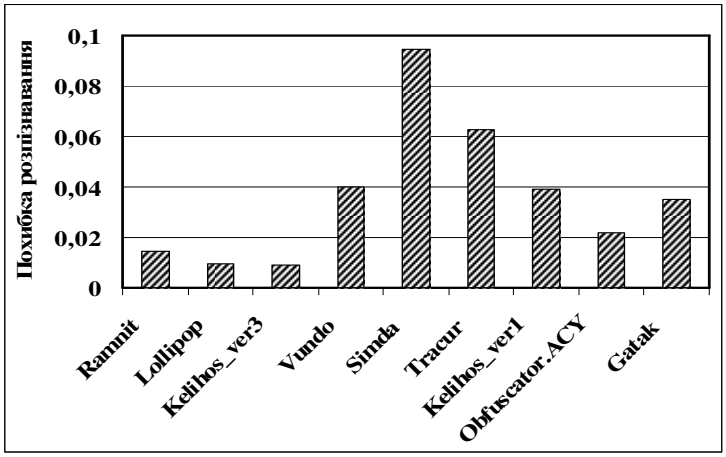

Рис. 6. Похибка розпізнавання

Результати проведених розрахунків вказують на те, що ефективність розробленого НМЗ приблизно в 1,14 рази вища ніж у подібних відомих засобів. Таким чином, результати досліджень підтверджують можливість підвищення ефективності розпізнавання комп'ютерних вірусів за рахунок застосування розроблених НММ та НМЗ.

У додатках вміщено акти впровадження результатів дисертаційної роботи та фрагменти кодів програм, що відображають практичну частину дисертаційного дослідження.

ВИСНОВКИ

Результатом виконаної дисертаційної роботи є розв'язання актуальної науково-практичної задачі підвищення ефективності протидії комп'ютерним вірусам, за рахунок розробки і дослідження нових нейромережевих моделей, методів і засобів розпізнавання комп'ютерних вірусів, здатних оперативно пристосовуватись до умов використання і реагувати на виникнення нових видів вірусів.

У процесі виконання дисертаційної роботи отримані такі вагомі результати:

1. Проведено аналіз сучасних нейромережевих моделей та методів розпізнавання комп'ютерних вірусів, що показав наявність низки недоліків, пов'язаних з високою потребою в обчислювальних ресурсах, низькою адаптованістю до проведення аналізу обфускованого програмного коду та недостатньою ефективністю розпізнавання. В результаті виконаного аналізу обґрунтована перспективність застосування в контурі розпізнавання систем антивірусного захисту нейромережевих засобів. При цьому показано можливість застосування нейромережевих моделей як в поведінкових аналізаторів, так і при використанні сигнатурного аналізу.

2. Вперше розроблено концептуальну модель оцінювання глибоких нейронних мереж, яка дозволяє визначити множину сучасних нейромережевих моделей для побудови ефективних антивірусних засобів.

3. Вперше розроблено модель формування параметрів навчальних прикладів глибокої нейронної мережі, яка дозволяє будувати засоби нейромережевого аналізу обфускованого програмного коду.

4. Вперше розроблено метод визначення архітектурних параметрів глибокої нейронної мережі, призначеної для розпізнавання комп'ютерних вірусів, що дозволяє сформувати набір величин, які забезпечують пристосованість такої мережі до визначених умов застосування. Використання розробленого методу проектування дозволяє приблизно в 1,5 рази зменшити

обчислювальні витрати, пов'язані з визначенням значень архітектурних параметрів.

5. Отримав подальший розвиток метод нейромережевого розпізнавання комп'ютерних вірусів, який забезпечує достатню похибку розпізнавання при різних умовах застосування з урахуванням обмежень щодо створення навчальної вибірки та обмежень щодо обчислювальних ресурсів системи антивірусного захисту. Показано, що ефективність методу приблизно в 1,14 рази вища, ніж у подібних методів розпізнавання.

6. На базі запропонованого методу нейромережевого розпізнавання комп'ютерних вірусів розроблено алгоритмічне забезпечення та відповідна програмна модель системи, що дозволила з достатньою точністю розпізнавати комп'ютерні віруси та забезпечити оперативність створення алгоритмів функціонування апаратно-програмних засобів захисту інформації.

7. Експериментальне дослідження програмної моделі системи, а також впровадження та успішне практичне використання відповідних розробок підтвердило достовірність теоретичних положень та гіпотез, практичних розробок і висновків дисертаційної роботи.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. I. Dychka, D. Chernyshev, I. Tereikovskyi, L. Tereikovska , V. Pogorelov, «Malware Detection Using Artificial Neural Networks», *Advances in Computer Science for Engineering and Education II. ICCSEEA 2019. Advances in Intelligent Systems and Computing*, Vol. 938. Springer, Cham, pp.3-12, 2019. (SCOPUS) DOI: *https://doi.org/10.1007/978-3-030-16621-2_1*.

2. I. Dychka, I. Tereikovskyi, L. Tereikovska, V. Pogorelov, S. Mussiraliyeva, «Deobfuscation of Computer Virus Malware Code with Value State Dependence Graph», *Advances in Computer Science for Engineering and Education. ICCSEEA 2018. Advances in Intelligent Systems and Computing,* Vol. 754. Springer, Cham, pp.370-379, 2018. (SCOPUS) DOI: *https://doi.org/10.1007/978-3-319-91008-6_37*.

3. Hu. Zhengbing, I. Tereykovskiy, L. Tereykovska, V. Pogorelov, «Determination of Structural Parameters of Multilayer Perceptron Designed to Estimate Parameters of Technical Systems», *International Journal of Intelligent Systems and Applications(IJISA)*, Vol. 9, No.10, pp. 57-62*)*, 2017. (SCOPUS)
DOI: *10.5815/ijisa.2017.10.07, ISSN 2074-9058.*

4. І. Терейковський, О. Заріцький, Л. Терейковська, В. Погорелов, «Метод розробки архітектури глибокої нейронної мережі, призначеної для розпізнавання комп'ютерних вірусів», *Захист інформації*, Т. 20, № 3, С. 188-199, 2018.

5. Л. Терейковська, Є. Іванченко, В. Погорелов «Метод адаптації глибокої нейронної мережі до розпізнавання комп'ютерних вірусів», *Науковий журнал «Комп'ютерно-інтегровані технології: освіта, наука, виробництво»*, Луцьк, Випуск № 35, С. 198-205, 2019.

6. В. Погорелов, «Дослідження нейромережевих засобів розпізнавання кібератак на мережеві ресурси інформаційних систем», *Системні технлогії,* №5, С. 61–69, 2017.

7. В. Погорелов, «Проблематика використання нейромережевих систем

розпізнавання кібератак», *Науковий журнал «Комп'ютерно-інтегровані технології: освіта, наука, виробництво»*, №27, С. 67–74, 2017.

8. O. Tereykovskiy, V. Pogorelov «Cyberattack recognition with radial basis function neural network», *Projekt interdyscyplinarny projektem XXI wieku*, Tom 2, pp. 255-262, 2017. (Колективна монографія)

9. Б. Айтчанов, И. Бапиев, А. Корченко, В. Погорелов, Л. Терейковская, «Концептуальная модель обеспечения эффективности нейросетевого распознавания кибератак», *Труды международной научно-практической конференции «Математические методы и информационные технологии макроэкономического анализа и экономической политики», посвященной празднованию 80-летнего юбилея академика НАН РК Абдыкаппара Ашимовича Ашимова*, *11-12*, С. 321–325, 2017.

10. B. Aitchanov, I. Bapiev, I. Terejkowski, L. Terejkowska, V. Pogorelov, «Calculation of expected output signal of neural network model for detecting of cyber-attack on network resources», *Information Technologies, Management and Society, The 15th International Scientific Conference Information Technologies and Management,* pp. 59–62, 2017.

11. V. Pogorelov, M. Karpinski, E. Ivanchenko, «Method of neural networks utilization for malware recognition», *The 10 th International Scientific Conference «ITSec» March*, pp. 58, 2020.

12. В. Погорєлов, «Використання графу залежностей значень і станів у задачі розпізнавання поліморфних комп'ютерних вірусів», *Стан та удосконалення безпеки інформаційно-телекомунікаційних систем (SITS' 2020)*, Миколаїв, 2020, С. 34-35.

13. В. Погорелов, «Нейромережевий метод розпізнавання комп'ютерних вірусів», *VI Міжнародна науково-практична конференція «Актуальні питання забезпечення кібербезпеки та захисту інформації*», 2020, С. 88-93.

14. I. Tereikovskyi, V. Pogorelov, O. Tereikovskyi, «Determination of structural parameters of a multilayer cyber threat detection perceptron», *Aviation in the XXI-st Century,* 2018, pp. 3.3.1 – 3.3.4.

АНОТАЦІЯ

Погорелов В.В. Нейромережеві моделі та методи розпізнавання комп'ютерних вірусів. – Рукопис.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.21 – системи захисту інформації, Національний авіаційний університет, Київ, 2020.

У роботі вирішено актуальну науково-прикладну задачу підвищення ефективності протидії комп'ютерним вірусам, за рахунок дослідження і розробки нових нейромережевих моделей, методів і засобів розпізнавання комп'ютерних вірусів, здатних оперативно пристосовуватись до умов використання і реагувати на виникнення нових видів вірусів. У дисертаційній роботі проведено аналіз сучасних нейромережевих моделей та методів розпізнавання комп'ютерних вірусів, що показав наявність низки недоліків, пов'язаних з високою потребою в обчислювальних ресурсах, низькою

адаптованістю до проведення аналізу обфускованого програмного коду та недостатньою ефективністю розпізнавання. Розроблено концептуальну модель оцінювання глибоких нейронних мереж. Розроблено модель формування параметрів навчальних прикладів глибокої нейронної мережі. Розроблено метод визначення архітектурних параметрів глибокої нейронної мережі, призначеної для розпізнавання вірусів. Отримав подальший розвиток метод нейромережевого розпізнавання комп'ютерних вірусів. Розроблене спеціалізоване програмне забезпечення, що базується на створених нейромережевих методах та моделях, дозволило забезпечити достатню точність розпізнавання комп'ютерних вірусів та забезпечити оперативність створення алгоритмів функціонування апаратно-програмних засобів захисту інформації.

Ключові слова: захист інформації, нейронна мережа, комп'ютерний вірус, антивірусний захист, шкідливе програмне забезпечення.

ABSTRACT

Pogorelov V.V. Neural models and methods for computer viruses recognition. – Manuscript.

Thesis for scientific degree of candidate of technical sciences, specialty 05.13.21 – Information Security Systems, National Aviation University, Kyiv, 2020.

In terms of the work, it is resolved the actual scientific and applied task of increasing the effectiveness of computer viruses detection by researching and developing new neural network models, methods and means of recognizing computer viruses that can quickly adapt to the conditions of use and respond to the emergence of new types of viruses. The prospects of application in the circuit of anti-virus protection systems and neural network tools for malware recognition are substantiated. The possibility of using of neural network model both in behavioral analyzers and when using signature analysis is shown. Also for domestic anti-virus protection systems, the set of expected conditions of application for the specified neural network means is defined. A conceptual model for assessing deep neural networks has been developed, which, due to the interrelated principles of permissibility of use, determining a set of effective types and evaluating the effectiveness of a type of deep neural network, makes it possible to determine a variety of modern neural network models for building effective antivirus tools. The model for construction of parameters of educational examples for a deep neural network was developed that is based on formal representation of encoded values of API-functions calls, bytes of sequence of N-grams, opcodes, the main registers of the processor, and also results of static analysis of samples of malicious and safe programs, two-dimensional interpretation of binary code, parameters of the values state dependence graph. The model allows to build means of the neural network analysis of the obfuscated code. A method for determining the architectural parameters of a deep neural network designed for virus recognition has been developed, which because of the use of the proposed conceptual model for

assessing deep neural networks and model for construction of training examples used to implement the stages of determining the basic conditions of application, neural network model and the most effective architecture, as well as the constrcution of parameters of educational examples and determining the parameters of the architecture of the most effective type of deep neural network, allows you to form a set of values that ensure the adaptability of such a network to certain conditions of use. The method of neural network recognition of computer viruses was further developed, which provides sufficient error of recognition under different conditions by determining the conditions of creation and application of neural network means, processes of forming portraits of viruses and secure programs, as well as determining architectural parameters of deep neural network, verification and evaluation of neural network means. The method takes into account the limitations related to creation of a training sample and the limitations related to the computing resources of the anti-virus protection system. The specialized software is developed that is based on the created neural network methods and models and allows providing sufficient accuracy of computer viruses recognition and providing efficiency of algorithms creation for hardware and software means of information protection.

The experiments conducted as a part of the work, the implementation and the successful practical tests of the system prove the validity of the theoretical positions, the hypotheses, and the conclusions that were made in the dissertation.

Keywords: information security, neural network, computer virus, antivirus protection, malware.