ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет програмування
та комп'ютерних і телекомунікаційних систем

Кафедра кібербезпеки та комп'ютерних систем і мереж

КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА

Метод розпізнавання кінцевих пристроїв корпоративної мережі за принципом свій/чужий

КРМКІ. 190178.01.02.00

Галузь знань 12 – Інформаційні технології

Спеціальність 123 Комп'ютерна інженерія

Виконав: студент 2 курсу, група КІ1м-19-1 _____ Т. М. Кисіль
Підпис

Керівник доц., к. т. н, доцент кафедри КбКСМ _____ Ю. П. Кльоц
Підпис

Нормоконтролер доц., к. т. н, доцент кафедри КбКСМ _____ І. В. Муляр
Підпис

До захисту допускаю:
Зав. кафедри КбКСМ, канд. техн. наук, доцент _____ Ю. П. Кльоц
Підпис

4 12 2020 р.

Хмельницький 2020

2

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет: ПРОГРАМУВАННЯ ТА КОМП'ЮТЕРНИХ І ТЕЛЕКОМУНІКАЦІЙНИХ СИСТЕМ
Кафедра: КІБЕРБЕЗПЕКИ ТА КОМП'ЮТЕРНИХ СИСТЕМ І МЕРЕЖ
Освітній рівень: МАГІСТР
Галузь знань: 12 ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ
Спеціальність: 123 КОМП'ЮТЕРНА ІНЖЕНЕРІЯ
Освітня програма: ОСВІТНЬО-ПРОФЕСІЙНА

ЗАТВЕРДЖУЮ
Зав. кафедри КбКСМ
_____ Кльоц Ю. П..
«_4_» ___09___ 2020 р.

ЗАВДАННЯ
НА ДИПЛОМНИЙ ПРОЕКТ (РОБОТУ)

Кисіль Тетяні Миколаївні
Прізвище, ім'я, по батькові студента

1. Тема проекту (роботи) Метод розпізнавання кінцевих пристроїв корпоративної мережі за принципом свій/чужий

Керівник проекту (роботи) Кльоц Юрій Павлович, к.т.н., доцент
прізвище, ім'я, по батькові, науковий ступінь, вчене звання

Затверджена наказом ректора університету від 01.09.2020 р. № 118

2. Строк подання студентом проекту (роботи) на кафедру 01.12.2020 р.

3. Вихідні дані до проекту (роботи). Наукові джерела, що стосуються систем виявлення зловмисного програмного забезпечення та штучних імунних систем.

4. Зміст пояснювальної записки (перелік питань, які потрібно розробити). 1. Проаналізувати ймовірнісні підходи до визначення рівня небезпеки корпоративної мережі. 2. Запропонувати структуру системи виявлення вторгнень, яка використовує генетичний алгоритм для еволюції шаблонів виявлення та їх запам'ятовування, що дозволить ідентифікувати кінцевий пристрій мережі як свій або чужий. 3. Запропонувати алгоритм процедури виявлення «чужого» пристрою як такого, що здійснює неправомірні дії в мережі, наприклад вторгнення 4. Запропонувати нечітку модель розпізнавання пристроїв за принципом свій / чужий.

5. Перелік графічного матеріалу (із зазначенням обов'язкових креслень).

_____
_____
_____

3

6. Консультанти розділів дипломного проекту (роботи)

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Нормоконтроль	Муляр І.В., доцент кафедри КбКСМ		

7. Дата видачі завдання «03» вересня 2020 р.

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів (розділів) дипломного проекту (роботи)	Строк виконання етапів проекту (роботи)	Примітка
1	Вибір напряму дослідження та узгодження тематики КРМ з керівником	01.09.2020 – 02.09.2020	Виконано
2	Ознайомлення з предметною областю; формулювання мети та задач дослідження; визначення об'єкта та предмета дослідження	03.09.2020 –08.09.2020	Виконано
3	*Написання 1 розділу КРМ*	*09.09.2020 – 20.09.2020*	*Виконано*
4	Написання 2 розділу КРМ	21.09.2020 – 27.09.2020	Виконано
5	Робота над науковою статтею	28.09.2020 – 7.10.2020	Виконано
6	*Написання 3 розділу КРМ*	*08.10.2020 – 13.10.2020*	*Виконано*
7	Написання 4 розділу КРМ	14.10.2020 – 05.11.2020	Виконано
8	Написання вступу, висновків, формування переліку джерел посилання та додатків	06.11.2020 – 08.11.2020	Виконано
9	Попередній захист дипломної роботи	09.11.2020 –10.11.2020	Виконано
10	Подача роботи на: кафедру, антиплагіат, рецензування, нормоконтроль	12.11.2020 – 3.12.2020	Виконано
11	*Захист дипломної роботи*	*4.12.2020 –15.12.2020*	

Студент _____ Т. М. Кисіль
Підпис

Керівник проекту (роботи) _____ Ю. П. Кльоц
Підпис

РЕФЕРАТ

Тема дипломної роботи: Метод розпізнавання кінцевих пристроїв корпоративної мережі за принципом свій/чужий

Автор роботи: Кисіль Тетяна Миколаївна

Керівник роботи: Кльоц Юрій Павлович

Загальний обсяг роботи: 110 сторінок, 25 рисунків, 7 таблиць, 2 додатки, 28 посилань.

Ключові слова: корпоративна мережа, свій/чужий, правило Хеммінга, правило r- послідовного збігу, штучні імунні системи.

Об'єктом дослідження є корпоративні комп'ютерні мережі.

Предметом дослідження є розпізнавання кінцевих пристроїв корпоративної мережі за принципом свій / чужий.

Рішення, сформульовані в магістерській роботі, базуються на методах теорії ймовірності, випадкових процесів, імітаційного моделювання, теорії прийняття рішень.

Наукова новизна одержаних результатів. В магістерській роботі отримані наступні результати, що характеризуються науковою новизною:

1. Структура системи виявлення вторгнень, яка використовує генетичний алгоритм для еволюції шаблонів виявлення та їх запам'ятовування, що дозволить ідентифікувати кінцевий пристрій мережі як свій або чужий.

2. Запропоновано нечітку модель розпізнавання пристроїв за принципом свій / чужий.

Практична значимість результатів магістерської роботи полягає в отриманих моделях і алгоритмах щодо ідентифікації пристроїв за принципом свій/чужий

ABSTRACT

a qualification work of Tetyana Kysil

entitled «The method of recognizing the end devices of the corporate network on the principle of own / foreign»

Mentor: Klots Yurii Pavlovich

Total volume of work: 110 pages, 25 figures, 7 tables, 2 appendices, 28 references.

CORPORATE NETWORK, SELF / NONSELF, HAMMING'S RULE, *R*-CONTIGUOUS MATCH RULE, ARTIFICIAL IMMUNE SYSTEMS.

The object of research is corporate computer networks.

The subject of the study is the recognition of end devices of the corporate network on the principle of self / nonself.

The solutions formulated in the master's thesis are based on the methods of probability theory, random processes, simulation modeling, decision theory.

Scientific novelty of the obtained results. The following results are obtained in the master's thesis, which are characterized by scientific novelty:

1. The structure of the intrusion detection system, which uses a genetic algorithm for the evolution of detection patterns and their storage, which will identify the final device of the network as self or nonself.

2. A fuzzy model of device recognition according to the principle of self / nonself is proposed.

The practical significance of the results of the master's work lies in the obtained models and algorithms for the identification of devices on the principle of self / nonself

Зміст

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ

ІС – імунна система;

ШІС – штучна імунна система;

ВШІС – штучна імунна система, що виявляє вторгнення;

АГ – антиген;

АТ – антитіло;

ІГ – імуноглобулін;

МГС – маркери гістосумісності;

РТК – рецептори Т-клітин;

ДНК – дезоксирибонуклеїнова кислота;

РНК – рибонуклеїнова кислота;

ВВ – виявлення вторгнень;

СВВ – система виявлення вторгнень;

МСВВ – мережева система виявлення вторгнень;

ППІ – подія, що представляє інтерес;

ОС – операційна система;

УРІБ – управління ризиками інформаційної безпеки.

ВСТУП

Актуальність роботи

Недоліками відомих способів організації взаємозв’язку компонентів розподілених систем для виявлення зловмисного програмного забезпечення в корпоративних комп’ютерних мережах є використання централізованої архітектури, що контролюється адміністратором. Це призводить до недостатньо високої достовірності виявлення і локалізації зловмисних дій, бо збір інформації про стан мережі, визначення присутності шкідливих дій та їх блокування здійснюється для обробки єдиним центром, що може бути сповільнено через передачу зібраних даних цьому центру, обчислювальні ресурси на яких він розміщений, а також вплив на його роботу адміністратора мережі [9].

Корпоративна комп’ютерна мережа складається з пристроїв, що під’єднані до неї на основі дротового та бездротового з’єднання. Стаціонарні пристрої мережі швидко ідентифікуються на основі MAC адреси [2], щодо тих, які використовують LAN мережу і бездротовий зв'язок - то така можливість ідентифікації втрачає сенс. Потрібно знати, які пристрої несуть потенційну загрозу для мережі, тобто є «чужими», а які ні – тобто є «своїми». Часом приналежність до класу «свій» чи «чужий» визначається поведінкою самого пристрою, його поведінка є типовою чи нетиповою, нормальною чи аномальною.

Вважається, що багато механізмів, що присутні в біологічній імунній системі, добре підходять для використання в області комп'ютерного виявлення вторгнень у вигляді штучної імунної системи [12, 16, 23]. В даній дипломній роботі представлені механізми біологічної імунної системи, наведені їх паралелі в штучній імунній системі та проаналізовано, як вони можуть застосовуватися для виявлення нетипової поведінки пристрою (зокрема, вторгнень) в комп’ютерному середовищі. Штучна імунна система може бути впроваджена та застосована для виявлення нав'язливої поведінки в реальних мережних даних в асимульованому мережевому середовищі.

Використання дозрівання за спорідненістю дозволяє проводити виявлення аномалій, використовуючи менші набори детекторів з високим рівнем специфічності, зберігаючи при цьому високий рівень покриття та різноманітності, що збільшує кількість справжніх позитивів, зберігаючи при цьому низький рівень фальшивих негативів [14, 20].

Об'єктом дослідження є корпоративні комп’ютерні мережі.

Предметом дослідження є метод розпізнавання кінцевих пристроїв корпоративної мережі за принципом свій / чужий.

Мета магістерської роботи полягає в розробці рекомендацій для прийняття рішень щодо розпізнавання пристрою за принципом свій/чужий.

Відповідно до вказаної мети в роботі поставлені, обґрунтовані і вирішені наступні завдання:

1. Проаналізувати ймовірнісні підходи до визначення рівня небезпеки корпоративної мережі.

2. Запропонувати структуру системи виявлення вторгнень, яка використовує генетичний алгоритм для еволюції шаблонів виявлення та їх запам'ятовування, що дозволить ідентифікувати кінцевий пристрій мережі як свій або чужий.

3. Запропонувати алгоритм процедури виявлення «чужого» пристрою як такого, що здійснює неправомірні дії в мережі, наприклад вторгнення

4. Запропонувати нечітку модель розпізнавання пристроїв за принципом свій / чужий.

Основні нові результати, отримані в роботі та виносяться на захист:

1. Структура системи виявлення вторгнень, яка використовує генетичний алгоритм для еволюції шаблонів виявлення та їх запам'ятовування, що дозволить ідентифікувати кінцевий пристрій мережі як свій або чужий.

2. Запропоновано нечітку модель розпізнавання пристроїв за принципом свій / чужий.

Практична цінність і реалізація результатів роботи.

Практична цінність результатів магістерської роботи полягає в отриманих моделях і алгоритмах, щодо ідентифікації пристроїв за принципом свій/чужий.

Достовірність наукових положень, висновків і обґрунтованість отриманих в магістерській роботі результатів підтверджується коректною постановкою завдань, коректністю використовуваного математичного апарату, результатами моделювання та апробацією отриманих результатів на конференціях. Отримані в ході виконання дослідження результати не суперечать раніше отриманим даним, описаним в літературі іншими авторами.

Особистий внесок. Всі дослідження, викладені в магістерській роботі, проведені автором в процесі наукової діяльності. Результати, які виносяться на захист, отримані автором особисто, запозичений матеріал позначений в роботі посиланнями.

Апробація роботи. За темою кваліфікаційної роботи ОКР «Магістр» опубліковано 3 наукові статті і 1 тези доповідей.

Структура і обсяг роботи. Кваліфікаційна робота ОКР «Магістр» складається зі вступу, основної частини, що містить 4 розділи, висновків і переліку джерел посилання. Загальний обсяг роботи з додатками складає 110 сторінок. Об’єм роботи без додатків становить 79 сторінок. Робота містить 25 рисунків та 7 таблиць. Список використаних джерел включає 28 найменувань.

1 ДОСЛІДЖЕННЯ ОРГАНІЗАЦІЇ ТА АНАЛІЗ ТЕХНОЛОГІЙ ПОБУДОВИ КОРПОРАТИВНОЇ МЕРЕЖІ

1.1 Аналіз технологій побудови корпоративної мережі

Глобальні мережі

Глобальні мережі (Wide Area Network, WAN) — віддалені один від одного, але взаємодіючі комп'ютери-вузли, задля передачі даних об’єднані комунікаційною мережею, також ця взаємодія забезпечується спеціальними програмами мережної операційної системи. Основу глобальної мережі складають потужні обчислювальні системи (Host — вузли) і спеціалізовані комп'ютери, на які покладені функції комунікаційних вузлів. Користувачі стають абонентами мережі після під'єднання свого пристрою до цих хостів. WAN охоплюються такими структурами телекомунікації, які об'єднують в собі локальні інформаційні мережі із загальним протоколом зв'язку, також методи під'єднання та протоколи обміну даними. Як правило, з WAN організовується для конкретних цілей, але в подальшому розширюється за рахунок під'єднання інших локальних мереж, які використовують її (WAN) ресурси і послуги [1].

Для глобальних мереж характерним є значний масштаб, як площа покриття (місто, країна, континент) так і число вузлів. Ще однією характерною рисою є неоднорідність архітектури та програмного забезпечення цих вузлів. Така мережа пов'язує користувачів або району, або області, регіону чи країни. Дана мережа є досить дорогою і поодинокі підприємства можуть не мати її, проте великі компанії, які розміщені на великих площах з віддаленими об’єктами, як правило використовують такий тип мереж. Як правило, це спеціалізована мережа. З огляду на масштабність, глобальними вважаються мережі, які охоплюють території в тисячі кілометрів. В WAN-мережах, як правило, існує центр управління мережею. На нього покладена відповідальність за ефективне і надійне функціонування

мережі, за оптимальний вибір маршрутів проходження повідомлень від одного абонента до іншого. Надійність мережі можна підвищити, якщо частина вузлів з'єднана, оминаючи центр. У вузлах мережі також розміщуються спеціалізовані обчислювальні комплекси, які використовуються для обробки інформації.

WAN об'єднують комп'ютери, віддалені один від одного на сотні, а то і тисячі кілометрів. Для побудови такої мережі можуть використовувати вже наявні лінії зв'язку, які можуть виявитися застарілими і неякісними. Саме через це та через значну віддаленість між вузлами WAN має меншу швидкість передачі даних аніж локальна мережа. Хоча якщо використати оптоволоконні технології, то WAN має можливість надавати таку ж швидкість, як і LAN [1].

Можливості сучасних корпоративних мереж

Вибір концепції побудови конкретної корпоративної мережі визначається цілою низкою чинників: затребувані інформаційні послуги, обсяги переданого трафіку, існуюча інфраструктура і т. д. Але існують і загальні вимоги до корпоративних мереж. Мережі підприємств повинні бути побудовані на основі перевірених технологій, що володіють такими якостями, як масштабованість, гнучкість, мультисервісність, і найголовніше – надійність [8].

Мережа сучасного підприємства, як правило, повинна підтримувати ряд найбільш затребуваних для бізнесу додатків і керованих сервісів. В першу чергу це:

- можливість високошвидкісного доступу до мережі Інтернет.
- створення віртуальних приватних мереж (VPN).
- передача голосу поверх IP.
- проведення відеоконференцій.
- захист інформації та зберігання даних

VPN

VPN (Virtual Private Network – віртуальна приватна мережа) – узагальнена назва технологій, що дозволяють забезпечити одне або кілька мережних з'єднань (логічну мережу) поверх іншої мережі (наприклад, Інтернет). VPN – це технологія створення зашифрованих каналів в інтернеті, що дозволяє організувати анонімний доступ в мережу і обхід різного роду блокувань. VPN, по суті, створюють тунель даних між локальною мережею та вузлом виходу в іншому місці, яке може бути за тисячі кілометрів А простими словами, VPN – це свого роду тунель між клієнтом і захищеним VPN-сервером. Усередині цього тунелю засобами VPN здійснюється захист, шифрування і зміна даних, якими обмінюється комп'ютер користувача і веб-сайти або веб-сервіси. Ця перевага забезпечує свободу в Інтернеті або можливість доступу до програм та веб-сайтів завжди.

Є й інша сторона конфіденційності. Без VPN ваш постачальник послуг Інтернету може знати всю історію перегляду. За допомогою VPN історія пошуку прихована. Це тому, що веб-діяльність буде пов’язана з IP-адресою сервера VPN. Постачальник послуг VPN може мати сервери по всьому світу. Це означає, що пошукова діяльність може з’являтися в будь-якому з них. Пошукові системи також відстежують історію пошуку, але вони пов’язуватимуть цю інформацію з не своєю IP-адресою.

Віртуальна приватна мережа (VPN) надає конфіденційність та анонімність через створення приватної мережі із загальнодоступного Інтернет-з'єднання. VPN маскує адресу інтернет-протоколу (IP), тому дії в Інтернеті практично неможливо простежити. Служби VPN встановлюють безпечні та зашифровані з'єднання, щоб забезпечити більшу конфіденційність, ніж навіть захищена точка доступу Wi-Fi.

Якщо ще простіше, то можна це уявити так: без підключення до VPN-сервісу Ваш комп'ютер (ноутбук, телефон, телевізор або будь-яке інший пристрій) при вході в мережу подібний приватному будинку, що необгородженийу парканом. У будь-який момент кожен може навмисно або випадково поламати дерева, потоптати грядки на городі. З використанням

VPN будинок перетворюється на неприступну фортецю, несанкціонований доступ до якої буде просто неможливий [1,8].

В залежності від потреб і призначення, VPN може забезпечувати з'єднання трьох видів: між двома вузлами, між вузлом і мережею, між мережами. Прицьому підключення є як правило одного із наступних виді (перших два є найбільш вживані):

1) PPTP (Point-to-point tunneling protocol) – це такий тунельний протокол типу «вузол-вузол», який дозволяє комп'ютеру користувача встановлювати захищене з'єднання з сервером за рахунок створення спеціального тунелю в стандартній, незахищеній мережі. Протокол тунельного з'єднання «вузол-вузол» є застарілим методом реалізації віртуальних приватних мереж. PPTP має багато відомих проблем безпеки. Специфікація PPTP не описує функції шифрування або автентифікації і покладається на протокол PP, який тунелює для реалізації будь-яких функцій безпеки. Призначення цього протоколу полягає у забезпеченні рівнів безпеки та рівнів віддаленого доступу, порівнянних із типовими продуктами VPN. Це найвідоміший і найпростіший в налаштуванні варіант підключення до VPN-сервісу. Але багато інтернет-провайдерів блокують роботу PPTP підключень [1].

2) OpenVPN –це система віртуальної приватної мережі (VPN), що реалізує методи створення безпечних з'єднань точка-точка або клієнт-сервер у маршрутизованих або мостових конфігураціях та засобах віддаленого доступу. Він реалізує як клієнтські, так і серверні програми. Дозволяє однолітkам автентифікацію за допомогою загальнодоступних секретних ключів, сертифікатів або імені користувача / пароля. При використанні в конфігурації мультиклієнт-сервера він дозволяє серверу випускати сертифікат автентифікації для кожного клієнта, використовуючи підписи та центр сертифікації. Він широко використовує бібліотеку шифрування OpenSSL, а також протокол TLS і містить багато функцій безпеки та контролю. Він використовує спеціальний протокол безпеки, заснований на SSL / TLS для обміну ключами. Він здатний обходити транслятори

мережевих адрес (NAT) та брандмауери. Використання, цієї технології – потребує встановлення додаткового програмного забезпечення для всіх операційних систем.

3) L2TP (Layer 2 Tunneling Protocol) – це мережевий протокол тунелювання канального рівня. Являє собою низку правил, що дозволяють провайдерам послуг Інтернету дозволяти VPN. Однак L2TP самостійно не шифрує дані, тому не забезпечує повну конфіденційність для користувачів. Цей протокол дозволяє створювати VPN із заданими пріоритетами доступу, однак не містить в собі засобів шифрування і механізмів аутентифікації (для створення захищеної VPN його використовують спільно з IPSec). За відгуками експертів, є найбільш захищеним варіантом VPN підключення, незважаючи на труднощі його налаштування. Весь пакет L2TP, включаючи корисне навантаження та заголовок L2TP, надсилається в рамках дейтаграми User Datagram Protocol (UDP). Перевагою передачі через UDP (а не TCP) є те, що вона дозволяє уникнути "проблеми розпаду TCP". [3] [4] Зазвичай передача сесій PPP проводиться в тунелі L2TP. L2TP не забезпечує конфіденційність або надійну автентифікацію сам по собі. IPsec часто використовується для захисту пакетів L2TP, забезпечуючи конфіденційність, автентифікацію та цілісність. Поєднання цих двох протоколів загальновідомо як L2TP / IPsec (обговорюється нижче). Дві кінцеві точки тунелю L2TP називаються LAC (L2TP Access Concentrator) і LNS (L2TP Network Server). LNS чекає нових тунелів. Після встановлення тунелю мережевий трафік між одноранговими мережами стає двонаправленим. Щоб бути корисними для роботи в мережі, протоколи вищого рівня потім запускаються через тунель L2TP. Щоб полегшити це, сеанс L2TP (або "виклик") встановлюється в тунелі для кожного протоколу вищого рівня, такого як PPP. LAC або LNS можуть ініціювати сесії. Трафік для кожного сеансу ізолюється L2TP, тому можна створити кілька віртуальних мереж через один тунель. Пакети, якими обмінюються в тунелі L2TP, класифікуються як контрольні пакети або пакети даних. L2TP забезпечує функції надійності для контрольних пакетів, але не забезпечує надійності для пакетів даних. За необхідності надійність

повинна забезпечуватися вкладеними протоколами, що працюють у межах кожного сеансу тунелю L2TP. L2TP дозволяє створити віртуальну приватну комутовану мережу (VPDN) для підключення віддаленого клієнта до його корпоративної мережі за допомогою спільної інфраструктури, якою може бути Інтернет або мережа постачальника послуг.

Не зважаючи на переваги, технології VPN мають декілька суттєвих недоліків: необхідність закупівлі невеликої кількості обладнання і програмного забезпечення; збільшення обсягів зовнішнього трафіку. Проте, ці витрати досить невеликі і з огляду на величезну кількість переваг VPN з ними цілком можна миритися.

Технологія ATM

ATM (Asynchronous Transfer Mode — асинхронний спосіб передачі даних) — мережева високопродуктивна технологія комутації та мультиплексування, заснована на передачі даних у вигляді мікропакетів фіксованого розміру (53 байти), з яких 5 байтів використовується під заголовок. Така технологія надає послуги канального рівня, використовуючи при цьому широкий спектр засобів зв'язку на фізичному рівні. Технологія реалізується як в локальних, так і в глобальних мережах. Допускається спільна передача різних видів інформації, включаючи відео, голос [1].

Ця технологія є технологією зі встановленням з'єднання. Вона краще пристосована для надання послуг передачі даних з дуже відмінним чи змінюваним бітрейтом.

Обсяги даних, що використовуються в ATM, менші в порівнянні з елементами даних, які використовуються в інших технологіях. Це дозволяє:

- передавати дані одними й тими ж фізичними каналами, причому як при низьких, так і при високих швидкостях;
- працювати з постійними і змінними потоками даних;
- інтегрувати будь-які види інформації: тексти, голос, зображення, відеофільми;

- підтримувати з'єднання типу точка-точка, точка-мережа, мережа-мережа.

Для передачі даних в мережі ATM створюються віртуальні канали трьох видів:

- постійний віртуальний канал, PVC (Permanent Virtual Circuit). Створюється між двома точками і існує протягом тривалого часу, навіть за відсутності даних для передачі. Канал PVC створюється шляхом статичного визначення конфігурації в рамках всієї інфраструктури провайдера і завжди перебуває в стані готовності.
- комутований віртуальний канал, SVC (Switched Virtual Circuit). Створюється між двома точками безпосередньо перед передачею даних і розривається після закінчення сеансу зв'язку.
- автоматично настроюваний постійний віртуальний канал, SPVC (Soft Permanent Virtual Circuit). В каналі SPVC з'єднання є статичним тільки від кінцевої точки (пристрій DTE) до першого комутатора ATM (пристрій DCE). А на ділянці від пристрою DCE відправника до пристрою DCE одержувача в межах інфраструктури провайдера з'єднання може формуватися, розриватися і знову встановлюватися на вимогу. Встановлене з'єднання продовжує залишатися статичним до тих пір, поки порушення роботи однієї з ланок каналу не викличе припинення функціонування цього віртуального каналу в межах інфраструктури провайдера мережі [8].

Для маршрутизації в пакетах використовують ідентифікатори пакета двох видів: VPI (англ. virtual path identifier) - ідентифікатор віртуального шляху (номер каналу) та VCI (англ. virtual circuit identifier) - ідентифікатор віртуального каналу (номер з'єднання)

Для надійного об'єднання локальних мереж і великих комп'ютерів у корпоративну мережу, крім мереж X.25 існують такі технології, як frame relay, SMDS і ATM. Ці технології, розроблені спеціально для глобальних

комп'ютерних мереж. Окрім них можна використати і територіальні мережі TCP/IP [8].

Таблиця 1.1 - Характеристики мереж з комутацією пакетів

Тип мережі	Швидкість доступу	Трафік	Примітки
X.25	1.2-64 Кбіт/с	термінальний	надмірна кількість протоколів, добре працюють на каналах низької якості
Frame Relay	від 64 Кбіт/с до 2 Мбіт/с	комп'ютерний	Відносно нові мережі, добре передають пульсації трафіку, в основному підтримують службу постійних віртуальних каналів
SMDS	1.544 –45 Мбіт/с	комп'ютерний, графіка, голос, відео	Порівняно нові мережі, поширені у великих містах Америки, витісняються ATM
ATM	1.544 – 155 Мбіт/с	комп'ютерний, графіка, голос, відео	Нові мережі, комерційна експлуатація почалась з 1996 року, на разі використовуються в основному для передачі комп'ютерного трафіку
TCP/IP	1.2-2.048 Кбіт/с	термінальний, комп'ютерний	Широко використовуються в некомерційному варіанті

У таблиці 1.1 приводяться характеристики цих мереж, причому в графі «Трафік» вказується тип трафіка, який найбільше підходить для даного типу мереж, а в графі «Швидкість доступу» - найбільш типовий діапазон швидкостей, надаваних постачальниками послуг цих мереж.

Підхід, реалізований у технології ATM, полягає в передачі будь-якого виду трафіка - комп'ютерного, телефонного або відео - пакетами фіксованої й дуже маленької довжини в 53 байта. Пакети ATM називають гніздами - cell. Поле даних гнізда займає 48 байт, а заголовок - 5 байт.

Технологія ATM розроблялася спочатку як «річ у собі», без обліку того факту, що в існуючі технології зроблені більші вкладення й тому ніхто не стане відразу відмовлятися від установленого й працюючого устаткування, навіть якщо з'являється нове, більш досконале. Ця обставина виявилася не

настільки важливою для територіальних мереж, які за потреби можуть надати свої оптоволоконні канали для побудови мереж ATM. Враховуючи, що вартість високошвидкісних оптоволоконних каналів, прокладених на більші відстані, часто перевищує вартість іншого мережного встаткування, перехід на нову технологію ATM, пов'язаний із заміною комутаторів, у багатьох випадках виявлявся економічно виправданим.

Для локальних мереж, у яких заміна комутаторів і мережних адаптерів рівнозначна створенню нової мережі, перехід на технологію ATM міг бути викликаний тільки досить серйозними причинами. Набагато привабливіше повної заміни існуючої локальної мережі новою мережею ATM виглядає можливість «поступового» впровадження технології ATM в існуючу на підприємстві мережу. При такому підході фрагменти мережі, що працюють за новою технологією ATM, могли б мирно співіснувати з іншими частинами мережі, побудованими на основі традиційних технологій, таких як Ethernet або FDDI, поліпшуючи характеристики мережі там, де це потрібно, і залишаючи мережі робочих груп або відділів у колишньому виді.

Застосування маршрутизаторів IP, що реалізують протокол Classical IP, вирішує цю проблему, але такий розв'язок не завжди влаштовує підприємства, що користуються послугами локальних мереж, тому що, по-перше, потрібна обов'язкова підтримка протоколу IP у всіх вузлах локальних мереж, а по-друге, потрібна установка деякої кількості маршрутизаторів, що також не завжди прийнятне. Чітко відчувається необхідність способи узгодження технології ATM із технологіями локальних мереж без залучення мережного рівня.

У відповідь на таку потребу ATM Forum розробив специфікацію, названу LAN emulation, LANE (тобто емуляція локальних мереж), яка покликана забезпечити сумісність традиційних протоколів і устаткування локальних мереж з технологією ATM. Ця специфікація забезпечує спільну роботу цих технологій на канальному рівні. При такому підході комутатори ATM працюють у якості високошвидкісних комутаторів магістралі локальної мережі, забезпечуючи не тільки швидкість, але й гнучкість з'єднань

комутаторів ATM між собою, що підтримують довільну топологію зв'язків, а не тільки деревоподібні структури.

Специфікація LANE визначає спосіб перетворення кадрів і адрес MACрівня традиційних технологій локальних мереж у гнізда, що й реалізують віртуальні з'єднання SVC технології ATM, а також спосіб зворотного перетворення. Усю роботу з перетворення протоколів виконують спеціальні компоненти, що вбудовуються у звичайні комутатори локальних мереж. Тому ні комутатори ATM, ні робочі станції локальних мереж не зауважують того, що вони працюють із далекими від них технологіями. Така прозорість була однією з головних цілей розробників специфікації LANE [13].

Через те, що ця специфікація визначає тільки канальний рівень взаємодії, то за допомогою комутаторів ATM і компонентів емуляції LAN можна утворювати тільки віртуальні мережі,які називаються тут емульованими мережами, а для їхнього з'єднання потрібно використовувати звичайні маршрутизатори.

Корпоративна мережа повинна передавати та екстраполювати інформацію з усіх куточків Землі. Це має і позитивні, і негативні сторони.

З одного боку, це дає можливість отримати доступ до всієї корпоративної інформації з будь-якого місця в будь-який час (можна перевірити електронну пошту з номера в готелі, отримати доступ до ресурсів компанії у будьякий час). Технологія VPN дозволила користувачам залишатися на зв'язку з життєво важливими ресурсами в будь-якому місці. Важко знайти функцію роботи, яку неможливо виконати за допомогою віддаленого доступу.

З іншого боку, чи є VPN саме такими - приватними?

Проектування захищених мереж - одна з найбільших проблем, з якою стикаються інженри. Користувач, який перевіряє свою електронну пошту з дому вночі, може бути зломщиком, який отримає доступ до ваших ресурсів. Зломщиків можуть використовувати ті самі лінії та порти, що є життєво важливими для бізнесу.

VPN не є єдиний ризиком для безпеки інфраструктури. Підключення до Інтернету, філій та служб електронної пошти представляє потенційну небезпеку. Архітектор мережі повинен знати про всі ці проблеми і відповідно планувати.

Перша задача, яка стоїть, – це оцінити рівень безпеки кожного елемента мережі. Оцінити, який кінцевий пристрій все-таки свій, а який чужий.

1.2 Ймовірнісний підхід визначення безпеки корпоративної мережі

Управління ризиками інформаційної безпеки (УРІБ) складається з трьох фаз: оцінки ризиків, зменшення ризиків та управління.

Ці фази об’єднані у нескінчений цикл, як показано на рисунку 1.1.

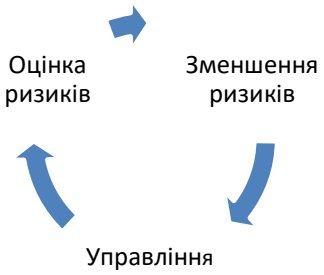


Рисунок 1.1 – Управління ризиками інформаційної безпеки

Початком є оцінка ризику. Розуміючи ризик та елементи, що сприяють цьому, ми робимо рекомендації щодо зменшення ризику там, де це необхідно. Нарешті, ми управляємо своїми контролем, моніторингом та вимірюванням, щоб забезпечити очікувані результати зменшення ризику. Періодично ми знову проводимо оцінку, щоб виявити нові загрози, вразливості чи інші зміни, які збільшили наш ризик понад допустимий поріг.

Таблиця 1.2 – Джерела загроз

Загроза	Мотивація	Зловмисна дія
Хакер, взломщик	Виклик, Его, Протест	-злом -соціальна інженерія -втручання в систему -неавторизований системний доступ
Кіберзлочинець	Знищення інформації, незаконне зберігання інформації, отримання грошей, неавторизована зміна даних	-компютерні злочини, - шахрайство, -інформаційний підкуп, -роздруківка несанкціонованих даних, -втручання в систему, -злом
Терорист	розсилання пошти з незаконним вмістом, деструктивна поведінка, використання недоліків системи, помста, політична помста	-бомби, -нформаційна війна, -атаки на електронні процеси (наприклад DDoS), -втручання в систему, -саботаж та зміна функцій системи, - інтелектуальні злочини
Промисловий шпіон	отримання перевги над суперником, економічний шпіонаж	-економічна експлуатація, - крадіжка цнтелектуальної власності, - крадіжка бізнес-інформації (фінансові дані, дані користувачів), -вторгнення в систему
Інсайдер	цікавість, его, знання, прибуток, помста, випадкові помилки/недогляд	-атака на робітників, -розсилання шкідливої пошти, -перегляд приватної інформації, -зловживання комп’ютером, -шахрайство та крадіжка, -інформаційний підкуп, -ввід в систему неправдивих або неправильних даних, -встановлення вірусних чи шкідливих програм, -продаж інформації користувачів чи їх інтелектуальної власності, -створення помилок в системі та некоректних налаштувань -вторгнення в систему, -саботаж, -неавторизований доступ

Шкідливі впливи на корпоративну мережу можуть бути як внутрішні так і зовнішні, під загрозою може бути не тільки інформаційна безпека, але й реальна фізична безпека людей і приміщень.

Виявити вразливості мережі або системи можна такими способами: скануванням та дослідженням відомих вразливостей до ймовірних загроз

(таблиця 1.2). Однак вони не дають повної картини при визначенні, наскільки ймовірно, що конкретний агент/дія загрози буде успішним.

Для кожного джерела загроз може бути кілька агентів / дій загроз. До прикладу, у таблиці 1.3 перелічені відповідні вразливості, виявлені та підтверджені поєднанням сканування та досліджень.

Таблиця 1.3 – Поєднання загроз / вразливостей

Джерело загроз	загрозлива дія	джерела вразливості
кіберзлочинець	Втручання в систему використовуючи різні інструменти та техніки	Відсутність розділення доступу
		Порт 1433 відкритий на фаєрволі
		операційна система серверу баз даних має вразливість до атак віддаленого доступу з відсутніми оновленнями
		Відсутність управління журналом подій та/або реакції на події
		відсутність систем виявлення втручань
		оновленя безпеки для віддаленого доступу MS SQL серверу не встановлені
		Оновлення безпеки для віддаленого доступу не встановлені

Табличний підхід – непоганий спосіб для демонстрації висновків. Однак, йому не вистачає глибини, необхідної для застосування існуючих засобів контролю для визначення фактичного ризику вразливостей у таблиці 1.3.

Оцінка контролю шляхів атаки

Непоганим інструментом для поточного аналізу управління є дерево атак [22, 24, 25, 28]. Дерево атак – це логічне представлення шляху, пристроїв та елементів керування загрозою, яке повинен пройти агент на шляху до цілі. Дерево атак для пари загроза / вразливість у таблиці 1.2 зображено на рисунку 1.2.

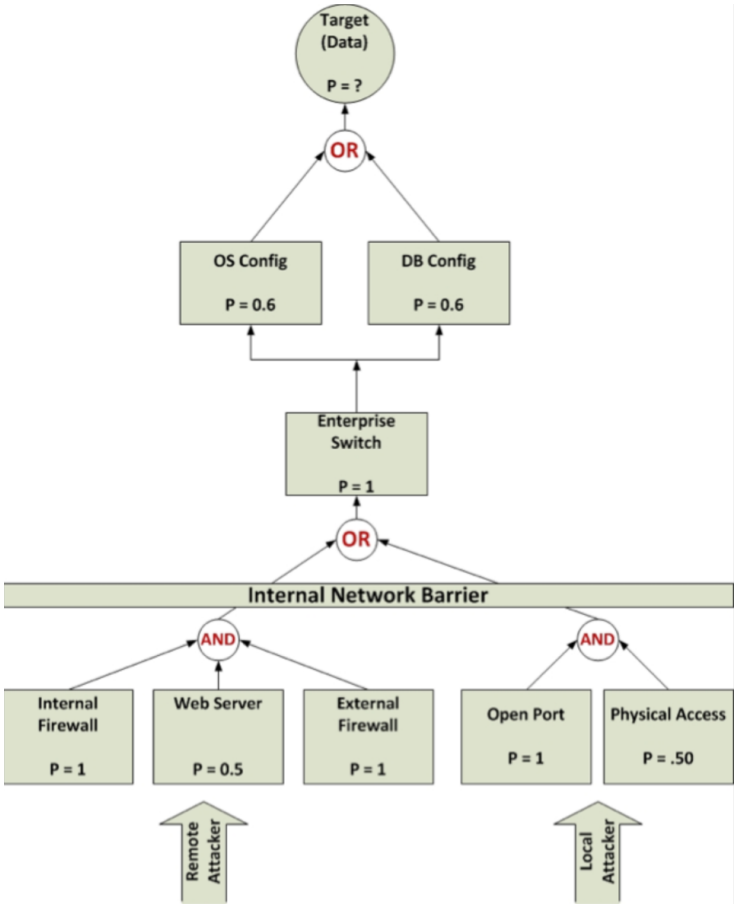


Рисунок 1.2 – Дерево мережевих атак

Імовірність того, що агент / дія може досягти цілі, обчислюється за допомогою кваліфікованих здогадок, зібраної інформації про оцінені загрози та елементів теорії ймовірності. На рисунку 1.2 “*P*” представляє ймовірність того, що зловмисник проникне через елемент керування / пристрій, представлений прямокутником.

Значення *P* в основному суб'єктивне. Занадто багато змінних, щоб розрахувати точне число. Тому використовується досвід та інформацію для наукового «здогадування». Часто для підвищення точності однієї або декількох оцінок ймовірності необхідна консультація із зовнішніми експертами.

Для кожного шару на рисунку 1.2 для переходу до наступного шару потрібен або набір умов (І), або одна (АБО) умова. Наприклад, віддалений хакер, що використовує веб-сервер як платформу для запуску своєї атаки, повинен пройти через внутрішній брандмауер І скомпрометувати веб-сервер І пройти через внутрішній брандмауер. Виходячи з поточних вразливостей,

мотивів джерела загрози та існуючих засобів контролю, ми оцінимо ймовірність 1,0 для порушення брандмауера та ймовірність 0,5 для компрометації веб-сервера. Якщо для переходу до наступного шару повинні існувати всі умови, ми обчислюємо ймовірність успіху як добуток, наприклад

$$P(\text{успіх}) = P(1)*P(0,5)*P(1)$$

Імовірність того, що зовнішній зловмисник досягне перемикача, становить 0,5. Якщо ми видалимо веб-сервер і припустимо пряму спробу підключення до бази даних через TCP-порт 1433, ймовірність проходження через цей шар зростає до P=1,0.

Внутрішній зловмисник повинен отримати фізичний доступ до будівлі та до відкритого, активного порту. Виходячи з нашого розрахунку, ймовірність успіху дорівнює P=0,5. Якщо неправомірний працівник намагається отримати доступ, перемикач дорівнює P=1,0. Аналогічним чином, P=1,0, якщо система працівника належить зловмиснику і керується з віддаленого місця.

Проходження через перемикач P=1,0. Не існує сегментів мережі із пов'язаними списками контролю доступу. На даний момент ми розраховуємо кожен шлях атаки окремо. Віддалена атака за допомогою веб-сервера має P=0,5. Оскільки зловмисник повинен пройти через периметр і перемикач, розрахунок $P(\text{успіх})=P(0,5)*P(1,0)$. Іншими словами, ймовірність потрапити із зовнішнього місця в місце за межами вимикача становить P=0,5. Така ж ймовірність існує і для внутрішньої атаки.

Ми можемо взяти два підходи до обчислення ймовірності останнього шару перед ціллю: обчислити ймовірність шляху через кожне поле або обчислити ймовірність шару та використовувати його в нашому розрахунку шляху атаки. Щоб все було простіше, ми використаємо останній підхід.

Наш розрахунок ймовірності змінюється на цьому шарі. Звернімо увагу, що зловмисник може вирішити зламати операційну систему або екземпляр бази даних – або обидва. При обчисленні умов I ми множимо. При

обчисленні АБО умов додаємо. Отже, ймовірність проходження через рівень операційної системи / бази даних становить P(успіх)=P(0,6)+P(0,6)=P(1,0). Хоча формула дає суму 1.2, значення понад 100% не має змісту.

Використовуючи останній рівень та попередні розрахунки, ймовірність досягнення цілі з віддаленого місця за допомогою веб-сервера становить P(успіх)=P(0,5)*P(1,0)=P(0,5). Імовірність внутрішньої атаки без фізичного доступу становить P=(1,0).

Також можна використовувати дерева атак для обчислення ймовірності того, що зловмисник подолає певне поле або компонент у нашому дереві мережевих атак. На рисунку 1.3 зображене дерево атак для визначення ймовірності того, що зловмисник може заволодіти веб-сервером. Використовуючи описані вище розрахунки ймовірності, ми отримуємо P=0,5. Формула (P(1,0)*P(1,0)*P(0,5))*P(1,0)=0,5. Цей підхід застосовується до будь-якого набору завдань, які повинен виконувати зловмисник для досягнення мети.

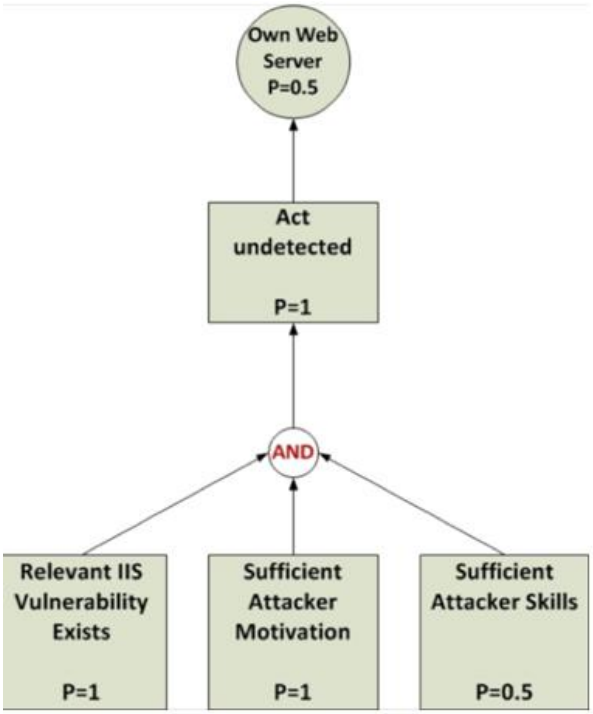


Рисунок 1.3 – Дерево атаки на веб-сервер

Інший підхід дерева атак – це просто позначати кожен вузол як можливий чи неможливий. На риснку 1.4 зображено дерево атак увімкнення /

вимкнення, працює за тим же принципом, що і логіка І та АБО. Щоб пройти через І, усі входи повинні бути увімкненими або “1”. Щоб пройти через АБО, повинен бути включений лише один із входів.

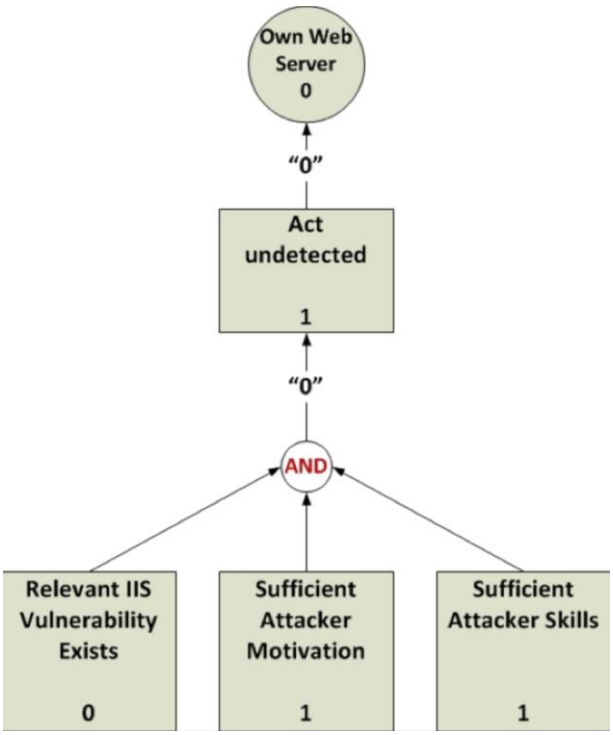


Рисунок 1.4 – Дерево атаки компонентів увімкнення / вимкнення

Для того, щоб «заволодіти» веб-сервером, потрібно «заволодіти» всіма компонентами нижчого рівня (рисунок 1.4). Іншими словами, вони повинні бути включеними.

Цей підхід, також застосовний до дерева мережевих атак вищого рівня, має одну головну перевагу перед імовірнісним підходом; це простіше. Якщо потенційну ціль класифікують як конфіденційну, можливо, не має сенсу розраховувати ймовірності на шляху атаки. Однак, підхід увімкнення / вимкнення має одну велику слабкість. Дерево увімкнення / вимкнення передбачає все або нічого. Оцінювач повинен вірити, що вразливість ніколи не існуватиме. але насправді завжди існує певна ймовірність того, що зловмисник може зламати рівень безпеки.

Даний підхід дає можливість оцінки ймовірність атаки мережі, але не дає відповідь на питання щодо того, який саме пристрій мережі є небезпечним і «чужим». З огляду на приклади, наведені у роботі, наімовірніше атака буде здійснюватись через один із «своїх» пристроїв,

поведінка якого є нетиповою, хоча і зовнішня атака з «чужого» пристрою також не виключена.

Дані приклади ілюструють дерева атаки фрагмента мережі. Зрозуміло, що масштаби корпоративної мережі передбачають величезну кількість пристроїв, кожен з яких може бути кінцевою ціллю зловмисника. Передбачити це є складною задачою, оскільки кожен «свій» пристрій в наступний момент часу може виявитись «чужим».

1.3 Постановка задачі

Сформулюємо задачу дослідження, як задачу розробки моделей і алгоритмів, що дозволять ідентифікувати кінцевий пристрій корпоративної мережі за принципом свій /чужий.

Перш за все слід зазначити, що пристрої, які під’єднані до корпоративної мережі за допомогою бездротового зв’язку, будуть вважатись підозрілими, і лише їх аномальна поведінка (як то спроба звернення до забороненої IP адреси, використання недозволених портів з’єднання тощо) буде вирішальною при ідентифікації за принципом свій/чужий.

Крім того «чужим» може виявитися також і стаціонарний пристрій мережі, який виявляє нетипову поведінку. А «своїм» може виявитися пристрій, який підєднався до мережі, але не виявляє зловмисних дій.

Відомі мережні рішення показують недостатньо високу достовірність виявлення загрозливих дій в локальних комп’ютерних мережах, зокрема, через використання централізованого способу організації взаємодії компонентів компонентів мережі. Остаточне рішення як правило приймається адміністратором мережі, якому необхідно кожного разу для кожного конкретного пристрою у разі нетипової поведінки приймати рішення – чужий чи свій – і відповідним чином реагувати.

Як правило «чужий» пристрій в мережі намагається здійснити вторгнення, причини і засоби були розглянуті в підрозділі 1.2 даної роботи. Тому необхідно перш за все виявляти такі вторгнення та приймати рішення щодо рівня безпеки такого вторгнення.

Метою даної роботи є розроблення методу розпізнавання кінцевих пристроїв за принципом свій/чужий. Для реалізації мети необхідно вирішити наступні задачі:

1. Проаналізувати існуючі підходи для виявлення вторгнень у мережу.
2. Запропонувати структуру системи виявлення вторгнень, що дозволить ідентифікувати кінцевий пристрій мережі як свій або чужий.
3. Запропоновати алгоритм процедури виявлення «чужого» пристрою як такого, що здійснює неправомірні дії в мережі.
4. Розробити нечітку систему керування, що дозволить негайно реагувати на загрозливі дії пристрою корпоративної мережі, та ідентифікувати його за принципом свій/чужий.

2 АЛГОРИТМИ ШТУЧНИХ ІМУННИХ СИСТЕМ ДЛЯ ВИЯВЛЕННЯ ЗАГРОЗ МЕРЕЖІ

2.1 Біологічні імунні системи

У біології імунітет – це здатність організму протистояти атакам інвазивних чужорідних речовин. Така чужорідна речовина називається патогеном і імунна система (ІС) розпізнає як антиген (АГ). Антиген може бути практично будь-яким видом великої чужорідної молекули в організмі, включаючи ті, що містяться в інфекційних агентах, отруті змії та скорпіонів, їжі та інших клітинах та тканинах різних видів, зокрема людських [14].

Існує два види імунітету; а саме, неспецифічний, вроджений імунітет та специфічний, набутий імунітет. Вроджений імунітет включає шкіру людини, оболонки дихальних та шлунково-кишкових шляхів, а також деякі інші захисні фактори. Ці імунні механізми пригнічують або вбивають найрізноманітніші мікроби, незалежно від того, чи вони раніше кидали виклик організму. Це неспецифічно, оскільки його механізми можуть діяти проти мікробів, які не обов’язково схожі один на одного.

Після того, як людина перехворіла хворобою і одужала, вона, як правило, не заражається цією хворобою знову. Це явище називається набутим імунітетом. Це специфічно тим, що його реакції пристосовані для дії проти певного мікроба або його продуктів. Такий імунітет набувається за рахунок того, що індивідуальні реакції надзвичайно збільшуються в результаті стимулювання попередньою присутністю даного мікроба або його продуктів.

Щоб мати можливість протистояти атакам антигенів, ІС повинна мати можливість розрізняти матеріали тіла та матеріали чужорідної речовини. Оскільки всі живі істоти складаються в основному з подібних будівельних блоків, здатність організму розрізняти молекули, з яких складається сам, тобто себе, від практично всіх інших, тобто не себе, є помітною. Ця здатність певною мірою присутня у всіх живих істот, але серед хребетних це особливість білих кров'яних клітин, які називаються лімфоцитами.

Лімфоцити та антигени.

Лімфоцити – це клітини, що відповідають за здатність організму розрізняти і реагувати на майже безмежну кількість різних АГ. Існує два основних типи лімфоцитів, тобто В- і Т-лімфоцити (також відомі як В- і Т-клітини). Стовбурові клітини як В-, так і Т-лімфоцитів беруть свій початок у кістковому мозку. Розпізнавання чужорідних антигенів в ІС здійснюється за допомогою рецепторів на поверхні як В-, так і Т-лімфоцитів.

Частина АГ, яку розпізнає рецептор, антигенна детермінанта, називається епітопом [14]. Отже, епітоп – це місце розташування на поверхні збудника або фрагмента білка, яке називається пептидом. Багато АГ мають різні епітопи на різних ділянках своєї поверхні. Таким чином, складні антигени можуть викликати відповіді з різних специфічних лімфоцитів.

Лімфоцити – це переважно спляча популяція, яка очікує відповідних сигналів для активізації. Вони рухаються лише мляво самостійно, але їх можна транспортувати по всьому тілу, переносячи в крові або лімфі. У будь-який час доросла людина має близько 2×10^{12} лімфоцитів, приблизно 1% з яких знаходиться в кров

В- лімфоцити.

В-лімфоцити – також звані імуноглобулінами (ІГ) - диференціюються в кістковому мозку (звідси і В). Кожен В-лімфоцит запрограмований на вироблення антитіла (АТ) однієї специфіки та розміщення його на зовнішній поверхні, щоб діяти як рецептор. Кожен В-лімфоцит має на своїй поверхні близько 105 однакових молекул АТ.

Коли АГ потрапляє в організм, він стикається з величезним набором лімфоцитів, які несуть різні антитіла, кожен із своїм індивідуальним місцем розпізнавання. АГ зв'язуватиметься лише з тими рецепторами, з якими він продукує добре [14].

Т- лімфоцити.

Багато мікроорганізмів, наприклад, віруси, користаються тим фактом, що якщо вони живуть усередині клітин свого хазяїна, гуморальне АТ – яке представлене на поверхні В-лімфоцитів – не може досягти їх. Таким чином,

існує субпопуляція лімфоцитів, що включає Т-лімфоцити, яка спеціалізується на дії проти внутрішньоклітинних організмів. На відміну від В-лімфоцитів, Т-лімфоцити диференціюються всередині вилочкової залози. Тимус – лімфоїдний орган у формі піраміди. У людини він знаходиться безпосередньо під грудною кісткою на рівні серця. Орган називають тимусом, оскільки його форма нагадує форму листя чебрецю. Т-лімфоцити відрізняються від В-лімфоцитів тим, що вони розпізнають АГ лише тоді, коли він знаходиться на поверхні клітини тіла. Тому рецептори Т-клітин (РТК), які відрізняються від молекул АТ, що використовуються В-лімфоцитами, розпізнають АГ плюс поверхневий маркер, що вказує на те, що він присутній на поверхні іншої клітини. Ці клітинні маркери належать до групи молекул, яка називається основним комплексом гістосумісності (МГС).

Кожна особина в популяції генетично здатна створити невеликий набір цих маркерів МГС (близько 10), але набір типів МГС варіюється від однієї особини до іншої. Отже, особини в популяції здатні розпізнавати різні профілі пептидів. Цей механізм забезпечує важливу форму різноманітності на рівні популяції [14, 15].

Розпізнавання антигенів.

Генетична перебудова, описана нижче, відбувається лише тоді, коли перші лімфоцити стають функціональними.

Зв'язування антитіла з антигеном або РТК з комплексом МГС-пептиду вимагає, щоб частини двох структур мали взаємодоповнюючі форми, які можуть наближатися одна до одної [14]. Ця взаємодоповнюваність форми дозволяє рецептору і АГ відповідати одне одному приблизно так само, як ключ, який вставляється в замок. Ця відповідність є наближеною, що дає можливість одному конкретному лімфоциту зв’язуватися з кількома різними видами структурно пов’язаних патогенів.

Ділянка на АГ, де вона контактує з антитілом, називається епітопом. Відповідна площа на АТ називається паратопом. Міцність зв'язування АГ з одним вузлом комбінування АТ залежить від спорідненості між ними.

Чим вища спорідненість, тим сильніше зв'язування.

Відомо, що ІС здатний розпізнати практично будь-який патоген, який існує або який може бути розроблений або природою, або наукою в майбутньому. Для виконання цього завдання ІС генерує мільйони різних специфічних рецепторів АГ, що, ймовірно, значно більше, ніж потрібно протягом життя окремих людей [14].

Молекули АГ-рецепторів – це білки, що складаються з кількох поліпептидних ланцюгів. Поліпептид – це пептид, що містить від 10 до більш як 100 амінокислот. Послідовність, у якій амінокислоти збираються для утворення певного поліпептидного ланцюга, позначається генами дезоксирибонуклеїнової кислоти (ДНК). Оскільки весь геном людини містить лише близько 30 000 – 40 000 генів, особи не можуть успадкувати один ген для кожного конкретного рецептора АГ. Отже, обмежений пул генних сегментів успадковується для кожного типу поліпептидного ланцюга, який кодує антитіла і РТК. По мірі дозрівання кожного лімфоцита сегменти гена з’єднуються, утворюючи по одному гену для кожного поліпептиду, який утворює специфічний рецептор. Ця перебудова генного сегмента здебільшого відбувається випадково.

Набір АТ та РТК у миші, за оцінками, містить близько 107 різних рецепторів, утворених набагато більшим потенційним набором кодованих зародкових ліній рецепторів [14].

Негативний відбір.

Коли попередники Т-лімфоцитів залишають кістковий мозок на шляху дозрівання в тимусі, вони байдужі до стимуляції АГ, оскільки вони ще не експресують рецептори. Коли вони потрапляють у тимус, їх називають незрілими лімфоцитами, а коли, або якщо вони виходять, їх називають зрілими лімфоцитами. Цей процес дозрівання часто називають толеруванням. Усередині тимусу Т-лімфоцити множаться багато разів, проходячи через сітку клітин тимусу. Розмножуючись, вони набувають рецепторів і диференціюються в різні підкласи Т-лімфоцитів [20].

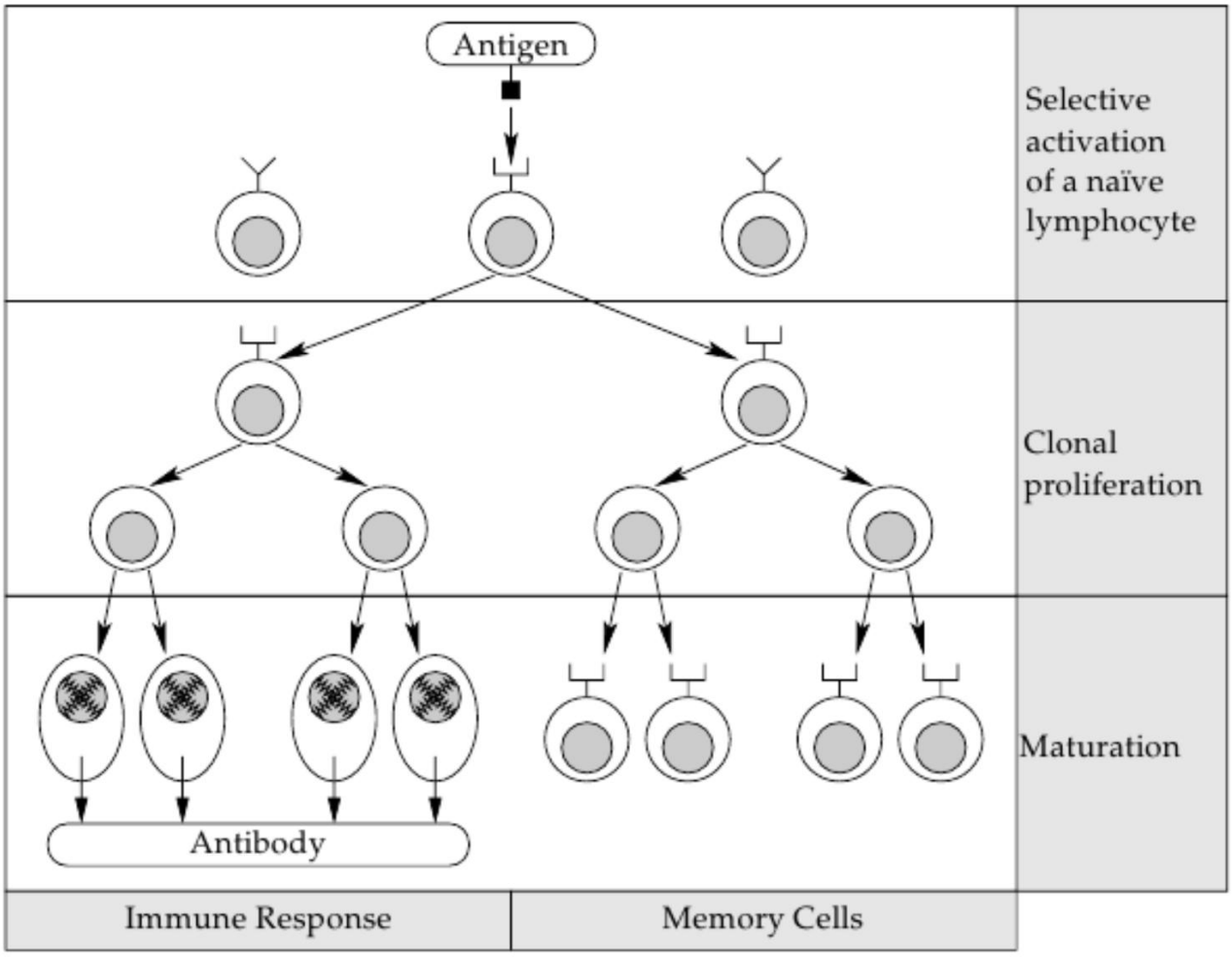


Рисунок 2.1 – Клітинна основа для генерації ефектора та клітин пам'яті шляхом клональної селекції після первинного контакту з антигеном

В організмі незрілих Т-лімфоцитів, що потрапляють у тимус, лише 2% завершують процес дозрівання і стають функціонуючими Т-лімфоцитами [14]. Це означає, що більшість Т-лімфоцитів, що потрапляють у тимус, також гинуть там, під час толерування. Це може здатися дуже марнотратним, але оскільки рецептори АГ створюються випадковим чином, багато з них розпізнають себе антигенами. Власні антигени – це молекули, що містяться у власних складових організму. Якщо аутореактивні лімфоцити, тобто вони реагують на себе, стають зрілими, вони атакують власні тканини організму. Тому більшість з них видаляються шляхом апоптозу в тимусі. Апоптоз є різновидом запрограмованої загибелі клітин. Цей механізм запобігання розвитку аутоімунних лімфоцитів називається негативним відбором. Вважається також, що негативний відбір В-лімфоцитів, що розвиваються,

відбувається, якщо вони стикаються з високим рівнем своїх АГ в кістковому мозку [14].

Клональний вибір.

Перше зіткнення між новим лімфоцитом та даним АГ називається первинною імунною відповіддю. Ця реакція є відносно слабкою порівняно із вторинною імунною відповіддю, яка є якісно та кількісно поліпшеною реакцією, яка виникає при другому зіткненні лімфоцитів з даним АГ [19].

Частина цієї поліпшеної реакції зумовлена процесом, який називається клональним відбором, який відбувається після того, як лімфоцит розпізнає специфічний АГ. Процес показаний на рисунку 2.1, де лімфоцит, відібраний специфічним АГ, зазнає багатьох поділів під час клональної проліферації, а потомство дозріває, утворюючи розширену популяцію клітин, що утворюють АТ. Частина нащадків вихідних АГ-реактивних лімфоцитів стає клітинами пам'яті, що не діляться.

Вторинна реакція в ІС сильніша, ніж первинна реакція. Оскільки лімфоцитів, які, можливо, зв’язуються з АГ під час вторинної відповіді, більше, інфекційний агент буде переможений швидше. Через цей механізм, коли перший контакт з АГ чітко вбиває якусь інформацію або надає деяку пам’ять, в ІС кажуть, що ІС розвиває набуту пам’ять [4]. Пам'ять, індукована одним АГ, не буде автоматично поширюватися на інший не пов'язаний АГ.

Під час клональної проліферації експоненційно зростає популяція лімфоцитів, здатних виявляти активуючий АГ. Найбільш імовірно, що лімфоцити з найбільшою долею між його рецепторами та епітопами збудника активізуються пізніше. Патогени, як правило, також розмножуються в цей проміжок часу, тому існує гонка стати найсильнішою популяцією.

Соматична гіпермутація.

Під час первинної відповіді В-лімфоцити, але, як правило, не Т-лімфоцити, зазнають точкових мутацій з високою швидкістю в генах змінної області. Цей механізм називається соматичною гіпермутацією і збільшує різноманітність антитіл та доцільність антитіл. Його називають соматичним, оскільки він займає місця в клітинах організму, а не в клітинах статевої лінії

(яйцеклітини та сперма). Мутації є результатом заміщення одного нуклеотиду і обмежуються змінною областю лімфоцитів, що означає, що ці мутації не впливають на константні області.

Нуклеотид - це основна структурна одиниця нуклеїнових кислот, таких як ДНК або рибонуклеїнова кислота (РНК). Послідовність нуклеотидів у ДНК або РНК кодує структуру білків, синтезованих у клітині. Швидкість мутації під час соматичної гіпермутації становить приблизно $10^{-4} – 10^{-3}$ на пару основ за покоління, що приблизно в мільйон разів вище, ніж для інших генів ссавців [14].

Вважається, що соматична гіпермутація - це спосіб ІС збільшити свої шанси у «гонці розповсюдження». Соматична гіпермутація в поєднанні з клональною експансією є адаптивним процесом, відомим як дозрівання аффінності [14].

Костимуляція.

Оскільки деякі самопептиди ніколи не експресуються в тимусі, зрілі лімфоцити, які були толеровані в тимусі, можуть зв’язуватися з цими білками і спричиняти аутоімунну реакцію [14]. На практиці цього не відбувається, оскільки на додаток до зв'язування з АГ, Т-лімфоцит повинен отримувати сигнал костимуляції, щоб бути активованим. Цей сигнал, як правило, є деяким хімічним сигналом, який утворюється, коли тіло якимось чином пошкоджене.

2.2 Виявлення вторгнення в мережу

Під виявленням вторгнень (ВВ) будемо розуміти виявлення невідповідної, неправильної або аномальної активності у комп'ютерній мережі або на хості шляхом дослідження певних видів даних. Такі заходи можуть бути розпочаті із зовнішніх зломщиків або через внутрішнє зловживання. Далі діяльність такого роду буде називатися подією, що представляє інтерес (ППІ). Зауважте, що ППІ не обов’язково має бути

навмисною атакою зовнішньої особи, а може бути наслідком аномальних дій або помилок законного користувача. Комп'ютерні програми, що виконують ідентифікацію, загалом називаються системами виявлення вторгнень (СВВ)

Підходи до виявлення вторгнень.

СВВ, яка працює на комп'ютері для виявлення шкідливої діяльності на цьому хості, називається СВВ на основі хоста, тоді як СВВ, яка намагається виявити цікаві події шляхом аналізу та моніторингу даних мережевого трафіку, називається мережевою СВВ або МСВВ система. Це два основні підходи до ідентифікації. СВВ на основі хоста зазвичай контролює та аналізує різні дані, що надходять від користувачів та виводяться з операційної системи (ОС). Наприклад, він може контролювати системні дзвінки. В ідеалі потрібно поєднати обидва вищезазначені підходи для збільшення ймовірності виявлення вторгнень.

Техніка виявлення вторгнень.

Дві основні категорії методів, що використовуються для виконання ідентифікації – як на основі хоста, так і на основі мережевого ідентифікатора – це виявлення на основі аномалій та виявлення на основі підписів. Остання техніка також відома як виявлення зловживання та виявлення відповідності шаблонів.

В [18] порівнюють завдання ВВ із загальною проблемою виявлення сигналу. У цьому випадку ППІ можна розглядати як сигнал, який слід виявити, а звичайну поведінку – як шум. У класичних підходах виявлення сигналу відома деяка інформація як про розподіл сигналу, так і про шум, і в процесі прийняття рішення використовується інформація про обидва розподіли, щоб визначити, чи належить дане спостереження до розподілу сигнал-плюс-шум або до розподілу шуму. З іншого боку, в ІД детектори зазвичай базують свої рішення або на характеристиках сигналу (виявлення на основі сигнатури), або на шумі (виявлення на основі аномалії). Кожен підхід має свої сильні та слабкі сторони, тоді як обидва страждають від труднощів характеристики розподілу.

Для успішного виявлення ППІ СВВ, що базується на підписах, покладається на наявність опису або підпису, який повинен відповідати події. Цей підпис може бути таким самим простим, як частина мережевого пакету, і таким складним, як опис нейронної мережі, який відображає кілька датчиків в ОС на абстрактне уявлення про атаку. Якщо для створення підпису використовується відповідна абстракція, СВВ на основі підпису може виявити вторгнення, які раніше не бачили, якщо вони абстрактно еквівалентні відомим підписам.

СВВ на основі аномалій базуються на припущенні, що незвична або ненормальна поведінка є нав'язливою. Використовуючи вищезазначене порівняння між ВВ та виявленням сигналу, СВВ, що базується на аномалії, виявлятиме вторгнення всякий раз, коли спостереження, здається, не є лише шумом, враховуючи, що існує повна характеристика розподілу шуму. Слід зазначити, що характеристика розподілу шуму для підтримки виявлення є нетривіальною

Обмеження спостережень

Причин, через які СВВ не може спостерігати – і, отже, не виявляти – ППІ, може бути кілька. Причини включають те, що ППІ може мати місце в іншій мережі, ніж та, що знаходиться під наглядом; ППІ може відбуватися прямо перед СВВ, але система не виявляє його, оскільки вона не працює; ППІ відбувається в протоколі, який СВВ не розуміє; або що ППІ трапляється протягом періоду, коли пропускна здатність СВВ перевищена.

Коли система виявлення вторгнень здійснює моніторинг та аналіз мережевих даних, результат в кожному випадку потрапляє в одну з категорій, показаних у таблиці 2.1.

Таблиця 2.1 – Категорії результатів моніторингу

ППІ	Результат виявлення	Категорія
Ні	Ні	Істинний негатив
Ні	Так	Хибний позитив
Так	Ні	Хибний негатив
Так	Так	Істинний позитив

Хибний негатив виникає, коли відбулася фактична нав'язлива дія, але СВВ розрізняє їй передати як ненав'язливу поведінку. Хибний позитивний результат виникає, коли система класифікує дію як аномальну, але насправді це є законною дією. Істинний негатив виникає, коли немає ППІ і не проводиться виявлення. Нарешті, істинний позитив виникає, коли СВВ правильно класифікує ППІ як нав'язливу поведінку.

Очевидно, що з точки зору користувача, хотілося б мінімізувати кількість хибних спрацьовувань та максимізувати кількість істинних спрацьовувань, виконуваних системою виявлення вторгнень. Це призведе до низької кількості помилкових тривог та великої кількості виявлених вторгнень відповідно. Зрозуміло, що велика кількість помилкових тривог неприваблива. Природно, що істинні та хибні негативи ніколи не помічаються в реальній системі

Локальні мережі.

Галсолл [13] повідомляє, що найпоширенішим типом локальної мережі є такий, що базується на багаторазовому доступі сенсора з контролером доступу (CSMA / CD). CSMA / CD більш відомий як Ethernet, який використовує Інтернет-протокол (IP) як основний протокол. IP не має жодних механізмів для підвищння наскрізної надійності даних, управління потоком, послідовності або інших служб, які зазвичай зустрічаються в протоколах від хосту до хоста, тому він часто супроводжується використанням управління передачею даних.

Протокол (TCP), який, згідно [8], призначений для використання як високонадійний протокол від хосту до хоста між хостами в мережах комп'ютерного зв'язку з комутацією пакетів та у взаємопов'язаних системах таких мереж.

2.3 Штучні імунні системи

У літературі про ШІС, що застосовується до ВВ, наприклад [12, 16, 19, 20, 23, 27], моделювання антитіл і лімфоцитів часто об'єднують у загальну сутність детектора. Ця практика також використовується при впровадженні ВШІС.

Здатність розрізняти своїх і чужих є, мабуть, найголовнішою рисою ІС. Це робиться шляхом розпізнавання лімфоцитами різних антигенів. Оскільки розпізнавання АГ в біологічному ІС відбувається, коли між рецепторами на поверхні імунних клітин та епітопами на поверхні патогенних мікроорганізмів встановлюються хімічні зв’язки, збіг на низькому рівні зводиться до узгодження білків або фрагментів білка, що називаються пептидами. Далі слово пептид буде використовуватися для представлення як штучних рецепторів, так і штучних агентів.

В ШІС пептиди часто представлені у вигляді рядків довжиною l, що складаються з символів з алфавіту, що містить m символів. Цей підхід найчастіше використовують для $m = 2$ (тобто бітових рядків).

Пептиди, що представляють АГ, будуть кодувати деяку інформацію, що стосується проблемного домену, до якого застосовується ШІС. Оскільки ІС повиннна розрізняти своїх та чужих на основі пептидів, ШІС повиннна розрізняти своїх та чужих на основі рядків фіксованої довжини l. Кожен такий рядок буде називатися агентом a. Сукупність усіх агентів утворює універсум, $U = \{a_1, a_2, ..., a_n\}$, який містить дві підмножини, що не перетинаються; тобто сукупність своїх, U_S, і сукупність чужих, U_N, так що $U = U_S \cup U_N, U_S \cap U_N = \varnothing$. Як зазначено в [15], АІС тоді стикається з проблемою класифікації; отримавши довільний рядок з U, класифікуйте його як свій чи чужий. Класифікація на своїх та чужих може також розглядатися як розподіл на нормальних та аномальних.

Ця модель пептидів дотримується вимоги про те, що вся відповідна інформація в проблемній області може бути представлена якимось чином і

що повинен існувати певний спосіб компактного кодування узагальнень цієї інформації.

Слід також зазначити, що коли реальні проблеми відображаються у таких уявленнях, свій та чужий не можуть бути роз'єднаними, оскільки два випадки можуть бути відображені в одному представленні.

Подібно до СВВ, ІС також може допускати два типи помилок розпізнавання (таблицю 2.1). Це справедливо і для ШІС. Хибно позитивний результат виникає, коли нормальний агент класифікується як чужий, а хибно негативний – коли аномальний агент класифікується як свій.

Описане вище кодування пептиду також використовується для моделювання рецепторів детекторів в ШІС. ШІС має сукупність детекторів D. Кожен детектор $d \in D$ має покриття C_D, яке описує кількість агентів, які він розпізнає. Якщо детектор d не розпізнає жодних агентів, його покриття $C_D = \varnothing$. З іншого боку, якщо d впізнає всіх інших агентів, його покриття є $C_D = U$ - всі агенти універсуму.

Це представлення пептидів дозволяє ШІС розпізнавати різні агенти за допомогою співставлення рядків. Але, одна з приємних особливостей ІС, розглянута з точки зору обробки інформації, полягає в тому, що вона здатна узагальнювати це співставлення. Узагальнення своїх та чужих, яке відбувається в ІС, здійснюється за допомогою наближеного збігу рядків.

У найбільш загальній формі проблема наближеного узгодження рядків полягає у пошуку тексту, де виникає заданий шаблон тексту, допускаючи обмежену кількість «помилок» у збігах. Кожна програма використовує іншу модель помилки, яка визначає, наскільки різними можуть бути рядки. Ці тексти можна розглядати як послідовності символів, складених з алфавіту довжини m.

Існує два таких правила – правило Хеммінга та правило *r*-послідовних збігів (рисунок 2.2 – 2.3). Тут основна увага буде зосереджена на правилі *r*-послідовних збігів, оскільки це правдоподібна абстракція зв’язування рецепторів в імунній системі. ІС дуже ефективна тим, що їй вдається розрізняти своїх і чужих, маючи відносно невеликий набір детекторів.

Використовуючи це правило для прогнозування оптимального розміру Т-клітини та області комбінування антитіл з урахуванням ефективного розпізнавання свій-чужий, результати узгоджуються з різними експериментальними визначеннями кількості контактних місць там, де поєднуються антитіла, та білки антигенів та розміром МГС-пептидного комплексу, який взаємодіє з РТК [15].

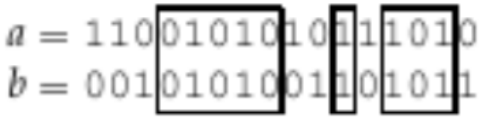


Рисунок 2.2– Співставлення рядків за правилом Хеммінга (між рядками довжиною в 16, що складаються з символів бінарного алфавіту, з відповідним обмеженням *r*=9. Два рядки *a* та *b* будуть збігатися для всіх *r*≥9)

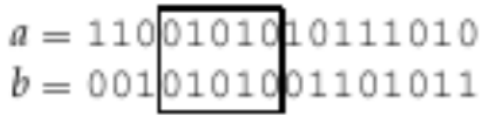


Рисунок 2.3 – Співставлення рядків за правилом *r*-послідовних збігів (між рядками довжиною 16, що складаються з символів бінарного алфавіту, з відповідним обмеженням *r*=5. Два рядки *a* та *b* будуть збігатися для всіх *r*≥5)

І правило збігу Хеммінга, і правило *r*-послідовних збігів контролюються пороговим параметром *r*, де $0 \le r \le l$. Якщо $r = 0$, покриттям *d* є всі рядки, $C_D = U$, а якщо $r = l$, то покриттям *d* є один рядок агента *a*, $C_D = \{a\}$. Чим вище значення *r*, тим конкретніше збіг. І конкретність збігу є аналогічною близькості зв'язування між АГ та лімфоцитом, або детектором.

Слід зазначити, що в ІС відповідність (або розпізнавання) між АГ та лімфоцитом базується на взаємодоповнюючих формах. В ШІС розглядатимуться бінарні рядки та їх «приблизна» рівність. Припущення Наприклад, у штучному пептидному збігу 1 на епітопі доповнює 1 на паратопі.

Правило поєднання Хеммінга базується на відстані Хеммінга між двома рядками. Якщо два рядки *a* і *b* мають однакові біти принаймні в *r* позиціях, вони збігаються. Це показано на рисунку 2.2. Згідно з правилом *r*-послідовного збігу, два рядки *a* і *b* збігаються, якщо вони мають однакові біти принаймні в *r* послідовних позиціях. На рисунку 2.3 наведено приклад такого збігу.

Імовірність збігу за допомогою правила Хеммінга та правила послідовного збігу:

Нехай $Hamming_{l,r}(a,b)$ є оператором, який визначає, чи збігаються дві рядки *a* і *b*, обидва довжиною l, використовуючи правило збігу Хеммінга, з обмеженням, що *r* біт попарно рівні.

Тоді ймовірність збігу між двома випадково вибраними рядками a та b становить

$$P(Hamming_{l,r}(a,b)) = 2^{-l}\sum_{i=r}^{l} C_l^i \qquad (2.1)$$

Імовірність отримують, з огляду на те, що 2^{-l} – це ймовірність одиничного збігу, а C_l^i - кількість рядків в U, які мають однакові біти в і позиціях.

Рисунок 2.4 показує, як вибір *r* і l впливає на ймовірність збігу між випадково вибраними бітовими рядками довжини l.

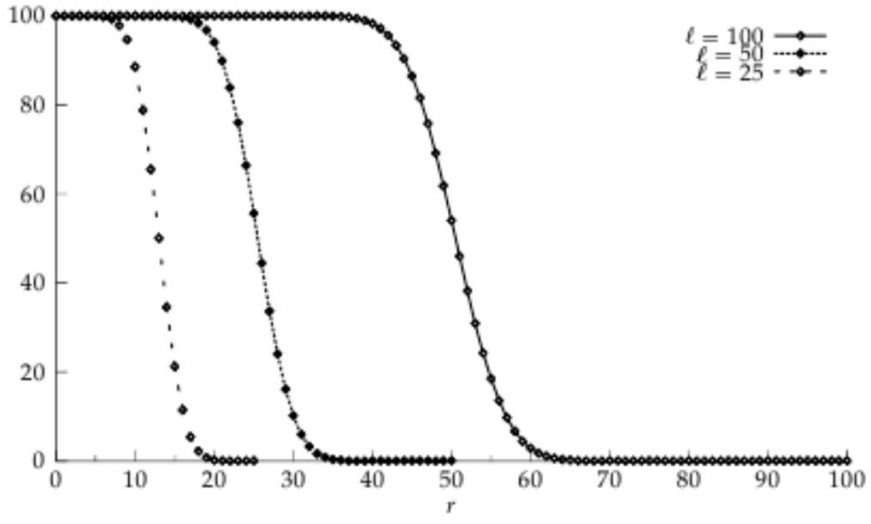


Рисунок 2.4 – Вплив величин *r* та *l* на ймовірність співпадіння (за правилом Хеммінга) двох випадкових рядків

Тексти про ШІС, які спираються на правило *r*- послідовного збігу, усі посилаються на формулу :

$$P_S = m^{-r}\left(\frac{(l-r)(m-1)}{m}+1\right) \tag{2.2}$$

де P_S - наближення значення ймовірності того, що випадковий рецептор розпізнає випадково вибраний АГ, у вигляді рядка фіксованої довжини, що складаються з символів алфавіту з потужності *m*, з відповідним порогом *r*, використовуючи *r*-послідовний збіг.

Але підставивши *m*=2, *l*=49 і *r*=4, (3.2) отримаємо значення

$$P_S = 2^{-4}\left(\frac{(49-4)(2-1)}{2}+1\right) = \frac{1}{16}\cdot\frac{47}{2} = 1.46875$$

Отже, цей приклад показує, що P_S не описує розподіл ймовірностей.

Тому імовірність збігу між двома випадково вибраними рядками фіксованої довжини за допомогою правила *r*-послідовного збігу потребує уточнення.

Механізм негативного відбору в ІС часто використовується в ШІС для проведення виявлення на основі аномалії. У [15] це змодельовано за умови, щоб дійсними детекторами були ті детектори, які не виявляють самоагентів під час толерування. Рисунок 2.5 ілюструє цей процес. По-перше, детектор генерується випадковим чином, що означає, що його рецептори можуть розпізнавати що завгодно. Якщо детектор щось розпізнає під час допуску, він гине. Якщо детектор переживає термін допуску, він стає зрілим і наївним детектором. Його називають наївним, оскільки він ще не виявив жодних збудників.

Рисунок 2.5 – Процес негативного відбору, який проходить детектор під час толерування

Це використання негативного відбору ґрунтується на припущенні, що детектор, який розпізнає що-небудь під час допуску, є своїм. Таким чином, ШІС дізнається, що все, що збігається з його зрілими детекторами та детекторами пам'яті, є чужим.

Поєднання клональної селекції та соматичної гіпермутації [19] є важливим фактором у процесі, відомому як дозрівання аффінності. Метою дозрівання аффінності є збільшення різноманітності в ІС та аффінності між детекторами та агентами в ІС.

Використання клональної селекції та соматичної гіпермутації для моделювання дозрівання аффінності в ШІС, застосованих до МСВВ, було запропоновано, але не реалізовано Хофмейром та Форестом [15]. Хоча було проведено деякі експериментальні роботи, що вивчають роль соматичної гіпермутації в ІС.

Оскільки детектори з часом контролюють кілька пептидів, це означає, що при низьких значеннях *r* детектори ШІС будуть відповідати практично будь-чому. З іншого боку, при високих значеннях *r* детектори будуть відповідати набагато меншому набору агентів.

Оскільки набір детекторів ШІС генерується за допомогою негативного відбору, менші значення *r* призводять до вищої ймовірності відповідності «свій» під час толерування, а більш високі значення *r* зменшують ймовірність відповідності «свій» під час толерування. Таким чином, чим нижче значення *r*, тим більше спроб потрібно ШІС для генерації кожного зрілого детектора. З вищими значеннями *r* необхідна менша кількість спроб

генерації детекторів, але для досягнення певного рівня покриття необхідний також більший набір детекторів.

Це призводить до ситуації компромісу, коли для нижчих значень *r* потрібен менший набір детекторів для досягнення певного покриття, тоді як ШІС потребує більше спроб для кожного дійсного детектора, який він генерує. Виходячи з ролі, яку поєднання клональної селекції та соматичної гіпермутації відіграє в ІС, передбачається, що такі механізми збільшать різноманітність детекторів та схожість між детекторами та агентами.

2.4 Висновки до розділу

В даному розділі проілюстровано роботу біологічної імунної системи. На прикладі такої системи проведено аналогію із штучною імунною системою. Описані підходи до виявлення вторгнень. Запропоновано розрізняти свої і чужі пристрої на основі певної інформації. Узагальнення своїх та чужих, яке відбувається в ІС, здійснюється за допомогою наближеного збігу рядків.

У найбільш загальній формі проблема наближеного узгодження рядків полягає у пошуку тексту, де виникає заданий шаблон, допускаючи обмежену кількість «помилок» у збігах.

3 АГЕНТНА МОДЕЛЬ ДЛЯ СИСТЕМИ ВИЯВЛЕННЯ ВТОРГНЕНЬ НА ОСНОВІ ШТУЧНОЇ ІМУННОЇ СИСТЕМИ

3.1 Запропонована структура ШІС та її компоненти

Для оптимальної роботи системи виявлення вторгнень необхідна правильна структура. Результатом неправильного проектування може бути зменшення ефективності СВВ. Запропонуємо структуру для СВВ – розподілена мережа, що використовує генетичний алгоритм для еволюції шаблонів виявлення та їх запам'ятовування.

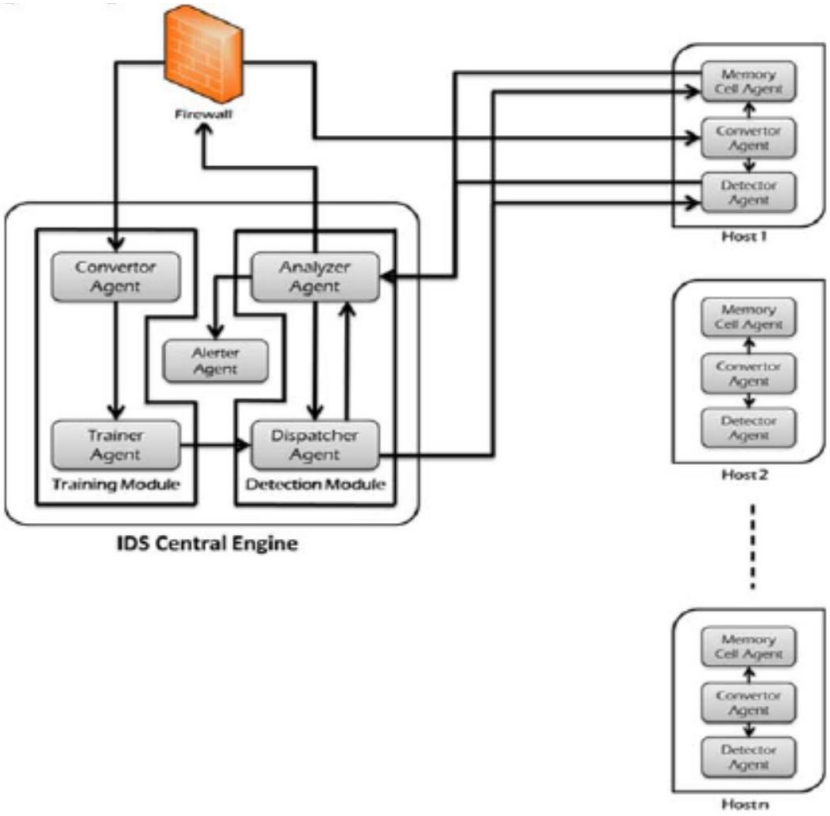


Рисунок 3.1 – Огляд архітектури запропонованої СВВ

Рисунок 3.1 показує основну структуру запропонованої СВВ на основі ШІС, яка складається з двох основних частин – основнне ядро ШІС та детекторів. Основне ядро ШІС розташоване на шлюзі кожної локальної мережі, а детектори - кожен користувач системи. Кожен з вказаних компонентів складається з агентів, що співставляють інформацію один від одного, щоб виявити аномалії та вторгнення. Ціль такого дизайну –

зменшити час виявлення для кожного з’єднання за допомогою надання можливостей детектора (аналіз трафіку та повідомлення про небезпеку) кожному користувачу. В результаті навантаження по обробці трафіку буде розподілятись на кожного користувача – кожен користувач сам відповідає за аналіз власного трафіку. Тому замість того, щоб аналізувати кожен пакет мережі (що створює велику потребу в обчислювальних можливостях та затримку в виявленні) центральне ядро буде обробляти сигнали небезпеки від користувачів мережі.

Основне ядро СВВ

Основне ядро складається з двох частин – модуль навчання та модуль детекторів-користувачів, обидві частини разом виконують чотири основні завдання:

- створення шаблонів ознак;
- аналіз повідомлень від користувачів;
- запам'ятовування робочих шаблонів;
- розподілення, синхронізація шаблонів ознак кожного детектора.

Кожен модуль є програмою на комп’ютері користувача чи шлюзові, що виконує одне із завдань (шлюз чи користувач може мати декілька програм одночасно).

Модуль навчання складається з програми-дешифратора та програми навчання, на ньому лежить відповідальність за створення основних випадкових шаблонів ознак на ранніх стадіях роботи системи. Модуль детекторів складається з програми аналізу та програми-диспетчера вторгнень. Перша програма обробляє сигнали від користувачів та в певних випадках запам’ятовує шаблони на які користувач зреагував і схрещує їх для виконання генетичного алгоритму, друга програма відповідає за розповсюдження та синхронізацію шаблонів між користувачами-детекторами.

Програма-дешифратор.

Перед оцінкою системи відбувається її попереднє навчання та налаштування параметрів. Попереднє навчання ШІС відбувається з

допомогою використання набору безпечних (своїх) даних та небезпечних (чужих) даних. Для обробки пакетів трафіку їх спочатку необхідно розшифрувати та перетворити в оброблену інформацію. Інформація містить такі поля як IP надсилача, IP отримувача, порт надсилача, порт отримувача, протокол, розмір пакету. Ця інформація дістається з пакетів та перетворюється в послідовності з 112 бітів. Таблиця 3.1 показує правильно розшифрований приклад елементів інформації, що отримується, та їх розмір (в бітах).

Таблиця – Приклад можливої інформації з пакету

Назва поля	Мінімум - максимум	Кількість бітів
IP отримувача	0.0.0.0 - 255.255.255.255	32 біта
IP надсилача	0.0.0.0 - 255.255.255.255	32 біта
порт отримувача	0 – 65535	16 бітів
порт відправника	0 – 65535	16 бітів
час продовження	0 – 4096 секунд	12 бітів
протокол	0 – 16	4 біта

Програма навчання.

Після дешифрування всіх тренувальних наборів інформації в бітові послідовності їх передають в програму навчання, що використовується для створення шаблонів для детекторів. Для створення першого покоління шаблонів використовується алгоритм негативного відбору.

Спочатку створюється і перевіряється "молодий" випадковий набір бітових послідовностей-шаблонів на базовому наборі тестових даних. Якщо шаблон спрацьовує на "своїх" пакетах, то він заміняється новим і так до тих пір, поки шаблон не перестане реагувати на свої пакети. Далі відбувається наступний крок алгоритму негативного відбору, що відсіює шаблон, що не реагує на жоден з "чужих" пакетів. Все, що не відсіялось, додається до результуючого набору шаблонів. Цей процес повторюється поки кожен з

чужих пакетів не співпаде хоча б із трьома шаблонами з результуючого набору шаблонів.

Для порівняння послідовностей використовується правило *r*-послідовного збігу (розділ 2.3 даної роботи).

Процес навчання основних детекторів показаний на рисунку 3.2.

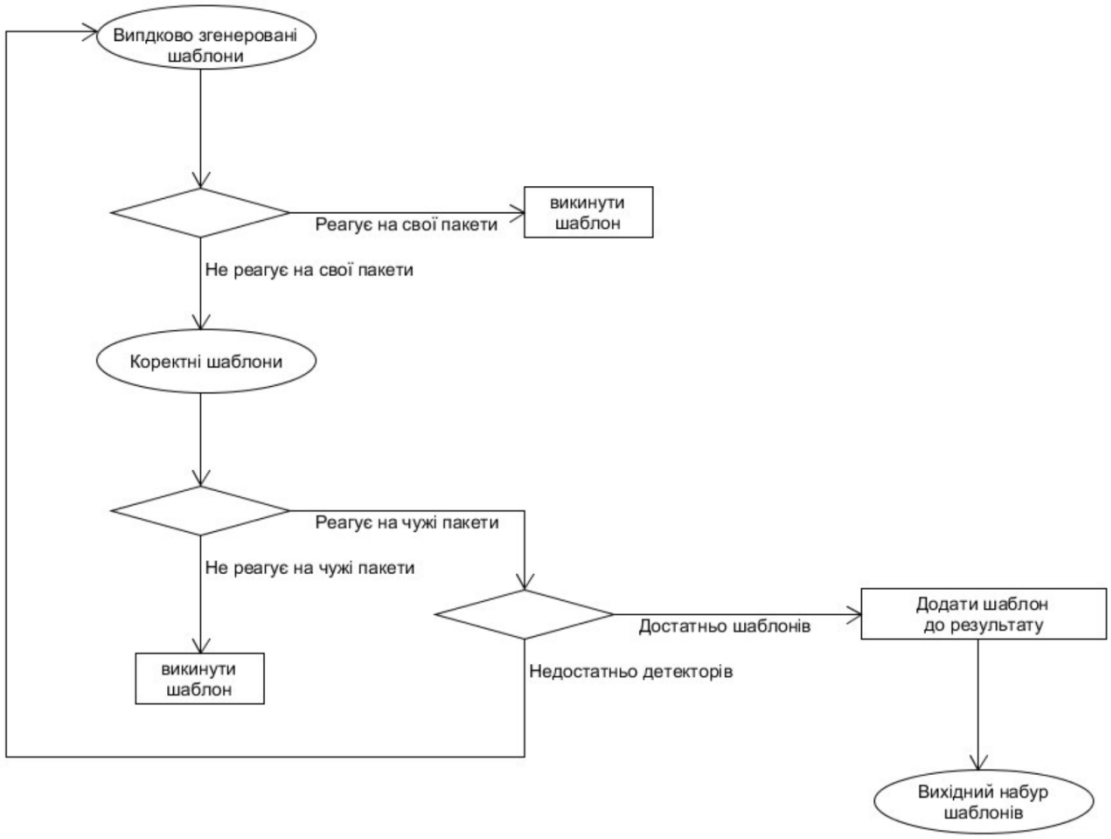


Рисунок 3.2 – Навчання детекторів

Програма-диспетчер.

Після навчання всі детектори мають отримати набір шаблонів. Цю роль бере на себе програма-диспетчер, що також синхронізує зміни шаблонів, а також встановлює для них шаблони ознак. Диспетчер також отримує сигнали небезпеки та перенаправляє їх до програми-аналізатора для обробки.

Програма-аналізатор.

Як тільки у користувача відбудеться вторгнення і користувач його помітить за шаблоном, то він відправить повідомлення, яке буде включати підозрілий пакет. Таким чином інформація про вторгнення, наприклад,

кількість детекторів, що помітили небезпеку, їх відношення до підозрілого пакету та їх профіль буде надіслано до основного ядра СВВ на аналіз та обробку. Програма-аналізатор з модуля-детектора цим і займається. Якщо кількість підозрілих пакетів виходить за межі норми, то відбувається покращення шаблонів. Для створення шаблонів, що краще виділяють певні види аномалій, варто реалізувати генетичний алгоритм. В той же самий час аналізатор спробує заблокувати пакети певного типу через фаєрволл. Також якщо кількість повідомлень про небезпеку менша за нижню границю небезпеки, то пакет все одно додається для подальшої обробки в пасивному режимі. Після створення нових шаблонів, вони будуть передані програмі-диспетчеру для розповсюдження між вузлами-користувачами.

Коли задіяні шаблони надсилаються в аналізатор, то генетичний алгоритм використовується для створення з них шаблонів, що будуть скопійовані для початкового покоління для відбору.

Формула яка визначає чи нам треба відбирати шаблон для генетичного алгоритму:

[Мінімальна оцінка принадності для відбору]=
=[сума оцінок шаблонів] / [кількість шаблонів]

Шаблони, оцінка яких вища за мінімальну, використовуються для створення наступного покоління з допомогою генетичного алгоритму. Кожен шаблон може копіюватись певну кількість разів – кількість визначається за формулою:

[кількість копій] =
= ціла частина(10 * [оцінка шаблону] / [сума оцінок шаблонів])

Після виконання клонувань виконується генетичний алгоритм – вибрані детектори проходять через операції кросоверу, мутації та репродукції певну кількість поколінь. В кожному поколінні визначається нова сума

оцінок шаблонів. В кожному поколінні вибирається новий кандидат на додавання. І якщо його оцінка менше ніж максимальна з початкового шаблону, то генетичний алгоритм зупиняється і кандидат на додавання розповсюджується між користувацькими вузлами-детекторами. Якщо через певну кількість поколінь не можна зробити кращий шаблон, то розповсюджується кращий із створених шаблонів.

Користувацькі вузли-детектори.

Для покращення механізму ССВ і збільшення ефективності шаблони розповсюджуються на всі вузли в мережі. Це також зумовлює простоту і розширюваність такої системи. В системі присутні два типи вузлів-детекторів: вузли пам’яті і активні детектори. Дешифратор використовується для перетворення пакетів для аналізу активними детекторами.

Вузли пам’яті

Вузли пам’яті дозволяють робити адаптивну відповідь ШІС на вторгнення. Вузли пам’яті містять набір шаблонів, що створюється і змінюється використовуючи генетичний алгоритм. Аналізатор виконує раніше вказаний генетичний алгоритм, що дозволяє детекторам краще ідентифікувати втручання. Використання вузлів пам’яті дозволяє зменшувати час відповіді і краще реагувати на раніше помічені види втручань. Також такий підхід збільшує ефективність ССВ, зменшуючи час обробки пакетів. Вузли пам’яті також добре працюють для зменшення кількості неправильних позитивних та неправильних негативних відкликів. Як тільки аномалія була помічена на вузлі і будь-який шаблон з вузлів пам’яті підійшов під трафік в мережі, підходящий трафік буде направлено на аналіз в основне ядро СВВ. Весь процес аналізу з серверної та клієнтської сторони показаний на рисунку 3.3.

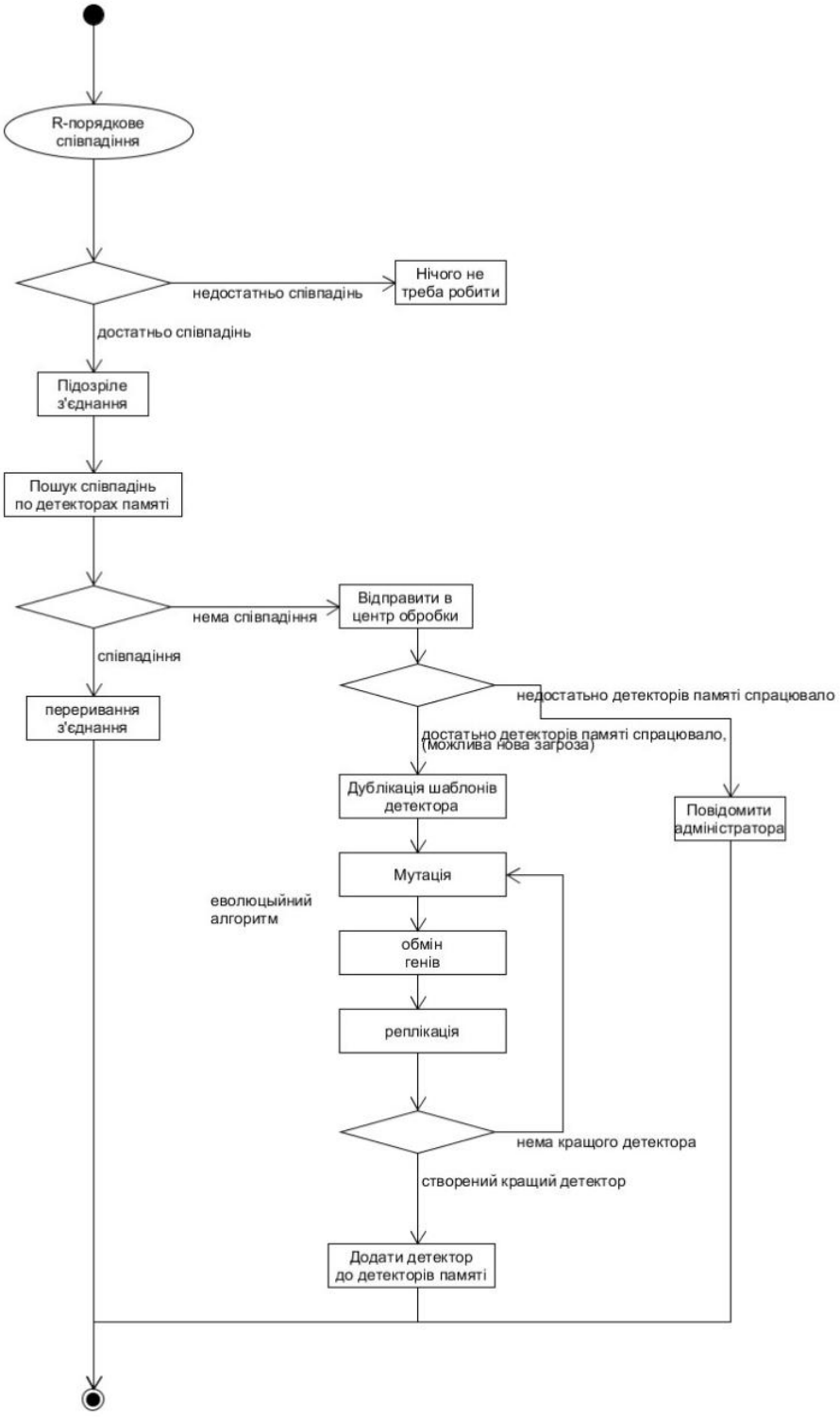


Рисунок 3.3 – Блок-схема процедури виявлення

Активний детектор.

Активний детектор містить набір шаблонів, що точно розрізняють трафік за схемою свій-чужий. Всі вхідні пакети перевіряються цими детекторами. Якщо будь-який пакет визнано аномальним за будь-яким шаблоном, то вказаний пакет передається далі на аналіз в основне ядро ССВ

для обробки. Кількість шаблонів, що були задіяні на підозрілому пакеті, оцінка кожного з задіяних шаблонів, властивості пакету – все це необхідно кожен раз передавати для аналізу в основне ядро. Межа підозрілості пакету – властивість, що дозволяє збільшувати точність визначення втручання і відсіяти неправильні позитивні спрацьовування. Якщо кількість шаблонів, що були зачеплені при аналізі, більша за мінімальну межу підозрілості, то сесія з даним пакетом буде примусово відключена фаєрволом.

3.2 Багатошарова структура запропонованої архітектури

Щоб отримати потрібну модель системи та розділити одну велику проблему на багато менших, що можуть бути вирішені окремо та дозволять СВВ працювати ефективно, аналіз на втручання відбувається в багатопотоковому режимі. Кожен рівень має власні завдання та механізм виявлення, що співпрацює з суміжними рівнями для автоматичної відповіді на втручання. Як вказано на рисунку 3.4, багатопотокова архітектура складається з 3 рівнів.

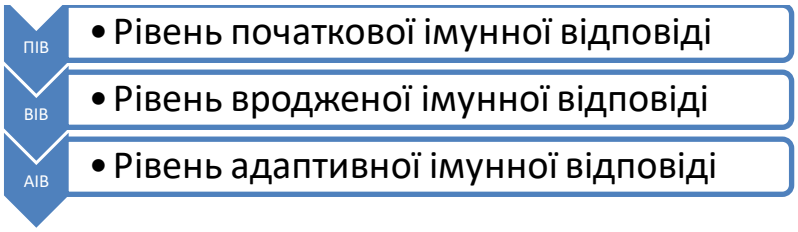


Рисунок 3.4 – Багатошарова структура СВВ

Рівень початкової імунної відповіді (ПІВ)

Мережевий фаєрвол – це зовнішній захисний механізм в даній моделі, також це один із перших бар'єрів, який захищає систему від втручання. Незважаючи на наявність фаєрволу, не всі втручання помічаються і припиняються ним. Фаєрвол розташований на шлюзі і має доступ до всіх пакетів, які проходять всередину і назовні з мережі. Фаєрвол контролює

вхідний і вихідний трафік відповідно до завдань, які йому були надані. Наприклад, згідно з одним завданням він може співпрацювати з іншими рівнями, які виявляють втручання та отримувати дані про пакети, які треба блокувати.

Рівень вродженої імунної відповіді (ВІВ)

Вроджена імунна система спрацьовує на ранніх періодах втручання. Вона використовується для визначення поведінки при втручанні і позначення атакуючих пакетів. ВІВ має неспеціалізований захист для негайної відповіді на підозрілі пакети. Цей рівень складається з дешифратора на клієнтській стороні та з диспетчера на серверній. Це дозволяє знизити затримку відповіді на втручання на локальних вузлах замість перекладання всієї роботи по фільтруванню трафіку на сервер/фаєрвол. В результаті такого підходу СВВ має високу швидкість виявлення втручання на клієнтських вузлах.

Рівень адаптивної імунної відповіді (АІВ)

Адаптивна імунна відповідь складається з спеціалізованих та загальних системних детекторів, що називаються модулі пам’яті. Цей рівень має надзвичайно високу ймовірність виявлення втручань, що траплялись раніше. Також цей рівень має надзвичайно високий спеціалізований захист від аномалій певного типу.

3.3 Висновки до розділу

1. Проаналізовано можливість використання штучної імунної системи для виявлення «чужих» пристроїв. Запропонована багатошарова структура такої імунної системи. Кожен рівень такої системи має власні завдання та механізм виявлення, що співпрацює з суміжними рівнями для автоматичної відповіді на втручання.

2. Запропонована структура системи виявлення вторгнень у вигляді розподіленої мережі, що використовує генетичний алгоритм для еволюції шаблонів виявлення та їх запам'ятовування.

3. Попереднє навчання ШІС відбувається з допомогою використання набору безпечних (своїх) даних та небезпечних (чужих) даних. Для обробки пакетів трафіку їх спочатку необхідно розшифрувати та перетворити в оброблену інформацію, яка містить такі поля як IP надсилача, IP отримувача, порт надсилача, порт отримувача, протокол, розмір пакету. Ця інформація дістається з пакетів та перетворюється в послідовності з 112 бітів. В подальшому дану інформацію використаємо для нечіткої системи управління, щоб дати відповідь на питання чи даний пристрій «свій» чи «чужий».

4 НЕЧІТКА СИСТЕМА ВИСНОВКУ ЩОДО ІДЕНТИФІКАЦІЇ ПРИСТРОЇВ КОРПОРАТИВНОЇ МЕРЕЖІ ЗА ПРИНЦИПОМ СВІЙ/ЧУЖИЙ

4.1 Методологія логіко-лінгвістичного моделювання

Нечітка множина – це пара (U, m), де U - це множина і $m : U \rightarrow [0,1]$ функція приналежності. Множина U називається універсальною, і для кожного елемента x з U, значення $m(x)$ називається ступенем приналежності x в (U, m). Функція $m = \mu_A$ називається функцією приналежності нечіткої множини $A = (U, m)$.

Нехай $x \in U$. Тоді скажемо, що *x:*

- не входить до нечіткої множини (U, m), якщо $m(x)=0$,
- повністю належить, якщо $m(x)=0$,
- частково належить, якщо $0< m(x)<1$ (нечіткий член).

Нечітка та лінгвістична змінні використовуються для опису об’єктів та явищ за допомогою нечітких множин.

Нечітка змінна — це набір $< \alpha, X, A >$, де

— α — назва змінної,

— X — область визначення α,

— A — нечітка множина на X, яка описує обмеження (тобто $\mu A(x)$) на значення нечіткої змінної α.

Лінгвістична змінна — це набір $< \beta, T, X, G, M >$, де

— β — назва лінгвістичної змінної;

— T — базова терм-множина значень, які представляють собою імена нечітких змінних, областю визначення, кожної з яких є множина X.

— G — синтаксична процедура, яка дозволяє оперувати елементами базової терм-множини T, зокрема, генерувати нові терми. Множина $G(T)$ — розширена терм-множина лінгвістичної змінної, тобто множина утворених термів;

— M — семантична процедура, яка дозволяє перетворити усі нові значення лінгвістичної змінної, які утворені процедурою G, у нечітку змінну, тобто вона дозволяє сформувати відповідну нечітку множину [7].

Терм-множина — це множина всіх можливих значень лінгвістичної змінної. Терм — це елемент терм-множини. У теорії нечітких множин терм визначається нечіткою множиною за допомогою використання функції приналежності. Нечіткий терм — це така нечітка множина, яка має властивість певного поняття [7].

Нечіткими висловленнями називається висловлення наступного виду:

1. Висловлення «$\beta \in \beta'$», де β — найменування лінгвістичної змінної, β' — її значення, якому відповідає нечітка множина на універсальній множині X. Означає, що змінна β має властивість β'. Наприклад, висловлення «сигнал слабкий» припускає, що лінгвістичній змінній «сигнал» надається значення «слабкий», для якого на універсальній множині X змінної «сигнал» визначена, відповідно даному значенню «слабкий», нечітка множина.

2. До висловлень можуть використовуватись підсилювачі, яким в природній мові відповідають слова «Дуже», «Більш-менш», «Набагато більше» та інші. Отримуються таким чином нові висловлення «$\beta \in m\beta'$», де m — один з цих модифікаторів. Наприклад: «сигнал дуже слабкий», «зріст нижче середнього» і т.д.

3. Інші висловлювання можуть утворюватись з вже існуючих за допомогою сполучників та конструкцій мови «І», «Або», «Якщо.., Тоді..», «Якщо.., Тоді.., Інакше». Такі висловлювання називаються складними [7].

Структура нечіткого логічного висновку

Нечіткий висновок займає очільне місце в нечіткій логіці (Fuzzy Logic) і системах нечіткого управління. Нечіткий висновок насправді є певним алгоритмом отримання нечітких висновків на основі нечітких умов. Системи нечіткого висновку служать концептуальним базисом всієї сучасної нечіткої логіки.

Оскільки розробка і застосування систем нечіткого висновку має міждисциплінарний характер, дана проблематика досліджень тісно взаємопов'язана з цілою низкою інших науково-прикладних напрямів, зокрема: нечітке моделювання, нечіткі експертні системи і т.д. [7] Нечіткі системи мають широке застосування в багатьох сферах науки і людського життя: і економіка, і прикладна математика, і педагогіка, і психологія, і медицина, і таке інше.

Основою нечіткого логічного висновку є база правил, яка містить нечіткі висловлювання вигляду *«Якщо — то»* і функції приналежності для всіх лінгвістичних термів. Але необхідне дотримання таких умов:

1) Для кожного лінгвістичного терма вихідної змінної існує хоча б одне правило.
2) Кожен терм вхідної змінної використовується щонайменше в одому з правил в якості передумови. В іншому випадку говорять, що має місце неповна база нечітких правил.

Нехай база правил містить в собі m правил виду:

$$
\begin{aligned}
&R_1\text{: Якщо } x_1 \in A_{11} \dots i \dots x_n \in A_{1n}\text{ , } TO\ y \in B_1\\
&\dots\\
&R_i\text{: Якщо } x_1 \in A_{i1} \dots i \dots x_n \in A_{in}\text{ , } TO\ y \in B_i \qquad (4.1)\\
&\dots\\
&R_m\text{: Якщо } x_1 \in A_{i1} \dots i \dots x_n \in A_{mn}\text{ , } TO\ y \in B_m,
\end{aligned}
$$

де $x_k, (k = \overline{1,n})$ — вхідні змінні,

y — вихідна змінна,

A_{ik} — деякі нечіткі множини разом із своїми функціями приналежності.

В результаті нечіткого виводу отримаємо чітке значення змінної y на основі заданих чітких значень $x_k, (k = \overline{1,n})$.

Процедура логічного висновку складається з чотирьох етапів: 1) фаззифікація - введення нечіткості; 2) нечіткий висновок; 3) композиція і 4) дефаззифікація - приведення до чіткості (рисунок 4.1).

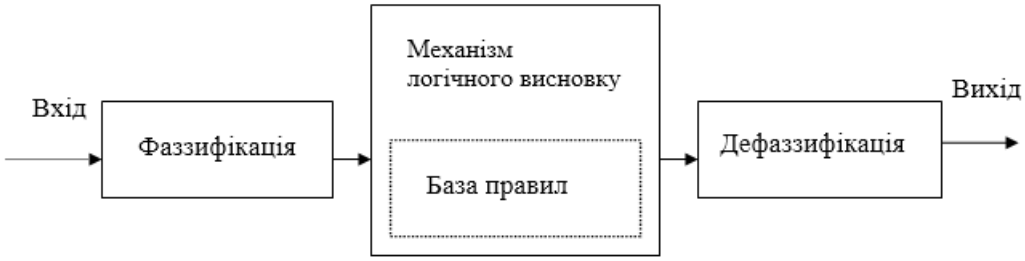


Рисунок 4.1 — Структура нечіткого логічного висновку

Всі відомі алгоритми нечіткого висновку розрізняються головним чином виглядом використовуваних правил, логічних операцій і різновидом методу дефаззифікації. Розроблено алгоритми нечіткого висновку Мамдані, Сугено, Ларсена, Цукамото. Найчастіше на практиці в задачах нечіткого моделювання застосовують алгоритм Мамдані [5].

Фаззифікація — зіставлення множини значень *x* з її функцією приналежності *M(x),* тобто переведення значень *x* в нечіткий формат.

Дефаззифікація — процес, обернений до фаззифікації.

Всі системи з нечіткою логікою працюють за одним принципом: показання вимірювальних приладів фаззифікуються (переводяться в нечіткий формат), обробляються, дефаззифікуються і у вигляді звичних сигналів далі подаються на виконавчі пристрої [7].

На сьогоднішній день існує декілька алгоритмів нечіткого виведення, працюючих по даній схемі, найпоширенішими є алгоритм нечіткого висновку Мамдані та Сугено.

Алгоритм Мамдані. Формально даний алгоритм можна розділити на такі кроки:

1. Формування бази правил системи нечіткого висновку. База правил являє собою скінченну кількість правил, які сформовані в таблицю.

2. Фаззифікація вхідних змінних — процедура знаходження значень вхідних змінних на основі звичайних вхідних даних. Етап фазиффікації

також називають введенням нечіткості. Після проведення цього етапу кожна змінна набуває свого значення.

3. Агрегування — процедура виявлення ступеня істинності умов за кожним з правил системи нечіткого виводу. Такий ступінь істинності умов для кожного з правил може бути обрахований за допомогою формули (4.2).

$$\mu(x_1 \cap x_2 \cap \ldots \cap x_n) = \min\big(\mu(x_1), \mu(x_2), \ldots, \mu(x_n)\big), \tag{4.2}$$

де n — кількість змінних.

4. Активація — процедура знаходження ступеня істинності кожного з підзаключень правил нечіткого висновку. Це значення можна знайти за формулою (4.3):

$$\mu'(y) = c_i \cdot \mu(y), \tag{4.3}$$

де $c_i, (i = \overline{1,n})$ — підзаключення в базі правил,

$\mu(y)$ — функція приналежності терма, який являється значенням деякої вихідної змінної y.

5. Аккумуляція — процес об’єднання всіх ступенів істинності для отримання функції приналежності вихідної змінної. Це можна знайти за формулою (4.4)

$$\mu(y) = \max\big(\mu_1(x_1), \mu_2(x), \ldots, \mu_n(x)\big). \tag{4.4}$$

6. Дефаззифікація вихідних змінних — процедура знаходження звичайного значення вихідної змінної. Використовується метод центра ваги, при якому значення вихідної змінної розраховується за формулою (2.5) .

$$y = \frac{\sum_{i=1}^{n} x_i \cdot \mu(x_i)}{\sum_{i=1}^{n} \mu(x_i)}. \tag{4.5}$$

Алгоритм Мамдані є одним з найпопулярніших алгоритмів, які застосовують на практиці в задачах нечіткого моделювання.

Алгоритм Сугено запропонований Сугено і Такагі. Даний алгоритм складається з таких кроків.

1. Формується база правил системи нечіткого виводу. Для цієї бази правил використовуються правила нечітких продукцій в формі:

ПРАВИЛО $<\#>$: Якщо "$\beta_1 \in \alpha'$" ТА "$\beta_2 \in \alpha''$"ТО "$w = \varepsilon_1 \cdot \alpha_1 + \varepsilon_2 \cdot \alpha_2$". (4.6)

Тут $\varepsilon_1, \varepsilon_2$ — деякі вагові коефіцієнти. У цьому випадку значення вихідної змінної w буде деяке дійсне число.

2. Фаззифікація вхідних змінних. Особливості фаззифікації збігаються з розглянутими вище при описі даного етапу.

3. Для знаходження ступеня істинності всіх правил нечітких продукцій використовується логічна операція min-кон'юнкції. Частина правил, ступінь істинності умов яких не нуль, вважаються активними і можуть бути використані для подальших розрахунків. Таким чином відбувається крок агрегування.

4. Активізація підзаключень в нечітких правилах продукцій: спочатку за допомогою методу (4.7) знаходять значення ступенів істинності всіх висновків правил нечітких продукцій.

$$\mu'(y) = \min\{c_i, \mu(y)\}, \tag{4.7}$$

Далі отримуємо значення вихідних змінних кожного правила, але це будуть звичайні, а не нечіткі, значення. Для цього застосовуємо формулу (4.6), в яку замість α_1 і α_2 підставляються значення вхідних змінних до етапу фаззифікації. Тим самим визначається множина значень $C = \{c_1, c_2, \dots, c_n\}$ і множина значень вихідних змінних $W = \{w_1, w_2, \dots, w_n\}$, де n — загальна кількість правил в базі правил.

5. Аккумуляція висновків нечітких правил продукцій. Фактично відсутній, оскільки розрахунки здійснюються із звичайними дійсними числами w_j.

6. Дефаззифікація вихідних змінних. Використовується модифікований варіант у формі методу центру ваги для одноточкових множин (4.8).

$$y = \frac{\sum_{i=1}^{n} c_i \cdot w_i}{\sum_{i=1}^{n} c_i}, \tag{4.8}$$

Якщо порівняти моделі нечіткого виведення по Мамдані і моделі типу Сугено, то можна відзначити, що вони відрізняються між собою форматом бази знань і процедурою дефаззифікації. Ці моделі є універсальними аппроксиматорами, але при великих об'ємах вибірки експериментальних даних ідентифікація за допомогою моделі типу Сугено забезпечує, як правило, більшу точність. Однак при цьому можуть виникнути труднощі з змістовною інтерпретацією параметрів нечіткої моделі і з рекомендаціями щодо логічного висновку. З моделлю типу Мамдані таких труднощів не виникає, її параметри легко інтерпретуються змістовно.

Можна зробити висновок, що для задач, де більш важливим є пояснення, обґрунтування прийнятого рішення, матимуть перевагу нечіткі моделі типу Мамдані, а для задач, де більш важливим є точність ідентифікації нелінійних залежностей, доцільним буде використання нечітких моделей типу Сугено.

Розглянуті теоретичні засади лежать в основі даної роботи.

Повідомлення детекторів, що поступають на центральний сервер корпоративної мережі, підлягають терміновому аналізу системним адміністратором. Пропонується розгляд даної інформації у вигляді бітового рядка, що розбивається на ділянки (таблиця 3.1). Інтерпретація змісту неспівпадаючих бітів в тому чи іншому місці цього двійкового ланцюга дає змогу перейти до математичного опису існуючих залежностей шляхом введення лінгвістичних змінних. В основу нечіткого логічного висновку покладемо алгоритм Мамдані.

4.2 Математична модель

У процесі формування структури моделі було виділено такі основні фактори впливу на ідентифікацію кінцевих пристроїв в корпоративній мережі (таблиця 4.1).

Таблиця 4.1 – Опис змінних моделі

Змінна	Позначення	Тип змінної
IP отримувача (32 біти)	x_1	Вхідна
IP надсилача (32 біта)	x_2	Вхідна
порт отримувача 16 бітів	x_3	Вхідна
порт відправника 16 бітів	x_4	Вхідна
час продовження 12 бітів	x_5	Вхідна
протокол	x_6	Вхідна
Частота	x_7	Вхідна
Несанкціонований доступ	y_1	Проміжна
Аномальна поведінка	y_2	Проміжна
Ідентифікація пристроїв	Y	Вихідна

На розпізнавання кінцевих пристроїв корпоративної мережі за принципом свій/чужий впливають 7 основних факторів (таблиця 4.1).

Велика кількість вхідних змінних значно ускладнює задачу опису причинно-наслідкових зв’язків за допомогою нечітких правил. Тому, при наявності великої кількості вхідних змінних, їх потрібно ієрархічно класифікувати.

Ієрархічні системи нечіткого висновку — це такі системи, в яких висновок однієї бази знань подається як вхідний параметр іншої, яка знаходиться на вищому рівні ієрархії. Зворотні зв’язки в таких системах відсутні.

Перевагою ієрархічних систем є компактність баз знань. Для проміжних змінних не виконуються фаззифікація та дефаззифікація. Логічний висновок підсистеми нижчого рівня одразу подається у вигляді нечіткої множини на вхід підсистеми вищого рівня ієрархії [3].

Після позначення вхідних, проміжних та результуючих змінних нечітка модель, матиме вигляд як на рисунку 4.1.

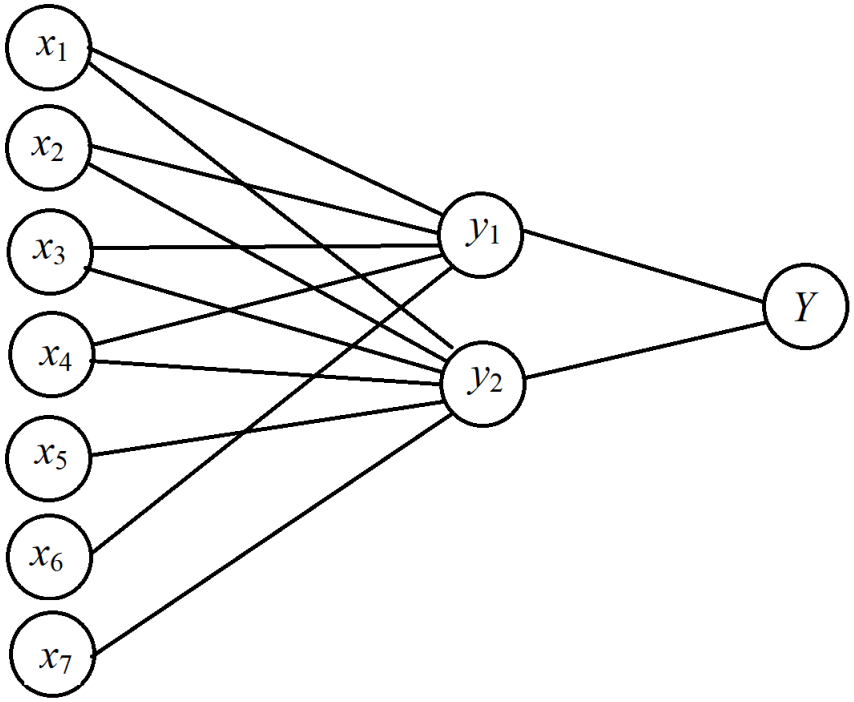


Рисунок 4.1 – Нечітка модель розпізнавання кінцевих пристроїв за принципом свій/чужий

Описані фактори впливу можна віднести до двох підсистем.

Підсистема «Несанкціонований доступ» описує спробу підключення до мережі користувача з недозволеного пристрою (IP отримувача, IP надсилача, порт отримувача, порт відправника, протокол):

$$y_1 = f(x_1, x_2, x_3, x_4, x_6). \tag{4.9}$$

Підсистема «Аномальна поведінка» описує спробу користувача в мережі отримати доступ до інформації, що не відповідає рівню доступу або

спроба передачі інформації нетипового об’єму (IP отримувача, IP надсилача, порт отримувача, порт відправника, тривалість, протокол):

$$y_2 = f(x_1, x_2, x_3, x_4, x_5, x_7). \quad (4.10)$$

Ідентифікація пристрою корпоративної мережі визначається як інтегральний показник вихідних параметрів описаних підсистем. Модель системи має вигляд:

$$Y = f(y_1, y_2). \quad (4.11)$$

Ідентифікація пристрою корпоративної мережі вимірюється від 0 до 1 за такою шкалою:

$0 \leq V \leq 0{,}3$ – свій (не потребує втручання системного адмінімтратора та охорони) ;

$0{,}3 < V \leq 0{,}6$ – свій, але підозрілий (потребує огляду системним адміністратором);

$0{,}60 < V \leq 1$ – чужий (рекомендовано заблокувати негайно даний пристрій).

Необхідно кожному показнику задати оцінку (рівень) вагомості $p_i \in [0; 10]$, $i = \overline{1{,}2}$ та скористатись лінійною згорткою, яка дозволяє отримати інтегральний показник в тих випадках, коли вхідними змінними є незалежні та рівноцінні величини:

$$Y = \sum_{i=1}^{2} y_i \cdot \beta_i, \quad (4.12)$$

де β_i – коефіцієнт вагомості, який характеризує відносну важливість показників, і обчислюється:

$$\beta_i = \frac{p_i}{\sum_{i=1}^{2} p_i}, \quad \sum_{i=1}^{2} \beta_i = 1. \tag{4.13}$$

З огляду на літературу та здогадки експертів з безпеки запропоновані коефіцієнти вагомості для усіх факторів та показників (таблиця 4.2).

Таблиця 4.2 – Важливість лінгвістичних змінних

Змінна	Вага	Змінна	Вага
$ЛЗ_1$	$\frac{10}{44}$	$ЛЗ_6$	$\frac{5}{44}$
$ЛЗ_2$	$\frac{10}{44}$	$ЛЗ7$	$\frac{4}{44}$
$ЛЗ_3$	$\frac{9}{44}$		
$ЛЗ_4$	$\frac{9}{44}$	$ПЛЗ_1$	0,4
$ЛЗ_5$	$\frac{7}{44}$	$ПЛЗ_2$	0,6

Опис лінгвістичних змінних

Загалом, ідентифікація пристрою корпоративної мережі залежить від 7 основних вхідних параметрів.

Множина T = { рідко (rarely), часто (often), дозволений (allowable), недозволений (unauthorized), частково дозволений (partially allowed), достатній (sufficient), недостатній (insufficient)} для всіх лінгвістичних змінних.

Введемо лінгвістичні змінні, визначимо множину X та базову терм-множину T, задамо функцію приналежності:

$ЛЗ_1$ – IP надсилача (Source IP Address), її область визначення $X = [0; 1]$, вимірюється за такою шкалою:

$ЛЗ_1 \leq 0{,}5$ – недозволений (unauthorized) – пристрій з таким IP неможливий у мережі;

$0{,}5 < ЛЗ_1 \leq 0{,}7$– частково дозволений (partially allowed) - пристрій з таким IP можливий у мережі за певних умов;

$ЛЗ_1 > 0{,}7$ – дозволений (allowable).

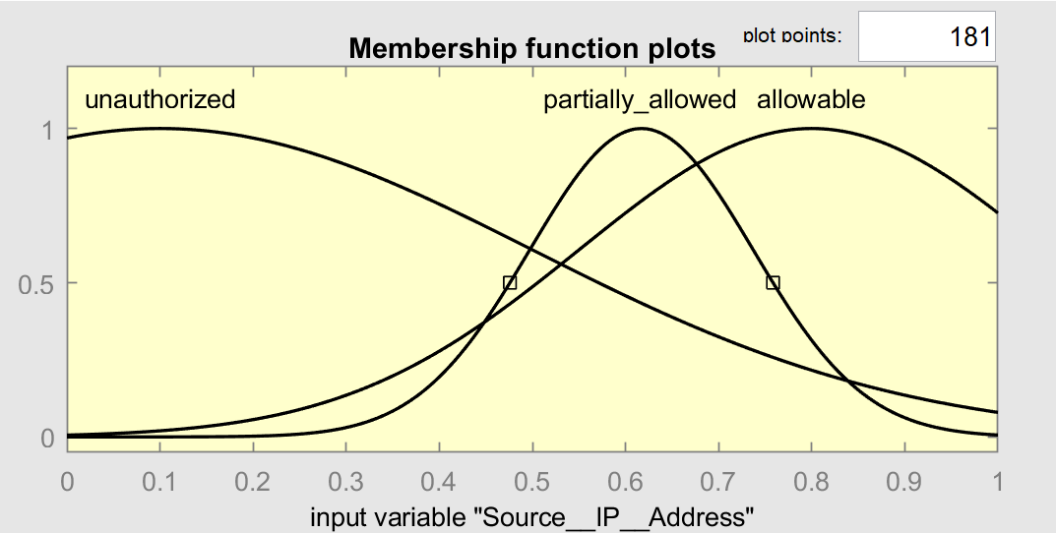


Рисунок 4.2 – Графік функції приналежності вхідної змінної $ЛЗ_1$

$ЛЗ_2$ – IP отримувача (Destination IP Address), її область визначення $X = [0; 1]$, вимірюється за такою шкалою:

$ЛЗ_2 \leq 0{,}5$ – недозволений (unauthorized) – пристрій з таким IP неможливий у мережі;

$0{,}5 < ЛЗ_2 \leq 0{,}7$– частково дозволений (PartiallyAllowed) - пристрій з таким IP можливий у мережі за певних умов.

$ЛЗ_2 > 0{,}7$ – дозволений (allowable).

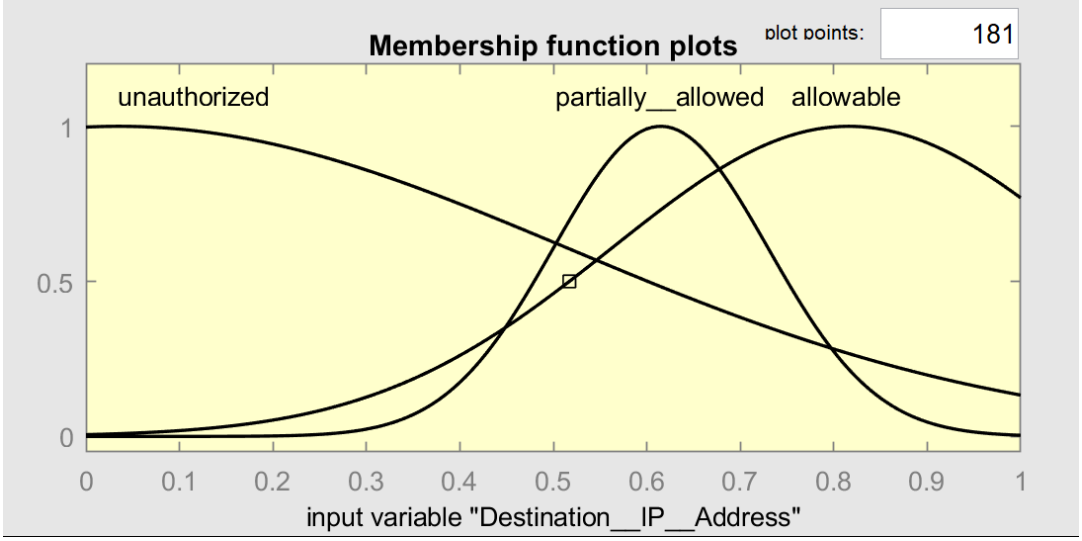


Рисунок 4.3 – Графік функції приналежності вхідної змінної $ЛЗ_2$

$ЛЗ_3$ – порт отримувача (Destination Port No), її область визначення $X = [0; 1]$, вимірюється за такою шкалою:

$ЛЗ_3 \leq 0{,}5$ – недозволений (unauthorized);

$0{,}5 < ЛЗ_3 \leq 0{,}7$– частково дозволений (PartiallyAllowed)

$ЛЗ_3 > 0{,}7$ – дозволений (allowable).

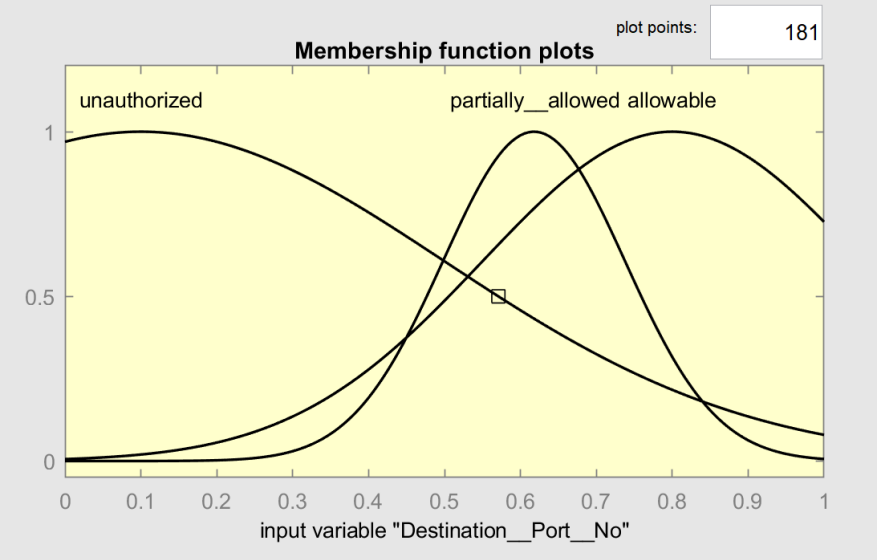


Рисунок 4.4 – Графік функції приналежності вхідної змінної $ЛЗ_3$

$ЛЗ_4$ – порт відправника (SourcePortNo), її область визначення $X = [0; 1]$, вимірюється за такою шкалою:

$ЛЗ_4 \leq 0{,}5$ – недозволений (unauthorized)

$0{,}5 < ЛЗ_4 \leq 0{,}7$– частково дозволений (PartiallyAllowed)

$ЛЗ_4 > 0{,}7$ – дозволений (allowable).

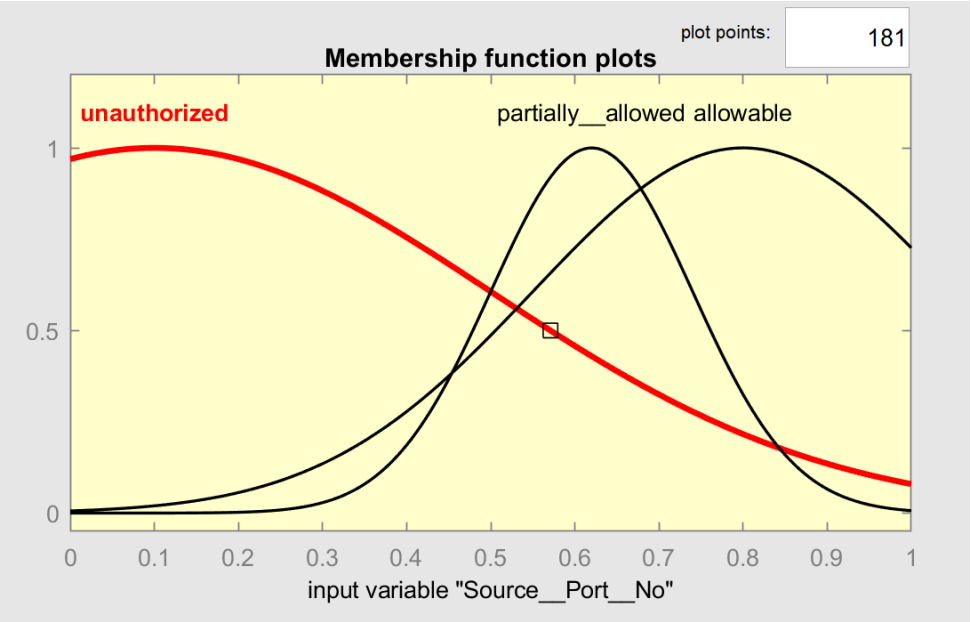


Рисунок 4.5 – Графік функції приналежності вхідної змінної $ЛЗ_4$

$ЛЗ_5$ – тривалість (Duration), її область визначення $X = [0; 1)$, вимірюється за такою шкалою:

$ЛЗ_5 \leq 0{,}3$ – недостатній (insufficient);

$ЛЗ_5 > 0{,}3$ – достатній (sufficient).

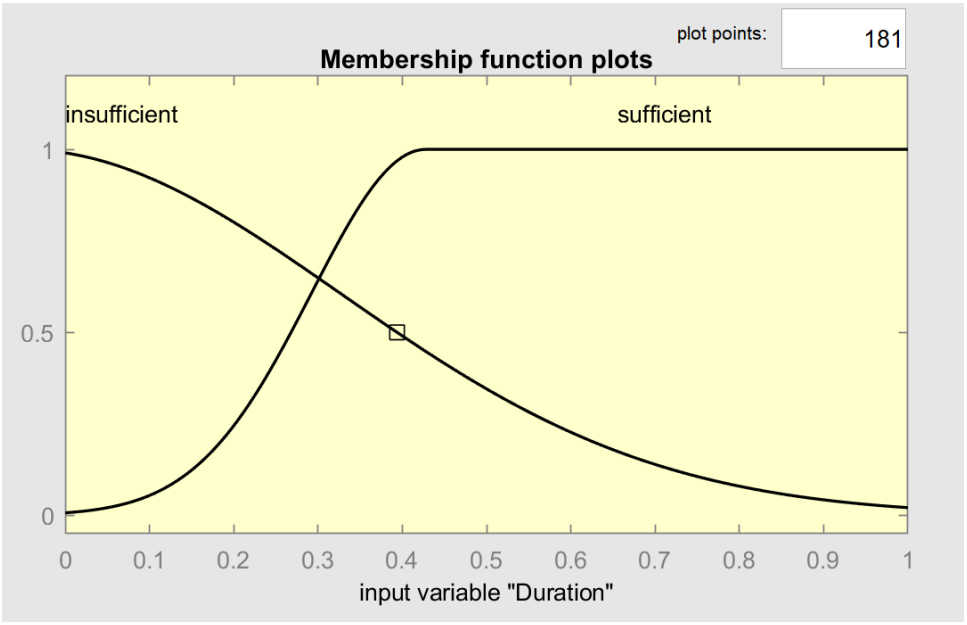


Рисунок 4.6 – Графік функції приналежності вхідної змінної $ЛЗ_5$

$ЛЗ_6$ – протокол (Protocol), її область визначення $X = [0; 1]$, вимірюється за такою шкалою:

$ЛЗ_6 \leq 0{,}5$ – недозволений (unauthorized);

$ЛЗ_6 > 0{,}5$ – дозволений (allowable).

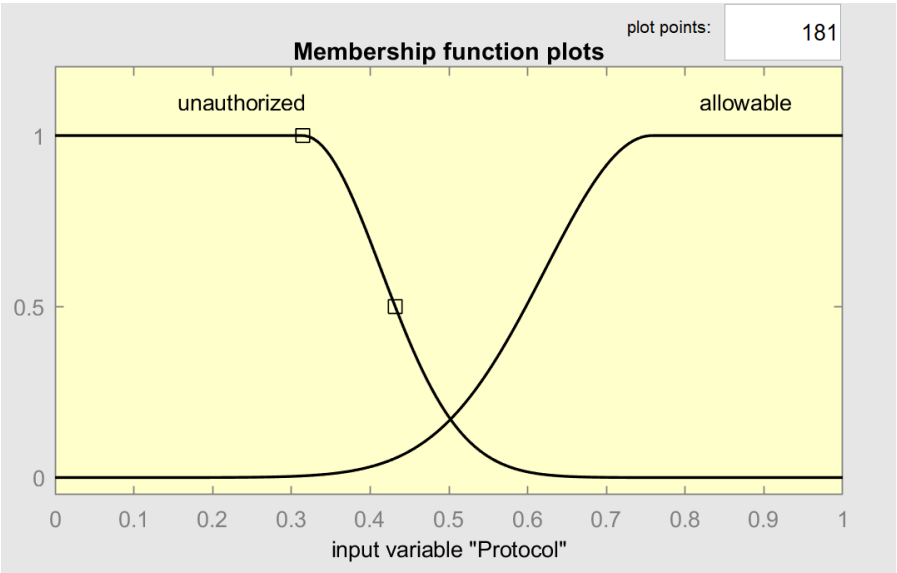


Рисунок 4.7 – Графік функції приналежності вхідної змінної $ЛЗ_6$

$ЛЗ_7$ – частота (frequency), її область визначення $X = [0; 10]$, вимірюється за такою шкалою:

$ЛЗ_7 \leq 2$ – рідко (rarely);

$ЛЗ_7 > 2$ – часто (often).

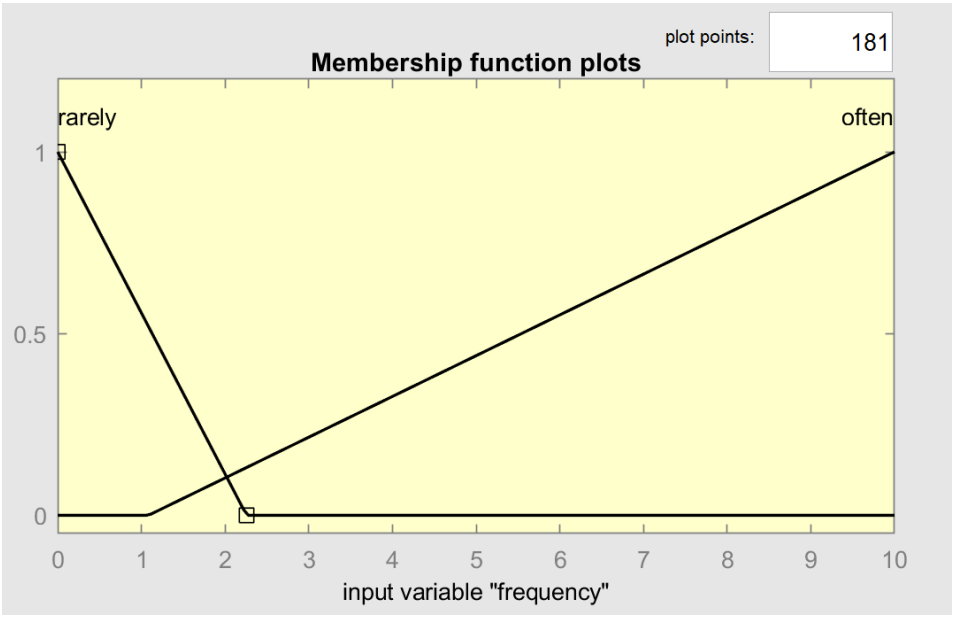


Рисунок 4.8 – Графік функції приналежності вхідної змінної $ЛЗ_7$

Окрім вхідних змінних модель складається з 2 проміжні лінгвістичні змінних.

$ПЛЗ_1$ – Несанкціонований доступ (UnauthorizedAccess), $X = [0; 1]$. Вхідні змінні: $ЛЗ_1$, $ЛЗ_2$, $ЛЗ_3$, $ЛЗ_4$, $ЛЗ_6$

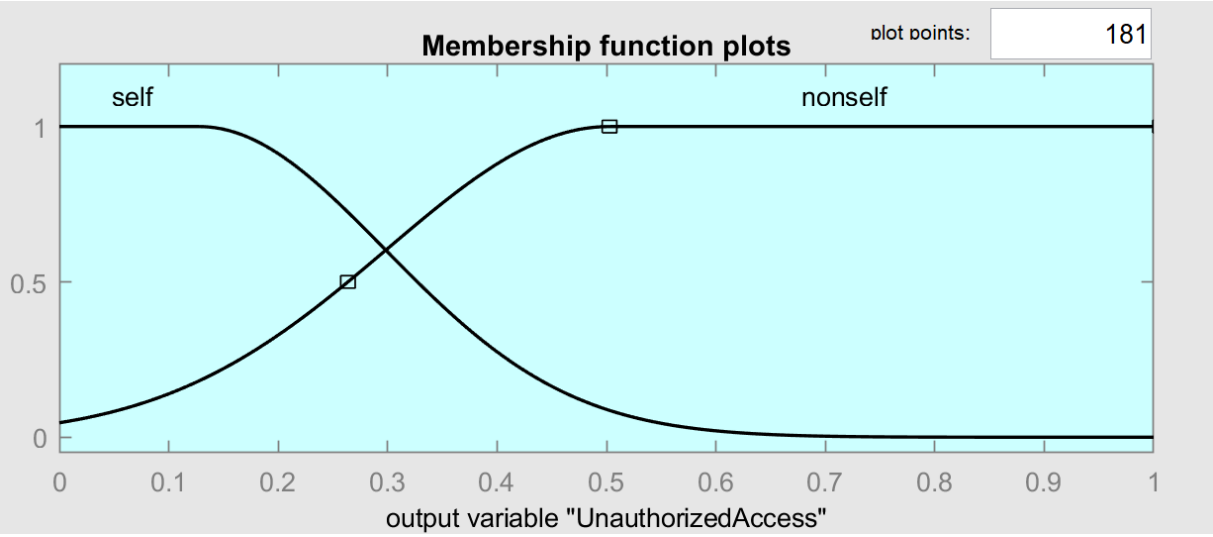


Рисунок 4.9 – Графік функції приналежності проміжної змінної $ПЛЗ_1$

$ПЛЗ_2$ – Аномальна поведінка (Abnormal behavior), $X = [0; 1]$. Вхідні змінні: $ЛЗ_1$, $ЛЗ_2$, $ЛЗ_3$, $ЛЗ_4$, $ЛЗ_5$, $ЛЗ_7$.

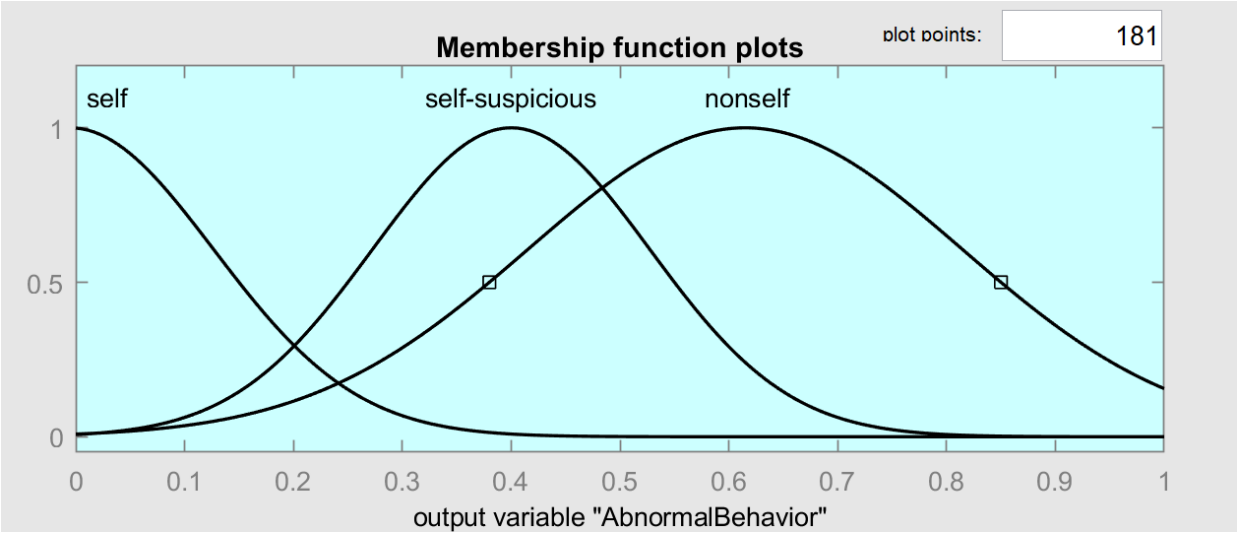


Рисунок 4.10 – Графік функції приналежності проміжної змінної $ПЛЗ_2$

Вихідна лінгвістична змінна – розпізнавання пристрою корпоративної мережі (Device identification), $X = [0; 1]$. Вхідні змінні: $ПЛЗ_1$ – Несанкціонований доступ (Unauthorized access), $ПЛЗ_2$ – Аномальна поведінка (AbnormalBehavior). Базова терм-множина має вигляд: T = { свій; свій, але

підозрілий; чужий}. Графік функції приналежності вихідної змінної наведено на рисунку 4.4.

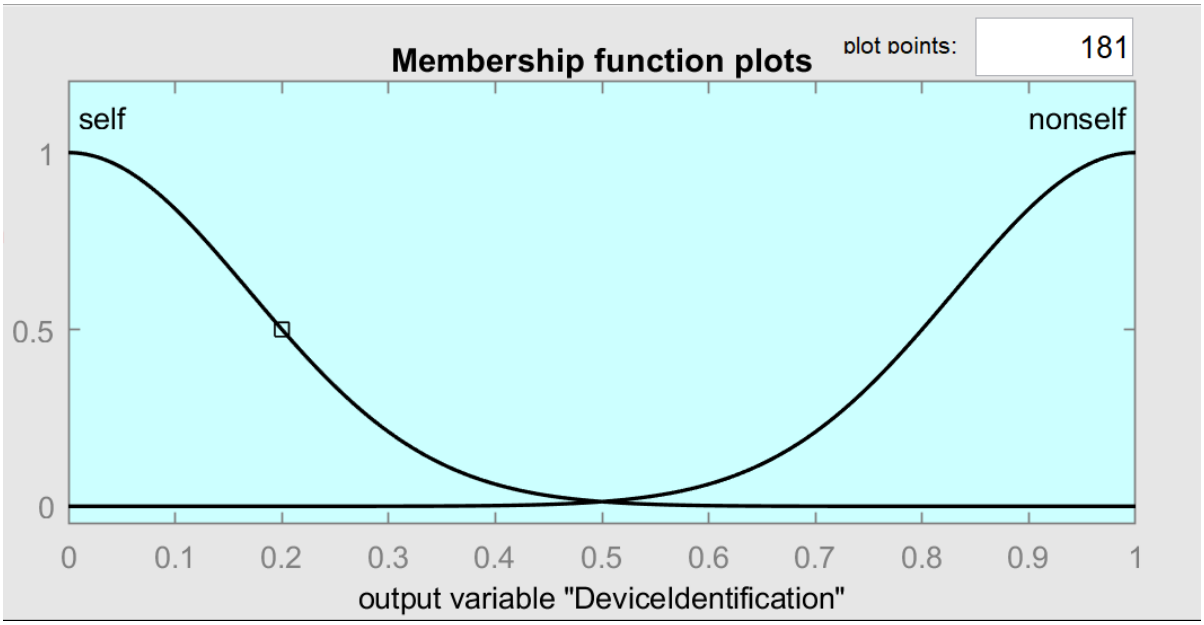


Рисунок 4.11 – Графік функції вихідної змінної *Y*

4.3 Побудова бази правил

Для будь-якого терма вхідної змінної є хоча б одне правило, в якому цей терм використовується в якості передумови.

Загалом, база правил складається з чотирьох блоків правил.

Блок правил для $ПЛЗ_1$ складається з 2 правил. Приклад нечітких правил:

If (SourceIPAddress is unauthorized) or (DestinationIPAddress is unauthorized) or (DestinationPortNo is unauthorized) or (SourcePortNo is unauthorized) or (Protocol is unauthorized) then (UnauthorizedAccess is nonself)

If (SourceIPAddress is allowable) and (DestinationIPAddress is allowable) and (DestinationPortNo is allowable) and (SourcePortNo is allowable) and (Protocol is allowable) then (UnauthorizedAccess is self)

Блок правил для $ПЛЗ_2$ складається з 12 правил. Приклад нечітких правил:

If (SourceIPAddress is PartiallyAllowed) or (DestinationIPAddress is PartiallyAllowed) then (AbnormalBehavior is self-suspicious)

If (SourceIPAddress is allowable) or (Duration is sufficient) then (AbnormalBehavior is self-suspicious)

If (DestinationIPAddress is allowable) or (Duration is sufficient) then (AbnormalBehavior is self-suspicious)

If (SourceIPAddress is allowable) and (DestinationPortNo is PartiallyAllowed) and (Duration is sufficient) then (AbnormalBehavior is self-suspicious)

If (DestinationIPAddress is allowable) and (DestinationPortNo is PartiallyAllowed) and (Duration is sufficient) then (AbnormalBehavior is self-suspicious)

If (DestinationIPAddress is allowable) and (SourcePortNo is PartiallyAllowed) and (Duration is sufficient) then (AbnormalBehavior is self-suspicious)

If (SourceIPAddress is allowable) and (SourcePortNo is PartiallyAllowed) and (Duration is sufficient) then (AbnormalBehavior is self-suspicious)

If (SourceIPAddress is allowable) and (DestinationIPAddress is allowable) and (DestinationPortNo is PartiallyAllowed) and (SourcePortNo is PartiallyAllowed) and (Duration is insufficient) and (frequency is rarely) then (AbnormalBehavior is self)

If (SourceIPAddress is allowable) and (DestinationIPAddress is allowable) and (DestinationPortNo is unauthorized) and (SourcePortNo is unauthorized) and (Duration is sufficient) then (AbnormalBehavior is nonself)

If (SourceIPAddress is allowable) and (DestinationIPAddress is allowable) and (Duration is insufficient) and (frequency is often) then (AbnormalBehavior is self-suspicious)

If (SourceIPAddress is allowable) and (DestinationIPAddress is allowable) and (Duration is sufficient) and (frequency is often) then (AbnormalBehavior is nonself)

If (SourceIPAddress is PartiallyAllowed) and (DestinationIPAddress is PartiallyAllowed) and (Duration is sufficient) and (frequency is often) then (AbnormalBehavior is nonself)

Таким чином, виділено 7 вхідних лінгвістичних змінних, 2 проміжних та 1 вихідна змінна. Побудовано базу правил. При побудові бази правил було використане середовище Fuzzy Logic Toolbox, це дозволило розробити модель на високому рівні та суттєво зекономити час [4,9].

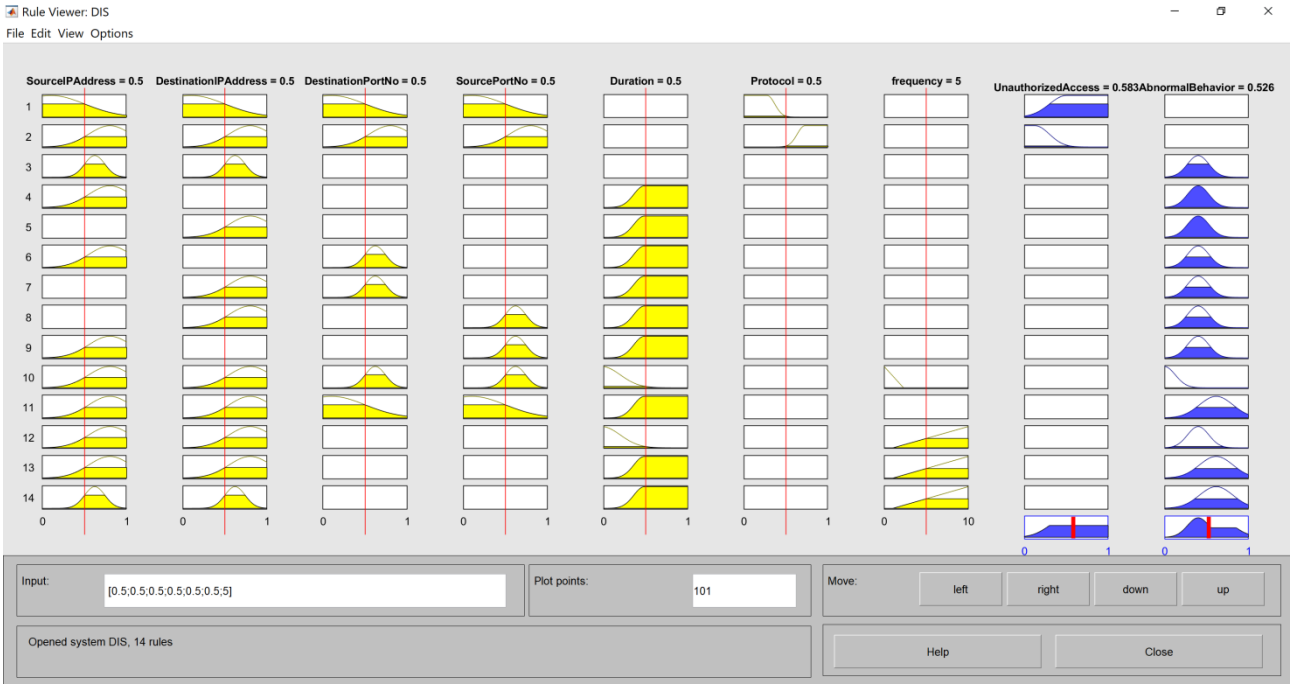


Рисунок 4.12 – Вікно огляду правил

4.4 Висновки до розділу

1. На основі послідовності з 112 бітів, що містить інформацію про IP надсилача, IP отримувача, порт надсилача, порт отримувача, протокол, тривалість доступу, частота розроблено лінгвіститичні змінні.

2. Для розробки нечіткої системи та побудови бази правил було використане середовище Fuzzy Logic Toolbox.

ВИСНОВКИ

В рамках дослідження отримано рішення задачі розробки моделей і алгоритмів розпізнавання кінцевих пристроїв корпоративної мережі за принципом свій / чужий.

1. Проаналізовано ймовірнісний підхід на основі дерева атак до визначення рівня небезпеки корпоративної мережі. З аналізу проглядається найбільша небезпека від інсайдерської атаки, тобто від «свого» пристрою, який здійснює нетипову поведінку і стає «чужим».

2. Запропоновано структуру системи виявлення вторгнень, яка потребує використання генетичних алгоритмів для еволюції шаблонів виявлення та їх запам'ятовування, що дозволить ідентифікувати кінцевий пристрій мережі як свій або чужий.

3. Запропонувати алгоритм процедури виявлення «чужого» пристрою як такого, що здійснює неправомірні дії в мережі, наприклад вторгнення

4. З огляду на централізовану структуру корпоративної мережі остаточне рішення щодо розпізнавання пристрою за принципом свій/чужий та відключенням пристрою від мережі приймає адміністратор мережі, тому запропоновано нечітку модель розпізнавання пристроїв за принципом свій / чужий.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Воробієнко П. П. Телекомунікаційні та інформаційні мережі : Підручник / П. П. Воробієнко, Л. А. Нікітюк, П. І. Резніченко. – К.: САММІТ-Книга, 2010. – 708 с.: іл..

2. Кисіль Т. М. Проблеми розпізнавання кінцевих пристроїв корпоративної мережі за принципом свій/чужий / Т. М. Кисіль, Ю. П. Кльоц, Т, В. Бондаренко, Є. С. Шаховал // Тези доповідей XVI Міжнародної науково-практичної конференції "Військова освіта і наука: сьогодення та майбутнє" Том 1 [Текст] / за заг. редакцією Ігоря Толока. – К. : ВІКНУ, 2020. С. 38-39.

3. Круглов В. В. Нечітка логіка й штучні нейронні мережі. / В. В. Круглов, М. І. Длит, Р. Ю. Голунь. – М.: Физматлит, 2001 – 221с.

4. Леоненков А. В. Нечеткое моделирование в среде MATLAB и fuzzyTECH / А. В. Леоненков. – СПб.: БХВ – Петербург, 2005 – 736 с.

5. Лисенко С. М. Метод виявлення кіберзагроз та ШПЗ для забезпечення живучості комп’ютерних систем в корпоративних мережах на основі самоадаптивності /С.М. Лисенко, Т. М. Кисіль, Ю. О. Нічепорук, А. В. Горошко // Вісник Хмельницького національного університету. Технічні науки. – 2020. – № 4, т. 1. – С. 39-43.

6. Литвиненко В.І. ПОБУДОВА ШТУЧНИХ ІМУННИХ СИСТЕМ // Наукові праці. Комп’ютерні технології. – 2010 р., Випуск 121, Том 134, С.166-178

7. Новак В. К. Математичні принципи нечіткої логіки / В. К. Новак, І. Т. Перфільєва, І. В. Мочкрож. – М.: Физматлит, 2006 – 352 с.

8. Попов А.Ф Комп’ютерні системи і мережі : Навчальний посібник / А.Ф. Попов.– Чернівці, 2010р. - 216с.

9. Патент на корисну модель 118663 Україна, МПК G06F 21/55 Спосіб ідентифікації бот-мереж у корпоративних комп'ютерних мережах на основі аналізу DNS-трафіку / Савенко О. С., Лисенко С. М., Бобровнікова К. Ю., Нічепорук А. О., Савенко Б. О.; власник Хмельницький національний

університет. — № u201612041; заявл. 28.11.2016; опубл. 28.08.2017, Бюл. № 16/2017

10. Штовба С.Д. Проектування нечітких систем в середовищі MATLAB // М.: Гаряча лінія – Телеком, 2007 – 288 с.

11. Al-Enezi J., Abbod M., Alsharhan S.: Artificial immune systems - models, algorithms and applications. Int. J. Res. Rev. Apps. Sci. **3**(2), 118-131 (2010).

12. Chao, Rui and Ying Tan.: A Virus Detection System Based on Artificial Immune System. International Conference on Computational Intelligence and Security 1: 6-10. (2009).

13. Halsall, Fred. 1996. Data Communications, Computer Networks and Open Systems. 4th ed. Electronic Systems Engineering Series. United States: Addison-Wesley Publishing Company. ISBN 0-201-42293-X.

14. Hightower Ron, Stephanie Forrest, Alan S. Perelson. The Baldwin Effect in the Immune System: Learning by Somatic Hypermutation Adaptive Individuals in Evolving Populations: Models and Algorithms, Addison-Wesley Publishing Company, Reading Massachusetts. 1996. P.159–167.

15. Hofmeyr S., Forrest S. Architecture for an Artificial Immune System Evolutionary Computation. 2000. 8 (4). P. 443–473.

16. Idris, I. Model and Algorithm in Artificial Immune System for Spam Detection. Int. J. Artif. Intell. Appl. 2012, 3, 83–94.

17. Lysenko S., Savenko O., Bobrovnikova K., Kryshchuk A., Savenko B.. Information technology for botnets detection based on their behaviour in the corporate area network. Communications in Computer and Information Science. 2017. – Vol. 702. – PP.83-95, ISSN: 1865-0929 (part scientometric SCOPUS).

18. John McHugh, , Alan Christie, and Julia Allen. 2000. “Defending Yourself: The Role of Intrusion Detection Systems.” IEEE Software 17 (5): 42–51 (September/October). ISSN 0740-7459.

19. V. T.Nguyen, N. P. Anh, M. T. Khang, N. H. Ngan, N. Q. Thai, and N. T. Quoc.: A combination of clonal selection algorithm and artificial neural

networks for virus detection, in Advances in Computer Science and its Applications. Springer, 2014, pp. 95–100. (2014).

20. V. T.Nguyen, T. T. Nguyen, M. T. Khang, and T. D. Le.: A combination of negative selection algorithm and artificial immune network for virus detection," in Future Data and Security Engineering. Springer, 2014, pp. 97–106. (2014).

21. Read, M., Andrews, P., Timmis, J.: Artificial immune systems, pp. 4–5 (2012)

22. Schneier, Bruce (December 1999). "Attack Trees". Dr Dobb's Journal, v.24, n.12. Retrieved on 2007-08-16

23. Shah S., H. Jani, S. Shetty, and K. Bhowmick.: Virus detection using artificial neural networks. International Journal of Computer Applications, vol. 84, no. 5. (2013)

24. Sheyner O., Jha S., Wing J. Two Formal Analyses of Attack Graphs. // IEEE Computer Security Foundations Workshop, Cape Brenton, Nova Scotia, Canada. – June 2002. – P. 49–63.

25. Sheyner Oleg, Jeannette Wing. Tools for Generating and Analyzing Attack Graphs. www.cs.cmu.edu/~scenariograph/sheynerwing04.pdf

26. Timmis J. Artificial immune systems: today and tomorrow. Natural Computing, 6(1):1-18, March 2007.

27. Vu Thanh Nguyen , Le Hoang Dung , Tuan Dinh: Le A Combination of Artificial Immune System and Deep Learning for Virus Detection. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 22 (2018) pp. 15622-15628

28. Todd Hughes, Oleg Sheyner: Attack scenario graphs for computer network threat analysis and prediction. Complexity 9(2) (ISSN: 1076-2787): 15-18 (2003)

ДОДАТОК А

(обов’язковий)

Публікації

ЗМІСТ

КОМП'ЮТЕРНІ НАУКИ, ІНФОРМАЦІЙНІ ТЕХНОЛОГІ, СИСТЕМНИЙ АНАЛІЗ ТА КІБЕРБЕЗПЕКА

DOI 10.31891/2307-5732-2020-287-4-39-43
УДК 004.93

С.М. ЛИСЕНКО, Т.М. КИСІЛЬ, Ю.О. НІЧЕПОРУК, А.В. ГОРОШКО
Хмельницький національний університет

МЕТОД ВИЯВЛЕННЯ КІБЕРЗАГРОЗ ТА ШПЗ ДЛЯ ЗАБЕЗПЕЧЕННЯ ЖИВУЧОСТІ КОМП'ЮТЕРНИХ СИСТЕМ В КОРПОРАТИВНИХ МЕРЕЖАХ НА ОСНОВІ САМОАДАПТИВНОСТІ

В роботі представлено метод забезпечення живучості комп'ютерних систем в умовах кіберзагроз на основі самоадаптивності, який дозволяє здійснювати адаптивне реконфігурування компонентів КС шляхом сценаріїв безпеки та забезпечує здатність системи до стійкого її функціонування в ситуації наявності кібератак. Живучість забезпечується адаптивним відновленням мережі. Ця реконструкція проводиться на основі сценарію безпеки, прийнятого на основі аналізу раніше зібраних ознак, притаманних кібератакам. Ознаки атак формуються як вектори ознак і підлягають класифікації. Результатом класифікації є віднесення об'єкту класифікації до відповідного класу, який відповідає певній кібератаці. Метою методу є вибір необхідного сценарію захисту мережевої реконструкції відповідно до кібератак. Експериментальні дослідження свідчать про високу достовірність запропонованого методу, зокрема достовірність виявлення кібератак до 99% та здатності забезпечення живучості КС в ситуації кібератак з рівнем до 70%.

Ключові слова: шкідливе програмне забезпечення, живучість, комп'ютерні системи, достовірність виявлення, кібератака, мережний трафік.

S. LYSENKO, T. KYSIL, Y. NICHEPORUK, A. GOROSHKO
Khmelnytskyi National University

METHOD FOR CYBER THREATS AND MALWARE DETECTION TO ENSURE THE COMPUTER SYSTEMS RESILIENCE OF IN CORPORATE NETWORKS BASED ON SELF-ADAPTIVITY

AThe paper presents a method for cyber threats and malware detection to ensure the computer systems resilience of in corporate networks based on self-adaptivity. The resilience is ensured by the adaptive reconfiguration of the network. Answer the question how the network has to be reconfigured is received by the means of the cluster analysis of the cyberattacks' features, which are observed in the network and network hosts. In order to choose the needed security scenarios, the proposed method uses SVM approach. The objects of classification are the feature vectors, which contain the set of the demonstrations, which may indicate the appearance of cyber threats on the in corporate networks. The purpose of the technique is to choose the network and network hosts' reconfiguration scenarios according to the cyber-attacks, performed by the botnets. The learning stage of the method consists of the following steps: a knowledge formation about the features that may indicate the cyberattacks performed by the botnet; presentation the knowledge about the cyberattacks as the set of feature vectors; a labeled data creation of the feature vectors of the cyberattacks based on knowledge. The monitoring stage of the method consists of the following steps: gathering of the inbound and outbound network traffic; gathering of the information about the hosts' network activity and reports of the hosts' antiviruses; construction of the feature vector, based on the information obtained from the network and hosts; implementation of the semi-supervised fuzzy c-means clustering for the choice of the security scenarios; implementation of the security scenarios for the corporate area network's infrastructure. Usage of the developed system makes it possible to detect known and unknown multi vector cyberattacks performed by the botnets. Experimental results demonstrated that the implemented principals of proposed technique into show the ability to ensure the resilient network functioning in the situation of the cyberattacks by botnets at the rate at about 70%.

Keywords: malware, computer systems, resilience, detection efficiency, network traffic, cyberattack.

Вступ

Сьогодні актуальною проблемою, яка призводить до негативних економічних та соціальних наслідків, є проблема боротьби із кіберзагрозами. З кожним роком результати їх впливу набирають значного масштабу, завдаючи шкоди усім сферам, де застосовуються комп'ютерні системи. Відомі методи та засоби не в змозі в повній мірі забезпечити належний рівень інформаційної безпеки. Одним з напрямків кібербезпеки, що сприяє своєчасному виявленню атак, запобіганню їх наслідків та зменшенню їх впливу є синтез живучих (резильєнтних) комп'ютерних систем – систем здатних продовжувати функціонувати в умовах здійснення кібератак [1, 2]. Тому метою роботи є підвищення достовірності виявлення атак та шкідливого програмного забезпечення (ШПЗ) з метою забезпечення живучості комп'ютерних систем в корпоративних мережах в комп'ютерних системах (КС) шляхом розроблення методу.

Пов'язані роботи

Сьогодні в наукових джерелах широко представлені різні методи виявлення кібератак в КС. Зокрема, в [3] запропоновано метод, що заснований на обробці подій для вирішення проблеми атак. У рамках цього підходу розроблено архітектуру IDS на основі моделі обробки подій (EPM). Це засновані на правилах IDS, в яких правила зберігаються в репозиторії шаблонів правил і приймають SQL і EPL Epser в якості посилання. В [4] представлено протокол аутентифікації атак КС, який використовує легкий метод шифрування, заснований на операції XOR для захисту від підробок і захисту конфіденційності. Для генерації синтезованих компонентів пропонованого полегшеного протоколу шифрування використовується САПР Quartus II. Існуючий механізм безпеки RFID-систем може бути посилений з акцентом на криптографічні протоколи. В [5] запропоновано метод виявлення Sibil атаки. Він представляє собою схему

захисту, включаючи виявлення Sibil атаки на основі соціальних графів (SGSD), а виявлення Sibil атаки на основі класифікації поведінки (BCSD). В [6] запропоновано IDS для виявлення атаки Warmhole атака за допомогою тренажера Cooja. Запропонована система використовує централізовану та розподілену архітектуру для розміщення IDS. В [7] подано технологію виявлення Sinkhole атака INTI (Intrusion Detection of Sinkhole attacks on 6LoWPAN). Інформаційна технологія INTI прагне зменшити негативний вплив атаки на КС, поєднує в собі стратегії спостереження, репутації та довіри для виявлення зловмисників шляхом аналізу поведінки пристроїв. В [8] описано інформаційну технологію радіочастотної ідентифікації в RFID – ключового протоколу шифрування, який забезпечує безпеку зв'язку, який може забезпечити аутентифікацію між міткою і сервером. В [9] описана інформаційна технологія захисту CloudEyes від атак мережного типу, яка надає ефективні та надійні служби безпеки для пристроїв з обмеженими ресурсами. CloudEyes виявляє підозрілу фільтрацію, заснована на структурі зворотних ескізів і забезпечує ретроспективне і точне наведення фрагментів злоякісної сигнатури. В [10] запропонували інформаційну технологію BMIDS (Behavioral Modeling IDS), яка використовує поведінкові шаблони. В [11] представлено нейромережний метод виявлення DDoS/DoS-атак. Виявлення було засноване на класифікації нормальних і небезпечних шаблонів. Модель ANN була перевірена на модельованій мережі Інтернету речей, що демонструє більше 99% точності. В [12] запропоновано інформаційну технологію на основі специфікації для захисту мережевої топології на основі RPL. Основна ідея полягає у вивченні станів, переходів і відповідної статистики на основі аналізу файлу трасування. Результати експериментів показують, що вище вказані методи здатні виявляти атаки, однак не забезпечують живучість КС в умовах здійснення атак.

Метод виявлення кіберзагроз та ШПЗ для забезпечення живучості комп'ютерних систем в корпоративних мережах на основі самоадаптивності

З метою реалізації принципів адаптивності та здатності до еволюції для забезпечення живучості (резильєнтності) КС в умовах кібератак розроблено метод забезпечення живучості (резильєнтності) комп'ютерних систем в умовах кіберзагроз на основі самоадаптивності, який дозволяє здійснювати адаптивне реконфігурування компонентів КС шляхом сценаріїв безпеки та забезпечує здатність системи стійкого її функціонування в ситуації наявності кібератак [13, 14]. Живучість (резильєнтність) забезпечується адаптивною відповіддю мережі на атаку шляхом її реконфігурації, яка здійснюється на основі застосування сценарію безпеки. Висновок щодо необхідного сценарію безпеки здійснюються на основі аналізу раніше зібраних ознак, притаманних кібератакам.

Ознаки атак формуються як вектори ознак і підлягають класифікації. Результатом класифікації є віднесення об'єкту класифікації до відповідного класу, який відповідає певній кібератаці.

Метою методу є вибір необхідного сценарію захисту мережевої реконструкції відповідно до кібератак. Метод включає кроки навчання та виявлення.

1. Навчання складається з наступних етапів:

1.1. формування знань на основі особливостей, які можуть вказувати на кібератаки;

1.2. презентація знань про кібератаки як сукупність функцій векторів;

1.3. позначення отриманих векторів кібератак з метою формування класів, де кожен клас відповідає певній кібератаці, і, в свою чергу, певний сценарій безпеки, який слід застосувати для пом'якшення кібератак.

2. Етап моніторингу складається з наступних етапів:

2.1. збір вхідних та вихідних мережних даних та збір інформації про діяльність хостів мережі та звіти про антивіруси хостів;

2.2. побудова функціональних векторів на основі інформації, отриманої від мережі та хостів.

3. Етап виявлення включає в себе класифікацію множини одержаних векторів ознак на основі застосування методу опорних векторів (SVM) з метою їх віднесення до одного з класів та вибору правильного сценарію безпеки.

4. Етап відновлення включає реалізацію сценарію безпеки інфраструктури корпоративної мережі.

Позначимо набір мережних компонентів, які зазнають атак, як $B = \{b_1, b_2, b_3\}$, де b_1 – хост мережі, b_2 – мережний пристрій, b_3 – сервер у мережі, $b_i \in B$. Потім позначимо набір кібератак, як $\mathbf{A} = \{\mathbf{a}_j\}_{j=1}^{N_A}$. Позначимо набір сценаріїв безпеки як $S = \{s_m\}_{m=1}^{N_S}$, де N_S – кількість сценаріїв безпеки, які слід застосувати залежно від типу атаки. Таким чином, функція вибору сценарію безпеки для відновлення мережі за наявності визначеного типу атаки f може бути представлена як $f: b_i \times a_j \rightarrow s_m$. Усі ознаки є основою набору векторів $\mathbf{X} = \{\mathbf{x}_k\}_{k=1}^{N}$, де кожен з вектора ознак x_k описує кібератаку, N – кількість векторів ознак. Нехай k позначає кількість попередньо визначених класів векторів ознак. Кожен клас відповідає визначеним кібератакам (і сценарій безпеки, який слід застосувати), а один клас відповідає відсутності атаки. Для здійснення класифікації множини одержаних векторів ознак в роботі застосовано метод опорних векторів (SVM) [15]. Для обчислення роздільної гіперплощини без явного проведення відображення в просторі функцій можна використовувати різні функції ядра [16]. Для проведення класифікації використовуються методи на основі «один проти всіх» та «один проти одного» SVM [17]. В роботі для здійснення класифікації векторів ознак було використано ядра: лінійне (1), поліноміальне (2), гауссове (3), експоненційне (4) та B-spline (5):

$$K(x, x_i) = x^T x_i + c, c \in R \quad (1)$$

$$K(x, x_i) = (\alpha x^T x_i + c)^p, \alpha \in R, c \in R, p \in N \quad (2)$$

$$K(x, x_i) = e^{\frac{1}{2\sigma^2 \|x - x_i\|^2}}, \sigma > 0 \quad (3)$$

$$K(x, x_i) = e^{\frac{1}{2\sigma^2 \|x - x_i\|^2}} \quad (4)$$

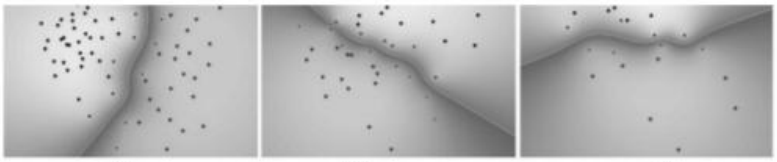
$$K(x, x_i) = B_{2p+1}(x - x_i), where\ p \in N\ with\ B_{i+1} \coloneqq B_i \otimes B_0 \quad (5)$$

З метою виявлення кібератак мережного типу проводиться моніторинг активності мережі, що може свідчити про появу кібератаки. З метою виявлення кібератаки типу хазяїна збирається інформація про мережеву діяльність хостів та звіти про антивіруси хостів. Зібрану інформацію надсилають класифікатору для подальшого аналізу. Далі, зібрані на попередньому етапі, потім аналізуються. Результатом аналізу є висновок про наявність або відсутність атаки та відповідний сценарій безпеки для відновлення мережі. В якості засобу вибору сценарію безпеки використовується SVM. На етапі виявлення об'єктами класифікації є вектори ознак x_k, отримані при аналізі корисного вхідного та вихідного трафіку. Результатом класифікації віднесення об'єкту класифікації до певного, який свідчить про необхідність застосування відповідного сценарію безпеки.

Виходячи з вибору, зробленого на попередньому етапі, слід застосувати сценарій безпеки. Кожен сценарій містить перелік дій з відновлення мережі.

Експериментальні дослідження ефективності роботи методу

Для того, щоб дослідити ефективність роботи методу, було використано різні функції ядра SVM. Приклад результатів класифікації за допомогою експоненційного ядра представлений на рис. 1. Процес класифікації поділяється на кілька ітерацій. У першій ітерації об'єкти класифікації поділяються на два класи: шкідливий і нешкідливий. Потім класифікатори поділяють об'єкти на інші два класи, наприклад: шкідливий трафік та spoofing трафік. Наступні ітерації розділяють шкідливий трафік та інші класи атак тощо, поки всі вони повністю не розділяться. Експериментальні результати різних класифікаторів SVM з'ясовували, що лінійні та поліноміальні ядра показали найгірші результати. Вони характеризувалися більш тривалими термінами виконання та вищими показниками загальної точності класифікації. Нелінійні класи класифікували кращі результати, де B-spline давав кращі результати, ніж інші. Таким чином, для експериментальних зразків оцінювання найефективнішим класифікатором, що використовує SVM, був B-spline, оскільки він забезпечував найбільшу відстань між гіперплощинами, найкоротший час оцінки та найкращу точність класифікації; таким чином, він був використаний як основна функція ядра в методі для прийняття рішення щодо застосування сценарію безпеки в залежності від класифікованої атаки.



a) b) c)

Рис. 1. Результати класифікації SVM за допомогою різних експоненційного ядра:
a) шкідливий трафік / не шкідливий трафік; b) шкідливий трафік / spoofing трафік; c) шкідливий трафік / smurf трафік

Для визначення достовірності запропонованим методом було проведено ряд експериментів. В експериментах було використано локальну мережу з 50 хостів (кожен з операційною системою Microsoft Windows), один виділений сервер (операційна система Linux OpenSusE з nginx HTTP-сервером). Експерименти тривали 24 години. Мережевий трафік захоплювався за допомогою утиліти tcpdump. Під час експериментів було здійснено 150 атак різних типів на хости, сервер та маршрутизатори. Метою було визначити, чи зможе корпоративна мережа функціонувати в ситуації атак (наприклад, якщо сервер, хости або мережний маршрутизатор зможуть надавати послуги з певними допустити характеристиками у визначений час). Як приклад, в даному підрозділі описано детальні результати експериментів із повільними DDoS, smurf та macflooding атаками [18–20]. Рис. 2 демонструє рівень мережного трафіку та часу відповіді сервера перед атакою, під час атаки та після застосування сценарію безпеки. Таким чином, видно, що під час атаки рівень трафіку залишається майже незмінним (рис. 2а), але час реакції сервера збільшувався, що спричиняло недоступність послуги (рис. 2б).

Застосування сценарію безпеки, отриманого за допомогою методу забезпечення живучості (резильєнтності) комп'ютерних систем в умовах кіберзагроз на основі самоадаптивності виявило незначно помітні зміни рівні трафіку, в той час як час відповіді сервера зменшився і сервер зміг надавати послуги. Під час smurf атаки значно сильно збільшувались рівень трафіку та час відповіді сервера. Застосування сценарію безпеки виявила значні зменшення рівня трафіку (рис. 2а), в той час як час відповіді сервера зменшився до нормального рівня і сервер також міг функціонувати (рис. 2б).

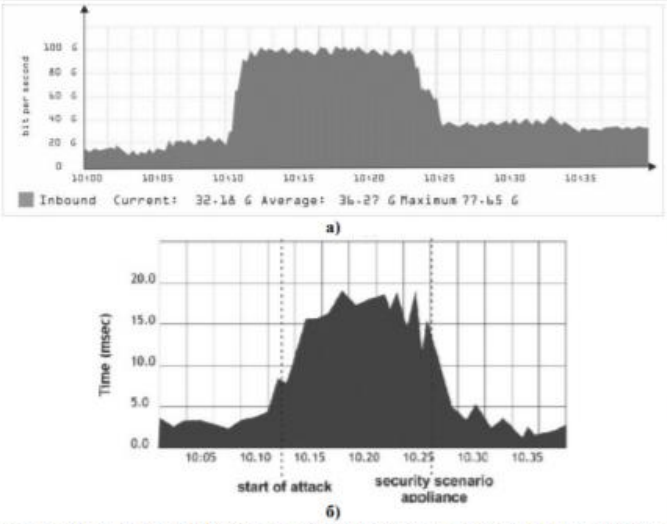


Рис. 2. Рівень трафіку (а) та часи відповіді сервера (б) до, під час та після smurf атаки

Результати показали на здатність забезпечення живучості КС в ситуації кібератак з рівнем до 70%, і продемонстрували, що метод досягає найкращих результатів для виявлення таких атак, як DDoS, ping-flooding, smurf, TCP SYN Flood, ping sweep, phishing тощо. У той же час, достовірність методу щодо Ampliation DNS, скидання TCP, RUDY, зашифровані SSL DDoS, XSS та DNS-атаки атаки нижча. Це пояснюється тим, що поведінка деяких атак дуже схожа на дії користувачів, а деякі функції атак не враховувались у процесі виявлення. Експериментальні дослідження продемонстрували високу достовірність виявлення атак запропонованим методом до 99%.

Висновки

У роботі представлено метод забезпечення живучості комп'ютерних систем в умовах кіберзагроз на основі самоадаптивності, який, дозволяє здійснювати адаптивне реконфігурування компонентів КС шляхом сценаріїв безпеки та забезпечує здатність системи до стійкого її функціонування в ситуації наявності кібератак. Живучість забезпечується адаптивним відновленням мережі. Ця реконструкція проводиться на основі сценарію безпеки, прийнятого на основі аналізу раніше зібраних ознак, притаманних кібератакам. Ознаки атак формуються як вектори ознак і підлягають класифікації. Результатом класифікації є віднесення об'єкту класифікації до відповідного класу, який відповідає певній кібератаці. Метою методу є вибір необхідного сценарію захисту мережевої реконструкції відповідно до кібератак. Експериментальні дослідження свідчать про високу достовірність запропонованого методу, зокрема достовірність виявлення кібератак до 99% та здатності забезпечення живучості КС в ситуації кібератак з рівнем до 70%.

Література

1. McAfee Mobile Threat Report Q1, 2020. URL: https://www.mcafee.com/content/dam/cons umer/en-us/docs/2020-Mobile-Threat-Report.pdf. – 9.12.2019p. (date of access: 10.07.2020).

2. 2020 State of Malware Report. URL: https://resources.malwarebytes.com/files/2020/0 2/2020_State-of-Malware-Report.pdf (date of access: 10.07.2020).

3. Jun C., Chi C. Design of complex event-processing IDS in internet of things. In Sixth International Conference on Measuring Technology and Mechatronics Automation (ICMTMA) (January 2014). 2014. P. 226–229.

4. Lee P. A., Clark L., Bushnell R., Poovendran A passivity framework for modeling and mitigating wormhole attacks on networked control systems, IEEE Trans. Autom. Control. 2014. Vol. 59. No. 12. Pp. 3224–3237.

5. Zhang J., Blum R.S., Lu X., Conus D. Asymptotically optimum distributed es- timation in the presence of attacks, IEEE Trans. Signal Process. 2015. Vol. 63. No. 5. P. 1086–1101.

6. Pongle P., Chavan G. Real time intrusion and wormhole attack detection in internet of things. International Journal of Computers and Applications. 2015. Vol. 121. No. 9.

7. Cervantes C., Poplade D., Nogueira M., Santos A. Detection of sinkhole attacks for supporting secure routing on 6lowpan for internet of things. In IFIP/IEEE International Symposium on Integrated Network Management (IM)(May, 2015). 2015. P. 606–611.

8. An R., Feng H., Liu Q., Li L. Three elliptic curve cryptography-based RFID authentication protocols for Internet of Things. In International Conference on Broadband and Wireless Computing, Communication and Applications. Springer International Publishing (November 2016). 2016. P. 857–878.

9. Sun H., Wang X., Buyya R., Su J. CloudEyes: Cloud - based malware detection with reversible sketch for resource - constrained internet of things (IoT) devices. Software, Practice & Experience. 2017. Vol. 47. No. 3. P. 421–441. doi:10.1002/spe.2420

10. Arrington B., Barnett L., Rufus R., Esterline A. Behavioral Modeling Intrusion Detection System (BMIDS) Using Internet of Things (IoT) Behavior-Based Anomaly Detection via Immunity-Inspired Algorithms. In 25th International Conference on Computer Communication and Networks (ICCCN) (August 2016). 2016. P. 1–6.

11. Hodo E., Bellekens X., Hamilton A., Dubouilh P.L., Iorkyase E., Tachtatzis C., Atkinson R. Threat analysis of iot networks using artificial neural network intrusion detection system. In International Symposium on

Networks, Computers and Communications (ISNCC)(May 2016). 2016. P. 1–6.

12. Le A., Loo J., Chai K. K., Aiash M.A. Specification-Based IDS for Detecting Attacks on RPL- Based Network Topology. Information. 2016. Vol. 7. No. 2. p. 25. doi:10.3390/info7020025

13. Lysenko S., Bobrovnikova K., Savenko O., Kryshchuk A. BotGRABBER: SVM-Based Self-Adaptive System for the Network Resilience Against the Botnets' Cyberattacks. Communications in Computer and Information Science, ISSN: 1865-0929. 2019. P. 127–143.

14. Lysenko S., Savenko O., Bobrovnikova K., Kryshchuk A. Self-adaptive system for the corporate area network resilience in the presence of botnet cyberattacks. Communications in Computer and Information Science, ISSN: 1865-0929. 2018. P. 385–401.

15. Weston J., Mukherjee S., Chapelle O., Pontil M., Poggio T. Vapnik Feature selection for SVMs. In: Advances in neural information processing systems. 2001. P. 668–674.

16. Hofmann T., Scholkopf B., Smola A. J. Kernel methods in machine learning. The annals of statistics. 2008. P. 1171–1220.

17. Foody G.M., Mathur A. A relative evaluation of multiclass image classification by support vector machines. IEEE Transactions on geoscience and remote sensing. 2004. Vol. 42. No. 6. P. 1335–1343.

18. Sergii Lysenko, Pomorova Oksana, Savenko Oleg, Kryshchuk Andrii, Bobrovnikova Kira. DNS-based Anti-evasion Technique for Botnets Detection. The IEEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications: Proceedings (Warsaw, Poland, September 24-26, 2015). Warsaw, 2015. Vol. 1. P. 453–458.

19. Лисенко С.М. Методи виявлення бот-мереж в комп'ютерних системах / С.М. Лисенко, К.Ю. Бобровнікова, В.С. Харченко // Сучасні інформаційні системи. – 2019. – Т. 3. № 4. – С. 87–95.

20. Canadian Institute for Cybersecurity. Botnet dataset. URL: https://www.unb.ca/cic/datasets/botnet.html (date of access: 10.07.2020).

References

1. McAfee Mobile Threat Report Q1, 2020. URL: https://www.mcafee.com/content/dam/cons umer/en-us/docs/2020-Mobile-Threat-Report.pdf. – 9.12.2019p. (date of access: 10.07.2020).

2. 2020 State of Malware Report. URL: https://resources.malwarebytes.com/files/2020/0 2/2020_State-of-Malware-Report.pdf (date of access: 10.07.2020).

3. Jun C., Chi C. Design of complex event-processing IDS in internet of things. In Sixth International Conference on Measuring Technology and Mechatronics Automation (ICMTMA) (January 2014). 2014. P. 226–229.

4. Lee P. A., Clark L., Bushnell R., Poovendran A passivity framework for modeling and mitigating wormhole attacks on networked control systems, IEEE Trans. Autom. Control. 2014. Vol. 59. No. 12. Pp. 3224–3237.

5. Zhang J., Blum R.S., Lu X., Conus D. Asymptotically optimum distributed es- timation in the presence of attacks, IEEE Trans. Signal Process. 2015. Vol. 63. No. 5. P. 1086–1101.

6. Pongle P., Chavan G. Real time intrusion and wormhole attack detection in internet of things. International Journal of Computers and Applications. 2015. Vol. 121. No. 9.

7. Cervantes C., Poplade D., Nogueira M., Santos A. Detection of sinkhole attacks for supporting secure routing on 6lowpan for internet of things. In IFIP/IEEE International Symposium on Integrated Network Management (IM)(May, 2015). 2015. P. 606–611.

8. An R., Feng H., Liu Q., Li L. Three elliptic curve cryptography-based RFID authentication protocols for Internet of Things. In International Conference on Broadband and Wireless Computing, Communication and Applications. Springer International Publishing (November 2016). 2016. P. 857–878.

9. Sun H., Wang X., Buyya R., Su J. CloudEyes: Cloud ‐ based malware detection with reversible sketch for resource ‐ constrained internet of things (IoT) devices. Software, Practice & Experience. 2017. Vol. 47. No. 3. P. 421–441. doi:10.1002/spe.2420

10. Arrington B., Barnett L., Rufus R., Esterline A. Behavioral Modeling Intrusion Detection System (BMIDS) Using Internet of Things (IoT) Behavior-Based Anomaly Detection via Immunity-Inspired Algorithms. In 25th International Conference on Computer Communication and Networks (ICCCN) (August 2016). 2016. P. 1–6.

11. Hodo E., Bellekens X., Hamilton A., Dubouilh P.L., Iorkyase E., Tachtatzis C., Atkinson R. Threat analysis of iot networks using artificial neural network intrusion detection system. In International Symposium on Networks, Computers and Communications (ISNCC)(May 2016). 2016. P. 1–6.

12. Le A., Loo J., Chai K. K., Aiash M.A. Specification-Based IDS for Detecting Attacks on RPL- Based Network Topology. Information. 2016. Vol. 7. No. 2. p. 25. doi:10.3390/info7020025

13. Lysenko S., Bobrovnikova K., Savenko O., Kryshchuk A. BotGRABBER: SVM-Based Self-Adaptive System for the Network Resilience Against the Botnets' Cyberattacks. Communications in Computer and Information Science, ISSN: 1865-0929. 2019. P. 127–143.

14. Lysenko S., Savenko O., Bobrovnikova K., Kryshchuk A. Self-adaptive system for the corporate area network resilience in the presence of botnet cyberattacks. Communications in Computer and Information Science, ISSN: 1865-0929. 2018. P. 385–401.

15. Weston J., Mukherjee S., Chapelle O., Pontil M., Poggio T. Vapnik Feature selection for SVMs. In: Advances in neural information processing systems. 2001. P. 668–674.

16. Hofmann T., Scholkopf B., Smola A. J. Kernel methods in machine learning. The annals of statistics. 2008. P. 1171–1220.

17. Foody G.M., Mathur A. A relative evaluation of multiclass image classification by support vector machines. IEEE Transactions on geoscience and remote sensing. 2004. Vol. 42. No. 6. P. 1335–1343.

18. Sergii Lysenko, Pomorova Oksana, Savenko Oleg, Kryshchuk Andrii, Bobrovnikova Kira. DNS-based Anti-evasion Technique for Botnets Detection. The IEEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications: Proceedings (Warsaw, Poland, September 24-26, 2015). Warsaw, 2015. Vol. 1. P. 453–458.

19. Lysenko S.M. Metody vyiavlennia bot-merezh v kompiuternykh systemakh / S.M. Lysenko, K.Iu. Bobrovnikova, V.S. Kharchenko // Suchasni informatsiini systemy. – 2019. – T. 3. № 4. – S. 87–95.

20. Canadian Institute for Cybersecurity. Botnet dataset. URL: https://www.unb.ca/cic/datasets/botnet.html (date of access: 10.07.2020).

Рецензія/Peer review : 17.10.2020 р. Надрукована/Printed :04.11.2020 р.

УДК 004.056.5

С.М. ЛИСЕНКО, Т.М. КИСІЛЬ, Р.В. ЩУКА
Хмельницький національний університет

ПРОЕКТУВАННЯ ТА РОЗРОБЛЕННЯ ІНТЕЛЕКТУАЛЬНОГО АГЕНТА ВИЯВЛЕННЯ КІБЕРЗАГРОЗ ТА ШПЗ В КОРПОРАТИВНИХ МЕРЕЖАХ

В роботі представлено інтелектуальний агент виявлення кіберзагроз та ШПЗ в корпоративних мережах, який представляє програмну систему із можливістю виявлення відомих та невідомих кібератак, ШПЗ мережного та хостового типу, а також здатністю продукувати множину сценаріїв безпеки для забезпечення резильєнтності КС в умовах кіберзагроз. Резильєнтність мережі та хостів забезпечується їх динамічною адаптивною реконфігурацією та множиною заходів, що дозволяють фунціонувати системам в умовах атак. Інтелектуальний агент виявлення кіберзагроз та ШПЗ BotGRABBER - це мультивекторна система захисту, оскільки вона поєднує аналіз як в мережі, так і в активності хостів. Комбінована інформація дозволяє не тільки виявляти кібератаки різного типу, але й автоматично застосовувати необхідний сценарій безпеки мережної реконфігурації та адаптації КС відповідно до типу виявленої кібератаки. Інтелектуальний агент забезпечує: можливість виявлення відомих та невідомих кібератак, можливість виявлення ботнетів, які використовують методи ухилення від DNS (циклічне відображення IP-адреси, "домен flux", "швидкий flux" та DNS-тунелювання), здатність самостійно застосовувати сценарії безпеки для пом'якшення кібератак, забезпечення резильєнтності корпоративних мереж в умовах кібератак, забезпечення мультивекторного захисту корпоративних мереж.

Ключові слова: шкідливе програмне забезпечення, інтелектуальний агент, кіберзагроа, кібератака, комп'ютерна мережа, сценарій безпеки

S. LYSENKO, T. KYSIL', R. SHCHUKA
Khmelnytskyi National University, Khmelnytskyi, Ukraine

DESIGN AND DEVELOPMENT OF AN INTELLECTUAL AGENT FOR DETECTION OF CYBER THREATS AND MALWARE IN CORPORATE NETWORKS

Abstract – The purpose of this paper is to develop an intellectual agent for detection of cyber threats and malware in corporate networks – BotGRABBER. It provides a novel botnet detection framework with the key features given below: ability to detect the most known botnets' cyberattacks; ability to detect the botnets that use the evasion techniques (cycling of IP mapping, "domain flux", "fast flux" and DNS-tunneling); ability to self-adaptive appliance of the security scenarios for the cyberattacks mitigation, performed by botnets; assuring the corporate area networks' resilience in the presence of botnets' cyberattacks; assurance of the multi vector protection for corporate area networks.

The main components of the intellectual agent are:

1. Knowledge base. Knowledge base provides the information storage concerning to the cyberattacks performed by a botnet in the network and in the hosts. Here, each cyberattack is presented as the feature vector, which consists of functional botnets' features.

To increase the efficiency of the botnet detection each stage of possible botnet's life cycle functioning (infection; initial registration or connection to C&C server; performance of the malicious activity; maintenance; its functioning termination) is presented by own feature vector.

2. Knowledge acquisition unit. Taking into account the increasing of the new ways to perform the cyberattacks proposed tool is provided by ability to update the knowledge about new botnets.

3. Network monitoring unit. This unit implements the network monitoring via gathering of the inbound and outbound network traffic. Collected information is converted into the feature vectors, and is sent to the SVM-based inference engine for further data processing.

4. Host monitoring unit. This unit implements the gathering the information about the hosts' network activity and reports of the hosts' antiviruses. It also converts the collected information into the feature vectors, and sends it to the SVM-based inference engine for further data processing.

5. SVM-based inference engine. This component provides an ability to classify the feature vectors obtained from the network. The main task of the SVM-based inference engine is to range obtained feature vector in a class, which will indicate whether it is cyberattacks, performed by botnet. If the attack is observed, the security scenario according to detected attack in order to mitigate it is to be applied.

6. Network reconfiguration unit. This unit applies produced by the SVM-based inference engine the security scenario for the CAN's infrastructure.

Keywords: malware, intellectual agent, cyberattacks, cyberthreats, computer network, security scenario.

Вступ

На сьогоднішній день з стрімким поширенням комп'ютерних систем та інформаційних технологій, а також їхньої інтеграції у глобальну мережу Internet, кібератаки та шкідливе програмне забезпечення (ШПЗ) є одним із основних видів кіберзлочинності. Збитки, заподіяні ними при інфікуванні хостів мережі, можуть бути від незначного збільшення вихідного трафіку до повного порушення працездатності мережі або втрати критично важливих даних. Причиною цього є те, що комп'ютерні системи не завжди характеризуються резильєнтністю – здатністю передбачати, протистояти, відновлюватись та пристосовуватися до атак. Відомі рішення цілісного підходу до забезпечення резильєнтного функціонування КС в умовах здійснення нових і невідомих атак [1-2]. Тому актуальною задачею є розроблення нових підходів до виявлення нових виді кіберзагроз на основі інтелектуального аналізу даних.

Пов'язані роботи

В [3] запропоновано виявлення на основі аналізу властивості асимптотичного рівнорозподіленості (АЕР) для програмного семантичного аналізу для вилучення семантично відповідних шляхів, забезпечуючи можливість семантично розуміти послідовності системних викликів. UNVEIL - інформаційна технологія, побудована на базі пісочниці для виявлення ШПЗ типу ransomware. За допомогою моделей поведінки можна виявити підозрілі дії файлової системи [4]. В [5] запропоновано метод статичного аналізу для автоматичного виявлення поведінки динамічного завантаження коду. Евристики реалізовані для пошуку викликів методів, пов'язаних із відповідними методами. MALT - це інформаційна технологія відлагодження, яка використовує режим управління системою для прозорого аналізу шкідливих програм, і здатна аналізувати та виявляти руткіти на основі гіпервізора та ядра операційної системи (ОС) [6]. TriggerScope – інформаційна технологія, заснована на методах аналізу програм для виявлення зловмисної логіки програми, яка виконується або спрацьовує за наявності механізмів логічної бомби [7]. Targetdroid - інформаційна технологія, яка може виявити цільове ШПЗ та викликати зловмисну поведінку. Стохастична модель, що викликає поведінку, розроблена на основі ланцюгів Маркова для вираження потоку керування [8]. В [9] запропоновано інформаційну технологію аналізу ШПЗ, яка використовує методику перевірки моделі для виявлення поведінки на високому рівні, наприклад, поведінки потоку інформації. Формальна поведінка визначається як нескінченна підмножина послідовності викликів бібліотек, а набір моделей поведінки з семантичним розумінням виражається за допомогою формул лінійної часової логіки (FOLTL) першого порядку. AppContext [10] - інформаційна технологія аналізу ПЗ, яка диференціює шкідливі та доброякісні форми поведінки ПЗ. AppContext отримує контексти поведінки, залежної від безпеки, та проводить статичний аналіз для визначення поведінки, залежної від безпеки.

Архітектура інтелектуального агента виявлення кіберзагроз та шпз в корпоративних мережах

Інтелектуальний агент виявлення кіберзагроз та ШПЗ в корпоративних мережах, який має назву BotGRABBER, представляє програмну систему із можливістю виявлення відомих та невідомих кібератак, ШПЗ мережного та хостового типу, а також здатністю продукувати множину сценаріїв безпеки для забезпечення резильєнтності КС в умовах кіберзагроз. Резильєнтність мережі та хостів забезпечується їх динамічною адаптивною реконфігурацією та множиною заходів, що дозволяють фунціонувати системам в умовах атак. Інтелектуальний агент виявлення кіберзагроз та ШПЗ BotGRABBER - це мультивекторна система захисту, оскільки вона поєднує аналіз як в мережі, так і в активності хостів. Комбінована

інформація дозволяє не тільки виявляти кібератаки різного типу, але й автоматично застосовувати необхідний сценарій безпеки мережної реконфігурації та адаптації КС відповідно до типу виявленої кібератаки.

Інтелектуальний агент забезпечує:

a. можливість виявлення відомих та невідомих кібератак;

b. можливість виявлення ботнетів, які використовують методи ухилення від DNS (циклічне відображення IP-адреси, “домен flux”, “швидкий flux” та DNS-тунелювання);

c. здатність самостійно застосовувати сценарії безпеки для пом'якшення кібератак;

d. забезпечення резильєнтності корпоративних мереж в умовах кібератак;

e. забезпечення мультивекторного захисту корпоративних мереж.

Компоненти архітектури інтелектуального агента представлено на рис.1.

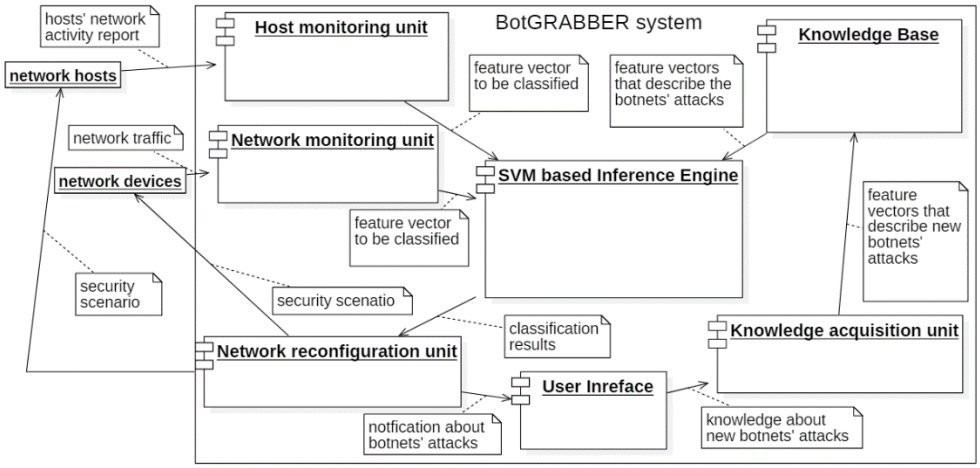


Рис.1. Архітектура інтелектуального агента BotGRABBER

Модуль бази знань. База знань забезпечує зберігання інформації про кібератаки та шкідливе програмне забезпечення, які виконуються в мережі та на хостах. Тут кожна кібератака та ШПЗ представлені вектором ознак та дій, з яких вони складаються. Для підвищення достовіності виявлення загроз кожен етап функціонування можливого життєвого циклу атаки чи ШПЗ (зараження, початкова реєстрація або підключення до сервера C&C, виконання шкідливого функціонування, обслуговування, припинення його функціонування тощо) представлений за допомогою власного вектора ознак та дій. Перелік атак, які аналізуються системою BotGRABBER: DDoS; ping flooding; smurf attack; TCP SYN Flood; Fragmented UDP Flood; DNS Amplification; TCP Reset; ICMP Flood; RUDY; SIP INVITE Flood ; Encrypted SSL DDoS; ping sweep attack; SQL /PHP injection; Cross-Site Scripting (XSS); DNS spoofing; TCP scan/UDP scan; Phishing; Port Binding; Connect-Back; Connection Availability Abuse; Legitimate Platform Abuse; Protocol/Port Listening; Custom DNS Lookup Use; Port Reuse; Common Service Protocol/File Header Abuse.

Модуль оновлення знань. Враховуючи все більшу кількість нових способів виконання кібератак, пропонований інструмент має можливість оновити знання про нові атаки.

Модуль моніторингу мережі. Цей пристрій реалізує мережний моніторинг шляхом збору вхідної та вихідної мережі. Зібрана інформація перетворюється у функціональні вектори та надсилається до модуля висновку для подальшої обробки даних.

Модуль моніторингу хостів. Цей блок реалізує збір інформації про діяльність мережі хостів та звіти про антивіруси хостів. Він також перетворює зібрану інформацію у функціональні вектори та надсилає її до системи висновку на основі SVM для подальшої обробки даних.

Модуль здійснення висновку. Основне завдання двигуна висновку на основі SVM - присвоїти векторний елемент xi, отриманий з мережі, класу в, де $x_i \in X$, $a_t \in A$, $A = \{a_t\}_{k=1}^{N_A}$ N_A- це кількість класів , де

кожному класу відповідає один заданий тип атак, виконуваних ботнетом. Двигун висновку на основі SVM робить висновок про наявність або відсутність кібератак та виявляє можливий тип атаки. Залежно від виявленого типу атаки a_t на сценарій безпеки s_q слід застосувати для відновлення мережі, $S = \{s_q\}_{q=1}^{N_S}$, де S - сукупність усіх сценаріїв безпеки, N_S - кількість захищених сценаріїв. Таким чином, функція f, вибираючи сценарій безпеки для відновлення мережі, визначається як: f: $d_u \times a_t \rightarrow s_m$, де $d_u \in D$, 1, $D = \{d_u\}_{u=1}^{N_D}$ де d_u - мережний компонент, атакований ботнетом , N_D - це кількість мережних компонентів.

Модуль відновлення. Якщо спостерігається атака, то слід застосувати сценарій безпеки, спродукований модулем здійснення висновку, щоб пом'якшити наслідки атаки. Цей модуль застосовує сценарій безпеки. Метою сценарію безпеки є відновлення мережевої інфраструктури залежно від типу атаки.

Інтерфейсі вікна програмної реалізації інтелектуального агента виявлення кіберзагроз та шпз в корпоративних мережах представлено на рисунках 2-5.

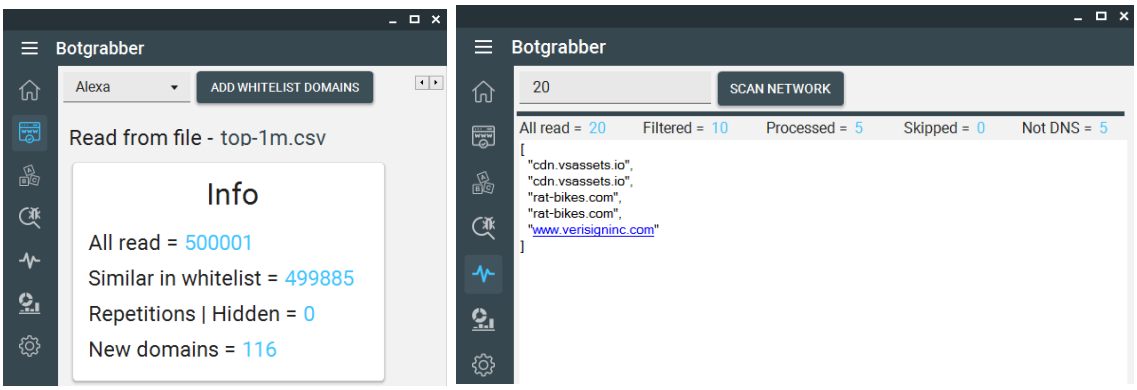


Рис.2. Наповнення бази даних біли списків доменних імен та сканування мережі

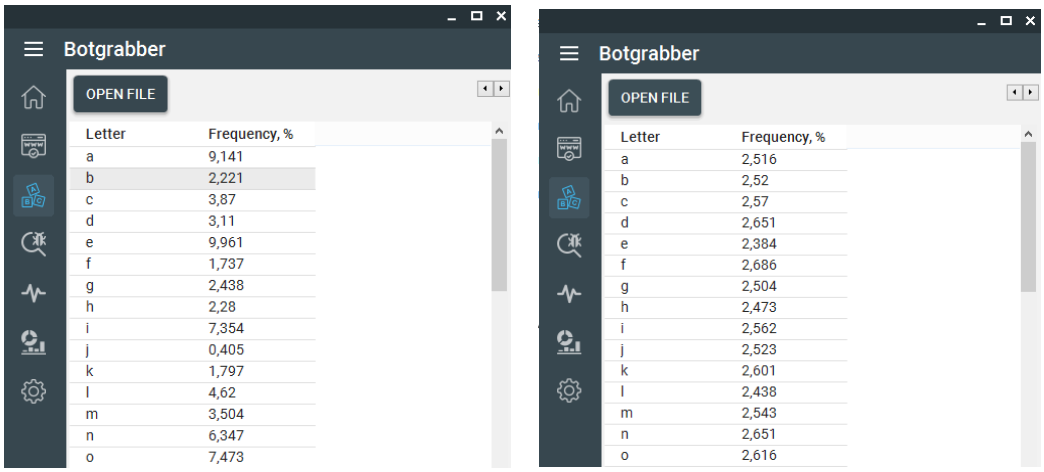


Рис.3. Частотний лексичний аналіз відомих доменних імен та частотний лексичний аналіз доменних імен сформованих алгоритмічно

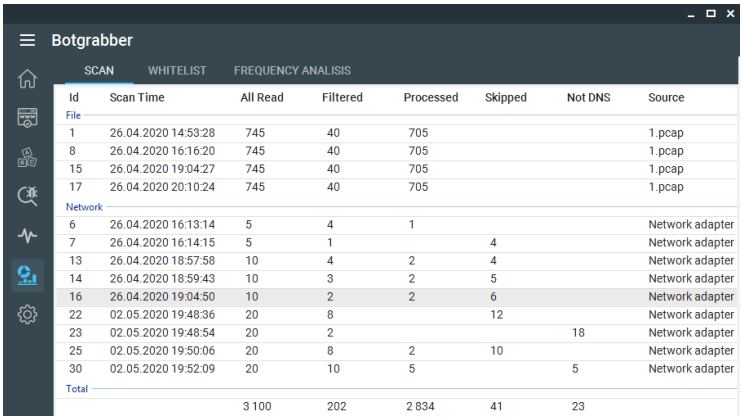


Рис.4. Статистика сканувань мережі та файлів

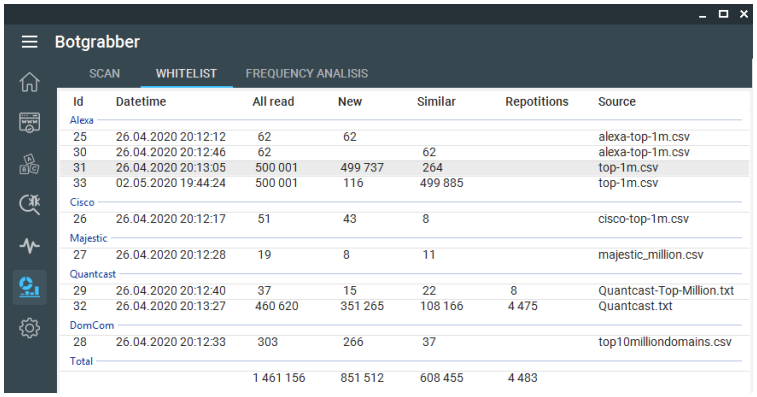


Рис. 5. Статистика наповнення бази даних білих списків доменних імен

Експерименти

Для того, щоб оцінити ефективність застосування інтелектального агента BotGRABBER, як реалізації інформаційної технології забезпечення резильєнтності КС в умовах кіберзагроз були проведені дослідження з використанням реального мережного трафіку. Для цього було використано набір даних [11], який поєднує загальність, реалістичність та репрезентативність. Набір даних містить як шкідливі (наприклад, сліди Storm, Zeus Neris, Rbot, Virut, NSIS, Menti, Sogou, і Murlo), так і нешкідливі набори (ігрові пакети, HTTP tracc і P2P програми, такі як bittorrent). Крім того, він містить сформований реальний трафік, який імітує поведінку користувачів (наприклад, SSH, HTTP та SMTP). Набір даних поділяється на навчальні набори T та оцінювання (тест) E, які включають ботнетів, які виконують атаки. Набір даних включає 19755 зразків, 49,56% з яких шкідливі, а нагадування містить нормальні потоки. Тестовий набір даних включає 18917 зразків, 55,77% з яких представляють шкідливі потоки.

Для проведення експериментів, університетська локальна мережа з 50 хостів (з операційною системою Microsoft Windows), один виділений сервер (операційна система Linux OpenSusE з nginx HTTP-сервером) та мережеві пристрої (MikroTik CCR1009-8G-1S-1S + Були застосовані маршрутизатори ПК). Мережевий трафік захоплювався за допомогою утиліти tcpdump. Усі експерименти були організовані в режимі реального часу та реальних мереж і тривали від декількох секунд (наприклад, фішинг, скидання TCP, ін'єкція SQL / PHP, XSS) до однієї години (наприклад, DDoS, ping-flooding, RUDY, фрагментований UDP Flood, TCP SYN Flood тощо) залежно від типу атаки.

Для оцінювання загальної достовірність виявлення кібератак різного типу системою BotGRABBER було використано метрики, що використовуються для оцінювання якості класифікації в теорії машинного навчання [1**Ошибка! Источник ссылки не найден.**]: чутливість, True Positive Rate (TPR) – відсоток зловмисних поведінок в КС, що класифіковані як зловмисні, $TPR = \frac{TP}{(TP+FN)}$; специфічність, True Negative Rate (TNR) – відсоток незловмисних поведінок в КС, що класифіковані як незловмисні, $TNR = \frac{TN}{(TN+FP)}$; достовірність виявлення кібератак системою BotGRABBER (Q): $Q = \frac{TP+TN}{TP+TN+FP+FN}$, де TP (true positives) – кількість шкідливих поведінок, класифікованих як шкідливі поведінки (атаки); TN (true negatives) – кількість нешкідливих поведінок, класифікованих як нешкідливі поведінки; FP (false positives) - кількість шкідливих поведінок (атак), класифікованих як нешкідливі поведінки (помилки першого роду, хибні спрацювання); FN (false negatives) - кількість класифікованих атак як нешкідливі поведінки (невиявлення, помилки другого роду).

Крім того, здатність системи BotGRABBER забезпечувати резильєнтність корпоративних мереж за наявності кібератак була оцінена за формулою: $GR = \left[R \times \frac{SRAP_{RP}}{SRAP_{DP}}\right] \times (TMPL)^{-1} \times RCAB$, де R – здатність до супротиву, яка вимірює продуктивність мережі між значеннями t_d і t_{ns}, $R \in [0,1]$, де 0 вказує на загальну втрату роботи та 1 - нормальне функціонування мережі; t_d - час, коли мережа є під впливом атаки; t_{ns} - час,

коли мережу було налаштовано відповідно до сценарію безпеки, обраного системою BotGRABBER; $SRAP_{DP}$ - значення швидкості під час фази атаки; $SRAP_{RP}$ - величина швидкості під час фази відновлення мережі; *TMPL* - усереднене в часі значення втрати продуктивності мережі, яке враховує час появи атаки до відновлення мережі; *RCAB* - здатність до відновлення мережі, яка описує ефективність роботи мережі, досягнута після застосованого сценарію безпеки.

Щоб отримати кількість успішних реконфігурацій мережі, необхідно обчислити міру резильєнтності *GR*. Це безрозмірна метрика, яка дозволяє оцінити резильєнтність різних систем під різними типами атак. Таким чином, вважатимемо, що значення метрики GR, що перевищує заданий поріг *(γ> 0,7),* означає, що стабільне функціонування мережі забезпечується. Досягнення необхідного значення показника *GR* після використання сценарію захисту означає, що відновлення мережі було успішним.

Результати тестування інтелектуального агента BotGRABBER для різних класів атак представлено в таблиці 1, з якої видно, що достовірність виявлення кібератак системою BotGRABBER знаходиться в межах від 90,40% до 98,42%. Більше того, чутливість TPR та специфічність TNR знаходяться в діапазоні 91,52–99,13% та 88,46–97,52% відповідно. Тому такий підхід вказує на здатність до забезпечення резильєнтного функціонування КС в умовах кіберзагроз. Інший аспект функціонування BotGRABBER - це можливість застосовувати сценарії безпеки для кібератак за допомогою реконструкції мереж. Для того, щоб з'ясувати можливість функціонування мережі під кібератаками, були імітовані різні типи атак на мережеві хости, сервери та доступні мережеві пристрої. Таблиця 1 демонструє, що кількість успішних реконфігурацій мережі знаходиться в діапазоні від 52,0% до 85%, середнє значення - 71,2%.

Таблиця 1

Результати тестування для різних класів атак

Тип атаки	*T*	*E*				Результат			
		Зловмисні		корисні		*SN*, %	*SP*, %	*Q*, %	*SR*, %
		TP	*FN*	*TN*	*FN*				
DDoS	574	661	16	489	14	97.64	97.22	97.46	73
Ping атака	564	585	11	465	13	98.15	97.28	97.77	76
death атака	364	568	15	429	21	97.43	95.33	96.52	76
TCP SYN	563	567	10	321	7	98.27	97.87	98.12	58
Fragmented UDP Flood	554	384	33	323	29	92.09	91.76	91.94	77
Ампліфікація DNS	421	435	38	553	41	91.97	93.10	92.60	73
Скидання TCP	671	575	31	644	19	94.88	97.13	96.06	85
ICMP attack	764	541	23	565	7	95.92	98.78	97.36	54
RUDY	198	764	36	548	39	95.50	93.36	94.59	77
SIP inv. Flood	611	434	21	561	22	95.38	96.23	95.86	79
secured SSL DDoS	571	554	41	464	35	93.11	92.99	93.05	77
Ping атака	521	494	8	198	8	98.41	96.12	97.74	69
SQL/PHP-ін'єкція	381	653	29	328	35	95.75	90.36	93.88	67
XSS	439	642	39	461	41	94.27	91.83	93.24	77
Фішинг	555	457	4	354	9	99.13	97.52	98.42	70
DNS spoofing	571	453	42	253	33	91.52	88.46	90.40	76
TCP-scan	345	451	21	326	12	95.55	96.45	95.93	67
UDP-scan	231	432	12	326	12	97.30	96.45	96.93	73
Smurf	237	344	15	433	8	95.82	98.19	97.13	68
MAC flooding	655	556	13	326	11	97.72	96.74	97.35	52

Ефективність застосування запропонованої інформаційної технології доводиться порівнянням з результатами виявлення відомими засобами виявлення атак різних типів, представленими авторитетними порталами, які розміщують аналізи останніх вірусних загроз, вивчають новітні розробки в боротьбі з вірусами та оцінки поточних антивірусних продуктів [13-15].

На рис. 6 представлено діаграму, яка демонструє результати порівняльного аналізу розробленого інтелектуального агента BotGRABBER з існуючим антивірусним програмним забезпеченням щодо найнижчих та найвищих значень достовірності виявлення атак, а також рівня резильєнтності.

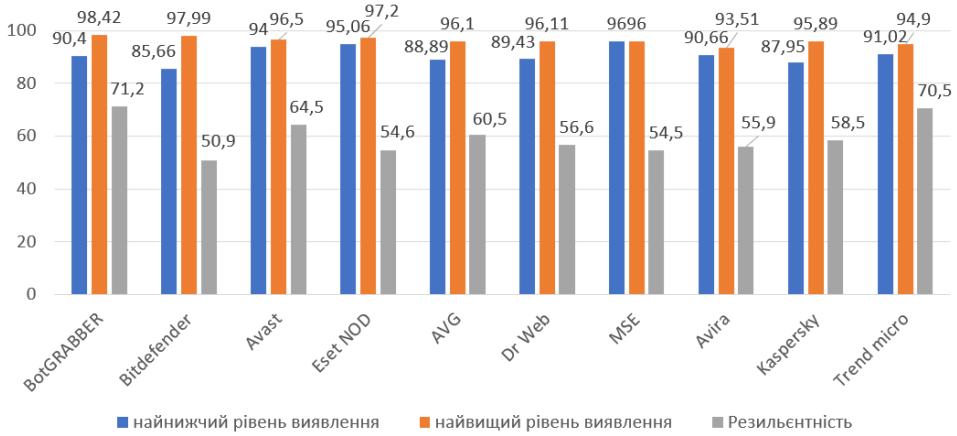


Рис. 6. Результати порівняльного аналізу інтелектуального агента BotGRABBER з існуючим антивірусним програмним забезпеченням

Висновки

Запропоновано інтелектуальний агент виявлення кіберзагроз та ШПЗ в корпоративних мережах, який представляє програмну систему із можливістю виявлення відомих та невідомих кібератак, ШПЗ мережного та хостового типу, а також здатністю продукувати множину сценаріїв безпеки для забезпечення резильєнтності КС в умовах кіберзагроз. Резильєнтність мережі та хостів забезпечується їх динамічною адаптивною реконфігурацією та множиною заходів, що дозволяють фунціонувати системам в умовах атак. Інтелектуальний агент виявлення кіберзагроз та ШПЗ BotGRABBER - це мультивекторна система захисту, оскільки вона поєднує аналіз як в мережі, так і в активності хостів. Комбінована інформація дозволяє не тільки виявляти кібератаки різного типу, але й автоматично застосовувати необхідний сценарій безпеки мережної реконфігурації та адаптації КС відповідно до типу виявленої кібератаки.

Інтелектуальний агент забезпечує: можливість виявлення відомих та невідомих кібератак, можливість виявлення ботнетів, які використовують методи ухилення від DNS (циклічне відображення IP-адреси, “домен flux”, “швидкий flux” та DNS-тунелювання), здатність самостійно застосовувати сценарії безпеки для пом'якшення кібератак, забезпечення резильєнтності корпоративних мереж в умовах кібератак, забезпечення мультивекторного захисту корпоративних мереж.

Експериментальні дослідження продемонстрували, що загальна достовірність виявлення кібератак системою BotGRABBER варіює від 90,40% до 98,42%. Більше того, чутливість та специфічність знаходяться в діапазоні 91,52–99,13% та 88,46–97,52% відповідно. Тому такий підхід вказує на здатність до забезпечення резильєнтного функціонування КС в умовах кіберзагроз.

Література

1. McAfee Mobile Threat Report Q1, 2020. URL: https://www.mcafee.com/content/dam/cons umer/en-us/docs/2020-Mobile-Threat-Report.pdf. – 9.12.2019р. (дата звернення: 10.07.2020).

2. 2020 State of Malware Report. URL: https://resources.malwarebytes.com/files/2020/0 2/2020_State-of-Malware-Report.pdf (дата звернення: 10.07.2020).

3. Naval S., Laxmi V., Rajarajan M., et al. Employing program semantics for malware detection. IEEE Trans Inform Forens Secur. 2015. Vol.10. No.12. Pp. 2591-2604. https://doi.org/10.1109/TIFS.2015.2469253

4. Kharraz A., Arshad S., Mulliner C., et al. UNVEIL: a large-scale, automated approach to detecting ransomware. Proc 25th USENIX Security Symp, 2016. Pp.757-772.

5. Poeplau S., Fratantonio Y., Bianchi A., et al. Execute this! Analyzing unsafe and malicious dynamic code loading in Android applications. Proc Network and Distributed System Security Symp, 2014. Pp.23-26. https://doi.org/10.14722/ndss.2014.23328

6. Zhang F.W., Leach K., Stavrou A., et al. Using hardware features for increased debugging transparency. Proc IEEE Symp on Security and Privacy, 2015. Pp.55-69. https://doi.org/10.1109/SP.2015.11

7. Fratantonio Y., Bianchi A., Robertson W., et al. Triggerscope: towards detecting logic bombs in Android applications. Proc IEEE Symp on Security and Privacy, 2016. Pp.377-396. https://doi.org/10.1109/SP.2016.30

8. Suarez-Tangil G., Conti M., Tapiador J.E. et al. Detecting targeted smartphone malware with behavior-triggering stochastic models. Proc 19th European Symp on Research in Computer Security, 2014. Pp.183-201. https://doi.org/10.1007/978-3-319-11203-9_11

9. Beaucamps P., Gnaedig I., Marion J. Y. Abstraction-based malware analysis using rewriting and model checking. Proc 17th European Symp on Research in Computer Security, 2012. Pp. 806-823. https://doi.org/10.1007/978-3-642-33167-1_46

10. Yang C., Xu Z.Y., Gu G.F., et al. DroidMiner: automated mining and characterization of fine-grained malicious behaviors in Android applications. Proc 19th European Symp on Research in Computer Security, 2014. Pp.163-182. https://doi.org/10.1007/978-3-319-11203-9_10.

11. Canadian Institute for Cybersecurity. Botnet dataset, (accessed January 10, 2019). https://www.unb.ca/cic/datasets/botnet.html.

12. Murphy K.P. Machine learning: a probabilistic perspective. 1 st edition. The MIT press., 2012. P. 1102.

13. AV Comparatives laboratories, (дата звернення 26.01.2020). URL: http://www.av-comparatives.org.

14. Virus Bulletin. URL: http://www.virusbtn.com (дата звернення 26.01.2020).

15. Comparative antivirus testing, (дата звернення 26.01.2020) URL: http:// www.av - comparatives .org.

УДК 004.94
Кисіль Т. М.
Кандидат фізико-математичних наук, доцент,
Доцент кафедри комп'ютерної інженерії та системного програмування
Хмельницького національного університету

РОЗПІЗНАВАННЯ КІНЦЕВИХ ПРИСТРОЇВ КОРПОРАТИВНОЇ МЕРЕЖІ ЗА ПРИНЦИПОМ СВІЙ/ЧУЖИЙ

Недоліками відомими способів організації взаємозв'язку компонентів розподілених систем для виявлення зловмисного програмного забезпечення в корпоративних комп'ютерних мережах є використання централізованої архітектури, що контролюється адміністратором. Це призводить до недостатньо високої достовірності виявлення і локалізації зловмисних дій, бо збір інформації про стан мережі, визначення присутності шкідливих дій та їх блокування здійснюється для обробки єдиним центром, що може бути сповільнено через передачу зібраних даних цьому центру обчислювальні ресурси на яких він розміщений, а також вплив на його роботу адміністратора мережі.

Імунна система є високо розподіленою, високоадаптивною, самоорганізованою за своєю суттю, зберігає пам'ять про минулі зустрічі та має можливість постійно дізнаватися про нові зустрічі. З обчислювальної точки зору, імунна система може надихати вчених та комп'ютерних інженерів. Оскільки обчислювальні проблеми ускладнюються, люди все частіше шукають нові підходи до цих проблем, часто звертаючись до природи за натхненням. Зараз велика увага приділяється імунній системі хребетних як потенційному джерелу такого натхнення, де існує думка, що можна отримати різні ідеї та альтернативні рішення, крім інших біологічно натхненних методів. З огляду на це підвищення уваги до імунної системи, представляється доцільним дослідити цю область досить детально. За аналогією як ІС розпізнає чужі молекули поаналізовано як штучна імунна система буде виявляти чужий пристрій на основі порівняння певної інформації із шаблоном за допомогою або правила Хеммінга або правила r-послідовних збігів.

На жаль, остаточне рішення щодо ідентифікації пристрою корпоративної мережі за принципом свій чужий спирається на досвід і думку адміністратора мережі, тому є необхідність розробити автоматизовану систему прийняття рішень, яка може ґрунтуватись на нечіткій логіці та спиратись на результати роботи вже існуючої СВВ. В даній роботі запропоновано аналізувати бітовий рядок інформації в якості основи для подальшої побудови нечіткої системи прийняття рішень.

***Ключові слова**: корпоративна мережа, свій/чужий, правило Хеммінга, правило r- послідовного збігу, штучні імунні системи*

Постановка проблеми. Вибір концепції побудови конкретної корпоративної мережі визначається цілою низкою чинників: затребувані інформаційні послуги, обсяги переданого трафіку, існуюча інфраструктура і т. д. Але існують і загальні вимоги до корпоративних мереж. Мережі підприємств повинні бути побудовані на основі перевірених технологій, що володіють такими якостями, як масштабованість, гнучкість, мультисервісність, і найголовніше - надійність.

Мережа сучасного підприємства, як правило, повинна підтримувати ряд найбільш затребуваних для бізнесу додатків і керованих сервісів. В першу чергу це:

- можливість високошвидкісного доступу до мережі Інтернет.
- створення віртуальних приватних мереж (VPN).
- захист інформації та зберігання даних.

Слід зазначити, що пристрої, які під'єднанні до корпоративної мережі за допомогою бездротового зв'язку, будуть вважатись підозрілими, і лише їх аномальна поведінка (як то спроба звернення до забороненої IP адреси, використання недозволених портів з'єднання тощо) буде вирішальною при ідентифікації за принципом свій/чужий.

Крім того «чужим» може виявитися також і стаціонарний пристрій мережі, який виявляє нетипову поведінку. А «своїм» може виявитися пристрій, який під'єднався до мережі, але не виявляє зловмисних дій.

Необхідно кожного разу для кожного конкретного пристрою у разі нетипової поведінки приймати рішення – чужий чи свій – і відповідним чином реагувати.

Як правило «чужий» пристрій в мережі намагається здійснити вторгнення, причинами якого можуть бути політичні, економічні, злочинні мотиви або навіть випадковість

Тому необхідно перш за все виявляти такі вторгнення та приймати рішення щодо рівня безпеки такого вторгнення.

Вважається, що багато механізмів, що присутні в біологічній імунній системі, добре підходять для використання в області комп'ютерного виявлення вторгнень у вигляді штучної імунної системи (ШІС).

Аналіз останніх досліджень і публікацій. У літературі про ШІС, і якій йдеться про системи виявлення вторгнень (СВВ), моделювання агентів і лімфоцитів часто об'єднують у загальну сутність детектора [2,3].

Здатність розрізняти своїх і чужих є, мабуть, найголовнішою рисою імунної системи (ІС). Це робиться шляхом розпізнавання лімфоцитами різних агентів. Розпізнавання агентів в біологічній ІС відбувається, коли між рецепторами на поверхні імунних клітин та епітопами на поверхні патогенних мікроорганізмів встановлюються хімічні зв'язки, збіг на низькому рівні зводиться до узгодження білків або фрагментів білка, що називаються пептидами. Далі слово пептид буде використовуватися для представлення як штучних рецепторів, так і штучних агентів.

Використання клональної селекції та соматичної гіпермутації для моделювання дозрівання аффінності в штучних імунних системах, застосованих до мережних систем виявлення вторгнень (МСВВ), було запропоновано, але не реалізовано Хофмейром та Форестом [4]. Хоча було проведено деякі експериментальні роботи, що вивчають роль соматичної гіпермутації в ІС [2].

Постановка завдання. Метою даної роботи є проаналізувати можливість використання штучної імунної системи для виявлення «чужих» пристроїв. Запропонувати структуру такої імунної системи.

Під «чужим» пристроєм будемо розуміти під'єднаний до мережі пристрій, що виявляє аномальні дії, як то спроба несанкціонованого доступу до забороненої частини мережі, виклик або запуск нетипових програм (використання нетипових портів доступу) і т.д. На основі налізу інформації такого роду розробити структуру такої штучної імунної системи

Виклад основного матеріалу дослідження.

В ШІС пептиди часто представлені у вигляді рядків довжиною l, що складаються з символів з алфавіту, що містить m символів. Цей підхід найчастіше використовують для $m = 2$ (тобто бітових рядків).

Пептиди, що представляють агентів, будуть кодувати деяку інформацію, що стосується проблемного домену, до якого застосовується ШІС. Оскільки ІС повиннна розрізняти своїх та чужих на основі пептидів, ШІС повиннна розрізняти своїх та чужих на основі рядків фіксованої довжини l. Кожен такий рядок буде називатися агентом *a*. Сукупність усіх агентів утворює універсум, $U = \{a_1, a_2, ..., a_n\}$, який містить дві підмножини, що не перетинаються; тобто сукупність своїх, U_S, і сукупність чужих, U_N, так що $U = U_S \cup U_N, U_S \cap U_N = \varnothing$. Як зазначено в [4], ШІС тоді стикається з проблемою класифікації; отримавши довільний рядок з U, класифікує його як свій чи чужий. Класифікація на своїх та чужих може також розглядатися як розподіл на нормальних та аномальних.

Ця модель пептидів дотримується вимоги про те, що вся відповідна інформація в проблемній області може бути представлена якимось чином і що повинен існувати певний спосіб компактного кодування узагальнень цієї інформації.

Слід також зазначити, що коли реальні проблеми відображаються у таких уявленнях, свій та чужий не можуть бути роз'єднаними, оскільки два випадки можуть бути відображені в одному представленні.

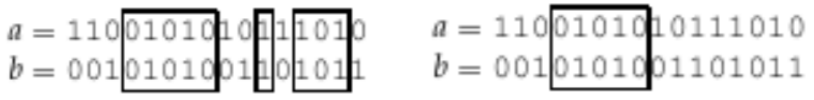
Подібно до СВВ, ІС також може допускати два типи помилок розпізнавання. Це справедливо і для ШІС. Помилково позитивний результат виникає, коли нормальний агент класифікується як чужий, а помилково негативний - коли аномальний агент класифікується як свій.

Описане вище кодування пептиду також використовується для моделювання рецепторів детекторів в ШІС. ШІС має сукупність детекторів D. Кожен детектор $d \in D$ має покриття C_D, яке описує кількість агентів, які він розпізнає. Якщо детектор d не розпізнає жодних агентів, його покриття $C_D = \varnothing$. З іншого боку, якщо d впізнає всіх інших агентів, його покриття є $C_D = U$; всі агенти у універсуму.

Це представлення пептидів дозволяє ШІС розпізнавати різні агенти за допомогою співставлення рядків. Але, одна з приємних особливостей ІС, розглянута з точки зору обробки інформації, полягає в тому, що вона здатна узагальнювати це співставлення. Узагальнення своїх та чужих, яке відбувається в ІС, здійснюється за допомогою наближеного збігу рядків.

У найбільш загальній формі проблема наближеного узгодження рядків полягає у пошуку тексту, де виникає заданий шаблон тексту, допускаючи обмежену кількість «помилок» у збігах. Кожна програма використовує іншу модель помилки, яка визначає, наскільки різними можуть бути рядки (Navarro 2001). Ці тексти можна розглядати як послідовності символів, складених з алфавіту довжини m.

Як правило використовують два таких правила відповідності - правило Хеммінга та правило *r*-послідовних збігів (рисунок 1). Тут основна увага буде зосереджена на правилі *r*-послідовних збігів, оскільки це правдоподібна абстракція зв'язування рецепторів в імунній системі. ІС дуже ефективна тим, що їй вдається розрізняти своїх і чужих, маючи відносно невеликий набір детекторів.



а) б)

Рисунок 1– Співставлення рядків а) за правилом Хеммінга (між рядками довжиною в 16, що складаються з символів бінарного алфавіту, з відповідним обмеженням *r*=9. Два рядки *a* та *b* будуть збігатися для всіх $r \leq 9$)

б) за правилом *r*-послідовних збігів (між рядками довжиною 16, що складаються з символів бінарного алфавіту, з відповідним обмеженням *r*=5. Два рядки *a* та *b* будуть збігатися для всіх $r \leq 5$)

І правило збігу Хеммінга, і правило *r*-послідовних збігів контролюються пороговим параметром *r*, де $0 \leq r \leq l$. Якщо $r = 0$, покриттям *d* є всі рядки, $C_D = U$, а якщо $r = l$, то покриттям *d* є один рядок агента *a*, $C_D = \{a\}$. Чим вище значення *r*, тим конкретніше збіг. І конкретність збігу є аналогічною близькості зв'язування між Ag та лімфоцитом, або детектором.

Слід зазначити, що в ІС відповідність (або розпізнавання) між агентом та лімфоцитом базується на взаємодоповнюючих формах. В ШІС розглядатимуться бінарні рядки та їх «приблизна» рівність.

Правило порівняння Хеммінга базується на відстані Хеммінга між двома рядками. Якщо два рядки *a* і *b* мають однакові біти принаймні в *r* позиціях, вони збігаються (рис. 1.а). Згідно з правилом *r*-послідовного збігу, два рядки *a* і *b* збігаються, якщо вони мають однакові біти принаймні в *r* послідовних позиціях (рис. 1.б).

Імовірність збігу за допомогою правила Хеммінга:

Нехай $Hamming_{l,r}(a,b)$ є оператором, який визначає, чи збігаються дві рядки *a* і *b*, обидва довжиною l, використовуючи правило збігу Хеммінга, з обмеженням, що *r* біт попарно рівні.

Тоді ймовірність збігу між двома випадково вибраними рядками a та b становить

$$P(Hamming_{l,r}(a,b)) = 2^{-l}\sum_{i=r}^{l} C_l^i \qquad (1)$$

Імовірність отримують, з огляду на те, що 2^{-l} - це ймовірність одиничного збігу, а C_l^i - кількість рядків в U, які мають однакові біти в і позиціях.

Механізм негативного відбору в ІС часто використовується в ШІС для проведення виявлення на основі аномалії. У [1, 4, 5] це змодельовано за умови, щоб дійсними детекторами були ті детектори, які не виявляють самоагентів під час толерування: По-перше, детектор генерується випадковим чином, що означає, що його рецептори можуть розпізнавати що завгодно. Якщо детектор щось розпізнає під час допуску, він гине. Якщо детектор переживає термін допуску, він стає зрілим і наївним детектором (його називають наївним, оскільки він ще не виявив жодних збудників).

Це використання негативного відбору ґрунтується на припущенні, що якщо детектор розпізнає що-небудь під час допуску, є своїм. Таким чином, ШІС імпліцитно дізнається, що все, що збігається з його зрілими детекторами та детекторами пам'яті, є чужим.

Оскільки детектори з часом контролюють кілька пептидів, це означає, що при низьких значеннях *r* детектори AIS будуть відповідати практично будь-чому. З іншого боку, при високих значеннях *r* детектори будуть відповідати набагато меншому набору агентів.

Оскільки набір детекторів ШІС генерується за допомогою негативного відбору, менші значення *r* призводять до вищої ймовірності відповідності «свій» під час толерування, а більш високі значення *r* зменшують ймовірність відповідності «свій» під час толерування. Таким чином, чим нижче значення *r*, тим більше спроб потрібно ШІС для генерації кожного зрілого детектора. З вищими значеннями *r* необхідна менша кількість спроб генерації детекторів, але для досягнення певного рівня покриття необхідний також більший набір детекторів.

Це призводить до ситуації компромісу, коли для нижчих значень *r* потрібен менший набір детекторів для досягнення певного покриття, тоді як AIS потребує більше спроб для кожного дійсного детектора, який він генерує. Виходячи з ролі, яку поєднання клональної селекції та соматичної гіпермутації відіграє в ІС, передбачається, що такі механізми збільшать різноманітність детекторів та схожість між детекторами та агентами.

СВВ на основі ШІС складається з двох основних частин - основне ядро ШІС та детектори. Основне ядро ШІС розташоване на шлюзі кожної локальної мережі, а детектори - кожен користувач системи. Кожен з вказаних компонентів складається з агентів що співставляють інформацію один від одного щоб виявити аномалії та вторгнення. Ціль даної структури - зменшити час виявлення для кожного з'єднання з допомогою надання можливостей детектора (аналіз трафіку та повідомлення про небезпеку) кожному користувачу. В результаті навантаження по обробці трафіку буде розподілятись на кожного користувача - кожен користувач сам відповідає за аналіз власного трафіку. Тому замість того щоб аналізувати кожен пакет мережі (що створює велику потребу в обчислювальних можливостях та затримку в виявленні) центральне ядро буде обробляти сигнали небезпеки від користувачів мережі.

Основне ядро складається з двох частин - модуль Навчання та модуль Детекторів-користувачів, обидві частини разом виконують чотири основні завдання:

- створення шаблонів ознак
- аналіз повідомлень від користувачів
- запам'ятовування робочих шаблонів
- розподілення, синхронізація шаблонів ознак кожного детектора

Кожен модуль є програмою на комп'ютері користувача чи шлюзові що виконує одне із завдань (шлюз чи користувач може мати декілька програм одночасно).

Модуль навчання складається з програми-дешифратора та програми навчання, на ньому лежить відповідальність за створення основних випадкових шаблонів ознак на ранніх стадіях роботи системи. Модуль детекторів складається з програми аналізу та програми-диспетчера вторгнень. Перша програма обробляє сигнали від користувачів та в певних випадках запам'ятовує шаблони на які користувач зреагував і схрещує їх для виконання генетичного алгоритму, друга програма відповідає за розповсюдження та синхронізацію шаблонів між користувачами-детекторами.

Перед оцінкою системи відбувається її попереднє навчання та налаштування параметрів. Попереднє навчання ШІС відбувається з допомогою використання набору безпечних (своїх) даних та небезпечних (чужих) даних. Для обробки пакетів трафіку їх спочатку необхідно розшифрувати та перетворити в

оброблену інформацію – цим займається програма-дешифратор. Інформація містить такі поля як ір надсилача, ір отримувача, порт надсилача, порт отримувача, протокол, розмір пакету. Ця інформація дістається з пакетів та перетворюється в послідовності з 112 бітів. Таблиця 1 показує правильно розширений приклад елементів інформації що отримується та їх розміром(в бітах).

Таблиця 1: Приклад можливої інформації з пакету

Назва поля	Мінімум - максимум	Кількість бітів
IP отримувача	0.0.0.0 - 255.255.255.255	32 біта
IP надсилача	0.0.0.0 - 255.255.255.255	32 біта
порт отримувача	0 – 65535	16 бітів
порт відправника	0 – 65535	16 бітів
час продовження	0 – 4096 секунд	12 бітів
протокол	0 – 16	4 біта

Після дешифрування всіх тренувальних наборів інформації в бітові послідовності їх передають в програму навчання що використовується для створеня шаблонів для детекторів. Алгоритм негативного відбору використовується для створення першого покоління шаблонів.

Спочатку створюється і перевіряється "молодий" випадковий набір бітових послідовностей-шаблонів на базовому наборі тестових даних. Якщо шаблон спрацьовує на "своїх" пакетах то він заміняється новим і так до тих пір поки шаблон не перестане реагувати на свої пакети. Далі відбувається наступний крок алгоритму негативного відбору що відсіює шаблон що не реагує на жоден з "чужих" пакетів. Все що не відсіялось додається до результуючого набору шаблонів. Цей процес повторюється поки кожен з чужих пакетів не співпаде з хоча б трьома шаблонами з результуючого набору шаблонів. Для порівняння послідовностей використовується *r*-бітове співпадіння.

Процесс навчання основних детекторів вказаний на рисунку 2.

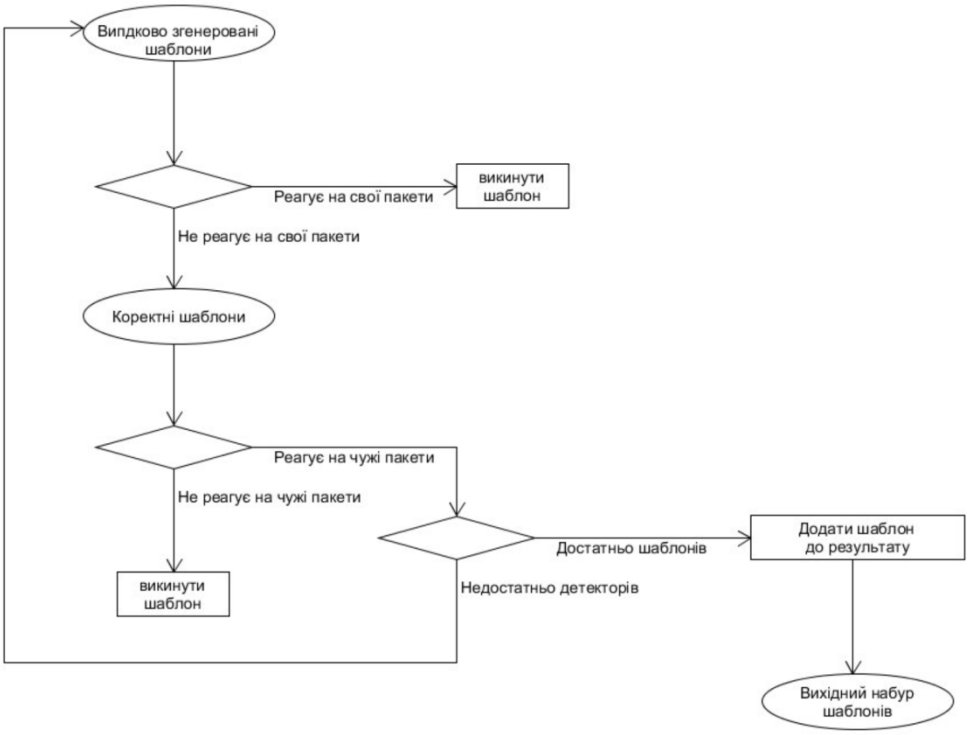


Рисунок 2 – Навчання детекторів

Після навчання всі детектори мають отримати набір шаблонів. Цю роль бере на себе програма-диспетчер що також синхронізує зміни шаблонів а також встановлює для них шаблони ознак. Диспетчер також отримує сигнали небезпеки та перенаправляє їх до програми-аналізатора для обробки.

Як тільки у користувача відбудеться вторгнення і користувач його помітить за шаблоном то він відправить повідомлення що буде включати підозрілий пакет. Таким чином інформація про вторгнення, наприклад, кількість детекторів що помітили небезпеку, їх відношення до підозрілого пакету та їх профіль буде надіслано до основного ядра СВВ на аналіз та обробку. Програма-аналізатор з модуля-детектора цим і займається. Якщо кількість підозрілих пакетів виходить за межі норми то відбувається покращення шаблонів. Генетичний алгоритм виконується для створення шаблонів що краще виділяють певні види аномалій. В той же самий час аналізатор спробує заблокувати пакети певного типу через фаєрволл. Також якщо кількість повідомлень про небезпеку менша за нижню границю небезпеки то пакет все одно додасться для подальшої обробки в пасивному режимі. Після створення нових шаблонів вони будуть передані програмі-диспетчеру для розповсюдження між вузлами-користувачами.

Коли задіяні шаблони надсилаються в аналізатор, то генетичний алгоритм використовується для створення з них шаблонів, що будуть скопійовані для початкового покоління для відбору.

Формула яка визначає чи нам треба відбирати шаблон для генетичного алгоритму:

[Мінімальна оцінка принадності для відбору]=
=[сума оцінок шаблонів]/[кількість шаблонів]

Шаблони, оцінка яких вища за мінімальну використовуються для створення наступного покоління за допомогою генетичного алгоритму. Кожен шаблон може копіюватись певну кількість разів - кількість визначається за формулою:

[кількість копій]=ціла частина(10*[оцінка шаблону]/[сума оцінок шаблонів])

Після виконання клонувань виконується генетичний алгоритм - вибрані детектори проходять через операції кросоверу, мутації та репродукції певну кількість поколінь. В кожному поколінні визначається нова сума оцінок шаблонів. В кожному поколінні вибирається новий кандидат на додавання. І якщо його оцінка менше ніж максимальна з початкового шаблону, то генетичний алгоритм зупиняється і кандидат на додавання розповсюджується між користувацькими вузлами-детекторами. Якщо через певну кількість поколінь не можна зробити кращий шаблон, то розповсюджується кращий із створених шаблонів.

Для покращення механізму СВВ і збільшення ефективності шаблони розповсюджуються на всі вузли в мережі. Це також зумовлює простоту і розширюваність такої системи. В системі присутні два типи вузлів-детекторів: вузли пам’яті і активні детектори. Дешифратор використовується для перетворення пакетів для аналізу активними детекторами.

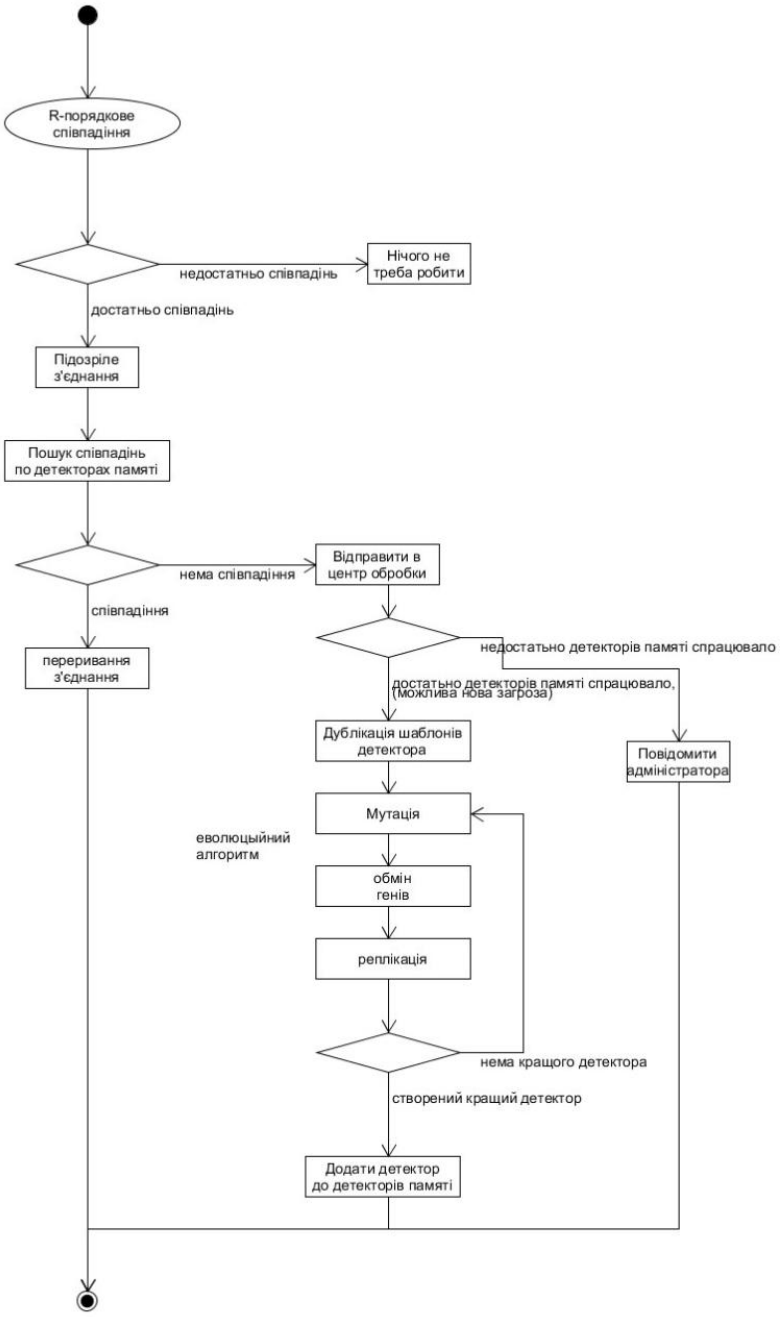


Рисунок 3 – Блок-схема процедури виявлення

Вузли пам’яті дозволяють робити адаптивну відповідь ШІС на вторгнення. Вузли пам’яті містять набір шаблонів що створюється і змінюється використовуючи генетичний алгоритм. Аналізатор виконує раніше вказаний генетичний алгоритм, що дозволяє детекторам краще ідентифікувати втручання.

Використання вузлів пам'яті дозволяє зменшувати час відповіді і краще реагувати на раніше помічені види втручань. Також такий підхід збільшує ефективність СВВ, зменшуючи час обробки пакетів. Вузли пам'яті також добре працюють для зменшення кількості неправильних позитивних та неправильних негативних відкликів. Як тільки аномалія була помічена на вузлі і будь-який шаблон з вузлів пам'яті підійшов під трафік в мережі, підходящий трафік буде направлено на аналіз в основне ядро ССВ. Весь процес аналізу з серверної та клієнтської сторони показаний на рисунку 3.

Активний детектор містить набір шаблонів, що точно розрізняють трафік за схемою свій-чужий. Всі вхідні пакети перевіряються цими детекторами. Якщо будь-який пакет визнано аномальним за будь-яким шаблоном, то вказаний пакет передається далі на аналіз в основне ядро СВВ для обробки. Кількість шаблонів, що були задіяні на підозрілому пакеті, оцінка кожного з задіяних шаблонів, властивості пакету - все це необхідно кожен раз передавати для аналізу в основне ядро. Межа підозрілості пакету - властивість що дозволяє збільшувати точність визначення втручання і відсіяти неправильні позитивні спрацьовування. Якщо кількість шаблонів, що були зачеплені при аналізі, більша за мінімальну межу підозрілості то сесія з даним пакетом буде примусово відключена фаєрволом.

Висновки. Попереднє навчання ШІС відбувається з допомогою використання набору безпечних (своїх) даних та небезпечних (чужих) даних. Для обробки пакетів трафіку їх спочатку необхідно розшифрувати та перетворити в оброблену інформацію, яка містить такі поля як ір надсилача, ір отримувача, порт надсилача, порт отримувача, протокол, розмір пакету. Ця інформація дістається з пакетів та перетворюється в послідовності з 112 бітів. В подальшому дану інформацію використаємо для нечіткої системи управління, щоб дасть відповідь на питання чи даний пристрій свій чи чужий

Список літератури:

1. Hightower Ron, Stephanie Forrest, and Alan S. Perelson. The Baldwin Effect in the Immune System: Learning by Somatic Hypermutation *Adaptive Individuals in Evolving Populations: Models and Algorithms,* Addison-Wesley Publishing Company, Reading Massachusetts. 1996. P.159–167.
2. Timmis J. Artificial immune systems: today and tomorrow. *Natural Computing*, 6(1):1-18, March 2007.
3. Литвиненко В.І. ПОБУДОВА ШТУЧНИХ ІМУННИХ СИСТЕМ // *Наукові праці. Комп'ютерні технології* 2010 р., Випуск 121, Том 134 С.166-178
4. Hofmeyr S., Forrest S. Architecture for an Artificial Immune System *Evolutionary Computation.* 2000. 8 (4). P. 443–473.
5. Lysenko S., Savenko O., Bobrovnikova K., Kryshchuk A., Savenko B.. Information technology for botnets detection based on their behaviour in the corporate area network. *Communications in Computer and Information Science*. 2017. – Vol. 702. – PP.83-95, ISSN: 1865-0929 (part scientometric SCOPUS).

Kysil T. M. RECOGNITION OF END DEVICES OF CORPORATE NETWORK ON THE PRINCIPLE OF SELF / NONSELF

The disadvantages of known methods of interconnecting distributed system components to detect malware in corporate computer networks are the use of a centralized architecture controlled by the administrator. This leads to insufficient reliability of detection and localization of malicious actions, because the collection of information about the state of the network, determining the presence of malicious actions and blocking them is carried out for processing by a single center, which can be slowed down by transmitting collected data to this center. also the impact on his work as a network administrator.

The immune system is highly distributed, highly adaptable, self-organized in nature, preserves the memory of past meetings and has the opportunity to constantly learn about new meetings. From a computational point of view, the immune system can inspire scientists and computer engineers. As computational problems become more complex, people are increasingly looking for new approaches to these problems, often turning to nature for inspiration. Much attention is now being paid to the vertebrate immune system as a potential source of such inspiration, where it is thought that different ideas and alternative solutions can be obtained in addition to other biologically inspired methods. Given this increase in attention to the immune system, it seems appropriate to explore this area in some detail. By analogy, how the IS recognizes foreign molecules analyzed as an artificial immune system will detect a foreign device based on comparing certain information with a pattern using either the Hamming rule or the r-sequence matching rule.

Unfortunately, the final decision on identifying a corporate network device based on one's own experience is based on the experience and opinion of the network administrator, so there is a need to develop an automated decision-making system that can be based on fuzzy logic and based on existing IDS. In this paper, it is proposed to analyze the bit string of information as a basis for further construction of a fuzzy decision-making system.

***Keywords**: corporate network, self / nonself, Hamming match rule, r-contiguous match rule, artificial immune systems.*

ВІЙСЬКОВИЙ ІНСТИТУТ
КИЇВСЬКОГО НАЦІОНАЛЬНОГО УНІВЕРСИТЕТУ
ІМЕНІ ТАРАСА ШЕВЧЕНКА

ТЕЗИ ДОПОВІДЕЙ

XVI Міжнародної науково-практичної конференції

"Військова освіта і наука: сьогодення та майбутнє"

ТОМ 1

27 листопада 2020 року

За загальною редакцією
к.пед.н., доц., заслуженого працівника освіти України
генерал-майора Ігоря ТОЛОКА

Київ – 2020

Сучасний бій ведеться із застосуванням передових технічних засобів розвідки та керування зброєю, що обумовлює необхідність об'єднання у систему заходів декількох видів бойового забезпечення, направлених на зменшення ефективності ведення противником розвідки та застосування високоточної зброї.

З метою замаскованого наближення до кораблів противника доцільно розглядати створення комбінованої системи захисту на основі засобів розробленого і вдосконаленого комплексу, який застосовується у складі корабельних комплексів оптико-електронного придушення, та засобів аерозольного захисту.

Аерозольне маскування, у комплексі з іншими заходами введення в оману, а в деяких випадках і самостійно, під час підготовки та у ході проведення операцій (бойових дій) надасть можливість:

- здійснювати протидію засобам розвідки і наведення зброї противника (в тому числі ВТЗ) шляхом постановки об'єктових і площинних димових завіс і екранів, ефективних у видимому і інфрачервоному (ІЧ) діапазоні спектра електромагнітного випромінювання;
- здійснювати протидію вражаючому впливу світлового випромінювання ядерного вибуху шляхом постановки площинних димових завіс, ефективних у видимому діапазоні спектра електромагнітного випромінювання.

к.фіз.-мат.н. Кисіль Т.М. (ХмНУ)
к.т.н. Кльоц Ю.П. (ХмНУ))
Бондаренко Т.В. (ВІТІ)
Шаховал Є.С. (ХмНУ)

Проблеми розпізнавання кінцевих пристроїв корпоративної мережі за принципом свій/чужий

Побудова корпоративних телекомунікаційних мереж ґрунтується на загальних принципах побудови сегментів фізичного канального й мережевого рівнів. Кількість рівнів структуризації в мережі є пропорційним до масштабу самої мережі.

Мережева безпека складається з положень і політики, прийнятої адміністратором мережі, щоб запобігти і контролювати несанкціонований доступ, неправильне використання, зміни або відмови в комп'ютерній мережі та мережі доступних ресурсів. Найбільш поширений і простий спосіб захисту мережевих ресурсів є присвоєння їм унікального імені та відповідного паролю.

Безпечний доступ для всіх типів клієнтів із використанням різноманітних механізмів доступу грає важливу роль для забезпечення доступу користувачів до потрібних даних, незалежно від їх місцезнаходження та використовуваних пристроїв.

Важливим етапом захисту комп'ютерної мережі є її захист по периметру, тобто мережа повинна ідентифікувати всі свої кінцеві пристрої, як дротові так і бездротові.

Існує два основних варіанти налаштування бездротової мережі:

– Ad-hoc - передача безпосередньо між пристроями;

– Hot-spot - передача здійснюється через точку доступу;

В Hot-spot мережах присутня точка доступу, за допомогою якої відбувається не тільки взаємодія всередині мережі, але і доступ до зовнішніх мереж. Hot-spot представляє найбільший інтерес з точки зору захисту інформації, бо зламавши точку доступу, зловмисник може отримати інформацію не тільки зі станцій, розміщених в даній бездротовій мережі.

Одним із методів обмеження доступу до мережі є фільтрація MAC-адреси: Фільтрацію можна здійснювати трьома способами:

– Точка доступу дозволяє отримати доступ станціям з будь-якою MAC-адресою;

– Точка доступу дозволяє отримати доступ тільки станціям, чиї MAC-адреси знаходяться в довірчому списку.

Найбільш надійним з точки зору безпеки є другий варіант, хоча він не розрахований на підміну MAC-адреси, що легко здійснити зловмисникові.

Список використаних джерел:

1. Воробієнко П. П. Телекомунікаційні та інформаційні мережі : Підручник / П.П. Воробієнко, Л.А. Нікітюк, П.І. Резніченко. – К.: САММІТ-Книга, 2010. – 708 с.: іл.

д.т.н., с.н.с. Комарова Л.О. (ОНАЗ)
к.т.н. Кльоц Ю.П. (ХмНУ)
Брітов О.В. (ХмНУ)
Нагребецький О.В. (ХмНУ)
Шаховал Є.С. (ХмНУ)

Тестування обладнання корпоративної мережі

Корпоративні мережі є подальшим етапом розвитку локальних мереж, однак специфіка їх побудови та використання значно відрізняються як від локальних так і від глобальних чи регіональних мереж. Вони характеризуються обмеженим розміром, як локальні мережі та розподілом на підмережі, як глобальні чи регіональні мережі.

Оскільки до корпоративних мереж висуваються більш жорсткі вимоги до обсягів даних, що передаються мережею, захисту цих даних та надійності інфраструктури, важливим етапом підтримання функціонування корпоративної мережі є тестування обладнання, на базі якого збудована мережа.

Процес тестування мережевого обладнання в загальному використовує стек протоколів TCP/IP та складається з двох етапів. Перший – встановлення типів обладнання та зв'язків між ними (Neighbor Discovery Protocol). Другий – безпосередньо тестування мережевого обладнання. Для корпоративної мережі

ДОДАТОК Б

(обов’язковий)

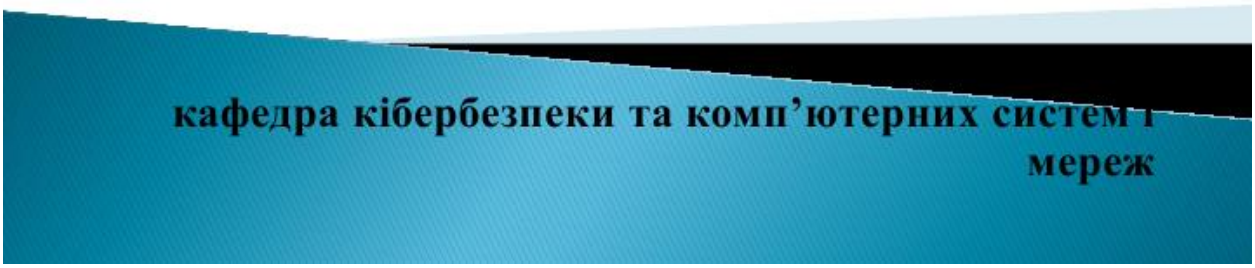
Презентаційні матеріали

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Кисіль Тетяна Миколаївна

МЕТОД РОЗПІЗНАВАННЯ КІНЦЕВИХ ПРИСТРОЇВ КОРПОРАТИВНОЇ МЕРЕЖІ ЗА ПРИНЦИПОМ СВІЙ/ЧУЖИЙ

Науковий керівник
к.т.н., доцент Кльоц Ю. П.



Предметом дослідження є метод розпізнавання кінцевих пристроїв корпоративної мережі за принципом свій / чужий.

Мета магістерської роботи полягає в розробці рекомендацій для прийняття рішень щодо розпізнавання пристрою за принципом свій/чужий.

Відповідно до вказаної мети в роботі поставлені, обгрунтовані і вирішені наступні **завдання**:

1. Проаналізувати ймовірнісні підходи до визначення рівня небезпеки корпоративної мережі.
2. Запропонувати структуру системи виявлення вторгнень, яка використовує генетичний алгоритм для еволюції шаблонів виявлення та їх запам'ятовування, що дозволить ідентифікувати кінцевий пристрій мережі як свій або чужий.
3. Запропонувати алгоритм процедури виявлення «чужого» пристрою як такого, що здійснює неправомірні дії в мережі, наприклад вторгнення
4. Запропонувати нечітку модель розпізнавання пристроїв за принципом свій / чужий.

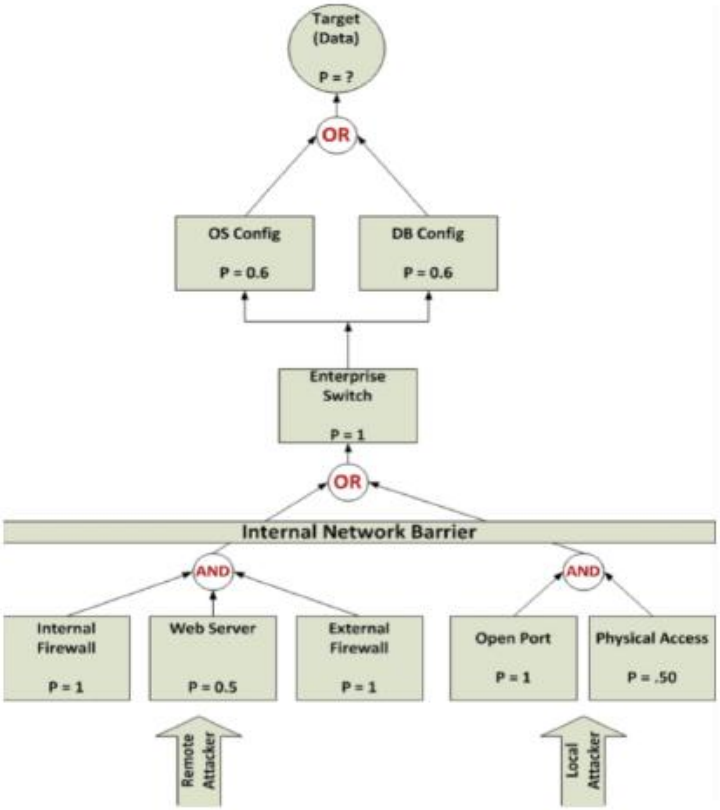
Основні нові результати, отримані в роботі та виносяться на захист:

1. Структура системи виявлення вторгнень, яка використовує генетичний алгоритм для еволюції шаблонів виявлення та їх запам'ятовування, що дозволить ідентифікувати кінцевий пристрій мережі як свій або чужий.
2. Запропоновано нечітку модель розпізнавання пристроїв за принципом свій / чужий.

Практична цінність результатів магістерської роботи полягає в отриманих моделях і алгоритмах, щодо ідентифікації пристроїв за принципом свій/чужий.

За темою кваліфікаційної роботи ОКР «Магістр» опубліковано 2 наукові статті і 1 тези доповідей.

Дерево атаки



$a = 1100101010010111010$
$b = 0010101001101011$

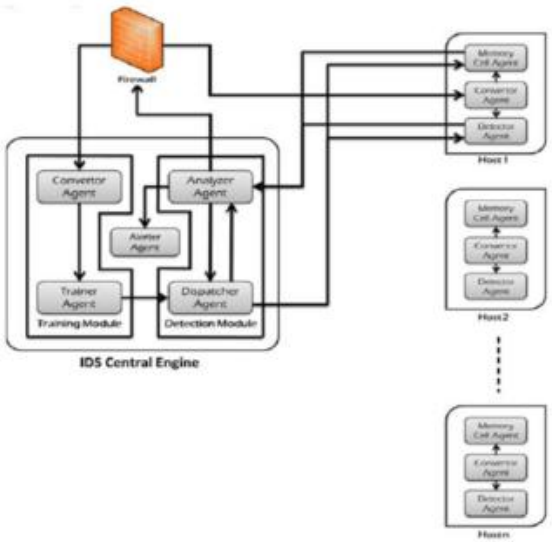
Рисунок – Співставлення рядків за правилом Хеммінга (між рядками довжиною в 16, що складаються з символів бінарного алфавіту, з відповідним обмеженням r=9. Два рядки a та b будуть збігатися для всіх $r \geq 9$)

$a = 1100101010111010$
$b = 0010101001101011$

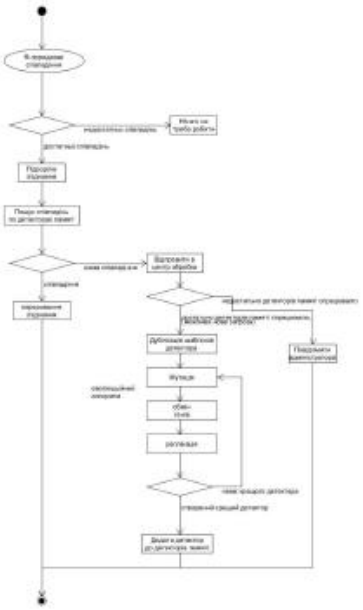
Рисунок – Співставлення рядків за правилом r-послідовних збігів (між рядками довжиною 16, що складаються з символів бінарного алфавіту, з відповідним обмеженням r=5. Два рядки a та b будуть збігатися для всіх $r \geq 5$)

Таблиця – Приклад можливої інформації з пакету

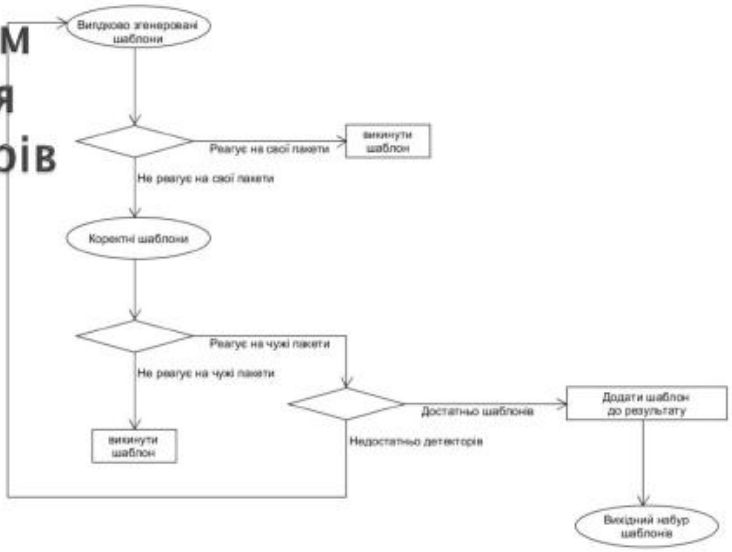
Назва поля	Мінімум - максимум	Кількість бітів
IP отримувача	0.0.0.0 - 255.255.255.255	32 біта
IP надсилача	0.0.0.0 - 255.255.255.255	32 біта
порт отримувача	0 – 65535	16 бітів
порт відправника	0 – 65535	16 бітів
час продовження	0 – 4096 секунд	12 бітів
протокол	0 – 16	4 біта



Блок-схема процедури виявлення



Алгоритм навчання детекторів



Таблиця – Опис змінних моделі

Змінна	Позначення	Тип змінної
IP отримувача (32 біти)	x_1	Вхідна
IP надсилача (32 біта)	x_2	Вхідна
порт отримувача 16 бітів	x_3	Вхідна
порт відправника 16 бітів	x_4	Вхідна
час продовження 12 бітів	x_5	Вхідна
протокол	x_6	Вхідна
Частота	x_7	Вхідна
Несанкціонований доступ	y_1	Проміжна
Аномальна поведінка	y_2	Проміжна
Ідентифікація пристроїв	Y	Вихідна

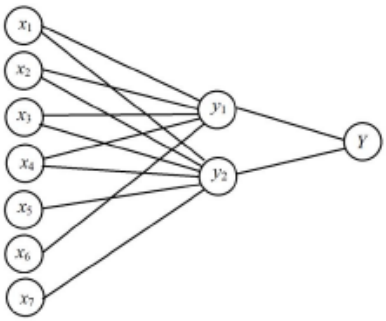
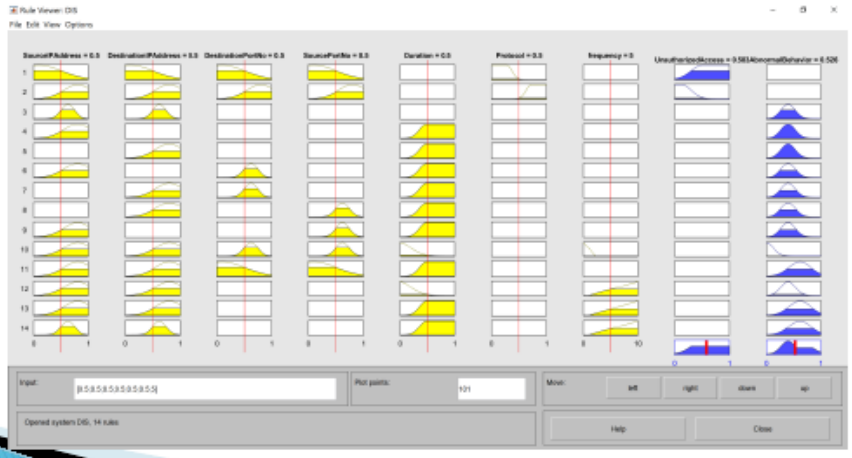


Рисунок – Нечітка модель розпізнавання кінцевих пристроїв за принципом свій/чужий



ВИСНОВКИ

1. Проаналізувано ймовірнісний підхід на основі дерева атак до визначення рівня небезпеки корпоративної мережі.
2. Запропонувано структуру системи виявлення вторгнень, яка потребує використання генетичних алгоритмів для еволюції шаблонів виявлення та їх запам'ятовування, що дозволить ідентифікувати кінцевий пристрій мережі як свій або чужий.
3. Запропоновати алгоритм процедури виявлення «чужого» пристрою як такого, що здійснює неправомірні дії в мережі
4. Запропоновано нечітку модель розпізнавання пристроїв за принципом свій / чужий.

UNICHECK

User name:
Кафедра кибербезпеки

Check ID:
1005451667

Check date:
14.12.2020 12:55:51 EET

Check type:
Doc vs Internet

Report date:
14.12.2020 12:56:33 EET

User ID:
100005590

File name: **Кисіль-3**

Page count: **75** Word count: **14362** Character count: **108349** File size: **2.21 MB** File ID: **1005741942**

9.28% Matches

Highest match: **1.83%** with Internet source (https://uk.wikipedia.org/wiki/%D0%90%D1%81%D0%B8%D0%BD%D1%85%D1%80%D

9.28% Internet sources 303 ..... Page 77

No Library search was conducted

0% Quotes

Exclusion of quotes is off

Exclusion of references is off

0% Exclusions

No exclusions

Modifind

Text modifications detected. Find more details in the online report.

Replaced characters 12

Anti-Plagiarism v-15.257

Максимальное совпадение с одним документом 5.0%

Словари проверки: en_US, ru_RU, ua_UA. **Ошибок в документах: 9%**

	Документ		Суммарное совпадение по Базе Данных	
ID: 82573 Название: Метод розпізнавання кінцевих пристроїв корпоративної мережі за принципом свій/чужий Добавлено в БД: 2020-12-05 Авторы: Кисіль Т. М. Руководители: Кльоц Ю. П. Консультанты: Опоненты:	Символы	Лексемы	Символы	Лексемы
	96959	800	11247 (12%)	116 (15%)

Источник плагиата

ID	Описание	Наличие плагиата в документе	
		Символы	Лексемы

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

РЕЦЕНЗІЯ НА ДИПЛОМНУ РОБОТУ

Дипломник студентка групи КІ1м-19-1 Кисіль Т. М..

Тема Метод розпізнавання кінцевих пристроїв корпоративної мережі за принципом свій/чужий

Спеціальність 123 – Комп'ютерна інженерія

Обсяг дипломної роботи:

Кількість листів креслень 0; кількість сторінок записки 110

1.Короткий зміст ДР та прийнятих рішень Представлена робота присвячена актуальній темі побудови системи виявлення вторгнень, яка дозволить ідентифікувати пристрій за принципом свій/чужий в залежності від поведінки пристрою в корпоративній мережі. Складається з наступних розділів: вступ, дослідження організації та аналіз технологій побудови корпоративної мережі, алгоритми штучних імунних систем для виявлення загроз, агентна модель для системи виявлення вторгнень на основі штучної імунної системи, нечітка система висновку щодо ідентифікації пристроїв корпоративної мережі за принципом свій/чужий, висновки, перелік джерел посилання, додатки

2. Висновок про відповідність ДР поставленому завданню Магістерська кваліфікаційна робота виконана у відповідності з завданням із дотриманням всіх вимог.

3. Характеристика виконання кожного розділу роботи, ступінь використання останніх досягнень науки і техніки і передових методів роботи: В першому розділі студентка провела дослідження ймовірнісних підходів до визначення рівня небезпеки корпоративної мережі. В другому розділі запропоновано структуру системи виявлення вторгнень на основі генетичного алгоритму для еволюції шаблонів. В третьому розділі запропоновано алгоритм процедури виявлення «чужого». пристрою В четвертому розділі запропоновано нечітку модель розпізнавання пристроїв за принципом свій/чужий

4. Позитивні сторони роботи До позитивних сторін роботи слід віднести актуальність даного направлення дослідження, деталізацію аналізу усіх розглянутих стратегій вирішення проблеми та поглиблене опрацювання всіх аспектів реалізації з практичним використанням запропонованого рішення.

5. Негативні сторони роботи До негативних сторін роботи слід віднести недоліки по оформленню представленого матеріалу, що були виправлені. Не вистачило програмної реалізації запропонованих алгоритмів, з ними робота була б більш повною

6. Оцінка графічного оформлення та пояснювальної записки роботи Дані матеріали роботи є структурованими у чіткій та логічній формі та відображають послідовність виконання поставлених завдань. І хоча й в них було знайдено декілька стилістичних та орфографічних помилок, вони були пізніше усунені. Тому дане виконання пояснювальної записки та графічного оформлення можна вважати прийнятним.

7. Відгук про роботу в цілому Загалом, зміст представленої роботи в повній мірі розкриває обрану тему. Дослідження, проведені в матеріалах є достатньо аргументованими. Прослідковуються високі теоретичні та практичні рівні у даному виконанні. Результатом проведення досліджень стали відповідні висновки і конкретні пропозиції щодо прийняття рішень в розпізнаванні пристроїв за принципом свій/чужий

8. Інші зауваження немає

9. Оцінка дипломної роботи Робота заслуговує оцінки ««відмінно»», а її автор – *присвоєння кваліфікації «магістра» з комп'ютерної інженерії.*

РЕЦЕНЗЕНТ (прізвище, ім'я, по-батькові, посада, місце роботи) Лисенко Сергій Миколайович, доктор технічних наук, доцент кафедри комп'ютерної інженерії та системного програмування ХНУ

"_____" _____ 2020 р. _____
(підпис)

РІШЕННЯ ЕКСПЕРНОЇ КОМІСІЇ
КАФЕДРИ КІБЕРБЕЗПЕКИ ТА КОМП'ЮТЕРНИХ СИСТЕМ І МЕРЕЖ
ПРО ДОПУСК КВАЛІФІКАЦІЙНОЇ РОБОТИ ДО ЗАХИСТУ

Підтверджуємо ознайомлення з результатом звіту подібності щодо роботи, генерованого системою виявлення текстових збігів/ідентичності/схожості:

Назва: Метод розпізнавання кінцевих пристроїв корпоративної мережі за принципом свій/чужий

Автор: Кисіль Тетяна Микоаївна

Спеціальність: 123 – Комп'ютерна інженерія

Освітня програма: Програмування та захист комп`ютерних систем і мереж

Науковий керівник: Кльоц Юрій Павлович, к.т.н., доцент

Після аналізу звіту подібності зроблено такий висновок:

№	Висновок	Позначка про відповідність
1	Запозичення, виявлені в роботі, є законними і не є плагіатом (далі – зазначаються підстави віднесення запозичень до правомірних). Робота приймається до захисту.	відповідає
2	Виявлені запозичення не є плагіатом, розміщені в розділах, які не описують безпосередньо авторське дослідження, але кількість цитат перевищує обсяг, виправданий поставленою метою роботи (далі – зазначаються детальні та аргументовані підстави віднесення запозичень до правомірних). Робота приймається до захисту, але має бути відкоригована. Відкоригований варіант має бути поданий на кафедру за 2 дні до захисту, разом із заявою щодо самостійності виконання письмової роботи та ідентичності друкованої та електронної версії роботи.	
3	Виявлені запозичення не є плагіатом, але частково розміщені в розділах, які описують безпосередньо авторське дослідження, а кількість цитат перевищує обсяг, виправданий поставленою метою роботи. В зв'язку з цим мета роботи та поставлені завдання не були досягнені. Робота може бути допущена до захисту (наступного року) після того як буде відкоригована та допрацьована і успішно пройде повторну перевірку на академічний плагіат.	
4	Робота містить навмисні текстові спотворення, передбачувані спроби укриття запозичень або інші прояви академічного плагіату. Робота містить фабрикацію або фальсифікацію даних. Робота не допускається до захисту.	
5	Інше:	

Підтвердження:

Запозичення, виявлені в роботі, є законними і не є плагіатом, оскільки:

1) запозичення розміщені в розділах аналізу існуючих аналогів та прототипів, які не описують безпосередньо авторське дослідження і не стосуються результатів роботи;
2) усі запозичення фрагментарні, або мають належним чином оформленні посилання;
3) окремі виявлені збіги є загальновживаними фразами або виразами, про що свідчить посилання системи на збіг з 10-40 джерелами на один фрагмент речення;
4) окремі виявлені збіги є усталеними термінами і фразами, що описують предметну область або відомі алгоритми

Сумарний обсяг всіх запозичень, визначений системою виявлення збігів/ ідентичності/схожості, складає 9.28% і адресується до 303 першоджерела, що, з урахуванням наведених обґрунтувань, відповідає характеру наукового дослідження і свідчить на користь кваліфікаційної роботи.

Керівник роботи _____ Ю.П. Кльоц

Завідувач кафедри КбКСМ, гарант ОП _____ Ю.П. Кльоц

Дата: 14.12.2020