Хмельницький національний університет
Факультет програмування
та комп'ютерних і телекомунікаційних систем
Кафедра кібербезпеки та комп'ютерних систем і мереж

КВАЛІФІКАЦІЙНА РОБОТА

бакалавр
Освітній рівень

Інтелектуальна система аналізу трафіку в складі комплексної системи захисту
Назва теми

КвРКБ.170146.17.01.08 ПЗ
Шифр

Галузь знань *12 «Інформаційні технології»*
Шифр, назва

Спеціальність *125 «Кібербезпека»*
Шифр, назва

Освітня програма *«Кібербезпека»*
Назва

Виконав: студент IV курсу, група КБ–17–1 _____ Підпис Д.В. Любінецький Ініціали, прізвище

Керівник _____ Підпис, дата О.С. Андрощук Ініціали, прізвище

Нормоконтролер _____ Підпис, дата І.В. Муляр Ініціали, прізвище

До захисту допускаю:
Зав. кафедри кібербезпеки та
комп'ютерних систем і мереж _____ Підпис Ю.П. Кльоц Ініціали, прізвище

« 7 » червня 2021 р.

Хмельницький 2021

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет Програмування та комп'ютерних і телекомунікаційних систем

Кафедра Кібербезпеки та комп'ютерних систем і мереж

Освітній рівень бакалавр

Галузь знань 12 Інформаційні технології

Спеціальність 125 Кібербезпека

Освітня програма Освітня програма «Кібербезпека»

ЗАТВЕРДЖУЮ

Завідувач кафедри Ю.П.Кльоц

"5" 01 2021 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

Любінецькому Денису Володимировичу

Прізвище, ім'я, по батькові студента

1. Тема проекту (роботи) Інтелектуальна система аналізу трафіку в складі комплексної системи захисту

Керівник проекту (роботи) Андрощук Олександр Степанович

Прізвище, ім'я, по батькові, науковий ступінь, вчене звання

доктор технічних наук, доцент

Затверджена наказом ректора університету від 05.02.2021 № 11 додаток №9

2. Строк подання студентом проекту (роботи) на кафедру 28.05.2021

3. Вихідні дані до проекту (роботи) Дослідження даних щодо мережевого трафіку та технологій його захисту

4. Зміст пояснювальної записки (перелік питань, які потрібно розробити)

Провести аналіз даних загроз мережевих атак, дослідити методи і способи захисту мережі середньостатистичного корпоративного підприємства, провести аналіз програмного та апаратного забезпечення, що використовують підприємства для захисту мережевого трафіку, побудувати підсистему захисту на основі використання інтелектуального модулю, практично впровадити і проаналізувати запропоновані рішення

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

Архітектура мережевої аналітики ExtraHop Reveal(x)

Архітектура NIDS

Архітектура реалізованого IDS

Структура реалізації власного програмного модулю

6. Консультанти розділів дипломного проекту (роботи)

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

7. Дата видачі завдання «05» 02 2021 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів (розділів) дипломного проекту (роботи)	Термін виконання етапів проекту (роботи)	Прим
1.	Вибір та затвердження теми кваліфікаційної роботи	Лютий - 2 декада	
2.	Отримання завдання на кваліфікаційну роботу	Лютий - 2 декада	
3.	Проектування та розробка загальної архітектури і структури системи захисту	Березень - 3 декада	
4.	Програмна реалізація запропонованого рішення та тестування системи	Квітень - 1 декада	
5.	Виконання розрахункової частини	Квітень - 3 декада	
6.	Формулювання висновків	Квітень - 3 декада	
7.	Погодження розділів з консультантом з нормоконтролю	Травень - 1 декада	
8.	Оформлення пояснювальної записки	Травень - 2 декада	
9.	Попередній захист кваліфікаційної роботи	Травень - 3 декада	
10.	Доопрацювання кваліфікаційної роботи	Червень - 1 декада	
11.	Подання роботи для перевірки на плагіат	Червень - 1 декада	
12.	Захист кваліфікаційної роботи	Червень - 1 декада	

Студент Підпис Д.В. Любінецький Ініціали, прізвище

Керівник проекту (роботи) Підпис О.С. Андрощук Ініціали, прізвище

АНОТАЦІЯ

Тема кваліфікаційної роботи: «Інтелектуальна система аналізу трафіку в складі комплексної системи захисту».

Автор роботи: Любінецький Денис Володимирович.

Керівник роботи: Андрощук Олександр Степанович.

Пояснювальна записка: 60 с., 27 рис., 3 табл., 2 дод., 25 джерел.

Графічна частина: 4 плакати.

Метою кваліфікаційної роботи: створити інтелектуальну систему аналізу трафіку, проаналізувати вже існуючі рішення, перевірити корисність та ефективність існуючих та власного рішення.

У кваліфікаційній роботі я проаналізував наявні проблеми та загрози підприємств та організацій, проаналізував методи та способи аналізу мережевого трафіку, переглянув існуючі апаратні та програмні рішення захисту мережевого трафіку, аналізував окреме програмне забезпечення, що вирішує питання мережевого захисту, переглянув різноманітні методи захисту мережевого трафіку, використавши попередньо визначену інформацію побудував надійну підсистему захисту, і надав деякі запропоновані рішення, а саме: використання надійного програмного забезпечення, використання інтелектуального модулю в існуючий засіб захисту та написав власний програмний модуль для виявлення аномалій трафіку. В завершенні роботи я визначив ефективність кожного із рішень. Опираючись на викладені рішення, я вважаю, що поставлена задача була виконана в повній мірі.

Підпис студента

03.06.21

Дата

Формат	Зона	Позиція	Позначення	Найменування	Кільк.	Прим.
А4		1		Завдання на дипломний проект	1	
А4		2		Анотація	1	
А4		3	КвРКБ.170146.17.01.08 ПЗ	Інтелектуальна система аналізу трафіку в складі комплексної системи захисту Пояснювальна записка	1	
А2		4	КвРКБ.170146.17.01.08 Е8	Архітектура мережевої аналітики в ExtraHop Reveal(x) Алгоритм роботи	1	
А2		5	КвРКБ.170146.17.01.08 Е8	Архітектура NIDS Алгоритм роботи	1	
А2		6	КвРКБ.170146.17.01.08 Е8	Архітектура реалізованого IDS Алгоритм роботи	1	

Зм.	Арк.	№ Докум.	Підп.	Дата	КвРКБ.170146.17.01.08 ВП		
Розробив		Любінецький Д.		25.06	Інтелектуальна система аналізу трафіку в складі комплексної системи захисту. Відомість проекту	Літера	Аркуш 1 / Аркушів 2
Перев.		Андрощук О.С		01.06		Н	
Н. контр.		Муляр І.В.				ХНУ, КБ-17-1	
Затв.		Кльоц Ю.П.					

Формат	Зона	Позиц	Позначення	Найменування	Кільк.	Прим.
A2		7	КвРКБ.170146.17.01.08 Е8	Структура реалізації	1	
				власного програмного		
				модуля		
				Схема структурна		

Зм.	Арк.	№ Докум.	Підп.	Дата	КвРКБ.170146.17.01.08 ВП	Арк. 2

ЗМІСТ

ВСТУП

На сьогодні питання кібербезпеки в світі є дуже актуальним. Досить очевидно, що час не стоїть на місці. З кожним днем ми можемо побачити все нові і нові винаходи, що мають полегшувати нам життя. Можна з повною впевненістю сказати, що сьогодні майже кожну людину оточує «комп'ютерний світ».

Зараз на Україні інтернет покриття є більш ніж на 40% території України та з ніж більш 6 млн. користувачів і ці цифри постійно збільшуються. Разом і тим збільшується ризик того, що твою інформацію вкрадуть, змінять чи використають проти тебе. Еволюція комп'ютерних систем дає змогу людям вигадувати все нові і нові способи для шахрайства, крадіжки і подібних видів протиправних дій.

Але для того, щоб безпечно користуватися інтернет-ресурсами вигадали термін інформаційна безпека – стан захищеності системи де досягнуті три основні критерії: конфіденційність, цілісність і доступність.

Через актуальність загроз цілісності [24] і конфіденційності інформації від компаній вимагається уважне ставлення до завдання її захисту. 20 років тому безпека інформації вирішувалася за допомогою засобів криптографічного захисту, встановлення міжмережевих екранів (файрволів), розмежування доступу. Зараз цих технологій недостатньо, будь-яка інформація, що має велику цінність для приватних і особливо для державних та світових органів влади має бути надійно захищена.

На мою думку, захист інформації є питанням номер один не тільки для окремих підприємств, а й для нашої держави загалом. Існує безліч методів і засобів захисту інформації. Технічні, програмні і організаційні засоби є невід'ємною частиною структури будь-якого підприємства яке хоч трохи турбується про безпеку.

Інформаційні технології застосовуються для рішення задач забезпечення національної, військової і економічної безпеки. І не дивно,

що для вирішення питання захисту цих технологій вигадали такий термін як кібербезпека.

Поняття кібербезпеки [14] представляє собою сукупність методів, технологій і процесів, призначених для захисту цілісності мереж, програм і даних від цифрових атак. В сучасному світі майже кожна із організацій (державні структури, фінансові, комерційні, медичні) збирають, обробляють і зберігають величезні об’єми даних. Серед яких є і особисті дані(персональні), дані про організацію, підприємство. Тому, я вважаю, що питання захищеності мережі таких мереж вартує особливої уваги з боку підприємств або організацій. Оскільки не можливо уявити, що може трапитись з компанією якщо її внутрішню інформацію вкрадуть, знищать чи змінять. Для випадків попередження таких подій найчастіше використовують моніторинг мережі або моніторинг мережевого трафіку.

Питання кібербезпеки [12] будуть завжди актуальні, адже інструменти злому постійно розвиваються. "Стандартна практика, коли зловмисники на крок попереду, лише тому що першими виявляють вразливість і використовують її до того, як вона стане відома розробникам", - кажуть фахівці.

Питання захищеності мережі та мережевого трафіку буде описано в цій роботі. Дана тема не є новою, але є досить актуальною для українських державних, приватних підприємств та організацій.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Аналіз мережевих атак

В сучасному світі, насиченому інформаційними технологіями, безпека мережі має досить високе значення. Компанії потребують забезпечення доступу пристроїв своїх співробітників до інформаційних ресурсів в будь-який час, для того і сучасна політика забезпечення інформаційної безпеки повинна враховувати ряд таких факторів як покращення надійності мережі, ефективне управління безпекою і захист від постійної еволюцій загроз та нових способів мережевих атак [19].

Згідно ДСТУ 3396.1-96 захист інформації – це комплекс заходів, які призначені для безпечного зберігання та захисту інформації від небажаних користувачів [12]. Безпека комерційних таємниць і обороту документів є головним завданням в захисті інформації. Інформацію охороняють методом технічного, програмного управління, передачею секретних даних і доступом.

Такі засоби майже в усіх підприємствах поділяють на:

- *фізичні* – механічні, електричні, електромеханічні, електронні, електронно-механічні та інші пристрої та системи, які функціонують автономно, створюючи різного роду перешкоди на шляху дестабілізуючих факторів.

- *технічні* – різноманітні електронні та електронно-механічні пристрої, які якимось чином мають схематично вбудовуватися в апаратуру системи обробки даних або з’єднуватися з нею спеціально для вирішення завдань захисту інформації [15].

У загальному випадку захист інформації технічними засобами забезпечується в наступних варіантах:

1. Джерело і носій інформації локалізовані в межах кордонів об'єкта захисту і забезпечена механічна перешкода від контакту з ними

зловмисника або дистанційного впливу на них полів його технічних засобів;

2. Співвідношення енергії носія і перешкод на вході приймача встановленого в каналі витоку таке, що зловмиснику не вдається зняти інформацію з носія разом з необхідною для її використання якістю;

3. Зловмисник не може виявити джерело або носій інформації;

4. Замість правдивої інформації зловмисник отримує неправдиву, яку він приймає як справжню [24].

- *програмні* – використовуються як спеціальні пакети програм або окремі програми, що включаються вже до встановленого програмного забезпечення для вирішення завдання захисту інформаційних потоків підприємства, організацій. Такі засоби захисту призначені спеціально для захисту комп'ютерної інформації та побудовані через використання криптографічних засобів захисту.

- *організаційні* - організаційно-технічні та організаційно-правові заходи, здійснювані при експлуатації або в процесі створення інформаційної системи з метою забезпечення захисту інформації. За своїм змістом такі організаційні заходи можна розподілити на такі групи:

1) заходи, здійснювані при створенні інформаційної системи (ІС);

2) заходи, що здійснюються в процесі експлуатації ІС: організація пропускного режиму, організація автоматизованої обробки інформаційних потоків, організація роботи працівників позмінно, розподіл повноважень та розмежування доступу (видача окремих паролів, профілів);

3) заходи спільного характеру: облік вимог захисту при підборі і підготовці кадрів, організація планових і превентивних перевірок механізму захисту, планування заходів щодо захисту інформації.

Але на мою думку, таких заходів не достатньо для забезпечення повної захищеності. Досить велика кількість атак і нападів відбувається саме на мережу. Гарним прикладом не здатності реагувати на атаки є

напад в 2017 році вірусом «Петя» або «Petya». Через це в Україні на значний проміжок часу були заблоковані величезна кількість підприємств та державних структур таких як: Ощадбанк, УкрЗалізниця, УкрПошта та інші., вихід з звичного режиму яких спричинило значну паніку серед населення. Головною причиною цього є неосвіченість людей та непідготовленість підприємств, чим і користуються зловмисники.

Тому, досить доречно наголосити, що на наших українських підприємствах не досить сильно приділяють увагу мережевій безпеці.

Для багатьох організацій проблема забезпечення мережевої безпеки стає все важчою і важчою, так як сьогоднішні «мобільні» працівники, які використовують особисті смартфони, ноутбуки і планшети для роботи, приносять нові потенціальні проблеми. При цьому, зловмисники також розвивають свої технології і створюють нові кіберзагрози.

Хочу розглянути основні мережеві загрози і атаки, які є небезпечними для комп'ютерної мережі організації, компанії, підприємства (таблиця 1.1).

Першим ми згадаємо сніфери пакетів (англ. – нюхати) – програми, які дозволяють проводити аналіз всього мережевого трафіку, який проходить через мережну карту комп'ютера або будь-якого іншого мережного пристрою користувача. Основну загрозу представляє можливість для зловмисника з допомогою цієї програми перехопити дані облікових записів – логіни, паролі, які передаються без видимого шифрування.

Далі IP spoofing, ARP spoofing – як вид хакерської атаки, який полягає в підміні IP- або ARP-адреси довіреного пристрою, яка має на меті обман систем безпеки. Даний вид атаки дозволяє отримати доступ до багатьох сегментів мережі, перехоплювати і підміняти мережний трафік, виконувати або запускати шкідливі команди на серверних сервісах. Так як протокол TCP має вбудовані засоби захисту під час встановлення захисту

(потрійне рукостискання), то основну загрозу даний вид вразливості представляє для UDP трафіку – основного протоколу передачі даних і відео для телевізійних мереж.

DoS і DDoS-атаки (від англ. – відмова служби) – найбільш часта в користуванні хакерами і найбільш проста для них атака. Проти даного типу атак до сих під не знайдено повністю захищеного рішення. Даний тип атаки не має на меті отримати доступ до мережі або на крадіжку інформації. Простіше кажучи, атака DoS робить мережу або окремі її сервіси недоступною для подальшого користування. При цьому, не використовуються проріхи систем безпеки або вразливості програмних пакетів, а використовуються загальні слабкості системної архітектури. Такий вид атак перевантажує серверні додатки і мережні пристрої величезною кількість з'єднань. Якщо така атака проводиться через велику кількість пристроїв, в тому числі через бот-мережу, то така атака називається розподіленою - DDoS. Такий вид, зазвичай, потребує величезної кількості пристроїв і рідко являється загрозою від внутрішньої мережі.

Атаки Man-in-the-Middle – атаки, при яких пристрій зловмисника підміняє мережні пакети, які передаються між двома адресатами таким чином, що ні однин з них навіть не здогадується про присутність посередника.

Парольні атаки – атаки, при якій використовуються велика кількість спроб доступу до ресурсів, з метою підібрати «грубою силою» (brute force) зв'язки логінів і паролів. Результатом атаки може слугувати користувацький або навіть привілейований доступ до системи.

Атаки на рівні додатків – різноманітні атаки, які використовують вразливості програмних пакетів, наприклад, SQL- , PHP- , XPath- ін'єкції, XSS атаки. Даний вид атаки дозволяє отримувати доступ до даних, викликати їх втрати, псувати або підміняти їх.

Підміна DHCP сервера – атаки з допомогою підміни DHCP-сервера, який видає IP адреси пристроям на основі їх Mac-адреси. Можна сказати, що це один з різновидів атаки Man-in-the-Middle яка в подальшому дозволяє емулювати нешифрований трафік з пристроїв, який отриманий з підмінених IP адрес.

DHCP starvation (голодування) – атака, при якій під'єднаний до мережі пристрій робить запит у DHCP-сервера для своєї MAC-адреси нову IP-адресу. Отримавши його на деякий час, він програмно змінює свою MAC адресу і відсилає новий запит. Тим самим роблячи пули адрес на DHCP сервері переповненими, що може викликати проблеми в роботі мережі. В комплексі з атакою підміни дозволяє переключити всіх клієнтів підприємства на отримування адрес тільки з DHCP сервера зловмисника.

Переповнення CAM-таблиць – атака на комутатор, при якому на одному пристрої генерується велика кількість MAC адрес і провокуються запити на ці адреси. В результаті комутатор, змушений заносити всі ці адреси в свою CAM-таблицю переповнюючи її. З такою переповненою таблицею комутатор змушений кожний новий запит перенаправляти на всі порти, дозволяє зловмиснику перехопити весь мережевий трафік комутатора на будь-якому з портів.

VLAN hopping – атака, при якій зловмисник може змусити порт комутатора працювати з його мережевим пристроєм в режимі trunk. Цей режим дає можливість доступу до всіх віртуальних мереж комутатора, а не тільки конкретного для цього комутатора. Така атака працює лише на порти комутатора, налаштованих в режим динамічного призначення режиму порта. Але така загроза не є дуже серйозною, так як зловмисник не може отримати відповідь в свою сторону.

MAC spoofing – так називають атаку з використанням підміни MAC-адреси пристрою і реєстрацій для подальшого його постійного

користування. Дозволяє перехоплювати частину фреймів, призначених для іншого адресата MAC.

Віруси, трояни та шпигунські програми – шкідливі програми, які зловмисники розповсюджують через мережу інтернет або змінні носії, що можуть бути джерелом для вище перерахованих загроз. Використовуючи програмні вразливості та недостатньо налаштовані системи безпеки комп'ютерів вони швидко «заражають» комп'ютери та пристрої по всій локальній комп'ютерній мережі. Такі програми можуть знаходитись в «режимі сну» досить великий проміжок часу і активуватись після здійснення якоїсь конкретної дії користувача. Зазвичай з їх допомогою викрадаються дані, створюються бот-мережі для DDoS атак, руйнуються великі і життєво необхідні інфраструктури і дані.

Якщо умовно розділити ці загрози безпеки мережі по можливості виникненню в середині або ззовні мережі, та вірогідності находження джерела атаки в внутрішній мережі чи ззовні, то ми отримаємо таблицю 1.

В більшості випадках джерелом загрози слугує внутрішня мережа організації. Це пов'язано з тим, що зазвичай периметр мережі є найбільш захищеною частиною цієї мережі. В найбільш рідких випадках, атака, виконана через загальнодоступні мережі загрожує мережним ресурсам лише для внутрішньої частини мережі. Найбільшу загрозу несуть атаки, зроблені всередині мережі, в тому числі з допомогою вірусних програм, принесених на особистих пристроях працівників і підключених до корпоративних мереж без злого наміру.

Тому, після обґрунтування вище перерахованого я можу зробити висновок, що потрібно досліджувати мережний трафік, щоб на майбутнє попередити подібні загрози і можливі їх наслідки.

Таблиця 1.1 – Класифікація загроз безпеки мережі по їх джерелу та цілі

Ціль	Джерело, що знаходиться ззовні мережі	Джерело, що знаходиться всередині мережі
Ресурси зовнішньої мережі	DoS, DDoS Man-in-the-Middle Парольні атаки На рівні додатків	Сніфери пакетів DoS, DDoS IP spoofing Парольні атаки Віруси, трояни і шпигунські програми
Ресурси внутрішньої мережі	Атаки на рівні додатків	Сніфери пакетів IP spoofing ARP spoofing Підміна DHCP сервера DHCP starvation Переповнення CAM-таблиць VLAN hopping MAC spoofing Man-in-the-Middle Парольні атаки Віруси і шпигунське ПЗ

1.2 Аналіз врахування трафіку для захисту інформації в державних установах

Приділимо увагу аналізу трафіку, що використовується в державних установах.

Мережевий трафік, що також називають трафіком мережі або трафіком даних, відноситься до даних, які переміщається через мережу в

будь-який конкретний момент часу. Мережеві дані складаються з пакетів – найменших фундаментальних одиниць даних, переданих по мережі. Дані про трафік мережі розбиваються на ці пакети для передачі і збираються в пункті призначення. Пакети складаються з корисного навантаження (the raw data) і заголовків (the metadata), що містять такі відомості, як IP адреси, його походження і призначення [9].

Існує чотири широкі категорії мережевого трафіку:

- Інтенсивний або зайнятий трафік, при якому використовується велика смуга пропускання.
- Трафік у режимі реального часу, який відноситься до пропускної здатності, витраченої протягом робочого часу.
- Інтерактивний трафік – трафік стикається з конкуренцією за пропускну здатність, що призводить до повільного часу відгуку, відбувається якщо пріоритети для трафіку та програм не встановлені
- Трафік чутливий до затримок, що також може призвести до поганого часу відгуку через конкуренцію за пропускну здатність

Найбільш часто застосовуються такі засоби моніторингу:

- аналізатори протоколів, або мережеві сніфери, дозволяють захоплювати трафік локальних мереж, представляти його в зручному для аналізу вигляді, але власне аналіз даних залишають адміністратору;

- маршрутизатори, що підтримують протокол NetFlow, збирають великі цифри даних про трафік глобальних мереж, передаючи його для аналізу програмних систем NetFlow, які автоматизують пошук атак і загроз;

- системи виявлення вторгнень (Intrusion Detection Systems, IDS) спеціалізуються на автоматичному розпізнаванні вторгнень і загроз в прослуховуванні трафіку локальних мереж;

- системи контролю трафіку і стану мережі адміністративного призначення.

Об’єктом, який містить інформацію про природу проблеми можна впевнено назвати мережевий трафік. Одним з найперших інструментів з відкритим кодом (open source), що допомагав у вирішенні цих задач був мережевий сніфер або мережевий аналізатор [10] Wireshark [11] створений одним із інженерів у 97 році минулого століття [13].

Стрімкий розвиток та поширення глобальних і локальних мереж змінило обчислювальні системи, які стали більш пов'язаними і менш захищеними від зловмисників, що володіють новими можливостями для своїх руйнівних цілей. Витрати на компенсацію збитків, завданих в результаті несанкціонованого доступу зловмисників до потокам даних в мережах постійно збільшуються. Автоматизація процесів обробки, зберігання та передачі інформації призводить до виникнення нових проблем, пов'язаних із забезпеченням її безпеки. В цей же час, сучасні обчислювальні системи стають все більш складними через динамічні зміни в конфігурації і програмному забезпеченні. Така ситуація створює практично необмежені можливості для зловмисників, які використовують програмні додатки і уразливості операційних систем для успішного проникнення в комп'ютерну систему. Складається стійка тенденція до збільшення кількості атак на обчислювальні системи і мережі. Технології та методи віддалених мережевих атак постійно вдосконалюються, і існуючі засоби захисту не дозволяють повністю припиняти зловмисний трафік. Ці обставини роблять розробку і впровадження нових методів та засобів захисту інформації в обчислювальних мережах дуже актуальними.

Задача класифікації мережевого трафіку

Завдання класифікації мережевого трафіку - це отримання на вхід певних характеристик мережевого трафіку і отримання на виході виду класу, до якого він належить. В якості вхідних характеристик можуть

прийматися дані пакетів, різні частотні характеристики. Як вихідні дані можуть бути ідентифікатори конкретних програм, які є джерелами генерації трафіку. Також можна класифікувати по ідентифікаторах виду трафіку. Класифікація трафіку необхідна для ідентифікації додатків і протоколів, які передаються по мережі.

Класифікація трафіку в відповідні класи вельми актуальна проблема, так як важлива для багатьох додатків, таких як контроль якості обслуговування (QoS), ціноутворення трафіку, планування використання ресурсів, виявлення шкідливих програм і вторгнень [18]. Через свою важливість, багато різних підходів були розроблені для вирішення цієї проблеми протягом останніх років, щоб пристосувати різноманітні зміни потреб різних сценаріїв застосування.

Методи аналізу мережевого трафіку [16]. На сьогоднішній день існує декілька способів для моніторингу і аналізу мережевого трафіку:

- З допомогою програм-аналізаторів (в тому числі з допомогою спеціальних протоколів в маршрутизаторах);
- Статистичні методи;
- Методи на основі нейронних мереж.

Самим простим і доступним способом аналізу мережевого трафіку являються *програми-аналізатори*. Програма-аналізатор або сніфер – програма або програмно-апаратне забезпечення, що використовується для захвату і подальшого аналізу захопленого трафіку або окремого сегменту мережі. В процесі захоплення всіх потоків, аналізатор захоплює і записує всі пакети, отримані від інтернет-трафіку. У випадку детального і більш інформативного аналізу відбувається декодування пакетів з зашифрованої форми представлення в таку, яку можна прочитати.

Є досить багато програм у вільному безкоштовному доступі, але через не дуже високу захищеність і вартість вони не підходять для великих компаній.

Статистичні методи. Існує декілька методів для аналізу трафіка допомогою математичних моделей. Серед них:

- Моделювання трафіку фрактальним броунівським рухом
- Аналіз з допомогою Марківської моделі
- Моделювання часових рядів

При будуванні моделі тимчасових рядів використовується експериментальна інформація, де потрібно менше припущень і більш адекватно відображається реальний об'єкт, тобто телекомунікаційна мережа. Даний метод є найбільш точним, оскільки в його основі лежить величезна кількість експериментальних даних.

Математична модель описує потік інформації в залежності від моменту Т. При статистичному аналізі часових потоків інформації необхідно здійснити виділення тренду, виділяти періодично складові коливання відносно тренду з деякою регулярністю, і проводити аналіз випадкового компоненту.

Математичний опис зазвичай включає в себе одну із подібних складових або суму деяких з них. Моделювання тренду може проводитись з допомогою гарно опрацьованих методів регресивного аналізу. Властивості і характеристики випадкової послідовності вивчаються з допомогою класичних методів математичної статистики і методів аналізу випадкових послідовностей.

Методи на основі нейронних мереж. Штучну Нейронну Мережу (ШНМ) можна назвати одним із великої кількості технологій створення інтелектуальних систем, що засновані на спробі імітації людського мислення, людського мозку, людської здатності приймати рішення. Існує велика кількість конфігурації нейронних мереж з різноманітними принципами функціонування. Для того, щоб реалізувати систему, яка зможе виявити атаки и аномалії, необхідно використовувати повнозв'язну нейронну мережу. Така нейронна мережа включає в себе декілька шарів,

де кожен нейрон довільного шару пов’язаний з іншими нейронами попереднього (прихованого) шару. На рисунку 1.1 представлена структура багатошарового перцептрону.

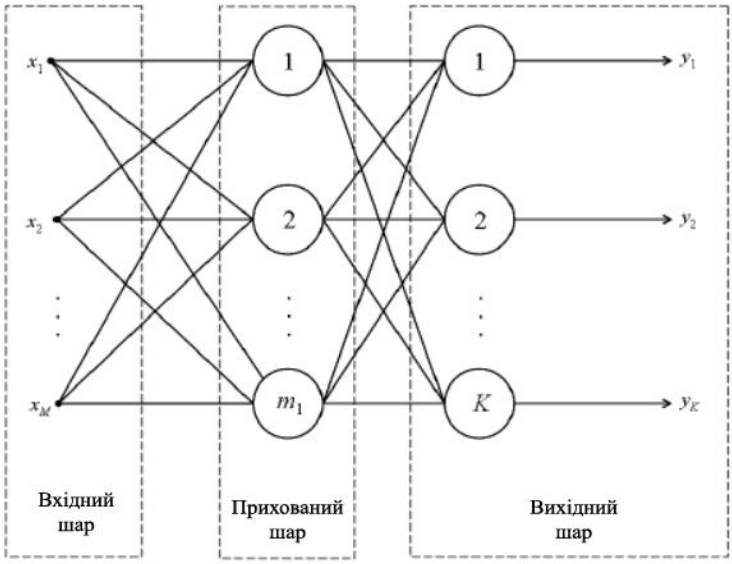


Рисунок 1.1 – Структура багатошарового перцептрону

Багатошаровий перцептрон має в собі три типи шару нейронів: вхідний, прихований і вихідний. Кожен нейрон має гладку нелінійну функцію активації. Багатошарові нелінійні нейронні мережі дозволяють формувати більш складні зв’язки між вхідними нейронами та вихідними, в порівнянні з одношаровими лінійними. Доведено, що трьохшарова мережа нейронних зв’язків з одним шаром, тільки прихованим, може бути навчена апроксимувати з довільно точністю будь-яку функцію, що є неперервною.

Також не буде зайвим згадати, що нейронні мережі застосовуються у багатьох сферах кібернетичної безпеки.

Прикладами будуть слугувати:

1. В системі виявлення та запобігання проникненням (IDS / IPS). Глибоке навчання, згорткові нейронні мережі та періодичні нейронні мережі (RNN) можуть бути застосовані для створення розумніших систем ID / IP, аналізуючи трафік з кращою точністю, зменшуючи кількість

помилкових сповіщень та допомагаючи командам безпеки диференціювати погану та хорошу мережеву діяльність.

2. В роботі з шкідливим програмним забезпеченням (ПЗ). Алгоритми глибокого навчання здатні виявляти звичайні та просунуті загрози не покладаючись на запам'ятовування відомих підписів та загальних моделей атак. Натомість вони вивчають систему і можуть розпізнавати підозрілу діяльність, яка може свідчити про наявність поганих факторів або шкідливого програмного забезпечення.

3. В системі виявлення спаму та соціальній інженерії. Обробка природних мов (NLP), техніка глибокого навчання, може допомогти вам легко виявляти спам та інші форми соціальної інженерії та боротися з нею. НЛП вивчає звичайні форми спілкування та мовні моделі і використовує різні статистичні моделі для виявлення та блокування спаму.

4. При аналізі мережевого трафіку. Поглиблене вивчення ANN показує багатообіцяючі результати в аналізі мережевого трафіку HTTPS для пошуку шкідливих дій.

5. Для аналізу підозрілої поведінки користувачів. Відстеження та аналіз діяльності та поведінки користувачів є важливою практикою безпеки для будь-якої організації. Це набагато складніше, ніж розпізнавання традиційних шкідливих дій проти мереж, оскільки воно обходить заходи безпеки і часто не піднімає жодних прапорів та попереджень. Наприклад, коли виникають інсайдерські загрози (напад працівників компанії), і працівники використовують законний доступ зі зловмисними намірами, проникаючи в систему НЕ ззовні, що робить багато інструментів кіберзахисту марними проти таких атак.

Після проведення аналізу, я думаю, що коли на потрібно обрати метод для аналізу мережі або трафіку, дуже важливо зупинитися на таких факторах, як об’єм аналізованого трафіку, надійність і інформативність обраного методу, доступність. Хочеться наголосити, що якщо необхідно

проаналізувати невеликий трафік, то для цього достатньо буде і безкоштовних програм-сніферів. Але якщо йдеться про великі об’єми даних, то логічніше буде застосувати статистичні методи або нейронну мережу. Крім того, аналіз з допомогою нейронних мереж – досить перспективний і дуже стрімко розвивається, також сьогодні досить гарно себе рекомендує, але на жаль методи і вартість таких систем не дозволяють використовувати їх всюди.

2 ІСНУЮЧІ РІШЕННЯ

2.1Програмне забезпечення щодо аналізу мережевого трафіку

Кожне підприємство обирає власні способи для вирішення проблем захисту мережевого трафіку. Одним з таких способів являється підбір програмного забезпечення. Більшість організацій і установ почали активний пошук такого захисту, який задовольнив би їх потребу в аналізі трафіку та захисті його мережі.

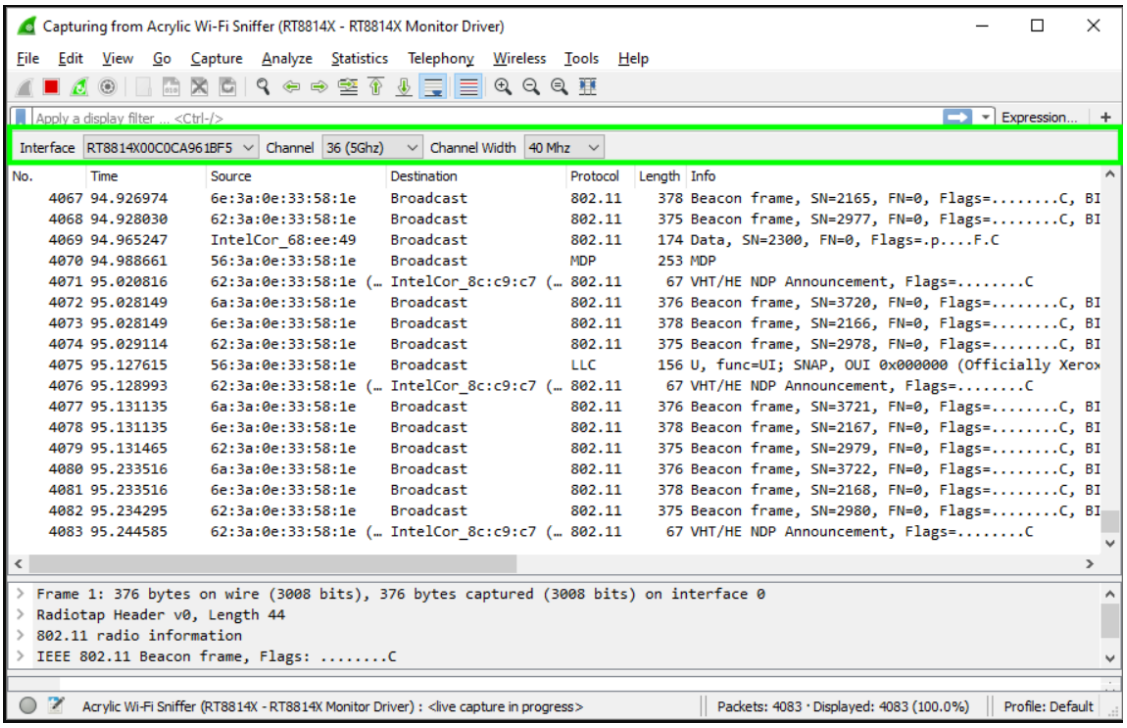


Рисунок 2.1 – Інтерфейс програми WireShark

Одним з таких прикладів є модуль розроблений на початку 2000-х років WireShark. Wireshark став вже певним стандартом в галузі мережевого аналізу (рисунок 2.1).

Досить сильно набув популярності аналізатор трафіку NetFlow (рисунок 2.2). NetFlow - це мережевий протокол, розроблений Cisco для збору інформації про IP-трафік та моніторингу потоку мережі. Для

ефективного моніторингу NetFlow пристрій, що працює як експортер потоків, збирає пакети даних у потоки та надсилає записи потоку на один або кілька серверів збору NetFlow.

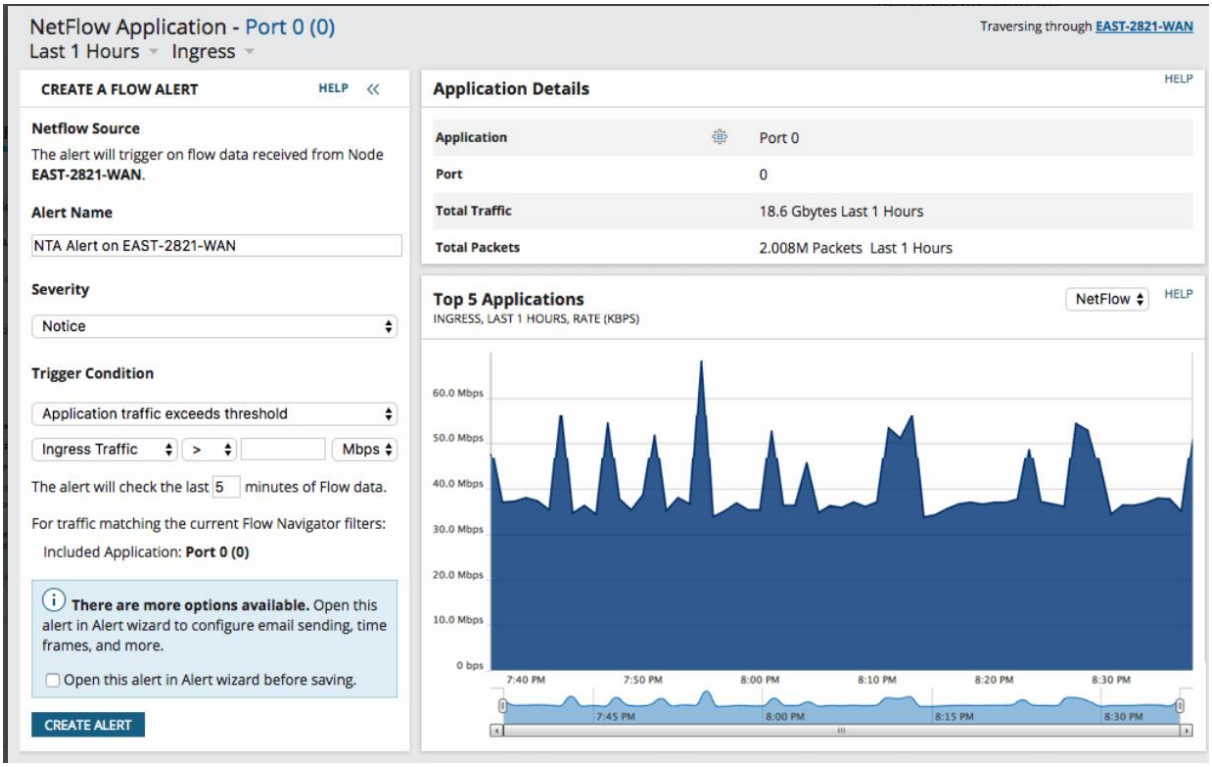


Рисунок 2.2 – Інтерфейс аналізатора трафіку NetFlow

Архітектура системи NetFlow будується на сенсорі, колекторі і аналізаторі:

- сенсор збирає статистику трафіку, що проходить через нього.
- колектор здійснює збір інформації від сенсорів. Отримані дані він скидає в файл для подальшої обробки. Різні колектори зберігають дані в різних форматах.

аналізатор, або система обробки, зчитує ці файли і генерує звіти у формі, більш зручної для людини. Ця система повинна бути сумісна з форматом даних, що надаються колектором. У сучасних системах колектор і аналізатор часто об'єднані в одну систему

Далі хочеться згадати аналізатор ManageEngine OpManager Plus. ManageEngine OpManager Plus – універсальний інструмент моніторингу IT-інфраструктури (рисунок 2.3). Він пропонує широкий спектр можливостей моніторингу, включаючи моніторинг стану пристроїв та аналіз потоку трафіку. Цей інструмент аналізу мережевого трафіку має систему фіксації потоку трафіку, здатну взаємодіяти з мережевими пристроями через AppFlow, IPFIX, J-Flow, NetStream, NetFlow та sFlow. OpManager Plus та відображає показники мережевого трафіку в реальному часі на динамічній інформаційній панелі, а пакети, захоплені системою, зберігаються у файлах для подальших цілей аналізу.

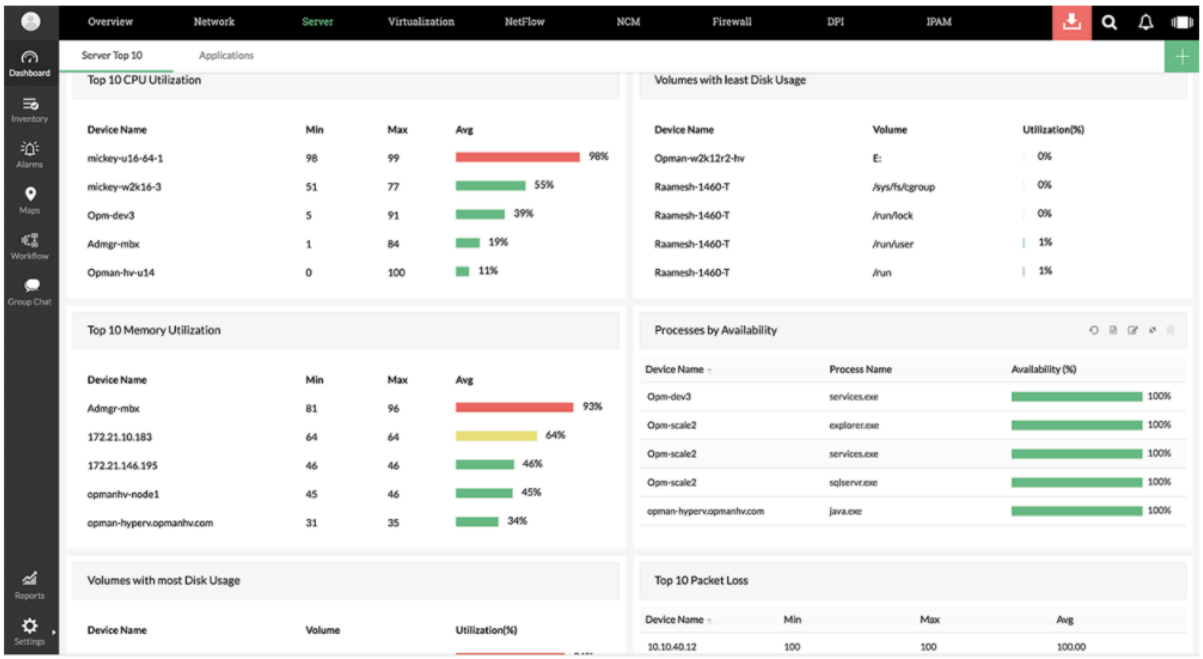


Рисунок 2.3 – Інтерфейс аналізатора ManageEngine OpManager Plus

Він використовує систему NBAR, який дозволяє розпізнавання мережевих додатків і є механізмом, що використовується деякими маршрутизаторами та комутаторами Cisco для розпізнавання потоку даних, перевіряючи деякі відправлені пакети. Це програмне забезпечення для аналізу мережевих потоків також включає допомогу у прогнозуванні для планування пропускної здатності [8].

Крім того, через простоту в використанні, підприємства використовують програмне забезпечення для глибокої перевірки пакетів NetFort (рисунок 2.4). Використовується воно для моніторингу, аналізу та звітування про цілий ряд інфраструктурних видів діяльності, включаючи користувачів, додатки та мережі. Це пасивне програмне забезпечення для аналізу мережевого трафіку, яке не впливає на продуктивність мережі. LANGuardian перевіряє вміст пакетів трафіку та заголовки та надає детальну та надійну інформацію про додаток та стан трафіку мережі.

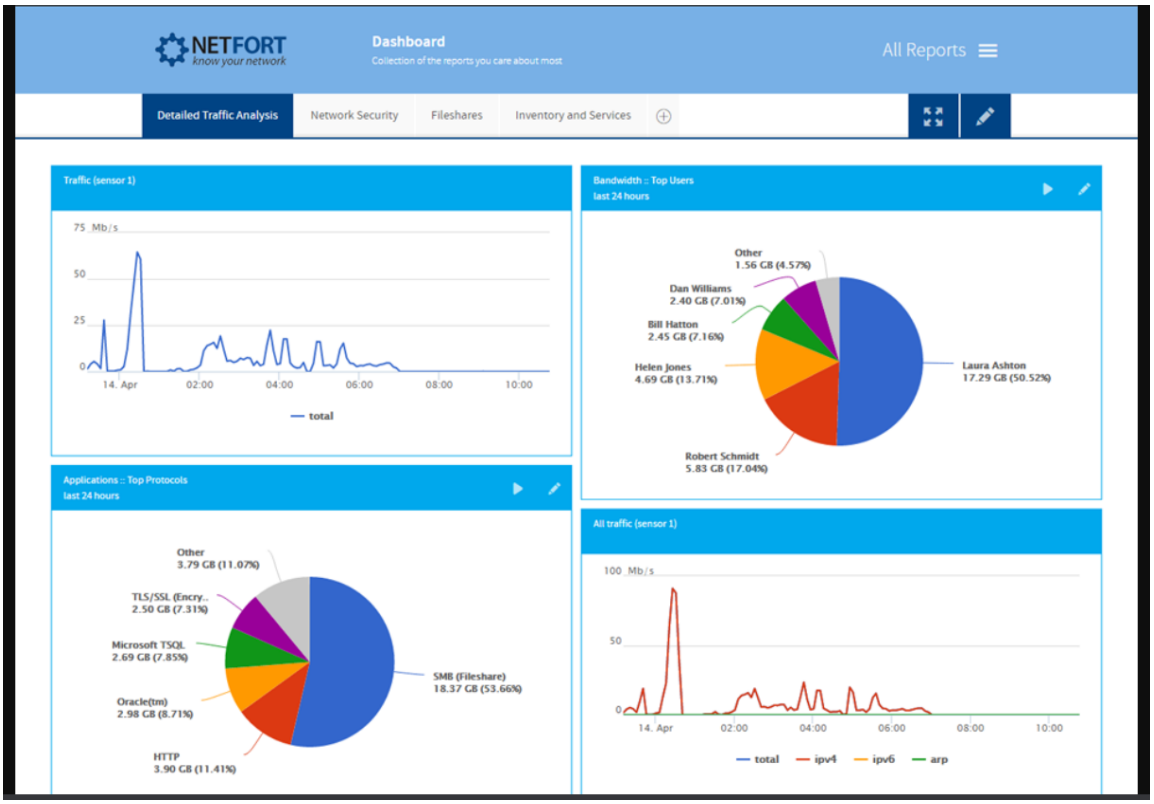


Рисунок 2.4 – Інтерфейс аналізатора NetFort

Аналіз мережевого трафіку використовує мережеві комунікації та їх протоколи для виявлення, ідентифікації та аналізу загроз кібербезпеки та потенційних операційних проблем. Вони пропонують робити це в режимі реального часу, використовуючи мережеві дані та аналітику з дротовою швидкістю, щоб задовольнити запити цифрового бізнесу та отримати

перевагу щодо скорочення середнього часу на виявлення (MTTD) / середнього часу на відповідь (MTTR).

Безперервна видимість, виявлення І здатність реагувати на мережевий трафік стає все більш важливою для захисту корпоративних активів у режимі реального часу, особливо через збільшення обсягу та швидкості передачі даних, що загрожує традиційним аналізом журналів та механізмом оповіщення, керованими групами SOC / NOC. Оскільки мережа є джерелом та механізмом транспортування підозрілого / зловмисного трафіку, чому б не використати нові технології та методи для не тільки виявлення та управління аномаліями, а й здійснення дій у реальному часі? Одним із методів є використання підходу з нульовою довірою за допомогою інспекції мережевого трафіку в режимі реального часу для аналізу та виправлення його в реальному часі.

Крім програмних існують і апаратні засоби аналізу мережевого трафіку. Хочеться навести приклад багатофункціональний тестер MTX150 від VeEX. Такий тестер - це повністю інтегроване самодостатнє рішення з тестування OTN, SDH, SONET, PDH, DSn, Ethernet, SyncE, Mobile Backhaul, і Fibre Channel (SAN). Цей міцний понад портативний тестер все-в-одному може бути налаштований з різними інтерфейсами і технологіями, необхідними для інженерів, які виробляють інсталяцію, повірку, обслуговування та пошук несправностей в мережевих з'єднаннях і сервісах комунікаційних мереж рівнів Transport, Metro, Access в тому числі застарілих різновидів [14].

Також крім тестерів варто згадати апаратний сервер Flowmon Networks [25]. Колектор випускається як у вигляді апаратного сервера, так і у вигляді віртуальної машини (VMware, Hyper-V, KVM) (рисунок 2.5). До речі, апаратна платформа реалізована на кастомізованих серверах DELL, що автоматично знімає велику частину питань з гарантією і RMA. Власною апаратною складовою є хіба що FPGA-плата захоплення трафіку,

розроблена дочірньою компанією Flowmon, що дозволяє здійснювати моніторинг на швидкостях до 100 Gbps.

Рисунок 2.5 – Зовнішній вигляд аналізатора Flowmon Collector

Центральною частиною всієї «екосистеми» Flowmon Networks є Collector, який одержує трафік з існуючого мережевого устаткування або власних зондів (Probe). Але для Enterprise рішення надавати функціонал виключно для моніторингу мережевого трафіку було б занадто просто. Це вміють робити і Open Source рішення, нехай і не з такою продуктивністю. Цінністю Flowmon є додаткові модулі, що розширюють базовий функціонал:

- модуль Anomaly Detection Security - виявлення аномальної мережевої активності, включаючи атаки «нульового дня», на підставі евристичного аналізу трафіку і типового мережевого профілю;
- модуль Traffic Recorder - запис фрагментів мережевого трафіку по набору визначених правил або по тригеру з модуля ADS, для подальшого траблшутінга і / або розслідування інцидентів ІБ;
- модуль DDoS Protection - захист периметра мережі від волюметричних атак відмови в обслуговуванні DoS / DDoS, в тому числі атак на додатки (OSI L3 / L4 / L7).

За результатами дослідження декількох основних програмних продуктів мною було виявлено, що дуже мала їх кількість використовує

аналіз мережі з допомогою інтелектуальних систем, тому я вважаю, що потрібно ближче вивчити питання інтелектуального захисту мережі.

2.2 Підсистеми захисту мережного трафіку

Для того, щоб вірно побудувати нашу підсистему захисту мережного трафіку визначимо які елементи будуть до нього входити (рисунок 2.6).

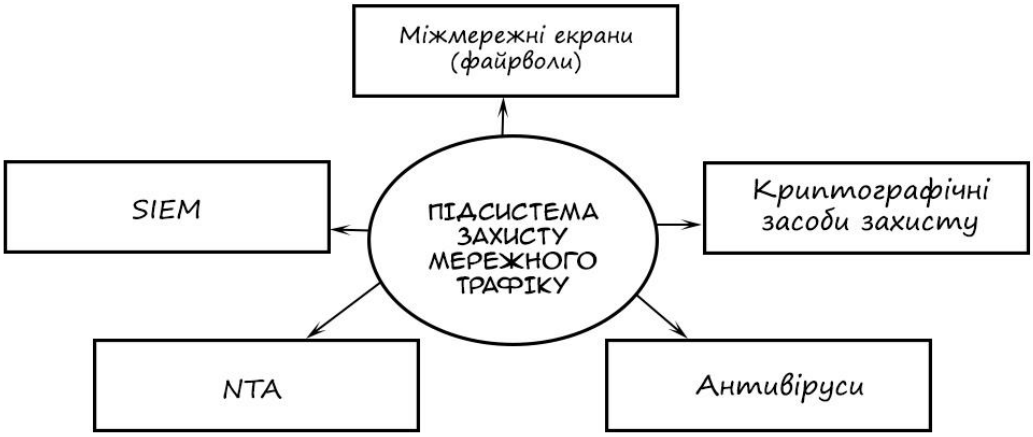


Рисунок 2.6 – Складові підсистеми захисту мережного трафіку

Невід'ємним елементом захисту мережі великої організації від вторгнення зловмисників є корпоративний *міжмережевий екран* (МЕ). Існує десяток різних пропозицій на ринку, що представлені десятками компаній, готових надати рішення для будь-яких середовищ: настільних систем, малого і домашнього офісу (SOHO), середнього і малого бізнесу, телекомунікаційних компаній.

Тому для прийняття правильного рішення про вибір брандмауера необхідно розуміння потреб бізнесу в забезпеченні мережевої безпеки і принципів дії цих продуктів.

Брандмауер (firewall, брандмауер) - це комплекс апаратних і / або програмних засобів, призначений для контролю і фільтрації трафіку, що

проходить через нього відповідно до заданих правил. Основним завданням цього класу продуктів є захист комп'ютерних мереж та вузлів від несанкціонованого доступу.

У загальному випадку, міжмережевий екран використовує один або кілька наборів правил для перевірки мережевих пакетів вхідного і / або вихідного трафіку. Правила брандмауера можуть перевіряти одну або більше характеристик пакетів, включаючи тип протоколу, адресу хоста, джерело, порт. Існує два основні способи створення наборів правил: «включаючі» і «виключаючі». Правила, створені першим способом, дозволяють проходження трафіку тільки за правилами і блокують все інше. Правила на основі виключаючого способу, навпаки, пропускають весь трафік, крім забороненого. Включаючі міжмережеві екрани зазвичай більш безпечні, ніж виключаючі, оскільки вони суттєво зменшують ризик пропуску фаєрволом небажаного трафіку.

Використання міжмережевих екранів може бути ефективно при вирішенні наступних завдань:

- Захист і ізоляція додатків, сервісів і пристроїв у внутрішній мережі від небажаного трафіку, що приходить з інтернету (поділ мереж);
- Обмеження або заборона доступу до сервісів мережі для певних пристроїв або користувачів;
- Підтримка перетворення мережевих адрес, що дозволяє використовувати у внутрішній мережі приватні IP-адреси або автоматично привласнюються публічні адреси.

Антивіруси

Комп'ютерні віруси залишаються в даний час найбільш актуальною проблемою інформаційної безпеки корпоративних систем. Через те, що більша частина вірусів розповсюджується через електронну пошту або вірусні пристрої, то міжмережні екрани не є досить ефективними.

Одним з методів, що застосовуються системними адміністраторами, поряд з використанням антивірусного програмного забезпечення, є фільтрація повідомлень, що містять вкладення певних форматів (найчастіше, виконувані додатки) [20].

Сучасні антивірусні програми, при всій їх різноманітності, використовують лише два принципово різні методи виявлення шкідливих програм:

- Пошук по сигнатурам;
- Евристичний аналіз.

Засоби криптографічного захисту інформації досить давно і широко використовуються в складі популярних мережевих технологій, таких як віртуальні приватні мережі (VPN) або Secure Shell (SSH). Однак з метою безпосереднього захисту особистої або корпоративної інформації застосування таких рішень до сих пір дуже обмежено. Так, приватне і ділове листування в більшості випадків ведеться відкрито, шифрування файлів і дисків теж мало поширене. У той же час, шифрування даних - це один з головних і найбільш надійних способів запобігання несанкціонованому доступу до інформації.

Мабуть, найширша сфера потенційного застосування криптографічних засобів - розмежування доступу до конфіденційної інформації і / або приховування існування такої інформації від нелегітимних користувачів. У масштабі корпоративної мережі ця задача досить успішно вирішується засобами AAA (автентифікація, авторизація та адміністрування). Однак при захисті локальних пристроїв вони найчастіше неефективні. Особливо гострою ця проблема постає у зв'язку зі збільшенням числа мобільних користувачів.

Рішення класу *Security Information and Event Management* (SIEM) призначені для моніторингу подій, що надходять в різних інформаційних

системах і додатках [22]. ІБ-рішення даного класу дозволяють виконати наступні завдання:

- збір та аналіз великих обсягів подій безпеки;
- моніторинг поточного стану засобів захисту ІТ-інфраструктури;
- виявлення комп'ютерних інцидентів в режимі реального часу;
- отримання повної картини того, що відбувається в ІТ-інфраструктурі;
- виявлення та реагування на збої в роботі ІТ- та ІБ-систем;
- побудова карти мережі для прогнозування ланцюжків атак;
- отримання даних для аналізу і оцінки ризику в реальному часі;
- виконання окремих вимог і нормативних актів законодавства в сфері моніторингу подій інформаційної безпеки (ІБ).

Часто зловмисники проникають та закріплюються в інфраструктурі і при цьому залишаються непоміченими протягом тривалого часу. Метою таких проникнень є непомітне для адміністраторів і фахівців з ІБ компроментування (злиття даних) або підготовка та організація атак на критичні вузли інфраструктури зсередини. Для більш ефективного виявлення інцидентів і боротьби з подібними проникненнями в цільову систему важливо, щоб рішення класу SIEM дозволяли проводити ретроспективний аналіз подій, використовуючи найактуальніші відомості про загрози (фіди, індикатори компрометації, експертні правила кореляції).

Принцип роботи рішень класу SIEM полягає в зборі всіляких логів (подій) від різних пристроїв як на рівні програмного забезпечення, так і на рівні апаратних компонентів. Далі всі події приводяться до єдиного формату для подальшого аналізу. Сукупність подій (кореляція), пов'язаних з одним і тим же елементом інфраструктури, може свідчити про кібератаку.

SIEM дозволяє побачити більш повну картину того, що відбувається в IT-інфраструктурі і, крім того, проаналізувати мережеву досяжність тих чи інших вузлів. Завдяки ретельному аналізу і кореляції даних з різних джерел, SIEM виявляє інциденти там, де традиційні засоби виявлення, що працюють окремо, діють не завжди ефективно.

Системи аналізу мережевого трафіку (N-network T-traffic A-analysis)

Рішення класу *Network Traffic Analysis (NTA)* призначені для виявлення мережевих атак, перехоплення і аналізу мережевого трафіку. Даний клас систем допомагає виявити присутність зловмисників на ранній стадії атаки, оперативно локалізувати загрози, а також забезпечити контроль дотримання регламентів ІБ.

На відміну від стандартних мережевих аналізаторів (IDS / IPS), NTA-системи аналізують трафік не тільки на периметрі, але і в IT-інфраструктурі. Даний клас рішень має можливість зберігати сесії мережевого трафіку для формування доказової бази і передачі в правоохоронні органи та Державну систему виявлення для попередження і ліквідації наслідків комп'ютерних атак на інформаційні ресурси. Крім того, з появою актуальних сигнатур, рішення класу NTA повинні мати можливість запускати аналіз мережевого трафіку, який зберігається в архіві (ретроспективний аналіз).

Рішення класу NTA можуть бути додатковим джерелом мережевих подій для рішень класу SIEM в ситуаціях з виявленням складних націлених атак.

На практиці такі рішення дозволяють, наприклад, виявити підозрілу спробу підключення з неавторизованого вузла на контролер домену, потім проаналізувати історичні дані по мережевим активностях вузла і перевірити, чи не було інших подібних спроб. Якщо вони траплялися, то це буде говорити про цілеспрямовану атаку або, по крайній мірі, спробах злому.

3 ПРОЕКТУВАННЯ ПІДСИСТЕМИ ЗАХИСТУ ШЛЯХОМ ВПРОВАДЖЕННЯ ВЛАСНИХ РІШЕНЬ

Великі або малі підприємства чи організації не завжди мають змогу придбати на постійне користування програмне забезпечення яке їм подобається. Тому для вирішення питання захищеності мережного трафіку я пропоную кілька рішень:

1. Вибрати програмне забезпечення, що використовує вже готовий, надійний інтелектуальний аналіз
2. Додати модуль аналізу інтелектуального аналізу трафіку
3. Написати власну програму виявлення шкідливого трафіку використовуючи ті самі засоби інтелектуального аналізу

3.1 Програмне забезпечення «ExtraHop Reveal(x)»

ExtraHop Reveal (x) - аналізатор мережевого трафіку, який об'єднує автоматичне детектування з технологіями машинного навчання і аналізу поведінки. Використовується воно з метою виявлення аномалій в роботі мережі і подальшого виявлення атак [3]. ExtraHop Reveal (x) наділяє можливістю оцінювати ризик і впевнено реагувати на загрози: немає помилкових спрацьовувань або сигналів тривоги, є тільки глибокий контекст і висока точність розуміння того, що необхідно захищати і що має найбільше значення (рисунок 3.1). Ціна такого рішення для нашого аналізу трафіку становить близько 650000 фунтів стерлінгів або 855000 доларів США на 3 роки.

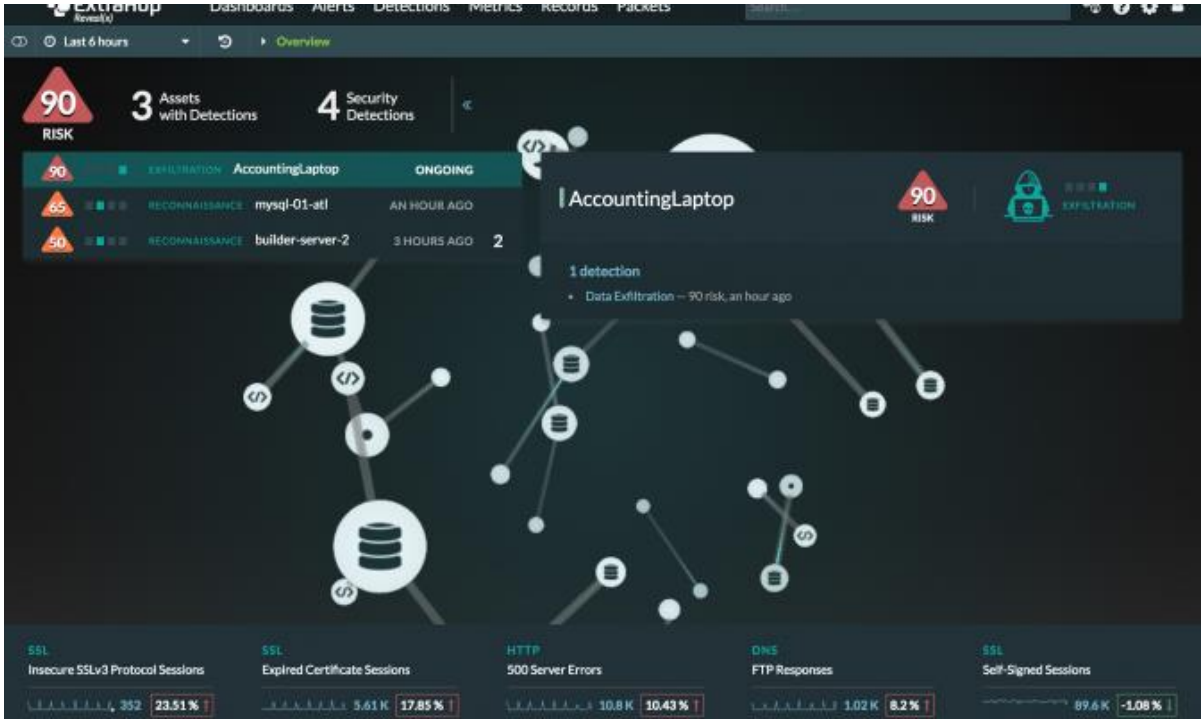


Рисунок 3.1 – Інтерфейс програмного забезпечення ExtraHop Reveal(x)

ExtraHop Reveal (x) використовує аналітичну обробку в режимі реального часу і машинне навчання по «провідним» даним (wire data) - найбільшому джерелу інформації, доступному в мережі, - для виявлення аномального поведінки, що впливає на критично важливі активи. Хмарна система попереджає про підозрілу поведінку і відображає активність по одному або декільком кроків в ланцюжку атаки: Command & Control, Reconnaissance, Exploitation, Lateral Movement або Action on Objective (рисунок 3.2).

Як видно з рисунку 2.3 програма виводить дані на екран через які ми можемо проаналізувати: завантаженість трафіку і з якого пристрою чи системи він надходить. При визначенні підозрілої активності він вказує конкретний пристрій, всі його параметри і все, що цей користувач встиг заподіяти (рисунок 3.3). Прикладом буде слугувати акаунт користувача AccountingLaptop.

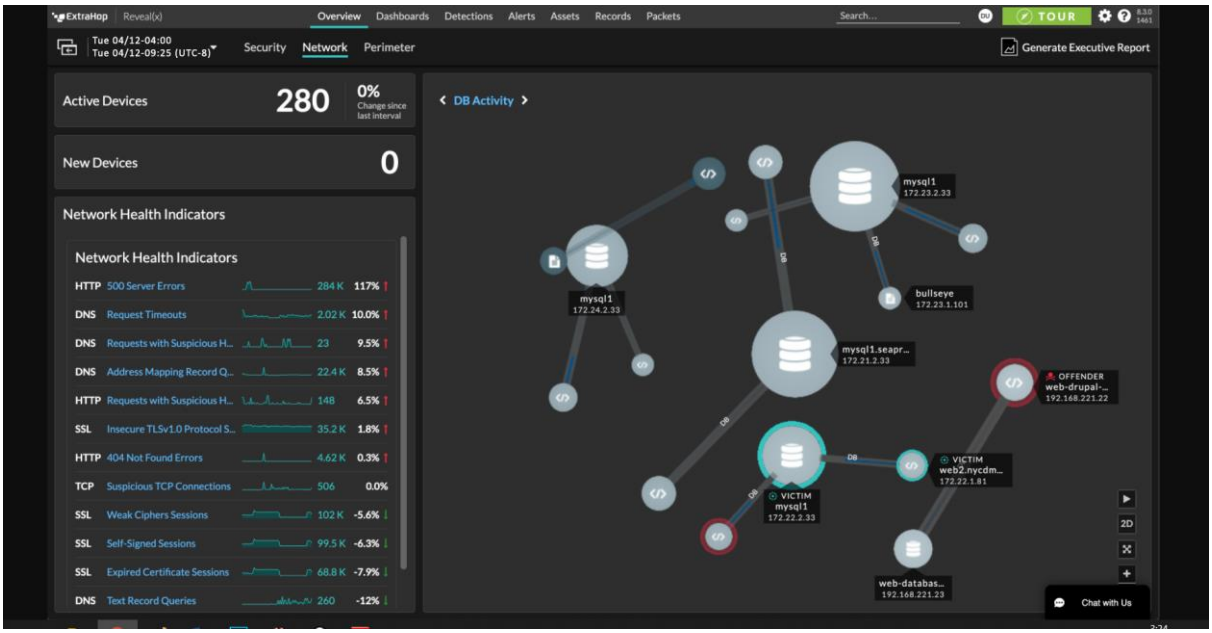


Рисунок 3.2 – Аналіз мережевої активності в ExtraHop Reveal (x)

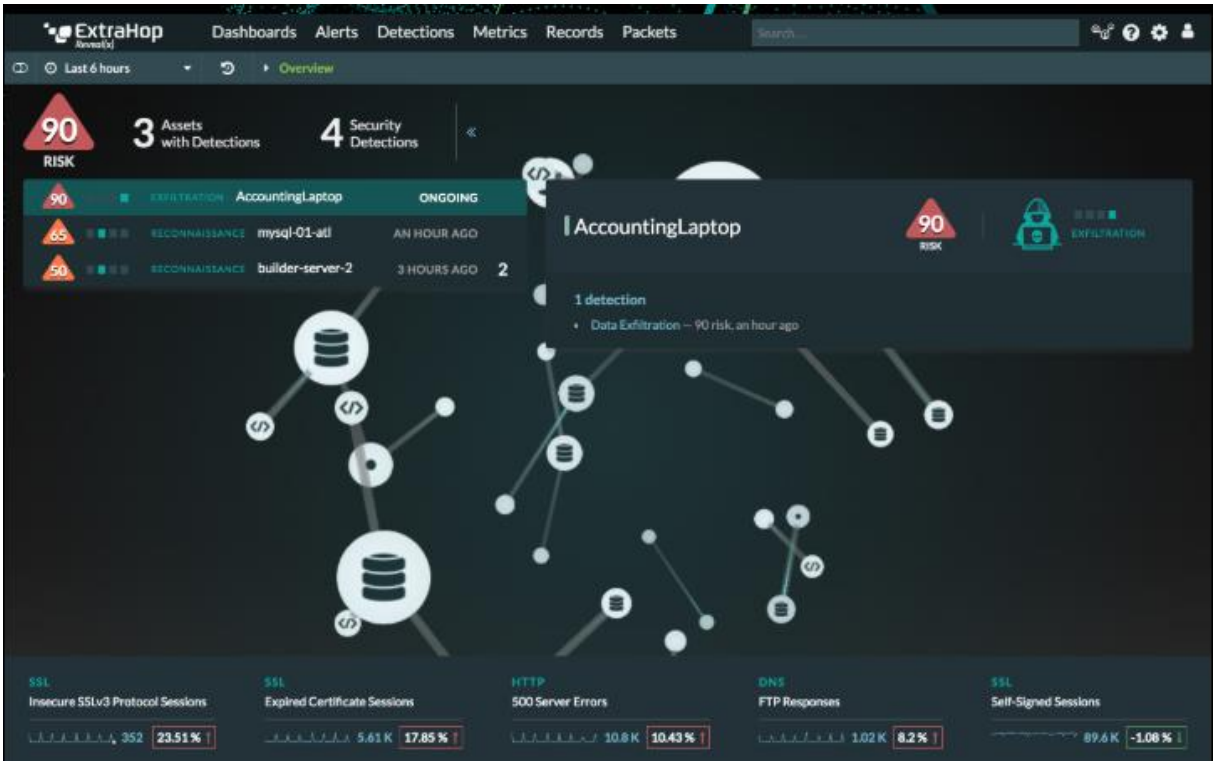


Рисунок 3.3 – Аналіз ризику користувача AccountingLaptop в ExtraHop Reveal (x)

Якщо виявлено підозрілу поведінку, Reveal (x) надасть повну інформацію про пакети, і може використовувати її для інтеграції з вашими

SIEM та інструментами управління, щоб ми могли карантинувати заражені системи, ініціювати обмеження та зосередитись на тому, що насправді потрібно до того, як відбудуться великі інциденти з мережевою безпекою (рисунок 3.4) [25].

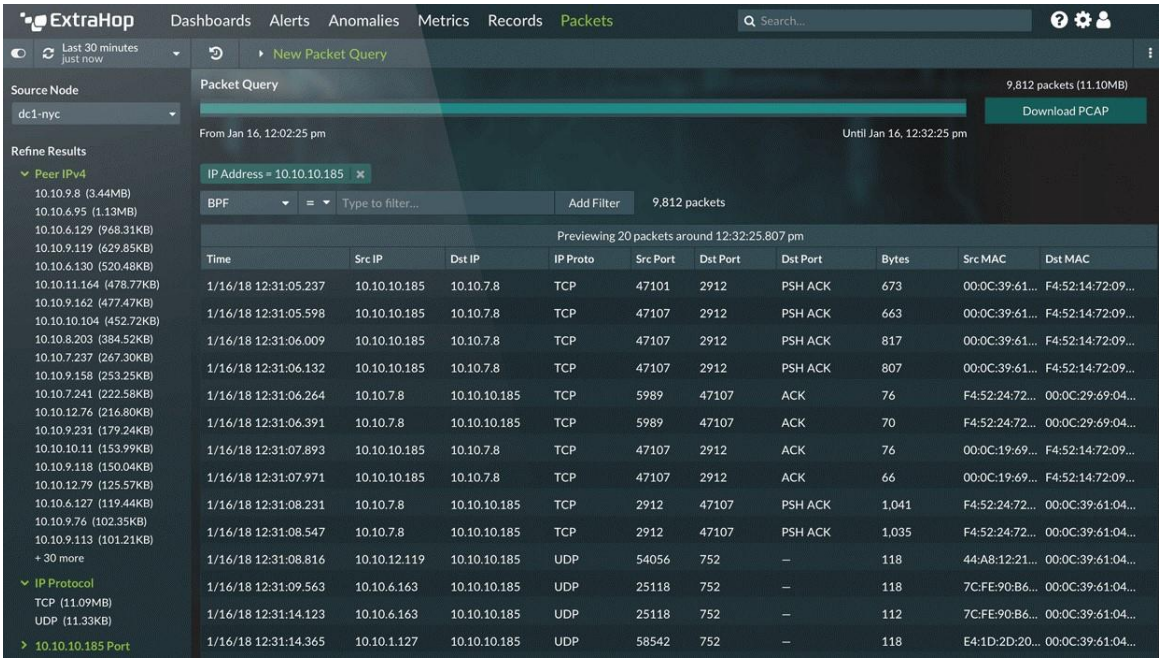


Рисунок 3.4 – Інформація про пакети користувача AccountingLaptop

Його ключові функції представлені розробником:

1.) Автоматичне знаходження і класифікація всієї техніки, яка пов'язана з мережею, включаючи BYOD, IoT і пристрої, які неможливо виміряти або зареєструвати.

2.) Легке фокусування на критично важливих активах, таких як бази даних, AAA- і DNS-сервери, ноутбуки керівників і адміністраторів, R & D-системи.

3.) Доступ до повного комплекту даних L2-L7 для всіх операцій, включаючи контекст і залежності між рівнями.

4.) Можливість аналізу більш ніж 50 протоколів ДЕШИФРОВАНОГО SSL- і PFS-трафіку.

Reveal (x) використовує підхід повного спектру виявлення загроз. За допомогою аналізу потоку в реальному часі зі швидкістю 100 Гбіт / с Reveal (x) миттєво виявляє нагальні загрози та CVE (Common Vulnerabilities and Exposures – база даних вразливостей) [25]. Потім він передає особливості поведінки та показники, захоплені з мережі, у хмарну систему ML, яка створює складні поведінкові моделі кожного пристрою та групи однолітків для виявлення прихованих, непомітних атак таким чином, що зловмисники не можуть уникнути власне розкриття. Reveal (x) застосовує мільйони моделей машинного навчання до 5000+ особливостей даних, отриманих із понад 4-х петабайт анонімізованої телеметрії загроз, зібраних з більш ніж 15 мільйонів пристроїв щодня, для ініціювання більш високої точності, що дозволяє командам охорони персоналу зосередитись на пошуку та вирішенні проблем пріоритетніших загроз.

Відповідно до звіту про загальний економічний вплив (TEI) за 2020 рік, проведеного Forrester Research, Reveal (x) зменшує час вирішення загрози на 84%. Reveal (x) безперервно збирає поведінковий зміст та криміналістичні дані в режимі реального часу, і ставить цю інформацію на передній план, крім того відправляє їх в центр «thread hunters».

Кожне виявлення містить вміст транзакцій, пов'язані з ними виявлення та доступ до повних розшифрованих пакетів, що використовують для криміналістики. Поки мережевий та хмарний трафік розшифровується він починає аналізуватися в режимі реального часу, Reveal (x) здатний розкривати повний спектр ризиків та усувати «сліпі зони» з повним його покриттям. Згідно з посібником Gartner Market для NDR, лише невелика група платформ має можливість власноруч розшифровувати трафік SSL / TLS, - ExtraHop можна вважати однією із них. Інші постачальники в першу чергу покладаються на методи аналізу зашифрованого трафіку, завдяки чому дані, які вони можуть

контролювати, є менш точними, а виявлення схильні до помилкових спрацьовувань.

Після аналізу вище перерахованого, я вирішив, що таке програмне забезпечення буде найбільш близьке до ідеального в моїй комплексній системі захисту. Але стає очевидним той факт, що не кожне українське підприємство може дозволити подібне програмне забезпечення. Лише великі компанії та організації які турбуються про свою мережеву безпеку повинні його використовувати через власну надійність та багатофункціональність. Тому, виходячи з цього, ми переходимо до наступного рішення.

3.2 Додавання модулю інтелектуального аналізу в систему IDS

Система виявлення вторгнень – це програмний чи апаратний засіб, призначений для виявлення вторгнень або шкідливих дій по відношенню до комп'ютерної системи. Алгоритми визначення загроз будуються на основі наперед заданих правил (сигнатур) або евристичного аналізу трафіку. Перші системи виявлення вторгнень використовували сигнатурний аналіз, який також називається аналізом на основі правил. Сучасний недолік таких систем показую непридатність захисту від нових або неврахованих раніше правил атак [2]. Через ці недоліки почали використовувати евристичний аналіз. В останні роки, для евристичного аналізу використовують нейронні мережі [6].

Штучна нейронна мережа (ШНМ) є одним з підходів технології створення інтелектуальних систем, заснованих на імітації поведінки людського мозку. Існує велика кількість різних конфігурацій нейронних мереж з різними принципами функціонування. Для реалізації системи виявлення атак будемо використовувати багатошарову повнозв'язну

нейронну мережу (багатошаровий перцептрон), в якій кожен нейрон довільного шару пов'язаний з усіма попередніми нейронами.

Трафік даних має певні характеристики. Ухвалення рішення про загрозу в трафіку здійснюється за певним законом на підставі наявності або відсутності в ньому сукупних ознак. Після отримання трафіку на вхід, система виділяє певні ознаки і передає їх для подальшої обробки. Всього було запропоновано виділити 48 ознак, в тому числі таких, як відсоток підключень до одного і того ж мережевого вузла, кількість схожих підключень з 100 останніх з'єднань, чи використовується http трафік, відсоток підключень до даного мережевого сервісу, кількість підключень до тієї ж служби з 100 останніх з'єднань, кількість підключень до того ж хосту з 100 останніх з'єднань, кількість переданих байт даних від джерела до місця призначення, кількість байт даних від адресата до джерела та інші.

Після перехоплення трафіку з нього виділяються необхідні для роботи ознаки і передаються на вхід нейронної мережі. Використання нейронної мережі дозволяє вирішити задачу класифікації трафіку на шкідливий (що містить загрозу) і не шкідливий.

Робота системи відбувається в двох режимах.

1. Режим навчання нейронної мережі.
2. Режим детектування загроз.

Навчання нейронної мережі вимагає наявності навчальних даних - трафіку, для якого заздалегідь відомо шкідливий він чи ні. «Навчальний» трафік надходить на перший шар нейронної мережі і за цими даними проводиться розрахунок початкових вагових коефіцієнтів.

На рисунку 3.5 зображена передбачувана архітектура майбутньої системи виявлення вторгнень.

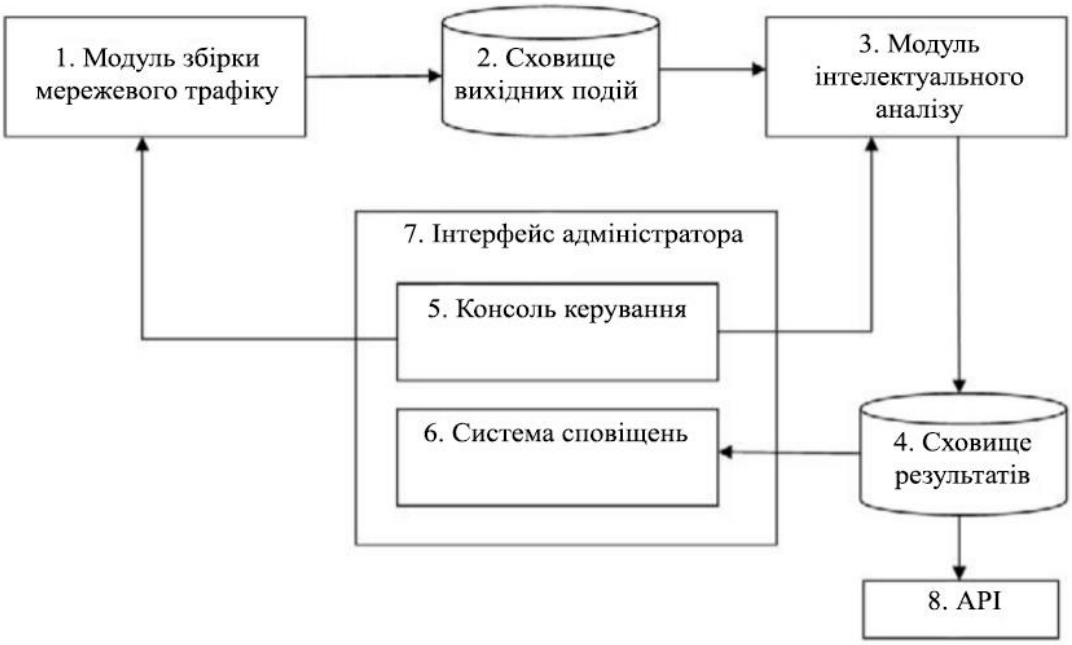


Рисунок 3.5 – Архітектура IDS з використанням модулю інтелектуального аналізу

Моя система буде складатися з наступних компонент:

1) Модуль збору мережевого трафіку необхідний для збору всієї інформації з мережевих пристроїв, що знаходяться на периметрі мережі, перетворення вихідного мережевого трафіку в необхідний вид (визначення і обчислення необхідних параметрів) і записи даних в сховище вихідних подій.

2) Місце вихідних подій призначене для зберігання інформації, яку необхідно проаналізувати на наявність мережевих атак. Кожен запис буде зберігати інформацію про потік, за параметрами, які потрібні для обробки модулем інтелектуального аналізу даних.

3) Модуль інтелектуального аналізу мережевого трафіку виконує аналіз кожного запису про потік, що зберігається в базі даних вихідних подій, за допомогою алгоритмів, що використовують методи машинного навчання. В результаті для кожного запису визначається вид з'єднання: нормальне або атака, а також вид атаки в випадку її виявлення. Результати аналізу будуть записані в базу даних результатів.

4) Місце результатів являє собою базу даних, де будуть зберігатися виявлені аномалії. Дана база буде використовуватися системою сповіщень, а також модулем сумісності для вилучення результатів аналізу. Зберігання результатів дозволить надалі виконувати тимчасовий аналіз стану безпеки системи.

5) Консоль управління системою призначений для загальної настройки всіх компонент програмного комплексу.

6) Система оповіщення про виявлені інциденти буде миттєво повідомляти адміністратора безпеки про виявлені інциденти, відображати в зручному для користувача вигляді виявлені аномалії, вказувати на тип виявлених аномалій, а також надавати можливості для побудови звітів.

7) Графічний інтерфейс адміністратора системи буде об'єднувати консоль управління і систему сповіщення. Він призначений для взаємодії користувача з системою.

8) Модуль інтеграції являє собою API для можливості інтеграції з системами реагування, інтерфейс для взаємодії з системою за допомогою http-запитів.

Розроблюваний модуль інтелектуального аналізу мережевого трафіку в якості механізму прийняття рішень про наявність атак буде використовувати методи індуктивного машинного навчання, а саме - штучні нейронні мережі (ШНМ).

При побудові модуля інтелектуального аналізу для навчання інтелектуальних моделей було прийнято рішення використовувати більш нову вибірку (базу даних атак), що містить відомості про велику кількість мережевих атаках UNSW-NB-15 [17, 18].

У базі UNSW-NB15 кожен запис містить 47 ознак мережевого трафіку п'яти типів: номінальні, цілочисельні, числові, тимчасові, бінарні. Повний перелік ознак сполучень, що використовуються в UNSW-NB1, представлений в [17].

Для кожного запису міститься інформація про те, до якого з класів відносять конкретне з'єднання: нормальні з'єднання (Normal) або один з дев'яти різних видів атак.

Таблиця 3.1 - Види атак, представлені в базі UNSW-NB15

№	Вид з’єднання	Кількість	Опис
1	Normal	2 218 761	Перехоплені дані транзакцій
2	Fuzzers	24 246	Спроба викликати призупинення програми або мережі шляхом подачі випадково генерованих даних
3	Analysis	2 677	Містить різні атаки сканування порту, спаму і проникнення html-файлів
4	Backdoors	2 329	Непомітний обхід доступу до даних комп’ютера (мережі)
5	DoS	16 353	Шкідлива спроба зробити сервер або мережевий ресурс недоступним для користувачів. Зазвичай це тимчасове переривання або припинення послуг хоста, підключеного до Інтернету
6	Exploits	44 525	Зловмисник знає про проблеми безпеки в системі і використовує дані уразливості в своїх цілях
7	Generic	215 481	Техніка працює проти всіх блокових шифрів (із заданим блоком і розміром ключа), незалежно від структури блочного шифру

Кінець таблиці 3.1

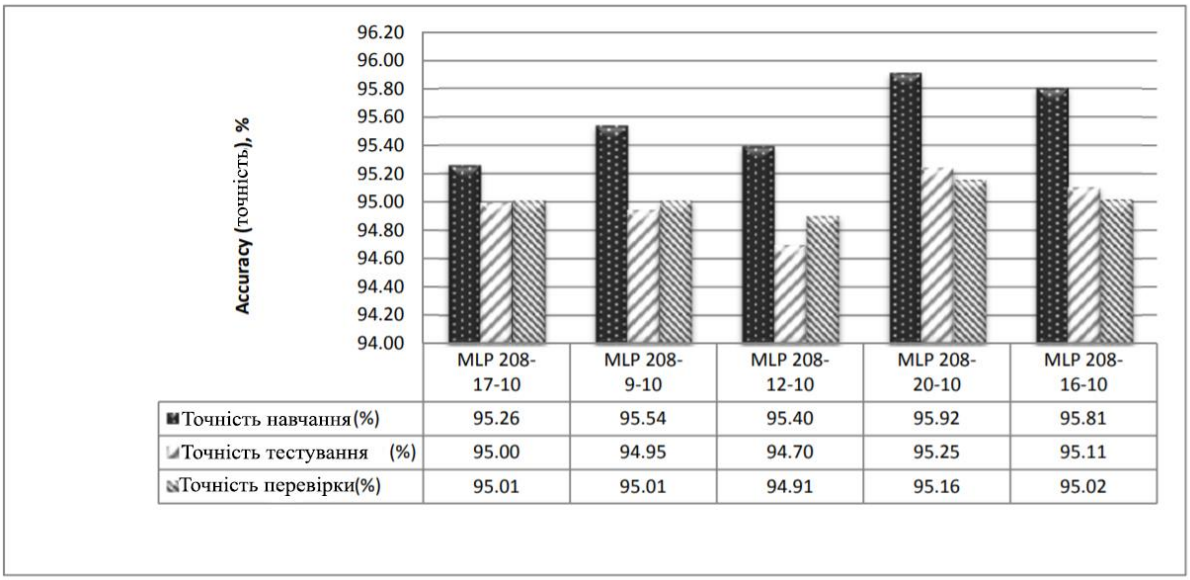
8	Reconnaissance	13 987	Містить всі типи атак, які збирають інформацію про мережі (з метою розвідки)
9	Shellcode	1 511	Невеликий фрагмент коду, який використовують як корисне навантаження при експлуатації вразливостей програмного забезпечення
10	Worms	174	Атакуючий реплікує себе, щоб поширитися на інші комп'ютери. Часто він використовує комп'ютерну мережу для поширення, покладаючись на збої в безпеці на цільовому комп'ютері для доступу до нього

В таблиці 3.1 перерахована деяка вибірка з'єднань, що зберігаються в базі UNSW-NB1, кількість представлених записів про них, а також короткий опис.

Проєктування нейронних мереж, що використовують набір даних UNSW-NB1, здійснювалося в пакеті *Statistica* з використанням інструментів *Automated Neural Networks*. Кількість нейронів на вхідному і вихідному шарах визначається набором даних UNSW-NB1, на прихованому шарі - випадковим чином, шляхом перебору різних варіантів, виходячи з мінімальної помилки навчання, тестування і перевірки. Діапазон кількості нейронів на прихованому шарі визначається по формулою, що є наслідком теорем Арнольда-Колмогорова-Хехт-Нільсена [13].

На першому етапі досліджень, було виконано побудову нейронних мереж з повним набором параметрів [17]: на вхід нейронної мережі подаються 45 ознак мережевого трафіку (за винятком ip-адреси відправника і одержувача).

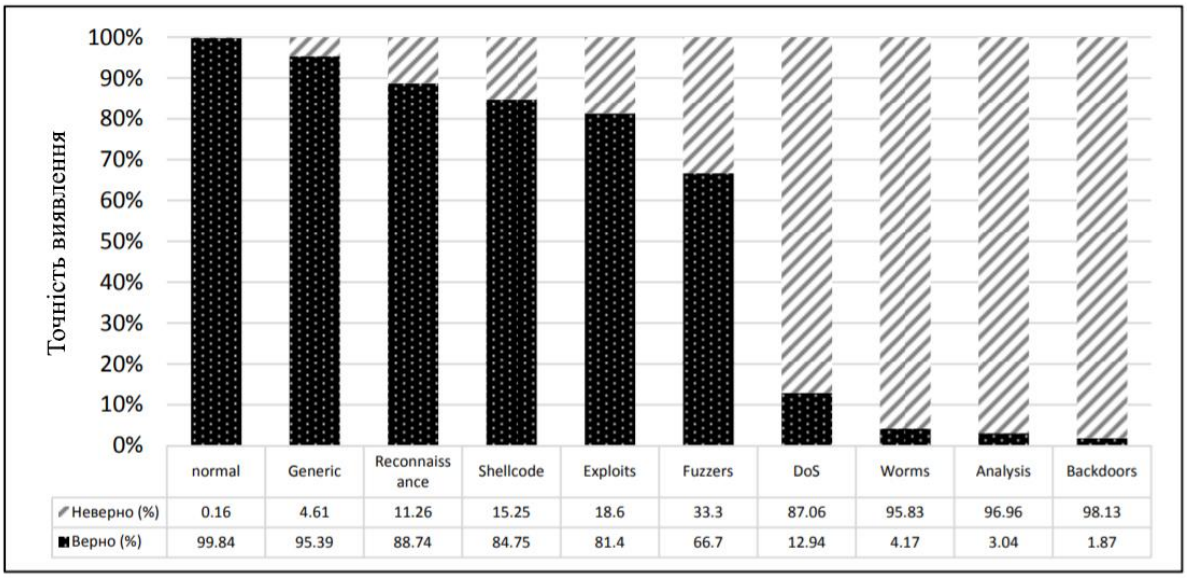
Були проведені дослідження на вхідних множинах на потужності 100000 записів. Вхідна множина автоматично розбивається на навчальну, тестову і перевірочну вибірку. Для навчання використовується 70% від вхідної множини, а для тестування і перевірки - по 15%. На рисунку 3.6 представлені 5 найкращих варіантів побудови нейронних мереж, змодельованих на вхідній множині потужністю 100000 записів. Як множини оцінки якості моделі використовується Accuracy (акуратність, точність) - частка від навчальної, тестової, перевірочної вибірки в відношенні якої класифікатор прийняв вірне рішення.



	MLP 208-17-10	MLP 208-9-10	MLP 208-12-10	MLP 208-20-10	MLP 208-16-10
Точність навчання(%)	95.26	95.54	95.40	95.92	95.81
Точність тестування (%)	95.00	94.95	94.70	95.25	95.11
Точність перевірки(%)	95.01	95.01	94.91	95.16	95.02

Рисунок 3.6 – Характеристики нейронних мереж з повним набором параметрів з використанням навчальної вибірки UNSW-NB1 потужністю 100 000 записів

Найкращі результати (найменшу помилку перевірки) показала нейронна мережа MLP 208-14-10 (208, 14 і 10 – це кількість нейронів на вхідному, прихованому і вихідному шарі відповідно), для якої на рисунку 3.7 представлені результати класифікації за видами атак.



	normal	Generic	Reconnaissance	Shellcode	Exploits	Fuzzers	DoS	Worms	Analysis	Backdoors
Неверно (%)	0.16	4.61	11.26	15.25	18.6	33.3	87.06	95.83	96.96	98.13
Верно (%)	99.84	95.39	88.74	84.75	81.4	66.7	12.94	4.17	3.04	1.87

Рисунок 3.7 - Результати перевірки MLP 208-14-10 за видами атак

Використовуваний інструмент нейромережевого моделювання після навчання нейронної мережі дозволяє вивантажити її параметри в файл формату * .xml, що дає можливість використовувати отриману нейромережу надалі при розробці додатків. Дана модель MLP 194-20-10 стала основою для розробки модуля інтелектуального аналізу.

Також, якщо ми хочемо спростити мережу, як вибірку ми можемо використати дані для конкретного виду загрози та їх опис.

Логіку роботи модуля інтелектуального аналізу мережевого трафіку зручно відобразити в вигляді блок-схеми, зображеної на рисунку 3.8.

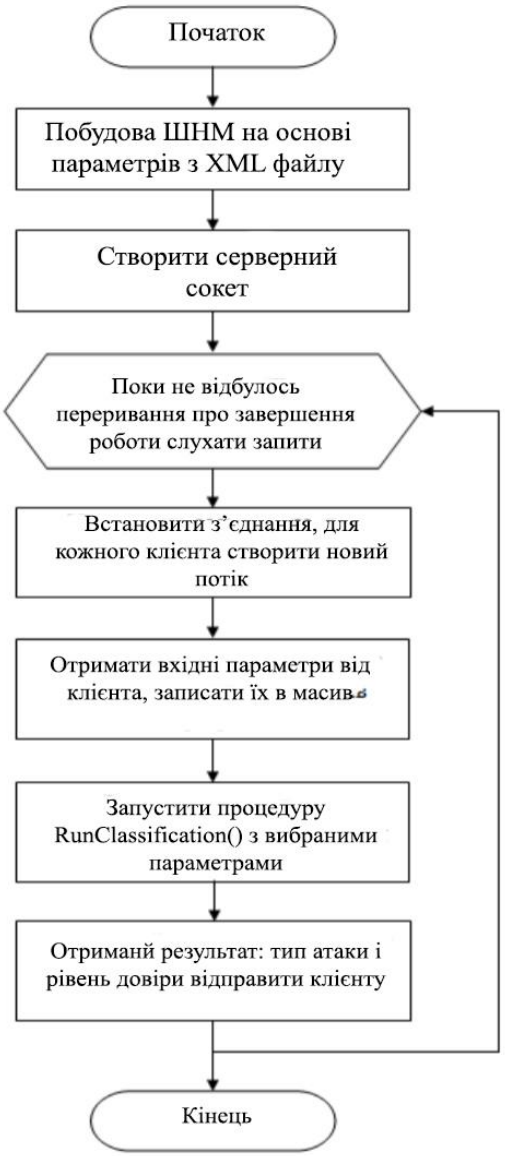


Рисунок 3.8 – Логіка роботи модуля інтелектуального аналізу

Одним із прикладів може служити DDoS attack scoreboard dataset [21]. Цей датасет є вільним для скачування. У ньому міститься пара srcip: srcport-dstip: dstport зі статистичними лічильником для кількості пакетів, ідентифікації протоколу, прапора (якщо це TCP) і кількості пакетів SYN (якщо це TCP). Кожна пара ip: port представляє один канал зв'язку. На сервісі Kagle також присутні тестові набори для навчання. Наприклад, в грудні 2018 року було опубліковано набір даних низько-інтенсивних DDoS атак [5]. Одним з найцікавіших наборів даних є CICID, який створили в

Канадському інституті кібербезпеки. У ньому містяться сучасні поширені атаки, які нагадують справжні реальні дані (PCAP). Він також включає в себе результати аналізу мережевого трафіку з використанням CICFlowMeter з позначеними потоками на основі мітки часу, IP-адрес джерела і призначення, портів джерела і призначення, протоколів і атак. Файли даного датасета збережені у форматі CSV. Канадський інститут кібербезпеки опублікував також і інші датасети для проведення різних досліджень, наприклад, даних CSE-CIC-IDS2018. В даному датасеті використовується поняття профілів для систематичної генерації наборів даних, які містять докладні описи вторгнень і абстрактні моделі поширення для додатків, протоколів або мережевих об'єктів нижчого рівня.

3.3 Власний програмний модуль для виявлення аномалій мережі

Загальноприйнята класифікація систем виявлення атак за способами включає системи пошуку аномалій. Однією з класичних робіт в області виявлення аномалій є звичайна робота, [2].

Загальний алгоритм виявлення мережевих аномалій може бути описаний таким чином. Даними для аналізу є мережевий трафік, представлений як набір мережевих пакетів, в загальному випадку фрагментованих на рівні IP. Зібрані сирі дані в подальшому будуть слугувати джерелом при формуванні необхідної інформації для подальшого аналізу.

На рисунку 3.9 представлена схема виявлення мережевих аномалій на основі показників мережевого трафіку.

Рисунок 3.9 – Алгоритм виявлення мережевих аномалій

Так, отримані дані можуть бути агреговані за певний часовий інтервал і нормалізовані з метою завдання визнаних атрибутів загального вигляду, які будуть потрібні при побудові поточного профілю активності.

Створений набір ознак порівнюється з набором характеристик нормальної діяльності об'єкта (користувача або системи) - шаблоном нормального поведінки. Якщо спостерігається суттєва розбіжність порівнюваних параметрів, то фіксується мережева аномалія. В іншому випадку відбувається уточнення шаблону нормальної поведінки за допомогою зміни параметрів його настройки з урахуванням поточного спостережуваного профілю мережевої активності. Описаний вище алгоритм може включати кілька варіантів виконання для реалізації підсистеми перевірки на відповідність шаблону нормальної поведінки.

Найпростішим з них є процедура порівняння з пороговою величиною, коли накопичені результати, що описують поточну мережеву активність, порівнюються з експертно заданою числовою планкою. У цьому підході випадок перевищення значень розглянутих параметрів зазначеної кордону є ознакою мережевої аномалії. Варто відзначити, що побудова шаблону нормальної поведінки є трудомістким завданням і часто не завжди здійсненним. Так, на практиці виявляється, що не кожна

аномальна поведінка є атакою, тому на цей випадок ми використаємо вже готову вибірку з вказаними атаками і їх ознаками (датасет CICIDS2017).

Для побудови нашої моделі я використав Python 3.6. Крім того, для візуалізації та навчання були використані наступні бібліотеки:

- Sklearn – для навчання;
- Numpy – для математичних операцій;
- Pandas – інструмент аналізу даних;
- Matplotlib – для графічної частини та візуалізації.

Опис завдання: Класифікація аномалій мережі

Реалізація складається з 5 етапів:

1) попередня обробка
2) статистика
3) фільтрація атак
4) вибір функцій
5) використання модулю та його впровадження;

1. Попередня обробка (preprocessing). Спочатку ми проводимо попередню обробку даних використовуючи файл .pcap (файл захоплених мережевих пакетів) з даними (рисунок 3.10).

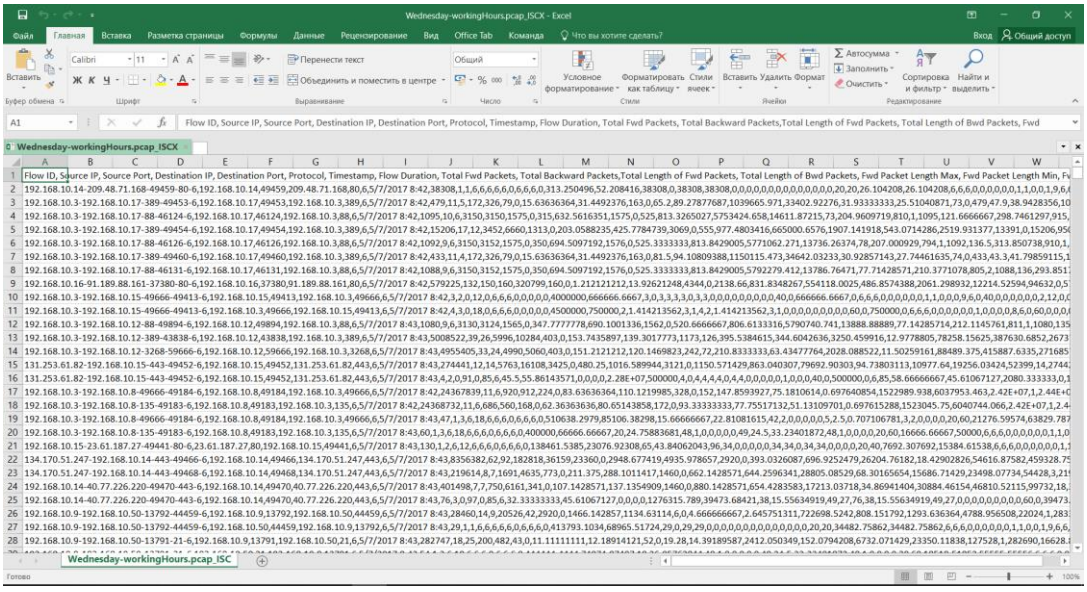


Рисунок 3.10 – Дані файлу .pcap Wednesday-workingHours

Такий файлів у нас буде 8. В них розписується статистика і напади за деякі конкретні дні. Наприклад файл Wednesday-workingHours.pcap_ISCX розписує : ідентифікатор потоку, IP джерела, порт джерела, порт призначення, протокол, мітка часу, тривалість потоку, загальна кількість пакетів FWD, загальна кількість зворотних пакетів, загальна довжина пакетів FWD, загальна довжина пакетів Bwd, , середня довжина пакета Fwd, довжина пакета Bwd/Макс., і т. д.

Таблиця 3.2 – Опис використовуваного датасету CICIDS2017

День запису потоку	Розмір файлу .pcap, ГБ	Тривалість	CSV розмір файлу, МБ	Назва атаки	Кількість потоків
Понеділок	10	Весь день	257	Без атак	529918
Вівторок	10	Весь день	166	FTP-Patator, SSH-Patator	445909
Середа	12	Весь день	272	DoS Hulk, DoS GoldenEye, DoS slowloris, DoS Slowhttptest, Heartbleed	692703
Четвер	7.7	Ранок	87,7	Web Attacks (Brute Force, XSS, Sql Injection)	170366
		Обід	103	Інфільтрація	288602
П’ятниця	8.2	Ранок	71,8	Bot	192033
		Обід	92,7	DDoS	225745
		Обід	97,1	PortScan	286467

Опис датасету можна побачити в таблиці 3.2.

В результаті попередньої обробки ми отримаємо all_data.csv де будуть міститися дані з цієї попередньої обробки.

2. Статистика. Вона є необов'язковою в даній ситуації. Вона перевіряє файл "all_data.csv" і друкує статистику атак та доброякісний реєстр на екрані (рисунок 3.11).

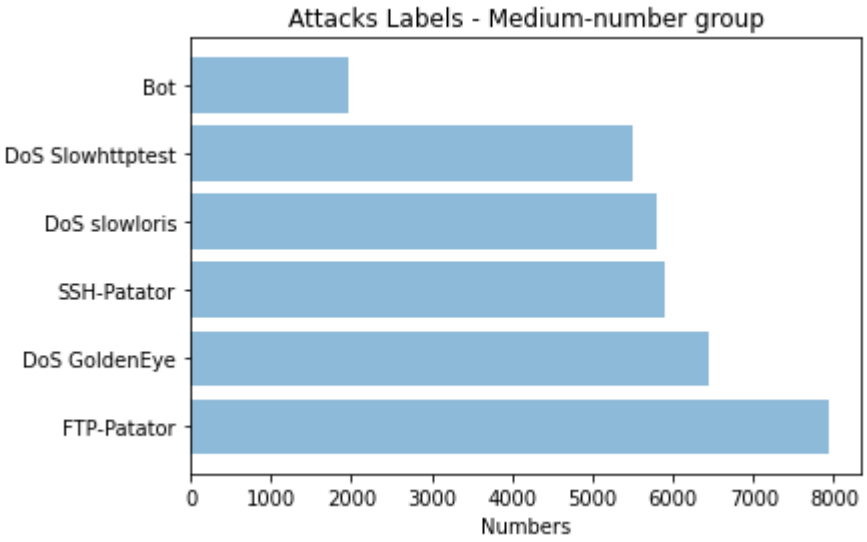


Рисунок 3.11 – Результат реалізації статистики

На рисунку 3.11 зображено результат статистики, що ми зібрали з наших .pcap файлів з попереднього пункту. Як видно з рисунку середня кількість атак припадає на:

- Ботнет-атака, що здійснюється за допомогою Ares, інструменту атаки, написаного в програмі Python. Це атака відбувається в п'ятницю вранці – в нашому датасеті.
- DoS SlowHTTPTest – ця атака зловживає функцією розміру вікна TCP, споживаючи файлові ресурси сервера жертв, щоб законні користувачі не могли скористатися послугою.
- DoS Slowloris – метою цієї атаки є споживання системних ресурсів жертви, що заважає законним користувачам отримувати послуги.

- Атака SSH-Patator так як і FTP-Patator здійснюється за допомогою Patator [28], багатопоточного інструменту, написаного на Python.
- DoS Goldeneye - метою цієї атаки є споживання системних ресурсів жертви, тим самим перешкоджаючи законним користувачам отримувати послуги. Goldeneye - це багатопотокова атака, яка може запустити атаку http Flood за допомогою багатопотокового процесора та обладнання пам'яті.

```
Bot file is completed
 attack:1966
 benign:3720


DDoS file is completed
 attack:76
 benign:79584


DoS GoldenEye file is completed
 attack:6453
 benign:19606


DoS Hulk file is completed
 attack:231073
 benign:477875


DoS Slowhttptest file is completed
 attack:5499
 benign:10709


DoS slowloris file is completed
 attack:5796
 benign:10933
```

Рисунок 3.12 – Результат реалізації етапу фільтрації атак

3. Фільтрація атак. використовує файл "all_data.csv". Створює файли атак, а потім зберігає їх у розташуванні "./attacks/" (рисунок 3.12). Набір даних містить 12 різних типів атак. Тому, для них створюються 12 різних файлів CSV. У кожному файлі є 30% атак та 70% доброякісного реєстру. Цей пункт є необхідною умовою для четвертого та п'ятого кроків.

4. Вибір функцій.

А) Спочатку визначимо які функції важливі для кожної атаки використавши алгоритм Random Forest Regressor для обчислення вагових коефіцієнтів об’єктів у наборі даних (рисунок 3.13).

Б) Створюємо ваги важливості функції, які діють для всього набору даних. Для цього використаємо файл "all_data.csv" та алгоритм Random Forest Regressor (рисунок 3.14).

SSH-Patator importance list:
SSH-Patator

Features	importance
Fwd Packet Length Max	0.000881
Flow Duration	0.000748
Flow IAT Max	0.000497
Total Length of Fwd Packets	0.000448
Flow IAT Mean	0.000425
Flow Packets/s	0.000423
Flow Bytes/s	0.000375
Fwd IAT Total	0.000329
Flow IAT Std	0.000177
Fwd Packet Length Mean	0.000158
Flow IAT Min	0.000111
Total Backward Packets	0.000100
Bwd Packet Length Min	0.000099
Fwd Packet Length Std	0.000070
Total Fwd Packets	0.000070
Fwd Packet Length Min	0.000040
Bwd Packet Length Max	0.000032
Total Length of Bwd Packets	0.000027
Bwd Packet Length Mean	0.000014
Bwd Packet Length Std	0.000008

Web Attack importance list:
Web Attack

Features	importance
Bwd Packet Length Std	0.007255
Total Length of Fwd Packets	0.006046
Flow Bytes/s	0.003366
Flow IAT Max	0.002102
Bwd Packet Length Max	0.001728
Flow IAT Mean	0.000760
Fwd Packet Length Max	0.000638
Fwd Packet Length Std	0.000616
Flow Duration	0.000541
Flow Packets/s	0.000526
Total Fwd Packets	0.000506
Fwd IAT Total	0.000505
Flow IAT Min	0.000499
Fwd Packet Length Mean	0.000490
Total Length of Bwd Packets	0.000454
Total Backward Packets	0.000177
Flow IAT Std	0.000140
Bwd Packet Length Mean	0.000102
Bwd Packet Length Min	0.000016
Fwd Packet Length Min	0.000008

Рисунок 2.14 – Результат виконання 4А

```
all_data importance list:
all_data
                              importance
Features
Bwd Packet Length Std           0.246627
Flow Bytes/s                    0.178777
Total Length of Fwd Packets     0.102417
Fwd Packet Length Std           0.063889
Flow IAT Std                    0.009898
Flow IAT Min                    0.006946
Fwd IAT Total                   0.005121
Flow Duration                   0.004150
Bwd Packet Length Max           0.004007
Flow IAT Max                    0.003579
Flow IAT Mean                   0.003266
Total Length of Bwd Packets     0.001305
Fwd Packet Length Min           0.000670
Bwd Packet Length Mean          0.000582
Flow Packets/s                  0.000541
Fwd Packet Length Mean          0.000526
Total Backward Packets          0.000169
Total Fwd Packets               0.000138
Fwd Packet Length Max           0.000125
Bwd Packet Length Min           0.000084
```

Рисунок 3.14 – Результат виконанання 4Б

5. Використання модулю та його впровадження

а) Спочатку використаємо файли атак у папці "./attacks/" як набір даних. Використані функції - 4 функції з найбільшою вагою для кожного файлу, створені файлом feature_selection_for_attack_files. Цей файл застосовує 7 алгоритмів машинного навчання до кожного файлу по 10 разів.

б) Далі реалізуємо метод машинного навчання для файлу "all_data.csv", використаємо функції з попереднього кроку. Набір функцій, що використовуватимуться, складається з поєднання 4-х функцій з найбільшою вагою, що досягається для кожної атаки в кроці «а» загалом. Таким чином, з кожного із 12 типів атак отримуються 4 функції, в результаті чого з’являється пул об’єктів, що складається з 48 атрибутів. Після видалення повторень кількість функцій становить 18.

в) Далі працюємо з файлом "all_data.csv". Він знаходить функцію, що дає найвищий коефіцієнт f для алгоритмів Naive Bayes, QDA та MLP

(багатошарового перцептрону), і друкує їх на екрані. Наївні методи Байєса (Naive Bayes) – це сукупність керованих алгоритмів навчання, заснованих на застосуванні теореми Байєса з „наївним” припущенням умовної незалежності між кожною парою ознак з урахуванням значення змінної класу. QDA – поведінковий сенсорний підхід до оцінки, який використовує описові панелі для вимірювання сенсорних характеристик. Якщо F-міра для кожної ознаки дорівнює або перевищує найбільше отримане значення, ця властивість додається до списку. В іншому випадку він вилучається зі списку. В результаті процесу цей крок надає найвищий отриманий коефіцієнт F та список властивостей, які для нього присутні. Метою цього пункту є пошук оптимального списку властивостей для Naive Bayes, а також алгоритмів QDA та MLP (рисунок 3.15).

```
Feature Number    Feature
1                 Bwd Packet Length Std
2                 Flow Bytes/s
3                 Total Length of Fwd Packets
4                 Fwd Packet Length Std
5                 Flow IAT Std
6                 Flow IAT Min
7                 Fwd IAT Total
8                 Flow Duration
9                 Bwd Packet Length Max
10                Flow IAT Max
11                Flow IAT Mean
12                Total Length of Bwd Packets
13                Fwd Packet Length Min
14                Bwd Packet Length Mean
15                Flow Packets/s
16                Fwd Packet Length Mean
17                Total Backward Packets
18                Total Fwd Packets
19                Fwd Packet Length Max
20                Bwd Packet Length Min

ML algorithm      Feature Name                F1-score    Accuracy   Feature List
Naive Bayes       Bwd Packet Length Std       0.86        0.88       [1,            ------> New feature found!!!
Naive Bayes       Flow Bytes/s                0.84        0.86       [1, 2,
Naive Bayes       Total Length of Fwd Packets 0.86        0.88       [1, 2,         ------> New feature found!!!
Naive Bayes       Fwd Packet Length Std       0.85        0.86       [1, 2, 3,
Naive Bayes       Flow IAT Std                0.85        0.86       [1, 2, 3,
Naive Bayes       Flow IAT Min                0.86        0.88       [1, 2, 3,      ------> New feature found!!!
Naive Bayes       Fwd IAT Total               0.82        0.83       [1, 2, 3, 4,
Naive Bayes       Flow Duration               0.82        0.83       [1, 2, 3, 4,
Naive Bayes       Bwd Packet Length Max       0.85        0.87       [1, 2, 3, 4,
Naive Bayes       Flow IAT Max                0.83        0.85       [1, 2, 3, 4,
Naive Bayes       Flow IAT Mean               0.85        0.87       [1, 2, 3, 4,
Naive Bayes       Total Length of Bwd Packets 0.07        0.18       [1, 2, 3, 4,
Naive Bayes       Fwd Packet Length Min       0.86        0.88       [1, 2, 3, 4,   ------> New feature found!!!
Naive Bayes       Bwd Packet Length Mean      0.85        0.86       [1, 2, 3, 4, 5,
Naive Bayes       Flow Packets/s              0.86        0.87       [1, 2, 3, 4, 5,  ------> New feature found!!!
Naive Bayes       Fwd Packet Length Mean      0.86        0.87       [1, 2, 3, 4, 5, 6,  ------> New feature found!!!
Naive Bayes       Total Backward Packets      0.13        0.21       [1, 2, 3, 4, 5, 6, 7,
Naive Bayes       Total Fwd Packets           0.52        0.48       [1, 2, 3, 4, 5, 6, 7,
Naive Bayes       Fwd Packet Length Max       0.84        0.84       [1, 2, 3, 4, 5, 6, 7,
Naive Bayes       Bwd Packet Length Min       0.85        0.85       [1, 2, 3, 4, 5, 6, 7,
F1= 0.86 Naive Bayes  The most efficient feature list = ['Bwd Packet Length Std', 'Total Length of Fwd Packets', 'Flow IAT Min',
ket Length Min', 'Flow Packets/s', 'Fwd Packet Length Mean']
```

Рисунок 3.15 – Список властивостей та результат впровадження алгоритму Naive Bayes

Ця програма була призначена для визначення аномалій мережі з використанням машинного навчання. У цьому контексті ми використали наш датасет як набір даних через його сучасність, широке розмаїття атак та різні мережеві атаки. Цей набір даних містить більше 80 функцій, що визначають мережевий потік. Було зроблено розрахунок ваги важливості за допомогою логаритму «Random Forest Regressor» , щоб вирішити, яка з цих функцій буде використана в методах машинного навчання. При проведенні цих розрахунків було використано два підходи. По-перше, ваги важливості обчислювалися окремо для кожного типу атаки. При другому методі всі атаки збираються в одній групі та розраховуються ваги важливості для цієї групи, тобто визначаються загальні властивості, важливі для всіх атак. Нарешті, до цих даних застосовано сім алгоритмів машинного навчання, які широко використовуються та мають різні якості. Ці алгоритми та досягнуті коефіцієнти продуктивності згідно з F-мірою такі (F-міра приймає значення від 0 до 1): Naive Bayes: 0,86, QDA: 0,86, Random Forest: 0,94, ID3: 0,95, AdaBoost: 0,94, MLP: 0,83 та K Найближчі сусіди: 0,97.

Слідуючи з вище всього сказаного, я роблю висновок, що наш модуль може використовуватися для виявлення аномалій в мережі. Подібними аномаліями можуть слугувати ідентифікатор потоку, IP адреса, порт джерела, кінцевий IP, порт призначення, протокол, мітка часу, тривалість потоку, загальна кількість пакетів FWD, загальна кількість зворотних пакетів, загальна довжина пакетів FWD, загальна довжина пакетів Bwd. Ми навчили нашу мережу визначати такі аномалії з допомогою алгоритму Naive Bayes. Після запуску і використання якось іншого .pcap або cvs файлу програма подає сигнал на конкретну ознаку як на рисунку 2.16. Тому, я вважаю, що поставлену задачу виконано.

Лістинг програми описано в додатку А

4 РЕАЛІЗАЦІЯ РЕЗУЛЬТАТІВ РОБОТИ

4.1 Практичне впровадження системи

Під час проходження переддипломної практики в АТ КБ «ПриватБанк» я отримав багато досвіду в розділі систем захисту мережевої безпеки. Крім того, я побачив роботу справжніх професіоналів які вирішують і усовують проблеми з мережевою безпекою в лічені хвилини. І саме там я зумів перевірити ефективність кожного з моїх рішень. Атаки які буди використані для перевірки були типу DoS та вірусні атаки. Результат першого рішення – програмного забезпечення Extrahop Reveal (x) зображено на рисунку 2.17 та 2.18. Для перевірки я використав вибірку з кількості атак 100, 500 та 1000 для подальшого розуміння ефективності.

Рисунок 4.1 – Ефективність рішення Extrahop Reveal (x) для DoS атак

Як видно з рисунку 2.11 захист програмним забезпечення Extrahop Reveal (x) від DoS здійснюється на 98,5%, що є досить ефективним в сучасних реаліях.

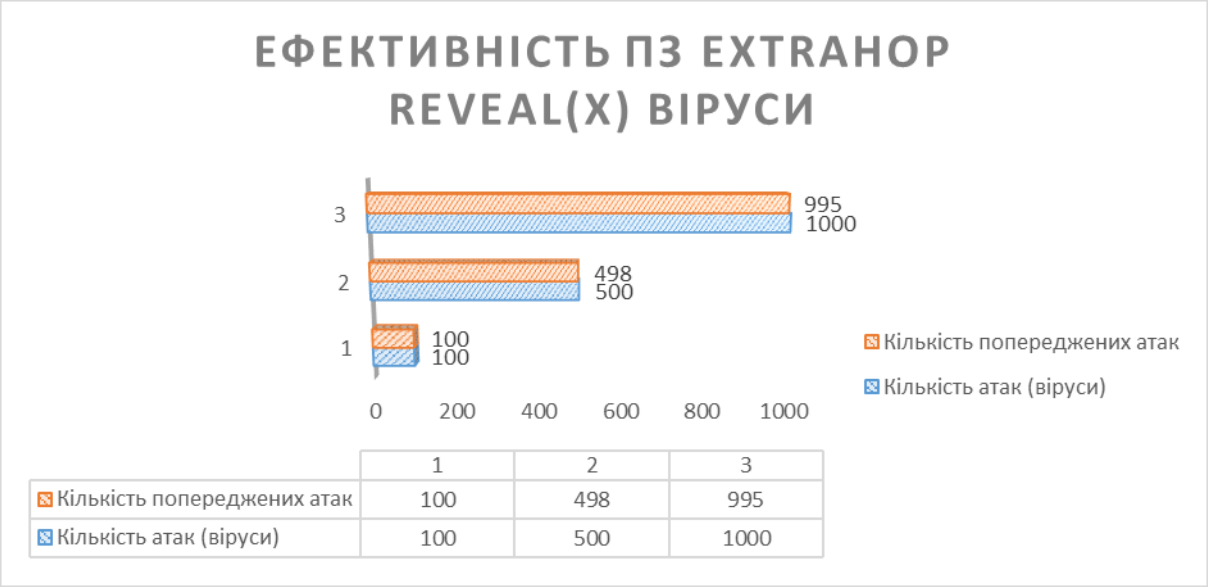
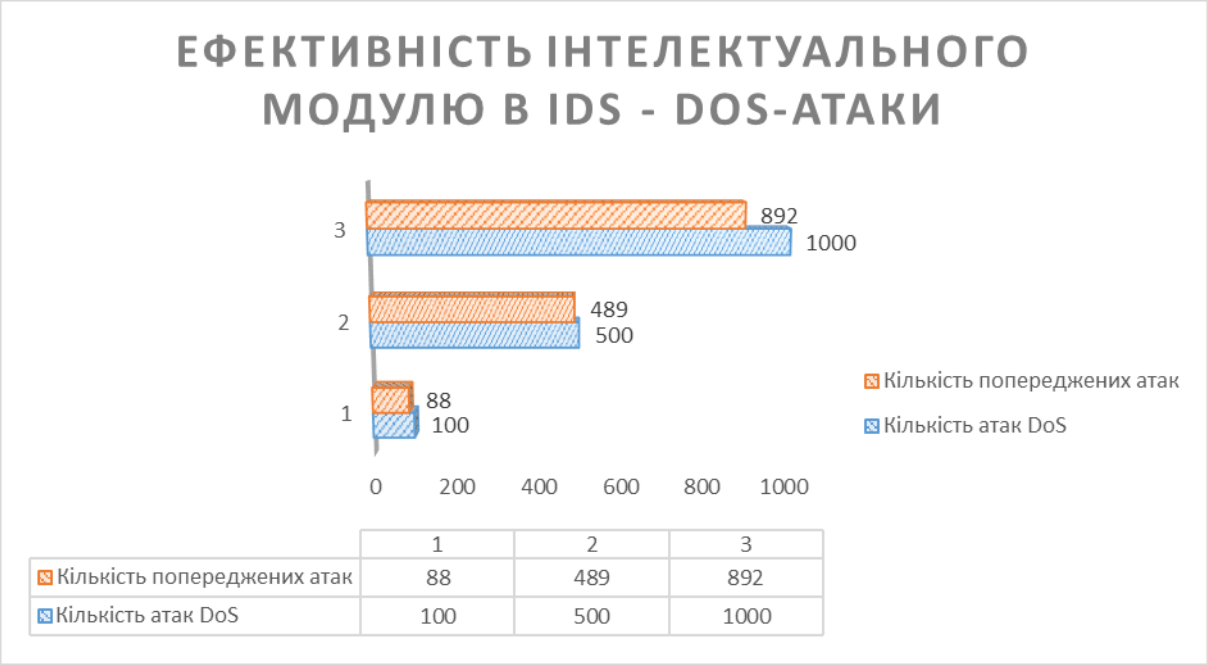


Рисунок 4.2 – Ефективність рішення Extrahop Reveal (x) для DoS атак

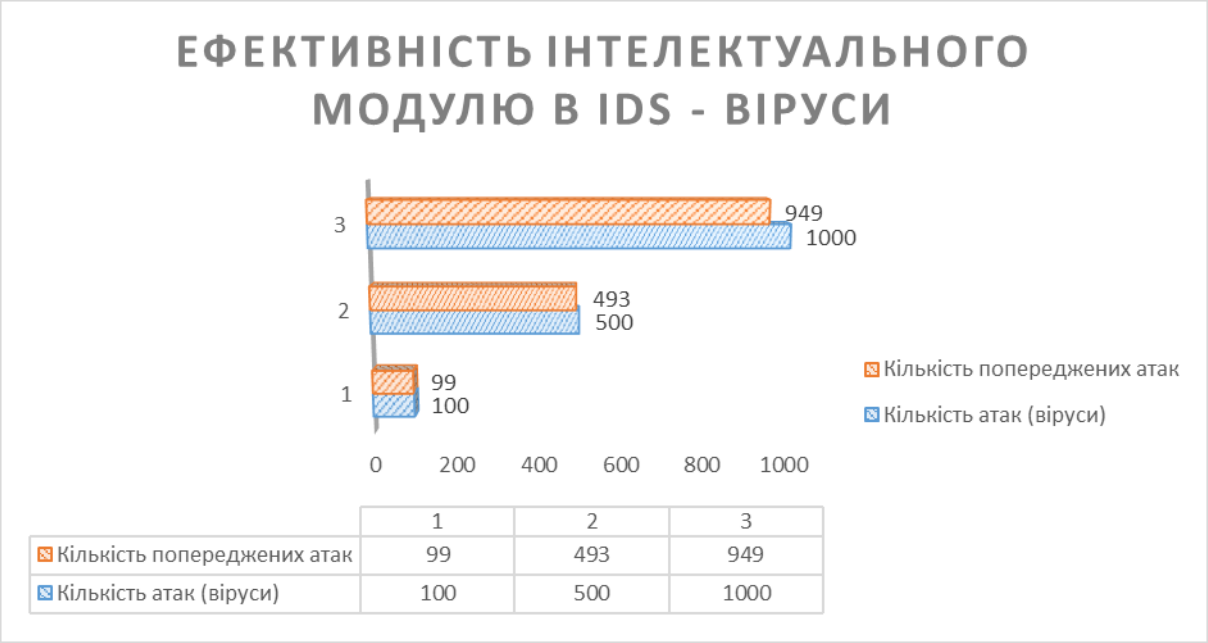
Далі я зробив перевірку нашого програмного забезпечення на захищеність від вірусів по типу троянів і шпигунського ПЗ. І подібний захист в ньому є навіть кращим, ніж з DoS атаками, оскільки інтелектуальна система ExtraHop Reveal(x) в режимі реального часу відслідковує можливі ознаки атак для подальшого їх попередження.

Потім на рисунку 2.19 я розглянув ефективність наступного рішення – впровадження інтелектуального модуля в систему виявлення вторгнень (IDS систему). Для перевірки я також використав попередню вибірку з кількості атак. Після аналізу стало очевидним, що визначеня DoS атак для такого рішення є ефективним на 90 %. На мою думку, на таку ефективність повпливала вибірка UNSW-NB-15 яка містить велику кількість інформації про різні типи мережевих атак.



	1	2	3
Кількість попереджених атак	88	489	892
Кількість атак DoS	100	500	1000

Рисунок 4.3 – Ефективність рішення інтелектуального модулю для IDS проти DoS атак

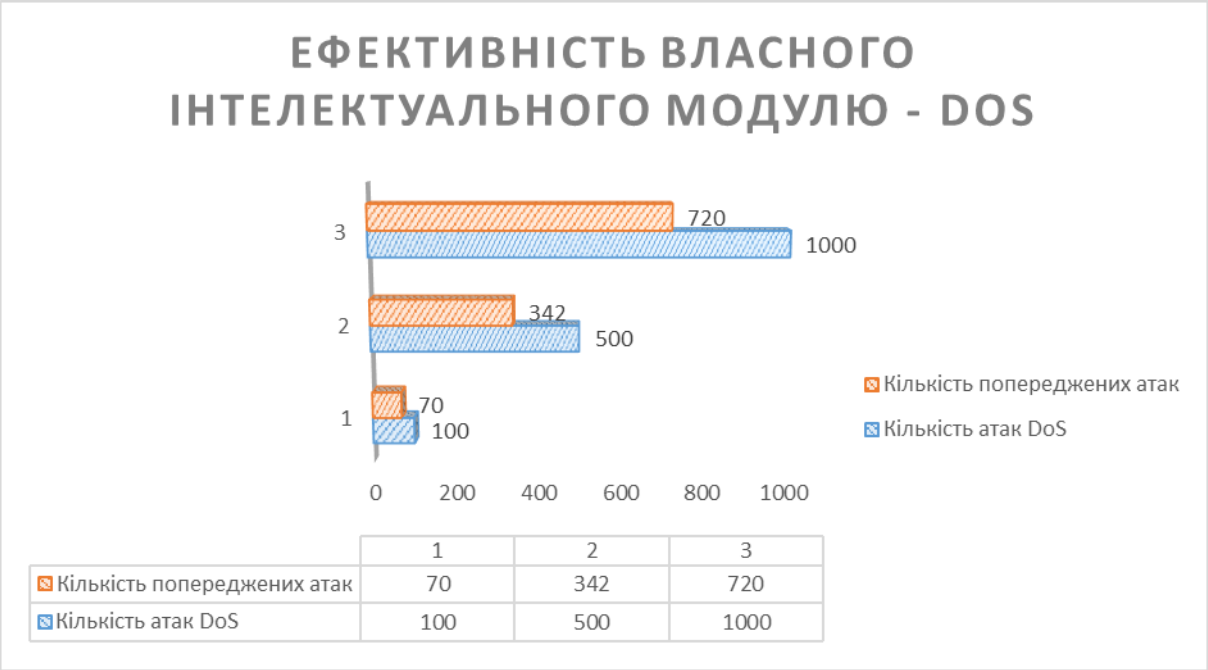


	1	2	3
Кількість попереджених атак	99	493	949
Кількість атак (віруси)	100	500	1000

Рисунок 4.4 – Ефективність рішення інтелектуального модулю для IDS проти вірусних атак

Ефективність рішення використання інтелектуального модулю в систему IDS в нашій системі захисту для вірусних атак предсталена на

рисунку 2.20. Захищеність та здатність системи для виявлення та знищення загрози вірусної атаки рівна 95%. Причиною цього є та сама вибірка UNSW-NB-15, що дає можливість виявити велику кільсть та величезне різноманіття вірусних атак.



	1	2	3
Кількість попереджених атак	70	342	720
Кількість атак DoS	100	500	1000

Рисунок 4.5 – Ефективність рішення власного інтелектуального модулю

Останнім запропонованим мною рішенням являється власний інтелектуальний модуль (рисунок 2.21). Навчений він був на власній вибірці, тому вона є не настільки потужною і точною як попередні рішення.

4.2 Оцінка запропонованих рішень

Для реалізації рішення захисту мережевого трафіку шляхом використання інтелектуальних систем я виділив 3 рішення.

Перше рішення – реалізація готового програмного забезпечення яке має власну потужну базу та сильну інтелектуальну систему. Таким програмним рішенням у мене стало ExtraHop Reveal(x). Це програмне

забезпечення наділене можливістю оцінювати ризик і впевнено реагувати на загрози. Не має помилкових спрацьовувань або помилкових сигналів тривоги. Є тільки глибокий контекст і розуміння того, що потрібно захищати і від кого. Розуміє система пріорітетні значення і використовує цей підхід для повного спектру виявлення загроз.

З допомогою потокового аналізу в режимі реального часу зі швидкістю 10Гбіт в сек. миттєво виявляє найбільш нагальні загрози та загрози які знаходяться в CVE – загальній баз даних вразливостей. Про ефективність говорить і статистика, відповідно звіту про загальний економічний вплив (TEI) за 2020 рік, проведеного Forrester Research, Reveal (x) зменшує час виявлення та вирішення проблем з загрозою на 84%.

Після впровадження такого програмного засобу в реальні системи з попереднього пункту, можна зробити висновок, що ця система визначає загрози рівня DoS і не тільки на рівні 98%. Крім того варто згадати вартість цього програмного засобу. Ця вартість коливаєть близько 1 млн. доларів на 3 роки, що є досить суттєвим для українського підприємця чи організації. Але, я думаю, що для забезпечення власної безпеки потрібно його використовувати незважаючи на вартість.

Другим моїм рішенням було інтеграція інтелектуального модуля в систему виявлення вторгнень IDS. Розроблюваний модуль інтелектуального аналізу мережевого трафіку в якості механізм прийняття рішень використав методи індуктивного навчання, а саме – штучні нейронні мережі. При подудові даної моделі було прийнято рішення використать вибірку, шо містить відомості про велику кількість мережевих атак - UNSW-NB-15. У базі UNSW-NB15 кожен запис містить 47 ознак мережевого трафіку п'яти типів: номінальні, цілочисельні, числові, тимчасові, бінарні. Використовуючи інструмент для нейромережевого модулювання Automated Neural Networks після навчання моделі дозволяє

вивантажити її параметри в файл формату * .xml, що дає можливість використовувати отриману нейромережу надалі при розробці додатків. Ефективність цього рішення, визначена на основі результатів дослідження з попереднього пункту дорівнює 95%, що є досить вагомим результатом в сфері мережевої безпеки.

На мою думку, така система є досить ефективною через високу потужність датасету і її реалізацію. Система виявлення вторгнень на основі інтелектуального модуля може використовуватися для захисту трафіку великих підприємств і вповною мірою буде забезпечувати безпеку цих мереж.

Останнім моїм рішенням була побудова власного програмного модуля для визначення аномалій в мережі. Даними для аналізу був наш мережевий трафік, представлений як набір мережевих пакетів, в загальному випадку фрагментованих на рівні IP. Для побудови основи нашої нейромережевої моделі я взяв датасет CICIDS2017. У цьому контексті ми використали наш датасет як набір даних через його сучасність, широке розмаїття атак та різні мережеві атаки.

Я вважаю, що ця програма здатна до виявлення аномалій в мережі. Подібними аномаліями можуть слугувати ідентифікатор потоку, IP адреса, порт джерела, кінцевий IP, порт призначення, протокол, мітка часу, тривалість потоку, загальна кількість пакетів FWD, загальна кількість зворотних пакетів, загальна довжина пакетів FWD, загальна довжина пакетів Bwd.

Цей програмний модуль був призначений для визначення аномалій мережі з використанням машинного навчання. Оцінюючи ефективність даного рішення і проаналізувавши результати впровадження, що це рішення є ефективним на 72%. Такий результат є досить непоганим, але не ідеальним. Таку системи можна розширити використавши інші алгоритми і методи оптимізації.

ВИСНОВКИ

В ході розробки даної інтелектуальної системи захисту, було досліджено предметну область, аналізовано теоретичну інформацію, різноманітні засоби і способи захисту мережевого трафіку, а також були освоєні навички планування та аналізу. З самого початку роботи я проаналізував актуальність розробки даного проєкту, тільки після цього розпочав над ним роботу. Дана система допоможе уникнути або навіть запобігти великої кількості загроз, що могли б спричинити негативні наслідки на роботу будь-якого підприємства чи організації.

На мою думку, збільшення використання мережевого трафіку може слугувати «індикатором» різних проблем. Адже будь-яка зайва діяльність в інтернеті, така як скачування заражених файлів, відкриття спам листів і подібної діяльності супроводжується стрибком мережевого трафіку, що може спричини величезну кількість загроз, зокрема, збій системи.

В роботі представлені кілька рішень для забезпечення безпеки мережі і керуванням трафіку, а саме: використання готового програмного забезпечення, що вже використовує власну потужну і надійну інтелектуальну систему; додавання інтелектуального модулю у вже існуючу систему захисту, у моєму випадку систему IDS; використання власного програмного модулю для ідентифікації аномалій мережі.

Я вважаю, що поставлена задача виконана в повному обсязі. В разі потреби, цю систему та рішення можна доповнити, розширити або удосконалити.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

1. 5 Best Network Traffic Analyzers [Електронний ресурс] : dnsstuff. – Режим доступу : https://www.dnsstuff.com/network-traffic-analyzers. – Назва з екрана, дата звернення 02.04.2021.(електронне джерело).
2. D. Bojovic, Petar; Basicevic, Ilija; Očovaj, Stanislav; Popovic, Miroslav (2017), “DDoS attack scoreboard da-taset”, Mendeley Data, v2;
3. ExtraHop Reveal(x) ExtraHop Networks [Електронний ресурс] : ANTI-MALWARE.– Режим доступу : https://www.anti-malware.ru/products/extrahop-revealx?width=500&height=480&inline=true#demonstration. – Назва з екрана, дата звернення 13.04.2021. (електронне джерело).
4. IDS [Електронний ресурс] : pnu.– Режим доступу : https://pnu.edu.ru/media/ejournal/articles-2017/TGU_8_339. – Назва з екрана, дата звернення 01.04.2021.(електронне джерело).
5. Low Rate DDoS Attack Dataset [Електронний ресурс] // 22.04.2019. 2018. URL: https://www.kaggle.com/keval17/ddosattackdetection;
6. Moustafa Nour, Jill Slay. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSWNB15 network data set). Military Communications and Information Systems Conference (MilCIS), 2015. IEEE, 2015. pp. 1-6.
7. Moustafa Nour, Jill Slay. The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set. // Information Security Journal: A Global Perspective (2016): 1-14.
8. NetFlow Traffic Analyzer [Електронний ресурс] : solarwinds.–Режим доступу : https://www.solarwinds.com/netflow-traffic-analyzer. – Назва з екрана, дата звернення 10.04.2021.(електронне джерело).

9. Network Traffic Monitor [Електронний ресурс] : solarwinds.– Режим доступу : https://www.solarwinds.com/network-bandwidth-analyzer-pack/use-cases/network-traffic-monitor. – Назва з екрана, дата звернення 30.03.2021.(електронне джерело)

10. Sniffer. https://www.opennet.ru/base/sec/arp_snif.txt.html, дата звернення 01.04.2021.

11. Wireshark. https://www.wireshark.org/, дата звернення 01.04.2021.

12. Аналіз мережевого трафіку для безпеки IT-операцій [Електронний ресурс] : eset.greycortex - network traffic analysis.–Режим доступу : https://www.eset.com/ee-ru/business/network-security-greycortex-mendel/. – Назва з екрана, дата звернення 08.04.2021.(електронне джерело).

13. Аналіз мережевого трафіку в режимі реального часу [Електронний ресурс] : ispras.– Режим доступу : https://www.ispras.ru/preprints/docs/prep_28_2015 – Назва з екрана, дата звернення 29.03.2021.(електронне джерело).

14. Багатофункціональний тестер MTX150 від VeEX [Електронний ресурс] : inkotel system.– Режим доступу : https://inkotel.com.ua/products/view/mnogofunkcionalnyy-tester-mtx150-ot-veex. – Назва з екрана, дата звернення 02.04.2021.(електронне джерело).

15. Безпечне місто [Електронний ресурс] : securitylab.– Режим доступу : https://www.securitylab.ru/news/520513.php. – Назва з екрана, дата звернення 28.03.2021.(електронне джерело)

16. Гладких А.М. Основні методи аналізу мережевого трафіку [Електронний ресурс] : Наукова електронна бібліотека «КіберЛенінка».– Режим доступу : https://cyberleninka.ru/article/n/osnovnye-metody-analiza-setevogo-trafika/viewer. – Назва з екрана, дата звернення 13.04.2021.(електронне джерело).

17. ГОСТ Р ІЗО/МЕК 17799-2005. Інформаційні технології. Практичні правила управління інформаційною безпекою [Електронний ресурс] :

2006. – 62с/ - Режим доступу: //www.consultant.ru (дата звернення 01.03.2021).

18. Дослідження методів класифікації зашифруваного трафіку [Електронний ресурс] : ЕЛЕКТРОННИЙ НАУКОВИЙ ЖУРНАЛ «ЩОДЕННИК НАУКИ».– Режим доступу : http://dnevniknauki.ru/images/publications/2020/3/technics/Vikhrov2 – Назва з екрана, дата звернення 30.03.2021.(електронне джерело).

19. Ерачитача [Електронний ресурс] : readera.– Режим доступу : https://readera.org/read/140129976. – Назва з екрана, дата звернення 02.04.2021.(електронне джерело).

20. ІБ: засоби захисту [Електронний ресурс] : tadviser.– Режим доступу : https://www.tadviser.ru/index.php. – Назва з екрана, дата звернення 02.04.2021.(електронне джерело).

21. ІТ.РАД.У5.ПЗ Методичний документ ФСТЕК Росії. Профіль захисту систем виявлення вторгнень рівня вузла п'ятого класу захисту (затв. ФСТЕК Росії 06.03.2012) [Електронний ресурс]: 2012. – 67с/ - Режим доступу: // fstec.ru /(дата звернення 12.03.2021).

22. Мережевий моніторинг і виявлення аномальної мережевої активності за допомогою рішень Flowmon Networks [Електронний ресурс] : securitylab.– Режим доступу : https://www.securitylab.ru/blog/company/ts-solution/346830.php. – Назва з екрана, дата звернення 12.04.2021.(електронне джерело).

23. Організаційний захист інформації [Електронний ресурс] : wikipedia.– Режим доступу : https://uk.wikipedia.org/Організаційний_захист_інформації. – Назва з екрана, дата звернення 01.04.2021.(електронне джерело).

24. Основи і методи захисту інформації [Електронний ресурс] : ЕУП.Інфоматика. – Режим доступу :

http://eos.ibi.spb.ru/umk/11_15/5/5_R1_T3.html. – Назва з екрана, дата звернення 02.04.2021. (електронне джерело).

25. Технології виявлення комп'ютерних атак [Електронний ресурс] : Безпека користувачів в мережі інтернет.– Режим доступу : https://safe-surf.ru/specialists/article/5274/656701 (NTA). – Назва з екрана, дата звернення 30.03.2021.(електронне джерело).

ДОДАТОК А

(Обов’язковий)

Програмна реалізація

1. Попередня обробка

```
import pandas as pd
import os
from sklearn import preprocessing
import time
seconds = time.time()
%matplotlib inline
number="0123456789"
# CSV назви наших пкапів:
csv_files=["Monday-WorkingHours.pcap_ISCX",
        "Tuesday-WorkingHours.pcap_ISCX",
        "Wednesday-workingHours.pcap_ISCX",
        "Thursday-WorkingHours-Morning-WebAttacks.pcap_ISCX",
        "Thursday-WorkingHours-Afternoon-Infilteration.pcap_ISCX",
        "Friday-WorkingHours-Morning.pcap_ISCX",
        "Friday-WorkingHours-Afternoon-PortScan.pcap_ISCX",
        "Friday-WorkingHours-Afternoon-DDos.pcap_ISCX",]
main_labels=["Flow ID","Source IP","Source Port","Destination IP","Destination Port","Protocol","Timestamp","Flow Duration","Total Fwd Packets"_Win_bytes_forward","Init_Win_bytes_backward","act_data_pkt_fwd","min_seg_size_forward","Active Mean","Active Std","Active Max","Active Min","Idle Mean","Idle Std","Idle Max","Idle Min","Label","External IP"]
main_labels2=main_labels
main_labels=( ",".join( i for i in main_labels ) )
```

```
main_labels=main_labels+"\n"
flag=True
for i in range(len(csv_files)):
    ths = open(str(i)+".csv", "w")
    ths.write(main_labels)
    with open("./CSVs/"+csv_files[i]+".csv", "r") as file:
        while True:
            try:
                line=file.readline()
                if  line[0] in number:# this line eliminates the headers of CSV files and incomplete streams .
                    if " – " in str(line): ##  if there is "–" character ("–", Unicode code:8211) in the flow ,  it will be chanced with "-" character ( Unicode code:45).
                        line=(str(line).replace(" – "," - "))
                    ths.write(str(line))
    ths.close()
     df=pd.read_csv(str(i)+".csv",low_memory=False)
    df=df.fillna(0)
    string_features=["Flow Bytes/s","Flow Packets/s"]
    for ii in string_features: #Some data in the "Flow Bytes / s" and "Flow Packets / s" columns are not numeric. Fixing this bug in this loop
        df[ii]=df[ii].replace('Infinity', -1)
        df[ii]=df[ii].replace('NaN', 0)
        number_or_not=[]
        for iii in df[ii]:
            try:  k=int(float(iii))
                number_or_not.append(int(k))
            except:
```

```
            number_or_not.append(iii)
        df[ii]=number_or_not
    string_features=[]
    for j in main_labels2: # In this section, non-numeric (string and / or categorical) properties (columns) are detected.
        if df[j].dtype=="object":
            string_features.append(j)
    try: string_features.remove('Label')#The "Label" property was removed from the list. Because it has to remain "categorical" for using with different machine learning approach.
    except: print("error!")
    labelencoder_X = preprocessing.LabelEncoder()
    for ii in string_features: ## In this loop, non-numeric (string and/or categorical) properties converted to numeric features.
        try:  df[ii]=labelencoder_X.fit_transform(df[ii])
        except:
            df[ii]=df[ii].replace('Infinity', -1)
    df=df.drop(main_labels2[61], axis=1) ## Column 61 is deleted because it is unnecessary, column 41 ("Fwd Header Length" feature) had be mistakenly rewritten.
    if flag:
        df.to_csv('all_data.csv' ,index = False)
        flag=False
    else: df.to_csv('all_data.csv' ,index = False,header=False,mode="a")
    os.remove(str(i)+".csv")
    print("The pre-processing phase of the ",csv_files[i]," file is completed.\n")
print("Total operation time: = ",time.time()- seconds ,"seconds")
```

2. Статистика

```
df=pd.read_csv('all_data.csv', usecols=["Label"])
```

```
print(df.iloc[:,0].value_counts())
a=(df.iloc[:,0].value_counts())
key=a.keys()
values=a.values
small_labels=[]
small_values=[]
big_labels=[]
big_values=[]
medium_labels=[]
medium_values=[]
attacak=0
benign=0
for i in range(0,len(values)):
    if values[i]>11000:
        big_labels.append(str(key[i]))
        big_values.append(values[i])
    elif values[i]<600:
        small_labels.append(str(key[i]))
        small_values.append(values[i])
    else:
        medium_labels.append(str(key[i]))
        medium_values.append(values[i])
    if str(key[i])=="BENIGN":
        benign+=values[i]
    else:
        attacak+=values[i]
        key =[benign,attacak]
labels=["BENIGN %"+str(round(benign/(benign+attacak),2)*100),
```

```
        "ATTACK %"+str(round(attacak/(benign+attacak),2)*100)]
graph(big_labels,big_values,"Numbers","Attacks Labels - High-number group")
graph(medium_labels,medium_values,"Numbers","Attacks Labels - Medium-number group")
graph(small_labels,small_values,"Numbers","Attacks Labels - Small -number group")
graph(labels,key,"Numbers","Attack and Benign Percentage")
```

3. фільтрація атак

```
benign=2359289
dict_attack={
"Bot":1966,
"DDoS":41835,
"DoS GoldenEye":10293,
"DoS Hulk":231073,
"DoS Slowhttptest":5499,
"DoS slowloris":5796,
"FTP-Patator":7938,
"Heartbleed":11,
"Infiltration":36,
"PortScan":158930,
"SSH-Patator":5897,
"Web Attack - Brute Force":1507,
"Web Attack - XSS":652,
"Web Attack - Sql Injection":21}
for i in dict_attack:
    a,b=0,0
    ths = open(".\\attacks\\"+i + ".csv", "w")
    ths.write(str(main_labels)+"\n")
```

```
    benign_num=int(benign/(dict_attack[i]*(7/3)))
    with open("all_data.csv", "r") as file:
        while True:
            try:
                line=file.readline()
                line=line[:-1]
                k=line.split(",")
                if k[83]=="BENIGN":
                    rnd=random.randint(1,benign_num)
                    if rnd==1:
                            ths.write(str(line)+"\n")
                            b+=1
                if  k[83]==i:
                    ths.write(str(line)+"\n")
                    a+=1
    ths.close()
    print(i ,"file is completed\n attack:%d\n benign:%d\n\n\n " %(a,b))
webs=["Web Attack - Brute Force","Web Attack - XSS","Web Attack - Sql Injection"]
flag=True
for i in webs:
    df=pd.read_csv(".\\attacks\\"+str(i)+".csv")
    if flag:
        df.to_csv('.\\attacks\\Web Attack.csv' ,index = False)
        flag=False
    else:
        df.to_csv('.\\attacks\\Web Attack.csv' ,index = False,header=False,mode="a")
    os.remove(".\\attacks\\"+str(i)+".csv")
```

4. вибір функцій

```
ths = open("importance_list_for_attack_files.csv", "w")
folder("./feaure_pics/")
for j in csv_files:
    df=pd.read_csv(".\\attacks\\"+j,usecols=main_labels)
    df=df.fillna(0)
    attack_or_not=[]
    for i in df["Label"]:#it changes the normal label to "1" and the attack tag to "0" for use in the machine learning algorithm
        if i =="BENIGN":
            attack_or_not.append(1)
        else:
            attack_or_not.append(0)
    df["Label"]=attack_or_not
    y = df["Label"].values
    del df["Label"]
    X = df.values
    X = np.float32(X)
    X[np.isnan(X)] = 0
    X[np.isinf(X)] = 0
    forest = sk.ensemble.RandomForestRegressor(n_estimators=250,random_state=0)
    forest.fit(X, y)
    importances = forest.feature_importances_
    std = np.std([tree.feature_importances_ for tree in forest.estimators_],
                 axis=0)
    indices = np.argsort(importances)[::-1]
    refclasscol=list(df.columns.values)
```

```
    impor_bars =
pd.DataFrame({'Features':refclasscol[0:20],'importance':importances[0:20]})
    impor_bars =
impor_bars.sort_values('importance',ascending=False).set_index('Features')
    plt.rcParams['figure.figsize'] = (10, 5)
    impor_bars.plot.bar();
    count=0
    fea_ture=j[0:-4]+"=["
    for i in impor_bars.index:
        fea_ture=fea_ture+"\""+str(i)+"\","
        count+=1
        if count==5:
            fea_ture=fea_ture[0:-1]+"]"
            break
    print(j[0:-4],"importance list:")
    print(j[0:-4],"\n",impor_bars.head(20),"\n\n\n")
    print(fea_ture)
```

5. використання модулю та його впровадження;

usecols=["Bwd Packet Length Std","Flow Bytes/s","Total Length of Fwd Packets","Fwd Packet Length Std","Flow IAT Std",
"Flow IAT Min","Fwd IAT Total","Flow Duration","Bwd Packet Length Max","Flow IAT Max","Flow IAT Mean","Total Length of Bwd Packets",
"Fwd Packet Length Min","Bwd Packet Length Mean","Flow Packets/s","Fwd Packet Length Mean","Total Backward Packets","Total Fwd Packets",
"Fwd Packet Length Max","Bwd Packet Length Min",'Label']

(ml_list).

ml_list={

```
"Naive Bayes":GaussianNB(),
"QDA":QDA(),
"MLP":MLPClassifier(hidden_layer_sizes=(13,13,13),max_iter=500),
"Random Forest":RandomForestClassifier(max_depth=5, n_estimators=10, max_features=1),
"ID3" :DecisionTreeClassifier(max_depth=5,criterion="entropy"),
"AdaBoost":AdaBoostClassifier(),
"Nearest Neighbors":KNeighborsClassifier(3)}
others=["Bwd Packet Length Std", "Flow Bytes/s", "Total Length of Fwd Packets", "Fwd Packet Length Std",
    "Flow IAT Std", "Flow IAT Min", "Fwd IAT Total"]
algorithms_features={"Naive Bayes":['Bwd Packet Length Std', 'Total Length of Fwd Packets', 'Flow IAT Min', 'Fwd Packet Length Min', 'Flow Packets/s', 'Fwd Packet Length Mean'] ,
"QDA":['Bwd Packet Length Std', 'Flow Bytes/s', 'Total Length of Fwd Packets', 'Flow IAT Min'],
"MLP":['Bwd Packet Length Std', 'Flow Bytes/s', 'Total Length of Fwd Packets', 'Fwd Packet Length Std',
'Flow IAT Min', 'Bwd Packet Length Max','Fwd Packet Length Min', 'Bwd Packet Length Mean',
'Total Backward Packets', 'Total Fwd Packets', 'Fwd Packet Length Max', 'Bwd Packet Length Min'],
"Random Forest":others,
"ID3" :others,
"AdaBoost":others,
"Nearest Neighbors":others}
seconds=time.time()#time stamp for all processing time
```

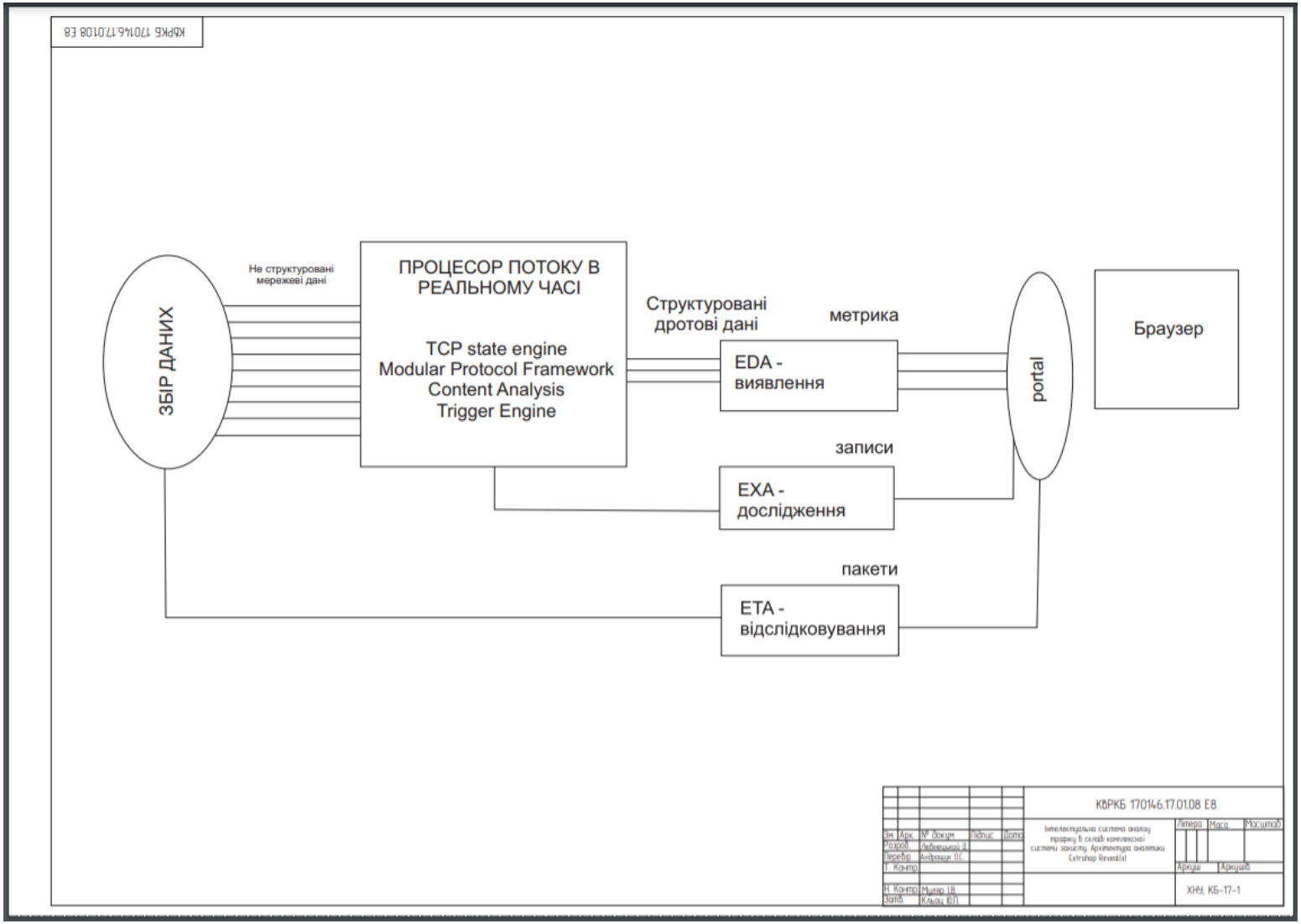
```
with open(result, "w", newline="",encoding="utf-8") as f:#a CSV file is created to save the results obtained.
    wrt = csv.writer(f)
    wrt.writerow(["File","ML algorithm","accuracy","Precision", "Recall" , "F1-score","Time"])
for j in csv_files: #this loop runs on the list containing the filenames.Operations are repeated for all attack files
    print ('%-17s %-17s  %-15s %-15s %-15s %-15s %-15s' % ("File","ML algorithm","accuracy","Precision", "Recall" , "F1-score","Time"))
    feature_list=usecols
    df=pd.read_csv(path+j,usecols=feature_list)#read an attack file.
    df=df.fillna(0)
    attack_or_not=[]
    for i in df["Label"]:
        if i =="BENIGN":
            attack_or_not.append(1)
        else:
            attack_or_not.append(0)
    df["Label"]=attack_or_not
    y = df["Label"]
     del df["Label"]
    feature_list.remove('Label')
    for ii in ml_list: #повторює 7 алгоритмів
        X = df[algorithms_features[ii]]
        precision=[]
        recall=[]
        f1=[]
        accuracy=[]
```

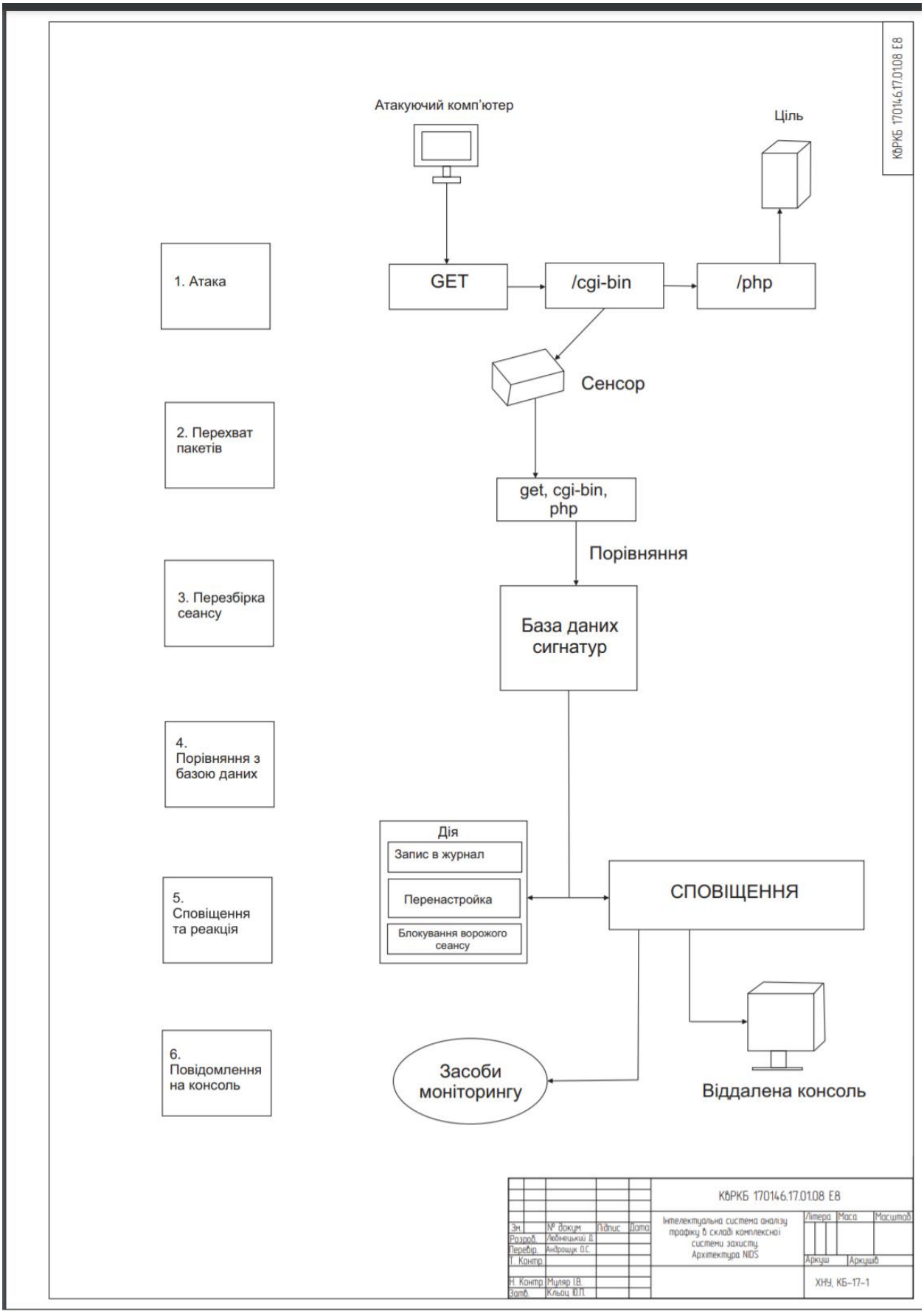
```
t_time=[]
for i in range(repetition): # повторює їх 10 разів
    second=time.time()#time stamp for processing time
    X_train, X_test, y_train, y_test = train_test_split(X, y,
       test_size = 0.20, random_state = repetition)
    #машинне навчання
    clf = ml_list[ii]#choose algorithm from ml_list dictionary
    clf.fit(X_train, y_train)
    predict =clf.predict(X_test)
          f_1=f1_score(y_test, predict, average='macro')
    pr=precision_score(y_test, predict, average='macro')
    rc=recall_score(y_test, predict, average='macro')
    precision.append(float(pr))
    recall.append(float(rc))
    f1.append(float(f_1))
    accuracy.append(clf.score(X_test, y_test))
    t_time.append(float((time.time()-second)) )
print ('%-17s %-17s  %-15s %-15s %-15s %-15s %-15s' % (j[0:-4],ii,str(round(np.mean(accuracy),2)),str(round(np.mean(precision),2)),
str(round(np.mean(recall),2)),str(round(np.mean(f1),2)),str(round(np.mean(t_time),4))))
with open(result, "a", newline="",encoding="utf-8") as f: # записуємо це все у файл.
    wrt = csv.writer(f)
    for i in range(0,len(t_time)):
       wrt.writerow([j[0:-4],ii,accuracy[i],precision[i],recall[i],f1[i],t_time[i]])
```

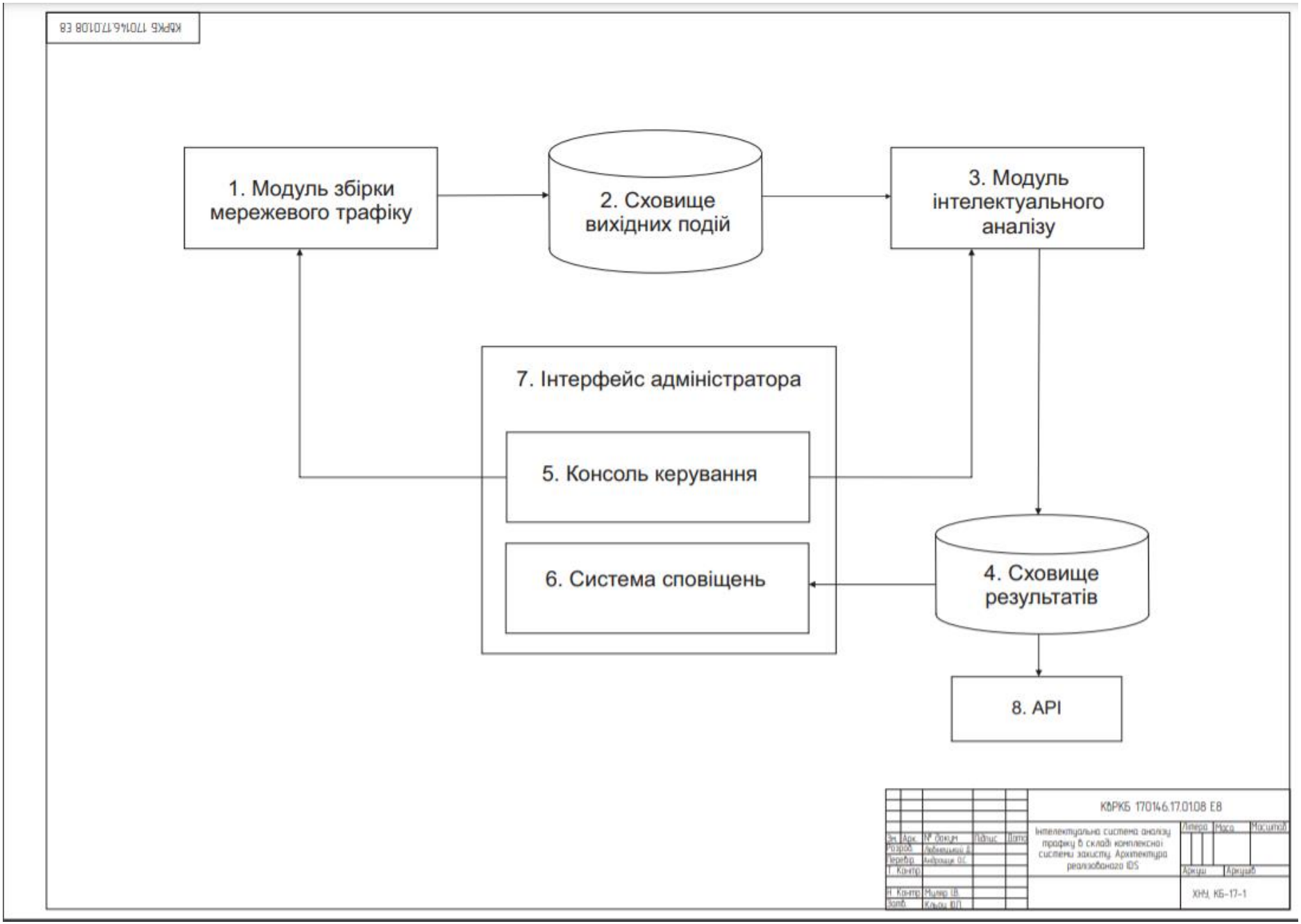
ДОДАТОК Б

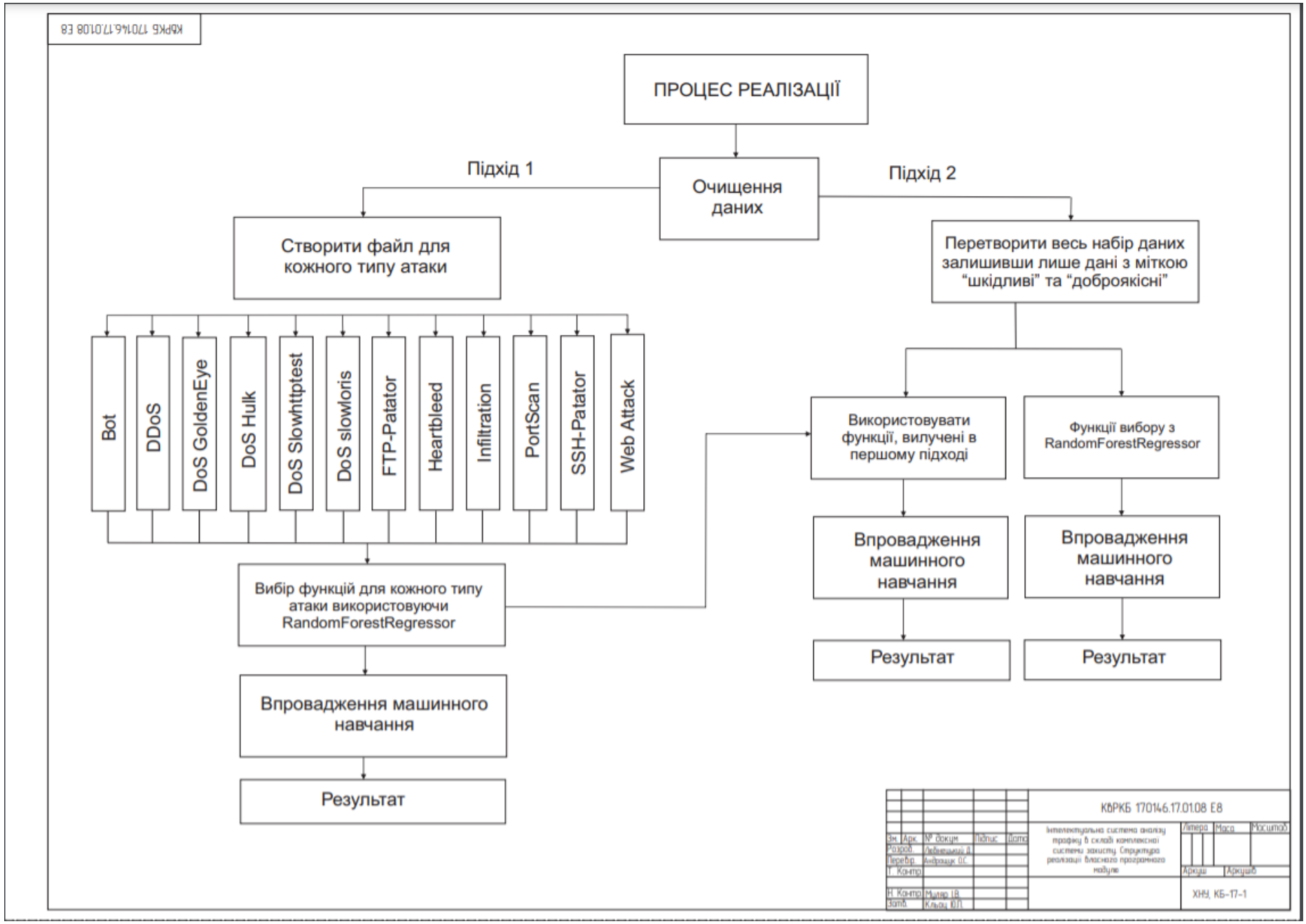
(Обов’язковий)

Копія графічної частини




КВРКБ 170146.17.0108 Е8
Атакуючий комп'ютер
Ціль
1. Атака
GET
/cgi-bin
/php
Сенсор
2. Перехват пакетів
get, cgi-bin, php
Порівняння
3. Перезбірка сеансу
База даних сигнатур
4. Порівняння з базою даних
Дія
Запис в журнал
Перенастройка
Блокування ворожого сеансу
5. Сповіщення та реакція
СПОВІЩЕННЯ
6. Повідомлення на консоль
Засоби моніторингу
Віддалена консоль
КВРКБ 170146.17.0108 Е8
Інтелектуальна система аналізу трафіку в складі комплексної системи захисту. Архітектура NIDS
Літера
Маса
Масштаб
Аркуш
Аркушів
ХНУ, КБ-17-1
Зм.
№ докум.
Підпис
Дата
Розроб.
Любинецький Д.
Перевір.
Андрощук О.С.
Т. Контр.
Н. Контр.
Муляр І.В.
Затв.
Кльоц Ю.Л.


КВРКБ 170146.17.0108 Е8
1. Модуль збірки мережевого трафіку
2. Сховище вихідних подій
3. Модуль інтелектуального аналізу
7. Інтерфейс адміністратора
5. Консоль керування
6. Система сповіщень
4. Сховище результатів
8. API
КВРКБ 170146.17.0108 Е8
Інтелектуальна система аналізу трафіку в складі комплексної системи захисту. Архітектура реалізованого IDS
ХНУ, КБ-17-1


ПРОЦЕС РЕАЛІЗАЦІЇ
Очищення даних
Підхід 1
Підхід 2
Створити файл для кожного типу атаки
Bot
DDoS
DoS GoldenEye
DoS Hulk
DoS Slowhttptest
DoS slowloris
FTP-Patator
Heartbleed
Infiltration
PortScan
SSH-Patator
Web Attack
Вибір функцій для кожного типу атаки використовуючи RandomForestRegressor
Впровадження машинного навчання
Результат
Перетворити весь набір даних залишивши лише дані з міткою “шкідливі” та “доброякісні”
Використовувати функції, вилучені в першому підході
Функції вибору з RandomForestRegressor
Впровадження машинного навчання
Впровадження машинного навчання
Результат
Результат
КВРКБ 170146.17.01.08 ЕВ
ХНУ, КБ-17-1

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: Інтелектуальна система аналізу трафіку в складі комплексної системи захисту

Автор: Любінецький Денис Володимирович

Спеціальність: 125 – Кібербезпека

Освітня програма: Кібербезпека

Керівник: Андрощук Олександр Степанович, д.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) всі зафіксовані системою ознаки модифікації тексту відносяться до комбінування латинських символів зі україномовними скороченнями індексів в формулах, що не є модифікацією тексту.

Сумарний обсяг всіх запозичень, визначений системою виявлення збігів/ ідентичності/схожості, складає 3.37% що, з урахуванням наведених обгрунтувань, відповідає характеру наукового дослідження і свідчить на користь кваліфікаційної роботи.

Керівник роботи О.С. Андрощук

Завідувач кафедри КбКСМ Ю.П. Кльоц

Доцент кафедри КбКСМ, Гарант ОП В.М. Чешун

Дата: 07.06.2021

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

РЕЦЕНЗІЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ

освітнього ступеня «бакалавр»

Студент Любінецький Денис Володимирович

Тема Інтелектуальна система аналізу трафіку в складі комплексної системи захисту

Спеціальність 125 – Кібербезпека

Обсяг кваліфікаційної роботи освітньо-кваліфікаційного рівня «бакалавр»:

кількість листів креслень 4 ; кількість сторінок записки 61 .

1. Короткий зміст роботи та прийнятих рішень У кваліфікаційній роботі розроблено та описано методи та способи аналізу мережевого трафіку .

2. Висновок про відповідність кваліфікаційної роботи завданню Кваліфікаційна робота у повній мірі відповідає поставленому завданню як в теоретичній, так і в практичній частині

3. Характеристика виконання кожного розділу роботи, ступінь використання останніх досягнень науки і техніки і передових методів роботи: У вступі подана загальна загальна характеристика поставленої задачі. В першому розділі описано загальні відомості та актуальність проблем захисту інформацію на підприємствах та організаціях. В другому розписано способи реалізації, власні рішення які повинні допомогти в знаходженні зловмисного трафіку. Крім того описаний власний інтелектуальний модуль з різними варіантами дата сетів. Також розписаний власний програмний модуль який допомагає в знаходженні та виявлені зловмисного трафіку. В завершенні представлені результати на конкретних числах.

4. Позитивні сторони роботи Кваліфікаційна робота має комплексну практичну та теоретичну цінність. Практична цінність полягає у розробці власного модуля інтелектуального аналізу з допомогою якого визначається аномалії мережевого трафіку. На основі даного аналізу в подальшому здійснюється подальші дії та способи протидії, контрзаходи. Практична цінність результатів кваліфікаційної роботи полягає у створені системи захисту від витоків даних, що може бути підлаштована для будь якої категорії даних, і у описі оптимальних моделей функціонування систем захисту подібного характеру. Теоретична цінність полягає в тому, що у власному програмномуу модулі який можна використовувати в реальній системі захисту. Такий модуль не потребує додаткового фінансування.

5. Негативні сторони роботи Розроблений програмний модуль є не дуже потужним, через невелику вибірку даних

6. Оцінка графічного оформлення та пояснювальної записки роботи Графічне оформлення виконане відповідно до теми кваліфікаційної роботи з дотриманням стандартів. В загальному графічне оформлення виконане якісно, пояснювальна записка відповідає нормам щодо її оформлення.

7. Відгук про роботу в цілому Робота студента, загалом заслуговує позитивної оцінки. Весь матеріал кваліфікаційної роботи структурований, чіткий та послідовний. Усі розділи роботи послідовні та логічні, що дозволяє чітко розуміти викладений матеріал в рамках тематики кваліфікаційної роботи. Розроблені методи є дуже актуальними на сьогоднішній день.

8. Інші зауваження Не враховані інші потужні модулі при реалізації власного модулю.

9. Оцінка кваліфікаційної роботи Враховуючи всі позитивні та негативні сторони представленої кваліфікаційної роботи, можна зробити висновок, що вона заслуговує оцінку «відмінно».

РЕЦЕНЗЕНТ (прізвище, ім'я, по батькові, посада, місце роботи)

Говорущенко Тетяна Олександрівна
Завідувач кафедри, д.т.н., професор

« 7 » 06 2021. (підпис)

Anti-Plagiarism v-15.257

Максимальное совпадение с одним документом 0.0%

Словари проверки: en_US, ru_RU, ua_UA. **Ошибок в документах: 13%**

ID: 92337 Название: Інтелектуальна система аналізу трафіку в складі комплексної системи захисту Добавлено в БД: 2021-06-04 Авторы: Любінецький Д.В Руководители: Андрощук О.С Консультанты: Опоненты:	Документ		Суммарное совпадение по Базе Данных	
	Символы	Лексемы	Символы	Лексемы
	72401	589	0 (0%)	0 (0%)

Источник плагиата

ID	Описание	Наличие плагиата в документе	
		Символы	Лексемы

UNICHECK

User name:
Кафедра кибербезпеки

Check ID:
1008178156

Check date:
04.06.2021 14:14:36 EEST

Check type:
Doc vs Internet

Report date:
04.06.2021 14:15:48 EEST

User ID:
100005590

File name: **Кваліф._робота_Любінец_к._Денис_КБ-17 на плагіат**

Page count: **61** Word count: **10666** Character count: **81443** File size: **7.35 MB** File ID: **1008256201**

3.37% Matches

Highest match: **0.61%** with Internet source (**https://github.com/naviprem/ids-deep-learning/blob/master/datasets/UNSW-NB15.m**

3.37% Internet sources | **46** ..... Page 63

No Library search was conducted

0% Quotes

Exclusion of quotes is off

Exclusion of references is off

0% Exclusions

No exclusions