НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ
Кафедра інформаційної безпеки

«До захисту допущено»
В.о. завідувача кафедри

_____ М.В.Грайворонський
(підпис)
“___” _____2019 р.

Дипломна робота

на здобуття ступеня бакалавра

з напряму підготовки 6.170101 «Безпека інформаційних і комунікаційних систем»
на тему: Модель профілювання поведінки користувачів за допомогою методу опорних векторів _____

_____

Виконав (-ла): студент (-ка) _4__ курсу, групи __ФБ-52_____
(шифр групи)

_____Перегудов Станіслав Вікторович_____ _____
(прізвище, ім’я, по батькові) (підпис)

Керівник доцент, к. т. н. Барановський Олексій Миколайович_____
(посада, науковий ступінь, вчене звання, прізвище та ініціали) (підпис)

Консультант _____ _____ _____
(назва розділу) (посада, вчене звання, науковий ступінь, прізвище, ініціали) (підпис)

Рецензент_____ _____
(посада, науковий ступінь, вчене звання, науковий ступінь, прізвище та ініціали) (підпис)

Засвідчую, що у цій дипломній роботі немає запозичень з праць інших авторів без відповідних посилань.
Студент _____
(підпис)

Київ - 2019 року

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ
Кафедра інформаційної безпеки

Рівень вищої освіти – перший (бакалаврський)

Напрям підготовки 6.170101 «Безпека інформаційних і комунікаційних систем»

ЗАТВЕРДЖУЮ
В.о. завідувача кафедри
_____ М.В.Грайворонський
(підпис)
«___»_____2019 р.

ЗАВДАННЯ
на дипломну роботу студенту

Перегудов Станіслав Вікторович
(прізвище, ім'я, по батькові)

1. Тема роботи Модель профілювання поведінки користувачів за допомогою методу опорних векторів,
науковий керівник роботи доцент, к. т. н. Барановський Олексій Миколайович ,
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)
затверджені наказом по університету від «____ » 2019 р. № _____

2. Термін подання студентом роботи 10 червня 2019 р.

3. Вихідні дані до роботи _____

4. Зміст роботи_____

5. Перелік ілюстративного матеріалу (із зазначенням плакатів, презентацій тощо) _____

6. Дата видачі завдання _____

РЕФЕРАТ

Робота обсягом 60 сторінок містить 17 ілюстрацій, 6 таблиць, 16 формул та 20 літературних посилань.

У роботі було розглянуто різноманітні механізми профілювання поведінки користувачів та методи, за допомогою яких можливо вдало створювати поведінкові профілі. В роботі було використано власну модель профілювання поведінки користувачів за допомогою методу опорних векторів, в основі якого лежить запропонована функція ядра.

Результати роботи можуть бути використанні для побудови системи профілювання поведінки користувачів, що значно підвищить рівень захищеності внутрішньої мережі інфраструктури будь-якої організації або компанії.

Ключові слова: профілювання поведінки користувачів, метод опорних векторів, аналіз поведінки користувачів, модель користувача.

ABSTRACT

A 60 pages' work contains 17 illustrations, 6 tables, 16 formulas and 20 references. The work examined various mechanisms for profiling user behavior and methods by which it is possible to successfully create behavioral profiles. In this work, we used our own model for profiling user behavior using the support vectors method, which is based on the proposed kernel function.

The results of the work can be used to build a user profiling system, which will significantly increase the level of security in internal network of any organization or company.

Keywords: user behavior profiling, the support vector method, user behavior analysis, user model.

ЗМІСТ

ВСТУП

Згідно дослідження проведеного Crowd Research Partner[1] у 2018 році 90 відсотків опитаних організації відчувають себе вразливими до атак в середині власної інфраструктури, а більше ніж 52 відсотки відкрито заявили, що впродовж року хоч раз ставали жертвою подібних атак.

На сьогоднішній день, склалась негативна тенденція – при побудові систем інформаційної безпеки більшість компанії не забезпечують достатній рівень захищеності внутрішньої мережі, зосереджуючи свою уваги на зовнішніх загрозах. Але наскільки не були б надійними механізми протидії несанкціонованим вторгненням, зловмисник все одно потрапить до внутрішньої мережі в тій чи інші точці, а традиційних методів виявлення потенційно небезпечних дії в середині інфраструктури вже не достатньо. Більшість порушників вже майже ідеально знають всі механізми захисту, призначені для виявлення інцидентів безпеки. SIEM системи зовсім не готові до виявлення нових векторів розвитку атаки, а системи аналізу поведінки користувачів, мають значні недоліки в принципах побудови профілів користувачів та не зовсім ефективно використовують новітні методи машинного навчання.

Все це та багато іншого змушує нас по новому розглянути механізми виявлення зловмисників у внутрішній мережі за допомогою аналізу поведінки користувачів та побудови профілю їх поведінки.

Актуальність роботи полягає в тому, що сучасні механізми профілювання поведінки користувачів не до кінця ефективні, через що впродовж тривалого часу зловмисники можуть перебувати у внутрішній мережі непоміченими.

Метою даної роботи є підвищення рівня захищеності внутрішньої мережі інфраструктури, шляхом розробки власної моделі профілювання поведінки користувачів за допомогою метода опорних векторів, в основі якого лежить запропонована функція ядра класифікатора. Для досягнення поставленої мети були поставлені та вирішені такі завдання:

1) Розглянути основні недоліки існуючих методів виявлення інцидентів безпеки.
2) Розглянути існуючі методи побудови поведінкових профілів користувачів.
3) Порівняти найпоширеніші алгоритми профілювання з методом опорних векторів.
4) Розробити власну модель профілювання поведінки користувачів за допомого метода опорних векторів.
5) Впровадити розроблену модель профілювання.

Об’єктом дослідження виступає механізм побудови поведінкових моделей користувачів.

Предметом дослідження є модель профілювання поведінки користувачів, в основі якої лежить метод опорних векторів.

Наукова новизна одержаних результатів полягає в тому, що в результаті роботи було розроблено власну модель профілювання поведінки користувачів, а також запропоновано функцію ядра класифікатора для вирішення задач профілювання поведінки.

Практичне значення одержаного результату є досить вагомим, оскільки дозволяє якісно профілювати поведінку користувачів, за допомогою чого можна заздалегідь виявити зловмисника у мережі.

1 ВИЯВЛЕННЯ ІНЦЕДЕНТІВ БЕЗПЕКИ У МЕРЕЖІ

Даний розділ присвячено сучасним системам виявлення інцидентів безпеки у внутрішній мережі організації, а також їх основним недолікам, що заважають їм своєчасно знаходити порушників, що вже потрапили до мережі.

1.1 Системи SIEM

В процесі свого функціонування, кожний пристрій, починаючи від персонального комп'ютера закінчуючи складними системами штучного інтелекту, записує усі події, що відбуваються у його системі, в журнали реєстру. Сама завдяки даними записаним в цих журналах, адміністратор може дати оцінку діяльності системи і сказати, які саме операції відбувалися в системі в певний момент часу. На сьогоднішній день, аналіз записів журналу є одним з провідних методів пошуку потенційно небезпечних дій у системі. Але через складність інфраструктури та різноманіття пристроїв, кількість журналів реєстрації події може досягати кількох десятків і аналізувати їх окремо один від одного майже не можливо. Саме для вирішення проблеми централізованого збору та аналізу всіх події були створені перші системи, що пізніше отримали назву SIEM.

Вперше термін SIEM було введено Gartner у 2005 році і являло собою злиття двох вже існуючих, на той час, термінів: SIM (Security Information Management) – управління інформаційною безпекою, SEM (Security Event Management) – управління подіями безпеки. Згідно з цим визначенням SIEM-система повинна вміти збирати та аналізувати інформацію з різноманітних мережевих пристроїв та пристроїв інформаційної безпеки, а також містити в собі системи управління доступом, інструменти управління вразливостями і базами даних. Тобто SIEM – програмний пристрій для централізованого збору даних журналів реєстрації подій з різноманітних робочих машин користувачів, мережевих пристроїв та пристроїв інформаційної безпеки, для їх подальшої категоризації, класифікації, та аналізу.

Кожне підприємство встановлює дані системи для вирішення власних цілей, але в загальному, перед системою встановлюється дві основні задачі:

- створення та надання звітів про інциденти інформаційної безпеки, таких як діяльність програмного забезпечення на серверах, автентифікація користувачів на пристроях інформаційної діяльності;
- аналіз інцидентів інформаційної безпеки та своєчасне попередження адміністратора про потенційну небезпеку, якщо діяльність порушує заздалегідь створень правила і набори політик безпеки.

Для виконання завдання, висунутих системі, робота з подіями проходить у декілька етапів:

- збирання даних – дані в системі збираються з усіх джерел, починаючи від операційної системи, встановленої на робочі машині користувача, закінчуючи пристроями інформаційної безпеки: системи протидії вторгненням, мережеві екрани, антивірусне забезпечення; збір даних відбувається за допомогою спеціальних агентів, хоч деякі сучасні рішення дозволяють збирати данні і без них;
- нормалізація даних – система збирає дані з величезної кількості журналів різних систем, що записують данні у різному форматі, тому на даному етапі система приводить до одного структурного формату;
- кореляція даних – на даному етапі система проводить аналіз отриманих даних, аналіз відбувається на основі правил, що надаються розробниками SIEM або створені та налаштовані адміністратором з інформаційної безпеки; правила кореляції визначає певну послідовність дій, що допомагає виділити більш важливі події, що відбуваються в системі
- формування звітів – можливість візуалізації даних або надання їх в більш зручній формі для обробки аналітиком;
- сповіщення – попередження адміністратора з інформаційної безпеки, про потенційно небезпечні дії, що відбуваються у системі.

Наведений вище список може доповнюватись іншими функціональними рішеннями, якщо цього вимагають цілі та задачі підприємства, на якому встановлена система, але він є основним і кожна сучасна SIEM дотримується даних етапів.

У сучасних системах безпеки та управління подіями аналіз даних, отриманих з журналів реєстру, може основуватись на двох різних принципах:

- Використання правил кореляції;
- Побудова моделей користувачів та активів.

Правила кореляції – логічний вираз, за допомогою якого система вирішує що саме вона повинна зробити. Наприклад, якщо система фіксує небезпечну мережеву активність, то вона повинна попередити адміністратора. Правила можуть приймати рішення на основі лише однієї події, але в більшості випадків правила об'єднуються в одне, більш складне, і робить свій висновок на основі декількох події. Складні правила використовуються для зменшення похибок першого та другого роду. Наприклад, в систему приходять дані, що свідчать про відключення механізмів антивірусного захисту на одній з користувацьких машин. Просте правило зробить свій висновок ґрунтуючись лише на даних про вимкнення механізмів захисту і тому одразу повідомить адміністратора. В той час як складне правило спочатку почекає дані, що будуть свідчити про вимкнення машини в цілому, і лише у випадку не надходження цих даних повідомить адміністратора.

Основним недоліком даного методу є те, що система основана на правилах кореляції не здатна приймати рішення на основі минулого досвіду. Тобто правила потребують постійної модифікації, яку виконує адміністратор з інформаційної безпеки.

Модель будується на основі нормальної поведінки користувача або активу. Основною відмінністю моделей від правил кореляції є те, що вони не завжди оцінюються, а працюють лише при відхиленні поведінки від звичайної. Зазвичай в моделях сотворяють правила, які класифікують різні типи поведінки, задля створення різних способів оповіщення.

Моделі формуються набагато легше за правила кореляції, бо їх логічні вирази менші та простіші.

Як можна побачити, моделі мають значні переваги перед правилами кореляції, але існують ситуації в яких простіше використовувати правила, наприклад:

- моніторинг відомих загроз – основні правила, які поставляються розробником, правила спроможні легко виявляти загальні загрози, що ні разу не використовуються;
- виявлення загроз на основі сигнатур – можливість задати вже відомі сигнатури до бази даних і відстежувати їх.

Моделі вибираються у випадку:

- неможливість однозначно визначити подію, що ідентифікує небажану поведінку;
- динамічні роботи умови, що роблять занадто складними або викликають велику кількість похибок першого і другого роду.

Загалом моделі використовують для виявлення загроз на основі поведінки користувачів. Більшість готових систем вже мають в своєму складі моделі для оцінки поведінки користувачів та активів, але вони мають низьку ефективність і велику похибку першого або другого роду. Тому завжди вимагають пере налаштування або створення власних моделей від адміністратора, що безпосередньо працює з SIEM.

1.2 Сучасні SIEM-рішення

На сьогоднішній день, існує велика кількість різноманітних SIEM-рішень від великої кількості розробників, адже ринок даних систем існує трохи менше п’ятнадцяти років. Але загальну більшість ринку займають лише декілька провідних кампанії. Згідно з квадрантом Gartner[2] це - IBM, Splunk, Micro Focus, Logrhythm, McAfee.

1.2.1 Qradar

Згідно інформації на сайті розробника IBM[3], Qradar – сучасне SIEM-рішення здатне реєструвати події більш ніж з тисячі кінцевих пристроїв. Ця система виконує миттєва нормалізацію і виявляє зв'язок між діями над необробленими даними, щоб відрізнити реальні загрози від помилкових спрацьовувань.

Рішення складається з декількох функціональних рішень, кожне з яких виконує окремі задачі:

- QRadar Log Manager – управління журналами, відповідає за збереження подій;
- QRadar SIEM – займається аналізом загроз та ризиків;
- QRadar Risk Manager – моделюю можливі загрози та ризики для системи, в якій встановлений;
- QRadar QFlow – пошук мережевих аномалій трафіку та його аналіз;
- QRadar vFlow – монітор рівня застосувань для фізичного та віртувльного середовища.

Переваги:

- виявлення неправильного використання додатків, внутрішнє шахрайство і невеликі загрози, які можна не помітити серед мільйонів подій;
- виконання миттєвої нормалізації подій і зіставлення їх з іншими даними, отриманими в результаті виявлення загроз, створення звітів про відповідність вимогам і проведення аудиту;
- Скорочення числа подій і потоків з мільярдів до невеликої кількості реальних порушень і визначення пріоритетів для них відповідно до загрозою для бізнесу;
- Використання опціонального ПО IBM Security X-Force Threat Intelligence для визначення дій, пов'язаних з підозрілими IP-адресами, наприклад, при підозрі у шкідливої активності;

- Доповнення у вигляді пристроїв IBM Security QRadar QFlow і IBM Security QRadar VFlow Collector для отримання глибокого розуміння і кращого відображення додатків (наприклад, додатків, які керують ресурсами підприємства), баз даних, продуктів для спільної роботи і соціальних мереж за допомогою аналізу мережевих потоків на рівні 7;
- Виконання об'єднаного пошуку у великих розподілених середовищах;
- Автоматичне виявлення більшості джерел надають журнали і моніторинг мережевих потоків для пошуку і класифікації комп'ютерів і серверів, відстеження додатків, протоколів, служб і портів, які вони використовують для істотної економії часу.

Недоліки:

- обмежені можливості персоналізації та налаштування;
- обмежена підтримка оренди одразу декількох систем;
- обмежена можливість розробки більш досконалих методів аналізу та кореляції;
- використання не досить досконалих правил аналізу поведінки користувача.

1.2.2 ArcSight

Одним з найкращих на даний момент рішень є SIEM ArcSight від компанії Micro Focus[4], згідно з інформацією наданою розробником, на офіційній сторінці, на сьогоднішній день система доступна в трьох різних варіантах:

- Платформа даних Arcsight Data Platform, яка забезпечує збір журналів, управління і генерацію звітів.
- Програмне забезпечення Arcsight Enterprise Security Management (ESM), призначене для розгортання широкомасштабного моніторингу безпеки.
- Програмне-апаратний комплекс Arcsight Express, заснований на пристроях «все в одному» і орієнтований на використання з попередньо

сконфігурованим моніторингом і звітністю, а також спрощеним управлінням даними.

Переваги даного рішення:

- Arcsight ESM надає повний набір можливостей SIEM, які можуть використовуватися для підтримки великомасштабного SOC, включаючи повний робочий процес розслідування інцидентів та управління, а також спеціальну консоль управління розгортанням;
- HPE User Behavior Analytics виявляє аномалії на основі аналізу поведінки користувачів і доповнює традиційну кореляцію, яка є базовою функцією arcsight;
- DNS Malware Analytics аналізує DNS-трафік і забезпечує повну видимість IT-інфраструктури, що допомагає виявити уразливі місця ще до того, як ними скористаються зловмисники. Ідея аналізу DNS-трафіку с метою виявлення зловмисної активності зародилася в дослідницькому підрозділі HP Labs близько п'яти років тому;
- Arcsight Threat Central містить інтерактивну базу знань загроз і дозволяє обмінюватися відомостями про способи їх виявлення та ліквідації. На порталі ArcSight Marketplace містяться правила (пакети безпеки) і додаткові додатки;
- розробники з HPE сподіваються, що до формування таких пакетів безпеки і створення додаткових додатків підключаться і партнери компанії. HPE arcsight має широкий вибір готових до використання сторонніх технологій і конекторів.

Недоліки:

  - велика складність розгортання та системи ті її подальшого налаштування;
  - не ефективність розгортання системи на підприємствах малих та середніх розмірів;

- потребує гарних навичок для управління і супроводу;
- висока складність навчання;
- не ефективні правила кореляції аналізу поведінки користувача.

1.2.3 Splunk

Splunk Enterprise[5] - це провідна в галузі платформа для операційної аналітики. Збирайте і індексуються будь машинні дані практично з будь-якого джерела в реальному часі. Головною особливістю є її можливість працювати з будь-якими даними як до неї надходять. Вся робота у системі виконується за рахунок запитів написаних на спеціальній мові SPL.

Дана система складається з чотирьох основних функціональних рішень:

- Incident Review - гнучкий інструмент переглядати та упорядковувати інцидентами, збагачений інформацією з зовнішніх джерел;
- Investigator - візуальний інструмент виявлення Kill Chain-атаки і створення нових кореляційних пошуків на базі отриманого досвіду;
- Glass Tables - наочне побудова логічних схем захищаємих ресурсів з вбудованим редактором. Можливість створення індивідуально налаштованих візуалізацій з ключовими показниками роботи SOC;
- Security Intelligence - великий набір налаштованих інтеграцій з зовнішніми джерелами інформації про загрози, включаючи інтеграцію з Facebook Threat Exchange.

Переваги:

- Splunk здійснює збір, пошук, моніторинг та аналіз за досить великим обсягом даних, як в режимі історичного пошуку, так і в реальному часі, видаючи швидкий результат і високу інтерактивність пошукових запитів на надзвичайно великих обсягах даних. Splunk є повноцінною Big Data платформою;

- Splunk є універсальною системою для машинних даних, яка забезпечує комплексний збір даних, їх обробку та аналіз. Таким чином система здатна об'єднати в собі машинні дані, бізнес дані, призначені для користувача дані і будувати аналітику в різних розрізах, що робить її вкрай універсальним;
- Splunk використовує технологію MapReduce, що забезпечує розподіл навантажень і швидку горизонтальну масштабованість системи. Також завдяки технології MapReduce зростає її продуктивність.

  Недоліки:
- система має суттєві обмеження при створені власних правил кореляції;
- складність налаштування функціоналу, безпосередньо пов’язаного з системою.

1.2.4 McAfee Enterprise Security Manager

McAfee Enterprise Security Manager – сучасне SIEM-рішення від одного з провідних розробників McAfee[6], постачаєтеся як програмне рішення так і апаратне. McAfee ESM включає джерела інформації про загрози, кореляцію, аналітику, профілювання, оповіщення про безпеку, презентацію даних та відповідність. Вона пропонує інтелект і інтеграцію для визначення пріоритетів, розслідування і реагування на загрози, в той час як вбудована система відповідності і вбудовані пакети вмісту безпеки спрощують операції аналітиків і відповідності. ESM є основним продуктом портфеля рішень SIEM компанії McAfee, яка включає в себе:

- McAfee Enterprise Log Manager (ELM) – займається збиранням та збереженням даних;
- McAfee Advanced Correlation Engine (ACE) – розширений механізм кореляції, спроможній до аналізу в режимі реального часу;
- McAfee Event Receiver (ERC) – додатковий збір даних журналу та потоку даних;

- McAfee Database Monitor Event (DEM) – монітор транзакцій бази даних та журналу;
- McAfee Application Monitor Data (ADM) – моніторинг подій прикладного рівня;
- McAfee Global Threat Intelligence (GTI).

McAfee ESM пропонує інтеграцію з десятками додаткових рішень з управління інцидентами та аналітикою, включаючи обмін інформацією McAfee Threat Intelligence.

Переваги:

- Enterprise Security Manager має хороші охоронні промислові системи управління (ICS) та пристрої диспетчерського управління та збору даних (SCADA);
- McAfee Data Exchange Layer (DXL) від Intel Security забезпечує інтеграцію зі сторонніми технологіями без використання API. Цей підхід дає можливість для використання ESM в якості платформ SIEM;
- McAfee Global Threat Intelligence дозволяє розширювати можливості SIEM-системи Enterprise Security Manager, доповнюючи поточну інформацію про загрози, що дозволяє швидко виявити події, які включають в себе небезпечні з’єднання.

Недоліки:

- невеликі можливості створення правил кореляції порівняно з іншими системами;
- обмеження в інтерфейсі користувача, складана навігація у системі;
- для ефективної роботи потрібно використовувати велику кількість агентів, що буду займатися пересиланням даних за журналів реєстру до SIEM систем;
- відсутня можливість аналізу великих об’ємів даних;
- відсутність підтримки при використані двох або більше рішень;
- обмежена можливість налаштування інтерфейсу.

1.2.5 Elastic stack

Elastic stack – рішення з відкритим кодом, що будується на чотирьох компонентах:

- Beats – агенті, що зазвичай розташовуються на кінцевих хостах і відповідають за відправку даних журналів до системи;
- Logstash – займається обробкою даних отриманих від агенті або безпосередньо з журналів реєстрації подій;
- Kibana – візуалізація даних, що зберігаються в системі;
- Elasticsearch – один з основних компонентів, в якому зберігаються оброблені дані.

Тобто, збирання даних відбувається за допомогою агенті, частіше за все це Filebeat, і Logstash, який нормалізує дані, що до нього надійшли і відправляє їх далі. Після обробки всі дані зберігаються у Elasticsearch.

Для створення ефективних правил кореляції події використовується Kibana, в якій створюються запити на основі аналізу проведеного у Logstash.

Переваги:

- система має можливість збирати та аналізувати будь-які дані з будь-яких журналів;
- рішення побудовано на використанні компонентів з відкритим кодом, що дозволяє будь-кому розвертати дану систему, незалежно від розмір підприємства;
- має надзвичайно потужний механізм візуалізації та аналізу у вигляді Kibana;
- надзвичайно зручний та швидкий пошук даних в базі даних.

Недоліки:

- Не налаштована певним чином, система не може називатися SIEM, адже не задовольняє основним характеристикам висунутими Gartner;
- Розробник не постачає жодних правил кореляції, тому всі потрібні правила потрібно буде створювати користувачу;

- Без додаткових компонентів не має механізму сповіщення адміністратора безпеки про потенційну загрозу.

1.3 Сучасні загрози

Як вже було сказано вище, системи SIEM використовують для збору та аналізу всіх події, що відбуваються в інфраструктурі підприємства. Тобто в ідеалі, система фіксує всі потенційно небезпечні події і завчасно попереджує адміністратора. Але кожне SIEM-рішення має свої недоліки, які дозволяють зловмисникам реалізовувати свої атаки.

Для прикладу, візьмемо вдалі атаки на банківські установи у період з 2017 по 2018 рік. Адже, на сьогоднішній день, кожна банківська установа використовує хоч одне з SIEM-рішень, описаних вище.

Згідно статистки[7] за період з 2017 по 2018 роки було зафіксовано більше декількох сотень спроб несанкціонованого доступу до внутрішньої мережі банківської установи, більшість з яких були вдалими. За неофіційними даними[8], лише на початок осені 2018 року загальна сума вкрадених грабіжниками грошей досягала одного мільярду доларів.

1.3.1 Порушення інформаційної безпеки банківської установи

Аналізуючи найбільші атаки за даний період, можна побачити, що, хоч атаки здійснені різними злочинними організаціями, але в загальному всі спроби крадіжки можна розглядати за одною загальною схемою:

- початкова розвідка та підготовка до вторгнення;
- проникнення у внутрішню мережу;
- закріплення та загальна розвідка в середині мережі;
- отримання безпосереднього доступу до основних банківських операцій у системі;

- знищення слідів несанкціонованого втручання.

При підготовці до вторгнення, зловмисники намагаються отримати якнайбільше різноманітної інформації про інфраструктуру організації, починаючи від домених імен пов'язаних з установою, закінчуючи специфікацією захисник систем. Для отримання даної інформації порушники користуються методами пасивної розвідки або знаходять недобросовісного працівника, що згоден за винагороду зібрати всю потрібну зловмисникам інформацію про банківську установу.

Після етапу збору інформації, зловмисники починають діяти. В першу чергу їм потрібно потрапити до внутрішньої мережі жертви-банку.

Аналізуючи вдалі атаки, виконані зловмисниками, можемо виділити два основних метода проникнення:

- Розсилання фішингових листів працівникам установи;
- Компрометація сторонніх організацій, що надають свої послуги банку-жертві або безпосередньо пов'язані з ним важливими бізнес-процесами.

При розгляді другого етапу, більш детально зупинимось на першому методі проникнення, адже його використовують більш ніж в 70 відсотках всіх атак, а також він більш цікавий при розгляді методів виявлення вторгнень за допомогою використання SIEM-рішень.

При проникненні у мережу, методом розсилання фішингових листів, деяким працівникам банківської установи на пошту починають надсилати листи з вкладеними у них різноманітними офісними документами. Більшість працівників, яким надходять дані листи, майже не в змозі відрізнити фішинговий лист від справжнього. Адже листи надсилаються з заздалегідь скомпрометованих поштових адресів і за формою та стилем зовсім не відрізняються від офіційної кореспонденції[9].

Після завантаження і відкриття, вкладеного у лист документу, на робочій станції жертви починав виконуватися макрос, за допомогою якого зловмисники отримували віддалений доступ до машини працівника. За приклад можемо взяти

метод, який експлуатували зловмисники Cobalt упродовж осені 2018 року. Згідно аналізу[10] даного методу, можемо сказати, що основною цілю макроса є виклик легітимних процесів операційної системи та подальше завантаження і встановлення бекдору. Даний метод використовує законі процеси, тобто не викликає спрацювань систем захисту, і гарно маскує свою присутність у системі. Тому виявити його, за допомогою стандартних механізмів захисту, майже не можливо.

Після проникнення у внутрішню мережу, порушники потрібно закріпитись в середині неї. Для цього він намагається отримати максимально доступні привілеї у системі. Отримання привілеїв відбувається за рахунок крадіжки даних локальних адміністраторів або занадто привілейованих користувачів.

Отримавши всю інформацію з машина користувача, зловмисник починає переміщатись по мережі. Задля цього використовуються правомірні системні процеси, що повсякчасно використовуються користувачами, наприклад PsExe або RAdmin. Хакер переміщається по мережі до тих пір поки не знайде цільові сервери або робочі станції. Після їх знаходження він переходить на четвертий етап.

На цьому етапі, зловмисник отримує доступ до основних банківських операції і намагається вкрасти кошти банку. На сьогоднішній день, використовують чотири основні способи крадіжки:

- видача грошей через банкомати;
- переведення коштів на критовалютні гаманці;
- контроль міжбанківських платіжних систем і переведення коштів на фіктивні рахунки;
- контроль банківських карт і рахунків.

Після вдалої крадіжки, зловмисники намагаються знищити всі сліди свого перебування у системі. Деякі порушники дуже старанно і ретельно підчищають свої сліди, видаляючи кожну подію пов’язану з ними. Але в більшості випадків, порушники повністю видаляться всі журнали і затирають усі диски у системі, що завдає надзвичайно великих збитків, іноді навіть більше чим саме пограбування.

Наприклад, спроба крадіжки в тайванському банку Far Eastern International Bank[11], в результаті якої зловмисникам не вдалося виконати крадіжку і вони зашифрували всі дані, тим самим знищивши можливі докази і ускладнивши розслідування.

1.3.2 Слабкі місця інфраструктури банківської установи

Серед величезної кількості різноманітних організації та підприємств, банківські установи відрізняються від інших великим рівнем організації інформаційної безпеки. Банки використовують майже всі новітні засоби забезпечення захищеності власної системи. Але це не робить їх абсолютно захищеними від хакарських атак. Згідно статистики, наведеної вище, та інших незалежних досліджень майже 70 відсотків банків не готові до атак[12].

Основними проблема, що найчастіше зустрічаються в інфраструктурі банківських установ, за допомогою яких зловмисник має змогу реалізувати власну атаку є:

- недостатня інформаційна обізнаність звичайних працівників;
- використання встановлених за замовченням паролів системних адміністраторів;
- відсутність двофакторної автентифікації для доступ до критичних систем інфраструктури;
- некоректне налаштування групових політик або надання надмірних привілеїв у системі звичайному працівнику;
- використання застарілого програмного забезпечення;
- недосконалості сучасних методів захисту.

Через недостатню обізнаність, користувачів завантажують та відкривають на робочих станціях небезпечні файли, що були вкладенні у фішинговий лист. Слабкі парольні політики та надмірні привілеї дозволять зловмиснику вільно використовувати будь-який функціонал. Відсутність автентифікації на критичних

системах дозволяє оперувати основними банківськими операціями. А недосконаліть сучасних методів – залишатись непомічними у системі на протязі всієї атаки.

Наведений список слабкостей систем не є повним. Кожна окрема установа має власні слабкі та сильні сторони. Один банк має гарну політику паролів, інший добре налаштоване розмежування доступу, а деякі зовсім не мають наведених проблем. Але залишається слабкість, притаманна всім банківським установам - недосконаліть сучасних методів захисту.

Однією з таких недосконалостей є складність завчасного виявлення зловмисника в середині мережі.

Сучасні атаки не обмежені у часі. Вони можуть тривати як декілька днів так і декілька місяців, впродовж яких зловмисник знаходиться у внутрішній мережі, ретельно її досліджуючи і вичікуючи вдалого моменту для початку активних дій.

Зловмисник майже ідеально імітує нормальну роботу системи, за рахунок чого тривалій час залишається не поміченим. Він не виконує жодних непередбачених системою дій, використовує лише дозволенні у системі процеси та, вже встановлене на робочій станції, програмне забезпечення. Тому, хоч всі його дії і фіксуються у журналах реєстрації події, а потім оброблюються системами SIEM, він все одно залишається невидимим для механізмів захисту і адміністратора з інформаційної безпеки.

Для вирішення даної проблеми, деякі розробники, такі як IBM, MicroFocus, Splunk, та інші, включають у свої рішення різноманітні механізми для аналізу поведінки користувачів. Адже, якщо оцінювати дії зловмисника в порівняні зі звичайним користувачем, можна одразу помітити основні відмінності і завчасно попередити адміністратора про наявність сторонніх осіб у мережі.

Але, на сьогоднішній день, механізми аналізу поведінки користувачів, що поставляють розробники разом зі свої SIEM-рішенням, недостатньо ефективні. Якщо поглянути на статистику атак на банківські установи у період з 2017 по 2018 рік, можна помітити, що системи фіксують зловмисників лише на останніх етапах

атаки, коли інфраструктурі вже завдано шкоди, або фіксують атаку вже постфактум.

Системи для аналізу поведінки користувачів мають ряд значних недоліків, більшість з них пов’язана з неефективністю використання ресурсів та методам профілювання поведінки користувачів:

- Система збирає та аналізує абсолютно всі дані, через що виникає потреба у великій розрахунковій потужності.
- Обробка великої кількості даних, що не мають жодної практичної цінності.
- Велика кількість помилок першого та другого роду, через некоректність отриманих моделей поведінки.
- В системах присутні лише профілі зловмисників, через що нові сценарії атак можуть залишитись не поміченими.
- Використання не раціональних методів машинного навчання, через що профілі поведінки визначені не правильно.

Висновки до розділу 1

В даному розділі розглянуто загальні принципи роботи систем SIEM. Описано ряд найпопулярніших сучасних SIEM-рішень, таких як IBM Qradar, Arcsight, McAfee Enterprise Security Manager, Splunk і Elastic stack та проаналізовано їх слабкі та сильні сторони.

Проведено аналіз атак на банківські установи у період з 2017 по 2018 роки, на основі якого визначено загальну схему атаки та основні слабкі місця інфраструктури. А також проаналізовано основні недоліки сучасних систем аналізу поведінки:

- Велика кількість помилок першого та другого роду, через некоректність отриманих моделей поведінки.
- В системах присутні лише профілі зловмисників, через що нові сценарії атак можуть залишитись не поміченими.
- Використання не раціональних методів машинного навчання, через що профілі поведінки визначені не правильно.
- Система збирає абсолютно всі дані всіх користувачів у інфраструктурі, через що потребує занадто великих розрахункових потужностей, а більшість.
- Більшість інформації, зібраної і проаналізованої системою, не мають жодної практичної цінності.

Звідси можемо зробити висновок, що системи аналізу поведінки користувачів потребують більш поглибленого аналізу та доопрацювання методів профілювання поведінки користувачів.

2 АНАЛІЗ ПОВЕДІНКИ КОРИСТУВАЧІВ

Даний розділ присвячено сучасним механізмам аналізу поведінки користувачів у внутрішній мережі та основним методам побудови поведінкового профілю: метод К-середніх; наївний Баєсовський класифікатор; “випадковий ліс”; метод опорних векторів.

2.1 Методи виявлення зловмисника

На жаль, сьогоднішні механізми забезпечення інформаційної безпеки швидко застарівають, і більш кваліфіковані хакери і порушники мають змогу обійти захисні засоби, які використовуються більшістю організацій та компанії.

У минулому ви були в безпеці, якщо ви використовували антивіруси, мережеві екрани та засоби запобігання вторгнення. Але це вже не стосується сучасного складного сценарію загрози, і особливо це стосується великих організацій, які мають дуже складні та великі інфраструктури.

Профілактичних заходів вже недостатньо. Ваші мережеві екрани не будуть на 100 відсотків надійними, а хакери та зловмисники потраплять у вашу систему в тій чи іншій точці. Саме тому, сучасні захисні засоби повинні забезпечувати не лише безпеку від зовнішніх нападників, а й контролювати внутрішні загрози.

Згідно, з визначенням Gartner, аналіз поведінки користувачів(UBA) – процес визначення базової поведінки користувачів у системі, для виявлення потенційно вторгнень або шкідливих дій.

На відміну від звичайних засобів захисту, системи аналізу поведінки користувачів не відслідковує події безпеки або пристрої, замість цього система відстежує лише користувачів та їх поведінку. Таким чином система фокусується на внутрішніх загрозах, таких як інсайдерські напади, тощо.

Вкрасти облікові дані користувача набагато легше, ніж зімітувати його поведінку у мережі. Наприклад, зловмисник заволодів обліковим записом одного з працівників. На відмінну від користувача, що в своїй роботі використовує лише

офісні програми, зловмисник починає виконувати різноманітні команди в операційній системі. Помітивши цю відмінність поведінки у поведінці облікового запису, система аналізу поведінки користувача одразу ж попередить адміністратора або заблокує потенційно небезпечний обліковий запис.

Загальна схема роботи робити систем аналізу(рисунок 2.1) складається з “трьох основних стовпів”:

- моделі порушників;
- дані, що надходять до системи;
- механізми для аналізу.

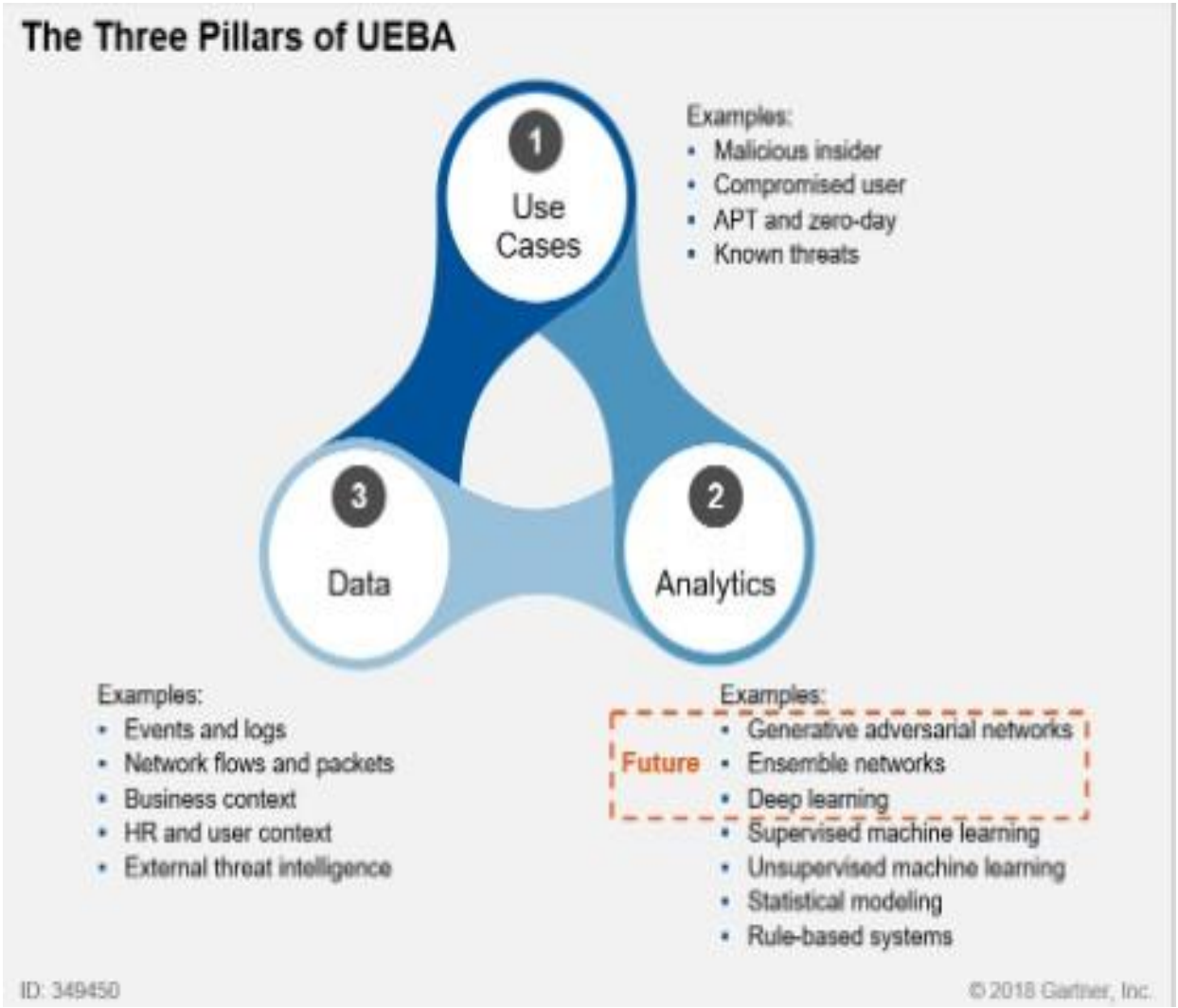

Рисунок 2.1 – основні “стовпи” UBA. Джерело: Gartner

Модель порушника, містить вже сформовані набори правил, які описують діяльність зловмисника або потенційно небезпечні дії працівників. Моделей може бути необмежена кількість, але основні з них:

- Недобросовісний працівник(інсайдер) - співробітник або підрядник з привілейованим доступом до інфраструктури, який має намір здійснити спробу несанкціонованого доступу до системи. Важко виміряти шкідливі наміри або виявити їх через файли журналів або звичайні події безпеки.
- Скомпрометовані облікові записи – майже завжди для проведення атаки, зловмисник намагається отримати у своє розпорядження облікові записи користувачів з великими привілеями або адміністратора. Традиційним засобам безпеки важко виявити такі скомпрометовані облікові записи, якщо дії порушника або сценарії атаки наразі невідомий, наприклад, при атаках “нульового дня”, або якщо атака переміщується через організацію шляхом зміни облікових даних або робочих станцій користувачів.
- скомпрометовані пристрої “інтернет речей” – сьогодні організації в своїй інфраструктурі використовують величезну кількість різноманітних пристроїв, що мають підключення до локальної мережі ,нехтують при цьому забезпеченням безпеки їх безпеки. Зловмисники можуть скомпрометувати пристрої та використовувати їх для крадіжки даних або отримання доступу до критичних систем або, ще гірше - використовувати їх у або інших атаках проти третіх сторін.

Для аналізу поведінки користувачів деякі системи UBA використовують стандартні методи пошуку підозрілої діяльності. Наприклад, визначені правила кореляції між подіями, що відбуваються, та відомими моделями порушників. Та традиційних методів недостатньо, тому що вони не здатні пристосовуватись до нових загроз. Сучасні механізми аналізу поведінки користувачів включають в себе більш новітні технології:

- Контрольоване машинне навчання - набори вже відомої звичайної і потенційно небезпечної поведінки подаються в систему, де вона навчається відрізняти одну поведінки від іншої;
- Байєські мережі - поєднують машинне навчання і правила для створення поведінкових профілів;

- Навчання без нагляду – система самостійно вивчає всі події у системі і на основі чого будує моделі;
- Посилене машинне навчання - гібридна модель, де основою є неконтрольоване навчання, але адміністратор самостійно визнає до якого класу потрібну віднести ту чи іншу модель;
- Глибоке навчання - система формує набори даних, що представляють сповіщення про безпеку та їх результати сортування, виконує самоідентифікацію функцій і здатна передбачати результати сортування для нових наборів оповіщення про безпеку.

Традиційні методи аналітики є статичными та однозначно визначеними: якщо певні умови були істинними, було сформовано попередження, а якщо ні, то система передбачала, що нічого потенційно небезпечного не відбувається. Перераховані вище методи аналізу відрізняються тим, що вони основані на евристичному підході аналізу інформації. Тобто, вони обчислюють оцінку ризику, яка є вірогідністю того, що подія являє собою аномалію або інцидент безпеки. Коли оцінка ризику перевищує значений рівень, система створює попередження про потенційну можливість порушення безпеки.

Системи аналізу поведінки користувачів здатні вирішувати завдання, які, нажаль, не спроможні вирішувати традиційні методи захисту, основані на сигнатурних методах. Одним з таких задач є:

- Виявлення інсайдерських загроз. Це не надто надумане уявити, що працівник, або, можливо, група працівників, може вийти з ладу, красти дані та інформацію, використовуючи свій власний доступ. UBA здатна виявити порушення даних, саботаж, зловживання привілеями та порушення політики, зроблені власним персоналом організації;
- Виявлення скомпрометованих облікових записів. Іноді облікові записи користувачів потрапляють до рук зловмисників. Системи аналізу поведінки користувачів дозволяють виявити такі записи і заблокувати їх, перш ніж вони скоять порушення;

- Виявлення змін у дозволах та створення користувачів з правами адміністратора або виявлення облікових записів, яким було надано непотрібні привілеї;
- Виявлення порушень доступу до критичних даних. Хоч більшість критичних даних і захищено від несанкціонованого доступу, але деякі працівники маю доступ до них. Тому важливо знати, коли працівник звертається до даних без жодних причин.

2.2 Модель користувача

Основна задача, при проведені аналізу поведінки користувача – побудова його моделі.

Кожен працівник, що має хоч один обліковий запис в інфраструктурі повинен мати сформовану модель, що складається з двох різних профілів:

- Загальна інформація про користувача;
- Поведінкова особливість користувача.

До загальної інформації про користувача, можна віднеси велику купи інформації, яка не змінюється з часом і є унікальною для інфраструктури:

- Ідентифікатор користувача в системі SIEM;
- ПІБ працівника;
- Різноманітні індикатори в інфраструктурі – облікові записи в операційних системах, робочі поштові адреса, облікові записи у базах даних, тощо;
- Робочий графік працівника.

Поведінка користувача:

- Програмне забезпечення, що використовує користувач;
- Пристрої, що використовуються для отримування доступу до внутрішньої мережі;
- Дані та ресурси, що використовуються під час роботи;
- Інформаційні потоки, що генерує користувач під час роботи;

- Тип підключення до мережі.

Кожна з частин моделі користувача відіграє значну роль при проведені аналізу поведінки користувачів: за допомогою загального профілю можливо однозначно прив’язати дії до певного користувача і розглядати його окремо від всієї системи, а за допомогою поведінкового профілю можна вдало порівнювати дії, що виконує користувач з вже відомим патерном його поведінки.

Побудування загального профілю користувача, майже не викликає складності, на відміну від поведінкового профілю. Побудова якого вимагає використання методів машинного навчання та використання потужних математичних апаратів.

2.3 Алгоритми побудови поведінкового профілю

Поведінковий профілі – одна з найважливіших частин моделі користувача. Саме за допомогою порівняння поведінки зловмисника з поведінкою класифікованого користувача система здатна виявити порушника.

На, сьогоднішній день існує безліч різноманітних алгоритмів класифікації, за допомогою яких системи UBA або новітні SIEM-рішення здатні створювати поведінкові профілі та виявляти відхилення від них.

Одними з найпоширеніших та найбільш потужніших алгоритмів класифікації, що використовуються, є:

- Метод К-середніх;
- Метод “випадкового лісу”(RF);
- Наївний Баєсовський класифікатор;
- Метод опорних векторів(SVM);

2.3.1 Метод K-середніх

Одним з найпопулярніших алгоритмів кластеризації великої кількості даних є метод К-середніх[13].

Головна ідея метода полягає в тому, що весь масив даних розбивається на к-кластерів. Потім, система вираховує початкове значення центра мас для кожного отриманого набору і у відношенні отриманих результатів, формуються нові кластери.

Наприклад, ми маємо дані, розподілені на площині випадковим чином(рисунок 2.2) і хочем класифікувати їх за 3 класами: 1 клас червоний, 2 клас зелений, 3 - синій. Спочатку всі об’єкти не належать до жодного класу, тому вони не забарвлені у жодний колір.

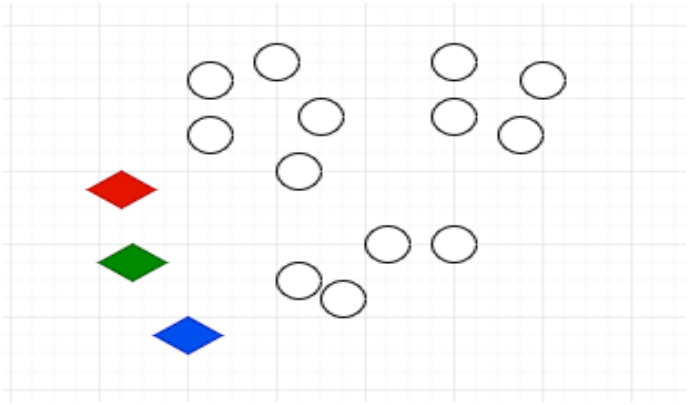

Рисунок 2.2 – початковий розподіл даних

Потім система випадковим чином розміщає початкові положення центрів мас, і будує класи(рисунок 2.3).

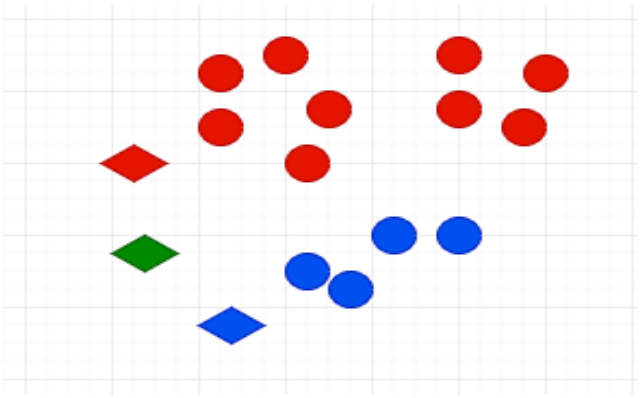

Рисунок 2.3 – перша ітерація алгоритму

Далі, згідно отриманих класів, вираховуються нові центри мас і об’єкти перерозподіляються по класам(рисунок 2.4).

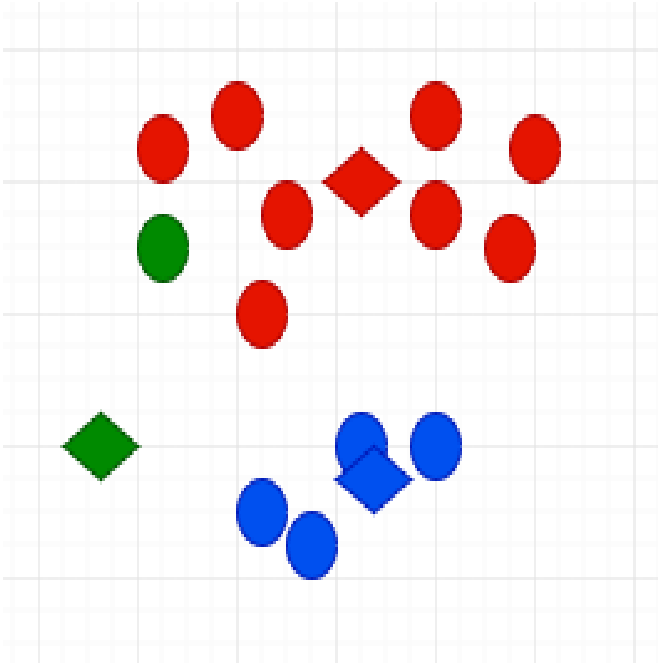

Рисунок 2.4 – друга ітерація алгоритму

Процес перерахування центру мас і перерозподілу об’єктів по клас, буду продовжуватись до ітерації, на якій жодних змін не відбудеться, тобто всі дані будуть класифіковані і визначені(рисунок 2.5).

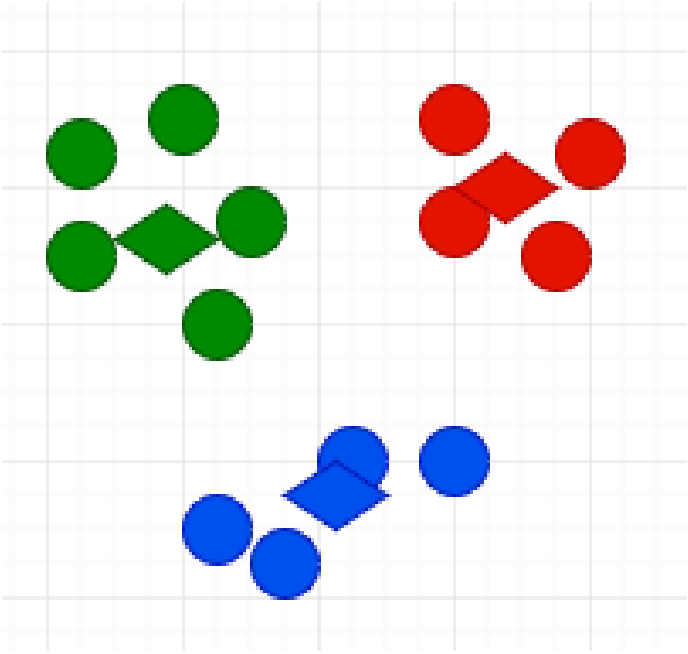

Рисунок 2.5 – розподілення об’єктів по класам

Переваги методу К-середніх:

- Висока швидкість класифікації;
- Висока ефективність при великих наборах даних;

Недоліки:

- Велика залежність від початкового вибору центрів мас;
- Значні проблеми при каутеризації дискретних даних.

2.3.2 Метод “випадкового лісу”

Метод “випадкового лісу” є одним з найпопулярніших та найсильніших методів класифікації великих обсягів даних. Даний метод є ансамблевим, тобто одночасно використовує декілька різних дерев рішень.

В загальному, дерево рішень - не самий гарний алгоритм класифікації, адже собою уявляє лише точку розбиття окремих атрибутів. Вершиною дерева визначається атрибут, згідно якого дані розділяться на групи. Гарним прикладом, буде використання бінарних дерев рішень(рисунок 2.6), де всі дані, в кожній вершині, розділяються на дві групи.

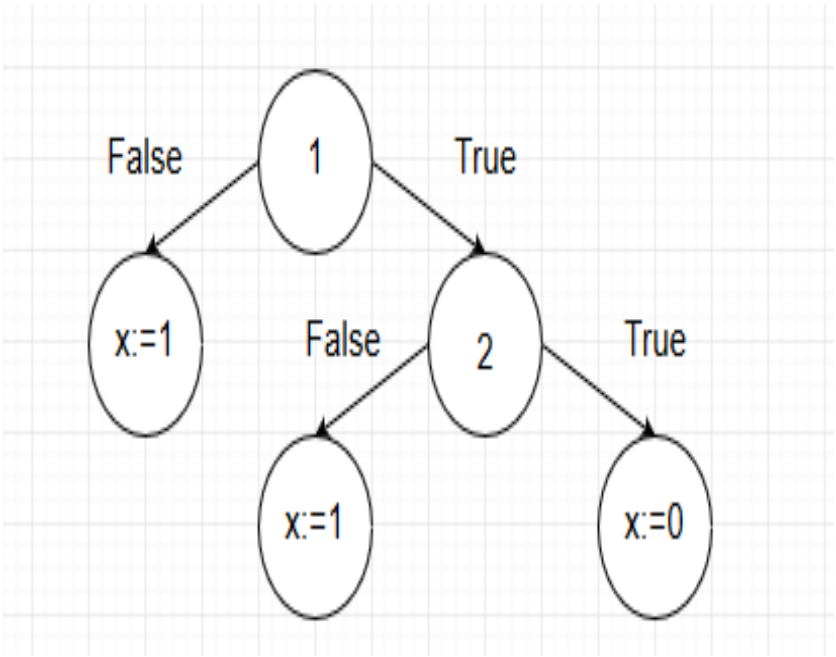

Рисунок 2.6 – бінарне дерево рішень

Як можна здогадатися, даний алгоритм класифікації має декілька основних недоліків:

- Параметр, згідно якого відбувається розбить даних у вершинах дерева;
- Нескінченість поділу даних на групи класів за атрибутами;
- Вибір порогового значення.

При використанні методу випадкового лісу, наведені проблеми майже нівелюються. Для один і тих же даних будується велика кількість дерев рішень, що мають різні параметри та порогові значення. А потім відбувається так зване “голосування”, для кожного об’єкту підраховується кількість входжень в різні класи. В результаті об’єкт потрапляє у класу за який віддано найбільше голосів.

Переваги використання методу випадкового лісу:

- Універсальність методу – здатний до вирішення майже 75 відсотків всіх задач машинного навчання;
- Вирішення великої кількості задач – класифікація, регресія, кластеризація;
- Легкість в реалізації – велика кількість вже готових моделей, написаних на мовах програмування Python та R.
- Можливість гарного масштабування.

Недоліки:

- Складність роботи з лінійними закономірностями;
- Потрібність у великому обсязі пам’яті – в результаті класифікації створюється величезна кількість дерев рішень, які потрібно зберігати.

2.3.3 Наївний Баєсовий класифікатор

Наївний Баєсовський класифікатор[14] - один з методів класифікації даних, в основі якого лежить теорема Баєса (2.1).

$$P(c|x) = \frac{P(c|x)P(c)}{P(x)} \tag{2.1}$$

де $P(c|x)$ – апостеріорна вірогідність;

$P(c)$ – апріорна вірогідність класу C;

$P(x)$ – апріорна вірогідність значення атрибуту.

В основі методу Баєсового класифікатору лежить припущення про незалежність різних атрибутів. Тобто наявність однієї з ознак у класі не залежить від наявності якоїсь іншої ознаки у цьому ж класі. Алгоритм робить припущення про належність об’єкта до якогось класну, звідси і наївність методу. Наприклад, фігура може бути квадратом, якщо вона має чотири рівні сторони та всі кути рівні 90 градусам. Дані ознаки незалежні, тому ймовірність належності до класу квадратів буде рівно 1. А якщо записати умови як: чотири сторони і чотири кути по 90 градусів, то результатом буде вірогідність приналежності об’єкту до класів квадрат та прямокутник.

Переваги методу:

- Швидка класифікація об’єктів;
- При незалежності атрибутів метод є одним з найкращих;
- Можливість зменшення вибірки;

Недоліки:

- Проблема “нульової частоти”;
- Допущення про незалежність ознак, що майже не зустрічається;
- Не однозначність класифікації об’єктів.

2.3.4 Метод опорних векторів

Метод опорних векторів[15] – лінійний алгоритм класифікації, що належить до групи граничних методів. Метод класифікує об’єкти за допомогою побудови гіперплощини, що максимально розділяє різні класи.

Гіперплощина - під-площина, розмірність якої на одиницю менша за площину, в якій знаходяться дані. Наприклад, якщо дані розміщено на площині розмірність 2, то гіперплощиною буде пряма, що має розмірність рівно 1 і розділяє дані на два класи(рисунок 2.7).

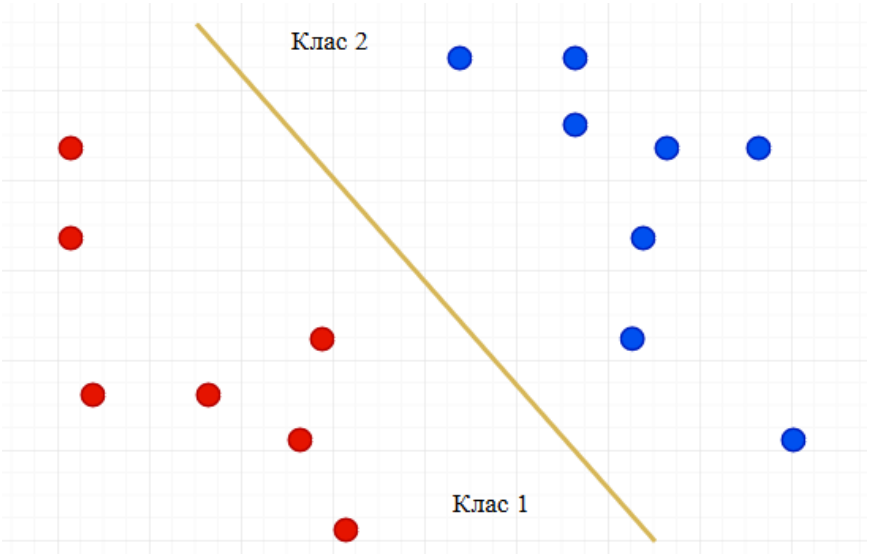

Рисунок 2.7 – приклад гіперплощини

Для своєї роботи алгоритм обирає не будь-яку під-площину, що відповідає визначенню гіперплощини. Обирається під-площина відстань від якого до кожного класу максимальна. Для побудови найкращої гіперплощини алгоритм шукає точки в класах, що знаходяться максимально близько один до одного, потім він будує площину між цими точками, таку що відстань від опорних векторів до цієї площини максимальна(рисунок 2.8), і приймає отриману площину за границю поділу між двома класами[15].

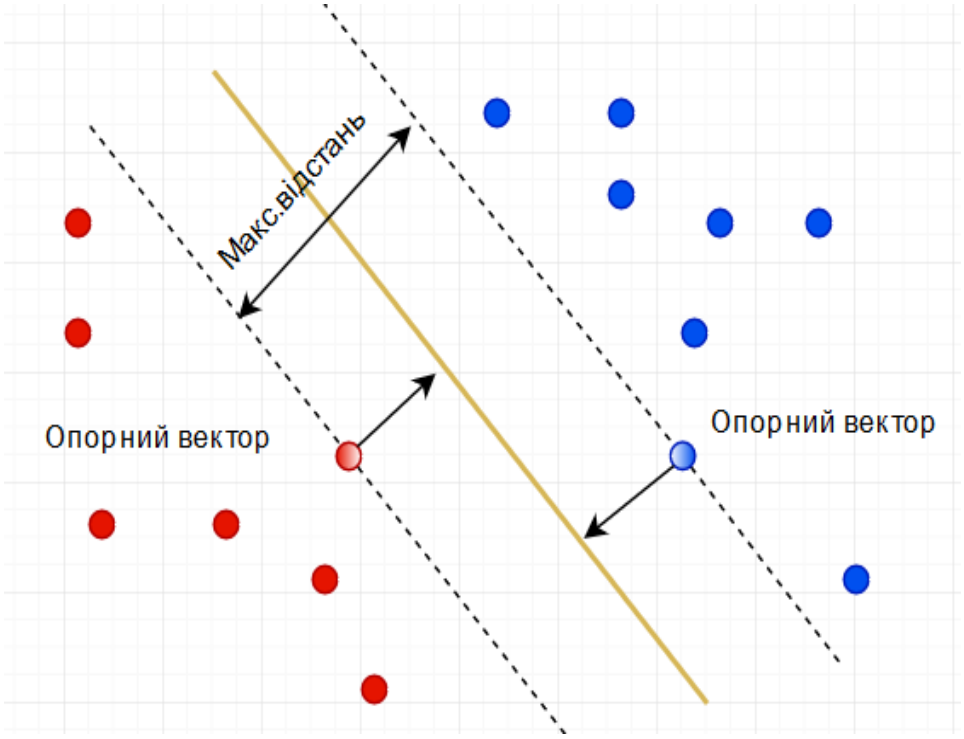

Рисунок 2.8 – вибір найкращої гіперплощини

Тепер представимо дану задачу в математичному вигляді. Нехай ми маємо вибірку $(x_1, y_1), (x_2, y_2) \dots (x_m, y_m), x_i \in R^n, y_i \in \{-1,1\}$, тоді алгоритм будує класифікуючи функцію (2.2).

$$F(x) = sign(\langle w, x\rangle + b) \tag{2.2}$$

де $\langle\ ,\ \rangle$– скалярний добуток;

w – вектор нормалі від опорного вектору до гіперплощини;

b – допоміжний параметр.

Легко побачити, що максимальна відстань між опорними векторами рівна $\frac{1}{||w||}$, тобто для пошуку потрібної гіперплощини треба вирішити задачу оптимізації (2.3).

$$\begin{cases} arg\min\limits_{w,b} \|w\|^2 \\ y_i(\langle w, x_i\rangle + b) \geq 1\ , i = 1, \dots, m \end{cases} \tag{2.3}$$

Даний метод, можливо використати, лише у випадку, коли задану вибірку можливо розбити лінійно. У протилежному випадку всі об’єкти за допомогою перетворення (2.4) переводять у площину X з більшою розмірністю. Причому перетворення (2.4) обирається таким чином, щоб у новій площині вибірка мала лінійне розбиття.

$$\varphi: R^n \to X \tag{2.4}$$

При переході в іншу площину, класифікуючи функція приймає вид (2.5).

$$F(x) = sign(\langle w, \varphi(x)\rangle + b) \tag{2.5}$$

А вираз виду (2.6) прийнято називати ядром класифікатора.

$$k(x, x') = \langle \varphi(x), \varphi(x')\rangle \tag{2.6}$$

Виділяють чотири основні ядра класифікатора:

- Радіальна базисна функція Гаусса;
- Радіальна базисна функція;
- Поліноміальне ядро;
- Сигмоїд.

Переваги методу опорних векторів:

- Найкраще теоретичне обґрунтування методу класифікації – метод базується на потужному математичному апарату вирішення задач оптимізації;

- Гарна класифікація, яка дозволяє класифікувати нові дані набагато швидше за інші методи;
- Однозначно визначає приналежність об’єкту до класу;

Недоліки:

- Потрібно ретельно нормалізувати та стандартизувати дані;
- Вибір ядра значно впливає на кінцевий результат класифікації;
- Занадто повільне навчання.

2.4 Оцінка методів побудови профілів поведінки

Вибір алгоритму побудови поведінкового профілю відіграє важливу роль в подальшій роботі системи аналізу поведінки користувача. При коректній побудові моделі користувача шанс відрізнити порушника від легально користувача набагато вищий.

Виходячи з основної задачі систем аналізу поведінки користувачів: виявлення порушників у внутрішній мережі установи. Можемо зробити висновок, що на виході ми можемо отримати лише дві різні відповіді – звичайна поведінка або потенційно небезпечна.

Для оцінки коректності роботи моделей користувачів, побудованих на алгоритмах К-середніх, RF та SVM будемо використовувати метрики[16] наведені у таблиці 2.1.

Таблиця 2.1 – Метрика оцінки алгоритмів

False Positive(FP)	True Negative(TN)
True Positive(TP)	False Negative(FN)

Де:

- FN – віднесення звичайної поведінки зловмиснику.
- TP – віднесення звичайної поведінки працівнику.
- FP – віднесення аномальної поведінки зловмиснику.
- TN – віднесення поведінки зловмисника працівнику.

З даної метрики можемо вирахувати основні показники оцінки. Кількість даних коректних відповідей(accuracy) (2.7).

$$accuracy = \frac{TP+TN}{TN+TP+FP+FN} \quad (2.7)$$

Точність роботи(precision) (2.8) – відсоток правильних відповідей, що видав алгоритм класифікації поведінки.

$$precision = \frac{TP}{TP+FP} \quad (2.8)$$

Повнота отриманої відповіді(recall) (2.9) – відсоток звичайної поведінки, що алгоритм виявив у порівняні до їх загальної кількості.

$$recall = \frac{TP}{TP+FN} \quad (2.9)$$

Для навчання і оцінки алгоритмів було вибрано дані[17], що місять дані діяльності користувача в операційних системах, а також розмічені дані про мережу діяльність облікових записів.

Навчання алгоритмів проводилось на серверах Amazon[18], що вже мають побудовані та запрограмовані алгоритми класифікації, описані вище у розділі.

В результаті навчання в цілому було задіяно 341749 файли діяльності облікового запису. Розміченні дані містили 173253 події створених в результаті звичайної діяльності працівника.

Було проведено два незалежних досліди, для яких всю загальну кількість даних про діяльність було розбито на два набори, таким чином що дані в них пересікалися не більше ніж на 20 відсотків.

Результатів дослідів було усереднено, та основі них підраховано показники точності роботи алгоритму та побудовано “матриці плутаниці”, для кожного методу окрім наївного Баєсовго класифікатора.

Метод К-середніх(рисунок 2.9) було отримано результати представлені в таблиці 2.2.

Таблиця 2.2 – Результати оцінки методу К-середніх

Accuracy	Precision	Recall
0,92	0,92	0,93

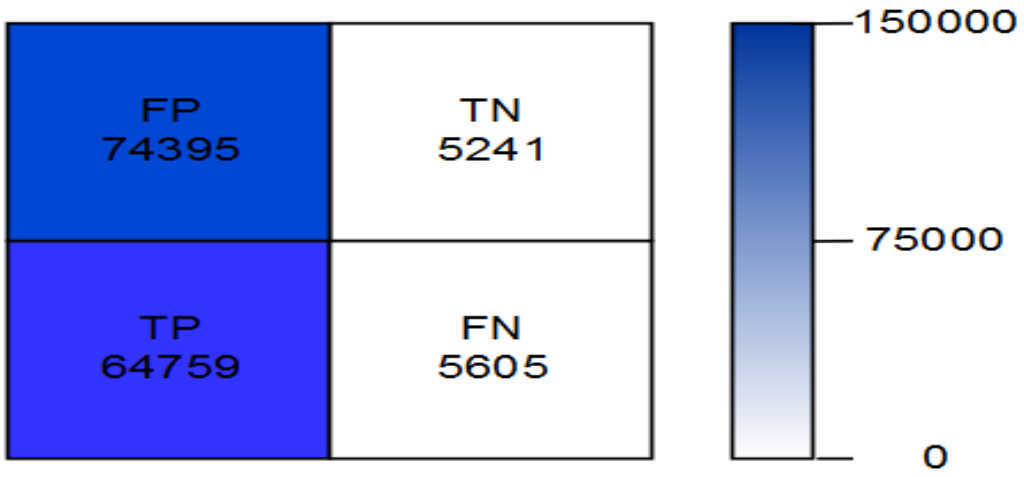

Рисунок 2.9 – Матриця плутаниці K-середніх

Для метод випадкового лісу(рисунок 2.10) було отримані результаті представлені в таблиці 2.3.

Таблиця 2.3 – Результат оцінку методу RF

Accuracy	Precision	Recall
0,92	0,91	0,92

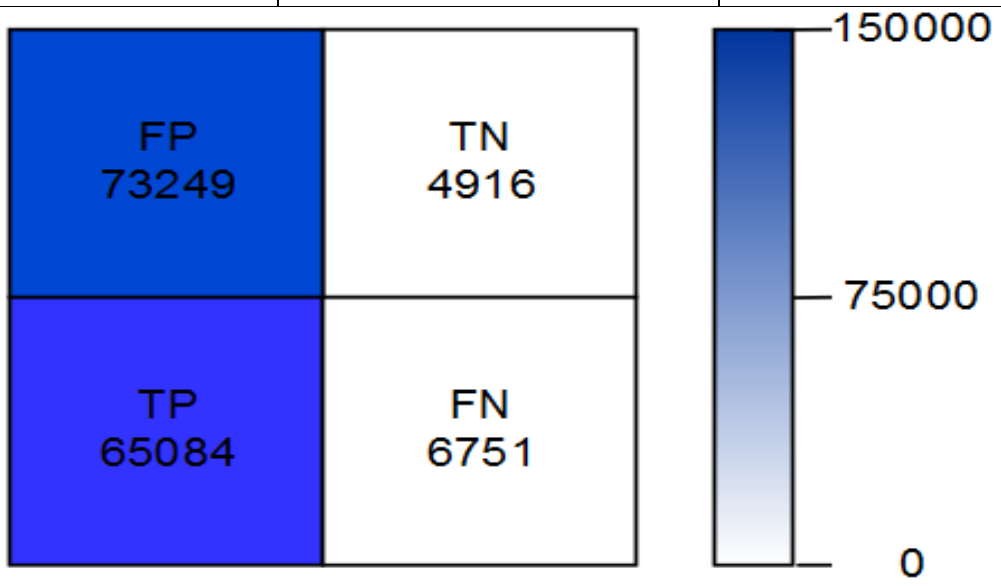

Рисунок 2.10 – Матриця плутаниці RF

Оцінка методу опорних векторів(рисунок 2.11) наведена у таблиці 2.4.

Таблиця 2.4 – Результат оцінки методу SVM

Accuracy	Precision	Recall
0,945	0,93	0,96

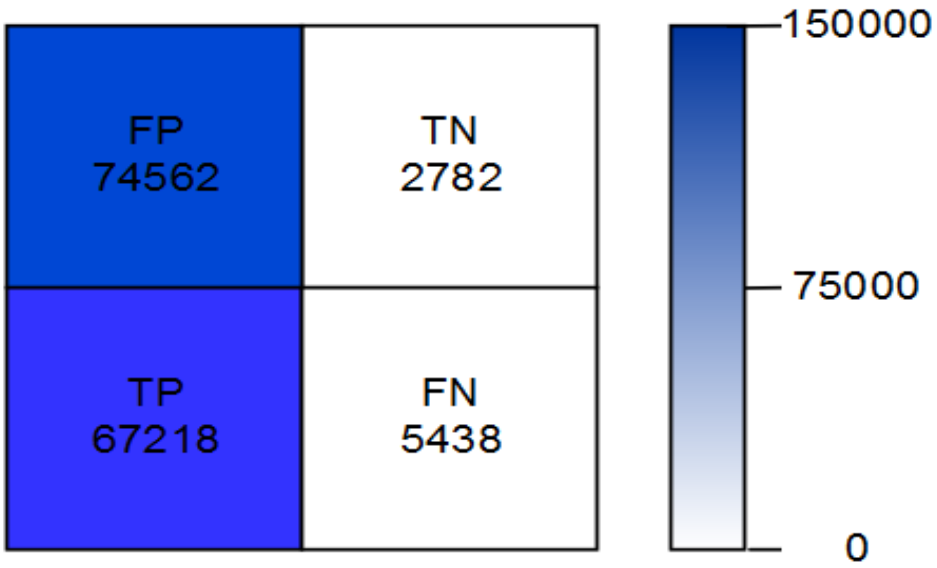

Рисунок 2.11 – матриця плутаниці SVM

Порівнюючи результати всіх дослід, що наведені у таблиці 2.5, можемо сказати, що метод опорних векторів демонструє найкращі показники серед всіх протестованих алгоритмів.

Таблиця 2.5 – Порівняння методів побудови поведінкового профілю

Дослід 1			
	Accuracy	Precision	Recall
К-середніх	0,9195	0,9235	0,9275
RF	0,9213	0,9131	0,9240
SVM	**0,9455**	**0,9334**	**0,9592**
Дослід 2			
	Accuracy	Precision	Recall
К-середніх	0,9234	0,9195	0,9135
RF	0,9203	0,9098	0,9213
SVM	**0,9447**	**0,9288**	**0,9635**

Висновки до розділу 2

В даному розділі було розглянуто систему аналізу поведінки користувачів(UBA) здатних до виявлення порушника в середині мережі шлях співставлення дій порушника з сформованими моделями користувачів. Описано основні принципи роботи, складові та завдання даного механізму.

Також було описано основні алгоритми побудови поведінкового профілю користувачів:

- Метод К-середніх.
- Метод “випадкового лісу”.
- Наївний Баєсовський класифікатор.
- Метод опорних векторів.

Методи К-середніх, “випадкового лісу” та опорних векторів було експериментально оцінено, а результати оцінки порівняно між собою та занесено до таблиці 2.5.

З результатів оцінки даних алгоритмів можемо зробити висновок, що метод опорних векторів демонструє найкращі результати: Accuracy – 94.5 відсотка, Precision – 0.93, Recall – 0.96.

Кожний з описаних вище методів можливо використовувати для побудови профілю користувача, але метод опорних векторів підходить для вирішення цієї задачі якнайкраще. Він не тільки продемонстрував найкращі результати, а й не має недоліків, що притаманні іншим методам: метод К-середніх вимагає оптимального вибору початкового розміщення центра мас, що доволі складна задача; для використання методу випадкового лісу потрібно більш ретельно передоброблювати дані; для наївного Баєсовського класифікатора потрібно заздалегідь вказувати апріорну ймовірність. Тому при побудові моделі профілювання поведінки користувачів бажано використовувати саме метод опорних векторів.

3 ПРОФІЛЮВАННЯ ПОВЕДІНКИ КОРИСТУВАЧА

Розділ присвячено вирішенню проблеми виявлення порушників у внутрішній мережі, шляхом побудови власної моделі профілювання поведінки користувачів, за допомогою методу опорних векторів.

3.1 Закономірність дій працівника у системі

Майже всі дії людини можливо типізувати і віднести їх то одного визначеного шаблону поведінки[19]. В більшій мірі це пов’язано з раціональним мисленням. Замість того щоб вигадувати нові способи дії, потрапляючи в одну й ту ж саму ситуацію, проще використати вже готову модель поведінки, що містить визначену у минулому послідовність дії.

В загальному, за одним працівником в організації чи на підприємстві закріплюється декілька бізнес-задач, вирішенням яких він займається тривалий проміжок часу. В процесі виконання встановлених завдань, він оптимізує всі свої дії і формує власну модель поведінки. Наприклад, складання щомісячної фінансової звітності бухгалтером. Для виконання даного завдання, бухгалтеру потрібно завантажити всю фінансову звітність за місяць з бази даних, відкрити одну з програм для роботи з даною інформацією, розрахувати всі параметри та сформувати звіт(рисунок 3.1).

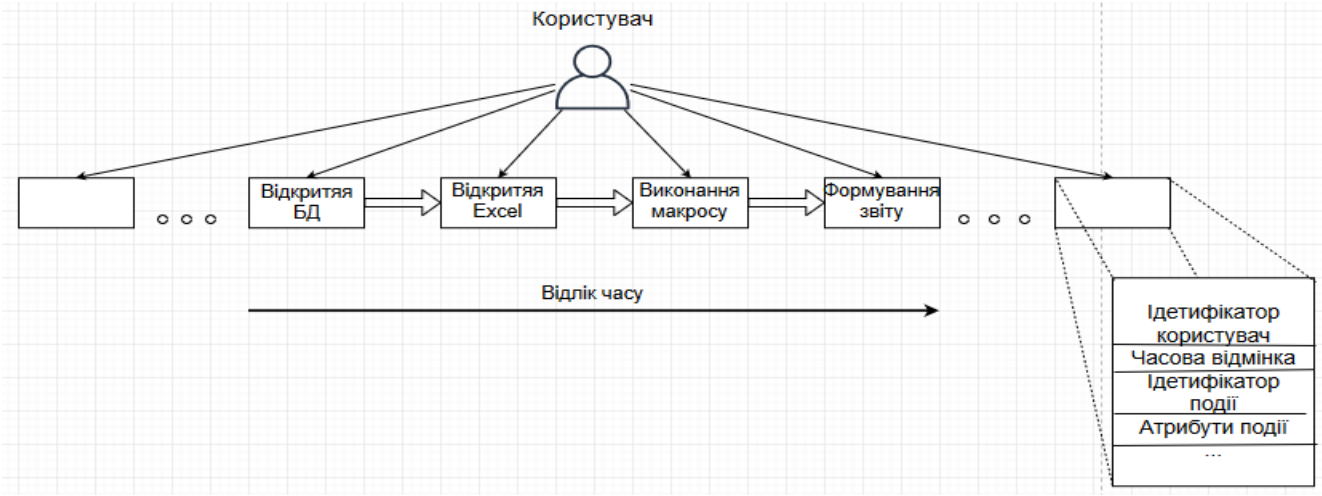

Рисунок 3.1 – Послідовність дії користувача

Аналізуючи дії, що виконує бухгалтер складаючи звіт, на протязі декількох місяців, можемо помітити, що, майже, кожного разу працівник виконує один і той же алгоритм дій(рисунок 3.2).

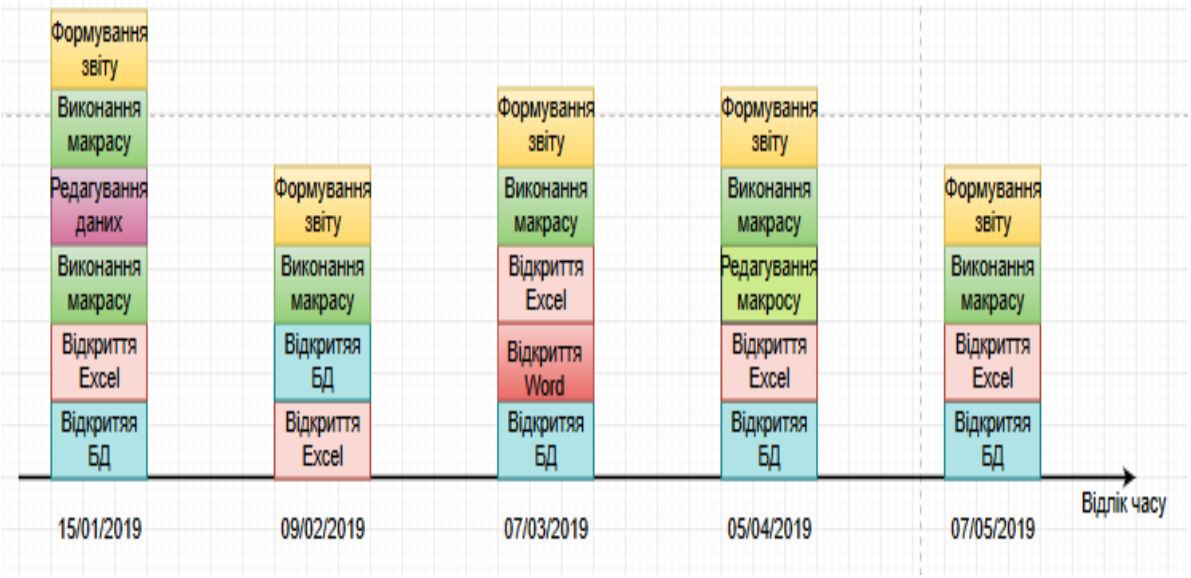

Рисунок 3.2 – Послідовність дій при складані звіту

Хоч послідовність дії працівника і не завжди однакова, але кожну його дію можна пояснити зі сторони бізнес-логіки: звіт 15/01/2019 – було виявлено помилку в даних, тому їх було відредаговано; 07/03/2019 – зміни в правилах банківської установи, стосовно формування звітів; 05/04/2019 – некоректне спрацювання макросу, через що його було відредаговано.

Основна бізнес-логіка зберігається, бухгалтер використовує фінансову інформацію, що зберігається у базі даних, лише для виконання встановлених перед ним бізнес-задач – складання фінансової звітності. Тобто, побачивши дану послідовність дій у системі, майже зі 100 відсотковою вірогідністю, можемо сказати, що під обліковим записом працює бухгалтер і він складає щомісячну звітність. І навпаки, якщо ми побачимо, що за допомогою облікового запису бухгалтера було відкрито базу даних, але загальна послідовність дії не спостерігається то можемо зробити висновок про аномалію в поведінки користувача.

Тобто, вірогідність що з облікового запису працює саме бухгалтер вже не 100 відсоткова.

3.2 Нормальна поведінка користувача

Кожна організація або підприємство мають власні критичні системи або дані, втрата або знищення яких завдасть величезних збитків. Саме ці дані і системи в першу чергу приваблюють зловмисника, що потрапив до внутрішньої мережі, або неправомірного співробітника. Тому вся діяльність порушників буде спрямована на знищення або крадіжку критичних систем та даних.

Більшість працівників в інфраструктурі не мають привілеїв на роботу з критичними системами або доступ до цих даних опосередкований. Не має сенсу збирати та аналізувати дії такого користувача, адже зловмисника не цікавлять облікові записи з низькими привілеями у системі бо за допомогою них він не зможе в повній мірі дослідити мережу, а непомітно підвищити привілеї облікового запису не вийде.

Щоб безпосередньо працювати з критичними даними, зловмиснику потрібно працювати з облікового запису привілейованого користувача:

- Адміністратор з інформаційної безпеки;
- Мережевий адміністратор;
- Керівники відділів;
- Уповноважені користувачі.

Так як, зазвичай, за кожним працівником в компанії закріплено один обліковий запис, з якого він і працює. То, для всіх дії, що виконуються цим працівником у відношенні до критичних систем або даних, можна побудувати характерні особливості використання та характерні послідовності.

Візьмемо приклад, наведений вище, і опишемо характерні особливості взаємодії бухгалтера і бази даних, в якій містить вся фінансова інформація установи.

Для початку потрібно виділимо користувача і всі облікові записи, що використовує працівник. В нашому випадку бухгалтер має лише один обліковий запис – fin_miol. Описуємо об’єкт взаємодії – система або дані, навколо

використання яких будується поведінковий профіль користувача: фінансова інформація, що зберігається у базі даних. Далі описуємо основні особливості взаємодії користувача з об’єктом: кожного разу, при роботі з базою даних працівник весь час використовує лише офісну програму Excel і кожен раз виконуються макроси “Work”, “Send E-mail”, “Test” і “Update”. І в кінці опишемо основні послідовності дій, що використовує працівник та занесемо всю інформацію до таблиці 3.1

Таблиця 3.1 – Поведінка користувача відносно фінансової інформації

Користувач	Обліковий запис	Об’єкт	Особливості поведінки	Послідовність дій
Бухгалтер	fin_miol	Фінансова інформація, що зберігається у базі даних	Разом з базою даних відкривається Excel і виконуються макроси: “Work”, “Send E-mail”, “Test” і “Update”	БД – Excel – макроси
				Excel – БД – макроси

Побудувавши, аналогічні таблиці для кожного об’єкту взаємодії, до яких користувач має безпосередній доступ, отримаємо модель нормальної поведінки користувача(рисунок 3.3).

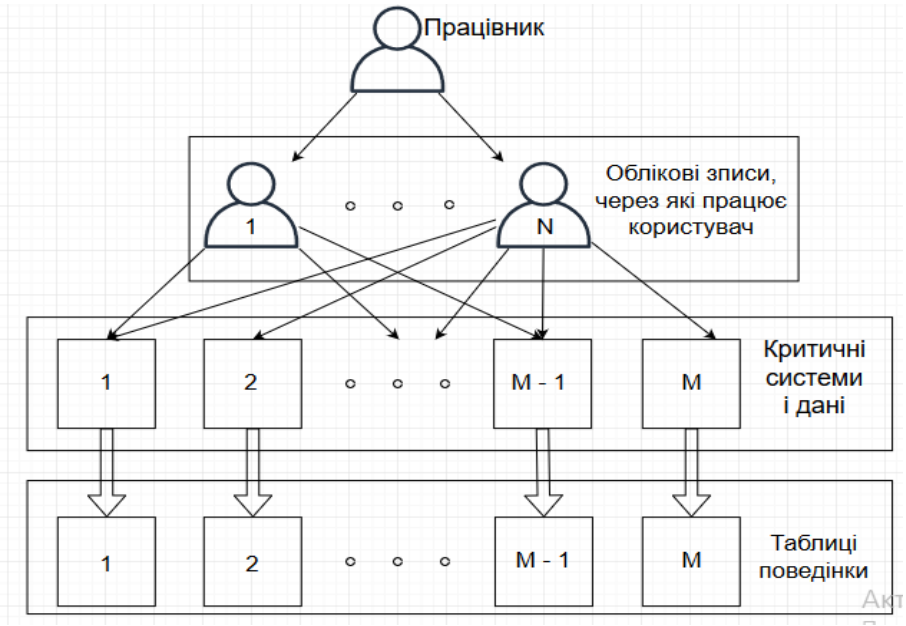

Рисунок 3.3 – Модель нормальної поведінки

Кожну отриману таблицю поведінки будемо називати нормою, а поведінка користувача, що повність вписується в норму – нормальна поведінка. І навпаки, поведінка, що не відповідає очікуванням і повністю суперечить нормі – аномальна поведінка.

3.3 Ядро класифікатора SVM

Згідно висновків, отриманих у другому розділі, профілювання поведінки користувачів буде здійснюватися за допомогою методу опорних векторів.

Насамперед, для того щоб використати метод опорних векторів потрібно визначитись з ядром класифікатора, так як дані поведінки користувачів не можливо розбити лінійно.

Згідно теореми Мерсера (3.1), функція двох змінних $K(x, y)$ називається ядром класифікатора, якщо вона симетрична і невід'ємно визначена.

$$\begin{cases} K(x,y) = K(y,x) \\ \iint K(x,y)f(x)f(y)dxdy \geq 0 \end{cases} \tag{3.1}$$

де $f(x)$ – будь-яка скінченна невід'ємна функція.

В загальному всі існуючи функція, що можуть бути використані як ядра класифікаторів можливо розділити на дві групи:

- Функція дистанції (3.2).
- Функції скалярного добутку (3.3).

$$K(x, y) = K(\|x - y\|), \tag{3.2}$$

$$K(x, y) = K(x * y) \tag{3.3}$$

Найвідомішим представником групи функцій дистанції є радіальна базисна функція (3.4).

$$K(\vec{x}, \vec{y}) = e^{\left(-\frac{\|x^a - y^b\|}{2\sigma^2}\right)} \tag{3.4}$$

де x – вектор атрибутів даних,

y – середній вектор класу.

У свою чергу поліноміальна функція(3.5) є одним з найяскравіших представників класу функцій скалярного добутку.

$$K(\vec{x},\vec{y}) = (x * y + c)^p \tag{3.5}$$

де p – визначає ступінь поліноміальної функції,

c – константа.

Для вирішення задачі профілювання поведінки користувачів за принципами, описаними вище у даному розділі, потрібно використовувати ядра класифікатора. Але при застосуванні даних функцій ми зіштовхуємось з рядом проблем:

- Точки, що знаходяться дуже близько один до одного, сильно корелюють у новому просторі.
- Точки, що знаходяться на межах просту, майже втрачають будь-які кореляційні ознаки.

Для вирішення цих проблем, функція, що є основою ядра класифікатора, повинна мати такі особливості:

- Зниження кореляційних ознак між точками в околі нуля;
- Помірне зменшення кореляційних ознак у точок, що знаходяться на межах площини.

Наприклад, гаусова радіальна базисна функція зовсім не задовольняє вимоги, а звичайна радіально базисна функція (3.3) лише першій. А функція KMOD (3.6) задовольняє обом вимогам.

$$K(\vec{x},\vec{y}) = K[e^{\left(-\frac{\gamma}{\|x-y\|+\sigma^2}\right)-1}] \tag{3.6}$$

Де K – константа нормалізації,

γ і σ – параметри, що контролюють розмір ядра і зменшення ознак у нулі.

Для побудови власної моделі профілювання поведінки користувачів, модифікуємо дану функцію (3.7).

$$K(\vec{x},\vec{y}) = \frac{K}{\left(\sqrt{\|x-y\|}+\sigma^2\right)-n} \tag{3.7}$$

Де K та n– константа нормалізації,

σ – параметри, що контролюють розмір ядра і зменшення ознак у нулі.

3.4 Профілювання поведінки користувачів

Профілювання поведінки користувачів буде здійснюватися за допомогою методу опорних векторів, основаного на запропонованій функції ядра класифікатора (3.7), в який будуть надходити дані діяльності користувача, що стосуються критичних систем або даних.

Виходячи з поставлених завдань, можемо запропонувати власну модель профілювання поведінки користувача(рисунок 3.4).

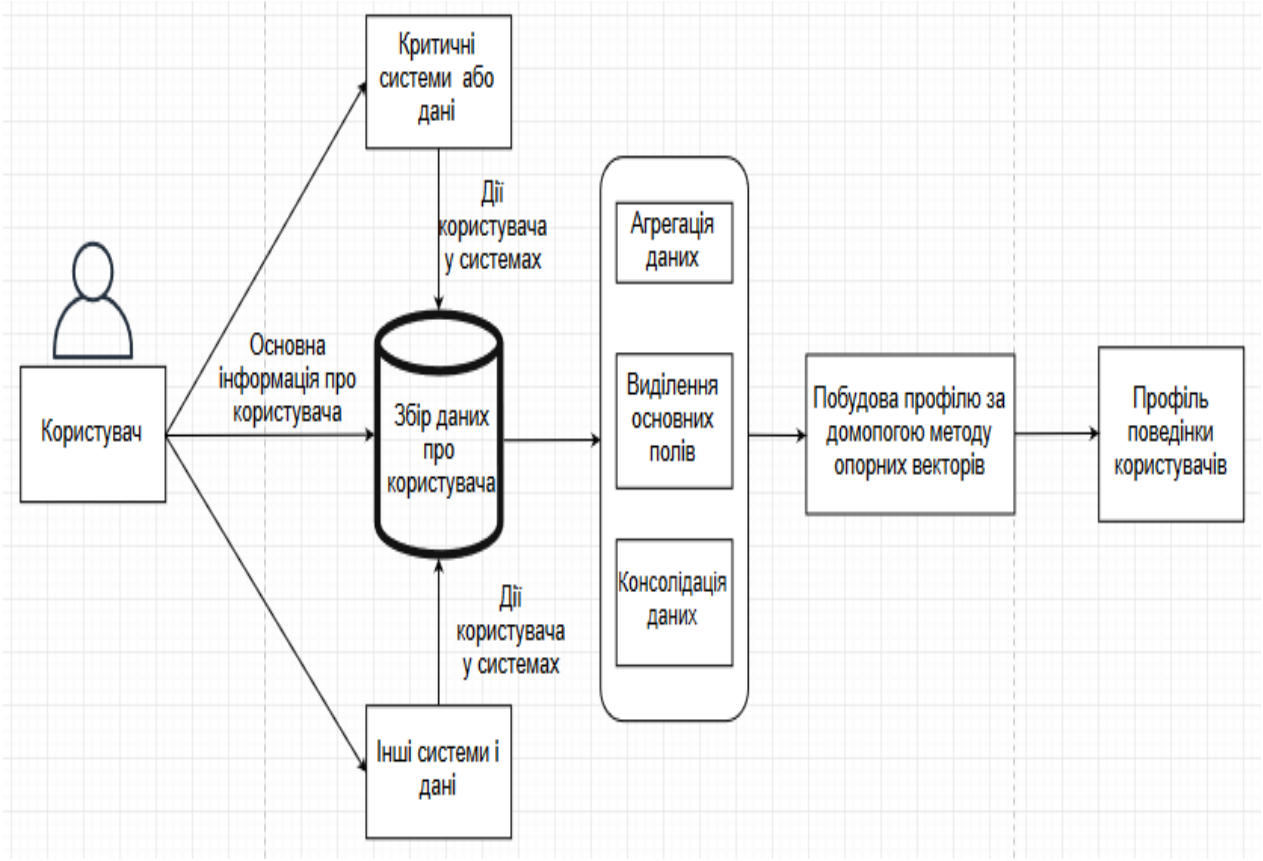

Рисунок 3.4 – модель профілювання поведінки користувачів

Користувач одразу взаємодіє з звичайними система і критичними системами і даними. Всі його дії потрапляють в колектор, де вони збираються, після чого потрапляють до механізму передобробки. У механізмі передобробки над даними проводиться три основні операції:

- Агрегація – поєднання інформації про діяльність користувача у критичних системах разом з діяльністю у звичайних системах.

- Виділення основних полів – більшість даних містить велику кількість непотрібних атрибутів, тому залишаємо лише потрібні для профілювання поля: ідентифікатор користувача, ідентифікатор дії, код ресурсу, час і інші.
- Консолідація – остаточно передобробка даних, в результаті якої маємо вже готові дані для профілювання методом опорних векторів.

Після передобробки дані надходять до методу опорних векторів, що за допомогою запропонованого ядра, розділяє всі ознаки на класи і формує остаточний профілю користувача(рисунок 3.3).

Отримані профілі поведінки можемо вже використовувати для виявлення аномальної активності користувача, за рахунок чого забезпечимо більший рівень безпеки внутрішньої мережі.

Дана модель вирішує всі проблеми сучасних механізмів аналізу поведінки користувача:

- Система не потребує великої кількості розрахункових потужностей.
- Система збирає інформацію лише про вповноважених користувачів, а всі не потрібні дані відсікаються на етапі передобробки, тому профілювання проходить швидше.
- У запропонованій моделі, згідно дослідів проведених у другому розділі, використовується найкращий метод профілювання поведінки.
- В даній моделі будуються профілі користувачів, через що системи захисту здатні виявляти не лише відомі сценарії атак, а й нові.

3.5 Реалізація запропонованої моделі

Як ми можемо побачити з запропонованої моделі(рисунок 3.4) всю її роботу можна розбити на декілька етапів:

- Збір всіх потрібних для аналізу даних.
- Попередня обробка отриманих даних.

- Профілювання поведінки користувачів за допомогою методу опорних векторів.

Розглянемо кожен етап окремо і опишемо механізми, що ми будемо використовувати для реалізації поставлених завдань.

Так як, більшість робочих станцій працівників та серверів працюють під управлінням операційних систем сімейства Windows, то для збору даних будемо використовувати Windows Event Forwarding(WEF). WEF – потужний механізм збору даних з журналів реєстрації подій, що не потребує використання додаткових агентів, а збирає всі дані на пряму (рисунок 3.5).

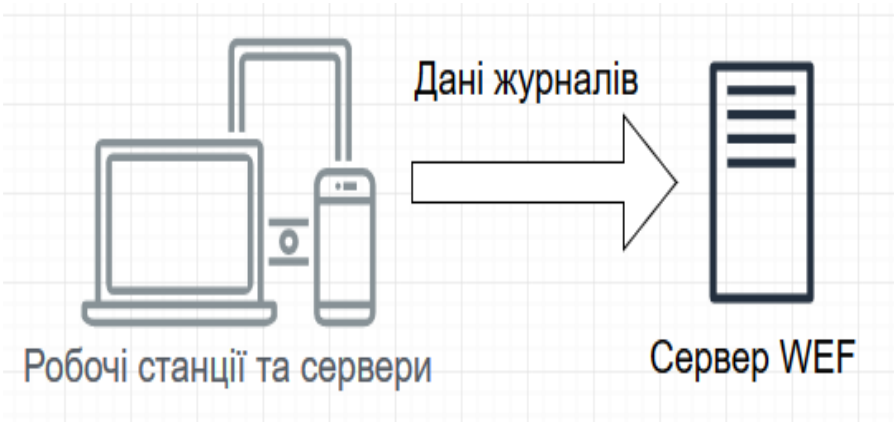

Рисунок 3.5 – Архітектура WEF

Для налаштування коректної роботи даного рішення потрібно налаштувати сервер, що буде збирати всі дані:

- Налаштувати підписку на отримання логів.
- Увімкнути службу Windows Remote Management.

А також налаштувати робочі станції та сервера, з яких будуть збиратись дані:

- Увімкнути відправку даних.
- Увімкнути службу WEF.
- Додати додаткові права на читання журналів службі WEF.

Налаштування роботи механізму Windows Event Forwarding є тривіальною задачею тому детально розглядати її не будемо і перейдемо до другого пункту завдання – попередня обробка отриманих даних.

Перше з чого почнемо попередню обробку даних – нормалізація. Для цієї задачі будемо використовувати два різні методи:

- Метод мінімакс – лінійне перетворення даних у діапазоні, наприклад від 1 до 10, де мінімальне значення відповідає 1, а максимальне 10.
- Метод Z-показника – масштабування даних на основі середнього значення та нормального відхилення; метод полягає в діленні різниці між даними та середнім значенням на нормальне відхилення.

Наступне, що потрібно зробити – агрегація даних. Даний етап потрібний, щоб зменшити кількість зібраних даних.

Останній етап – консолідація даних. На даному етапі ми будемо використовувати “пакетну аналітику”, що полягає в аналізі даних згідно заданого проміжку часу.

Для виконання завдання передобробки даних будемо використовувати програму Apache Hive, що повністю задовольняє нашим потребам.

Наступний етап – профілювання поведінки користувачів. Для його виконання будемо використовувати програмне забезпечення Amazon Web Service, додавши до стандартного методу опорних векторів запропоновану функцію ядра класифікатора.

Поєднавши всі елементи отримаємо систему для побудови профілю поведінки користувачів (рисунок 3.6), що працює згідно запропонованої моделі.

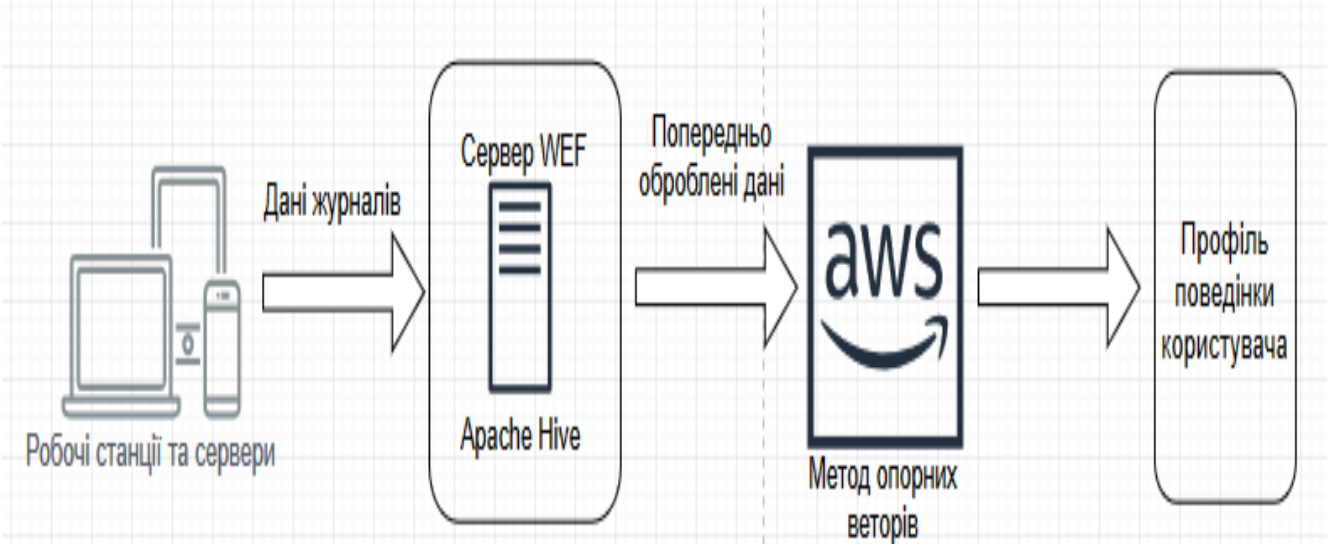

Рисунок 3.6 – Архітектура системи профілювання

Наведена архітектура має ряд значних переваг, що робіть її досить ефективною при використані в невеликих та середніх за розмірами організаціях або компаніях:

- Значна простота в налаштуванні.
- Легкість в адмініструванні даної системи.
- Економічні – більшість елементів, на яких побудована система, є безкоштовними.
- Невибагливість до апаратної частини – система потребує лише один сервер, на якому проводяться не значні за складністю операції, а більшість складних операцій проводяться на хмарних серверах Amazon.
- Можливість додавання окремих нових модулів, що будуть виконувати сумісні задачі.
- Надійність кожного модулю – постійна підтримка розробників, можливість власноруч змінювати конфігураційні файли.

Висновки до розділу 3

В даному розділі розглянуто закономірності дій працівника у системі, а також наведено приклад сформованого шаблону дій, при виконані встановлених перед працівником бізнес-задач.

Також, більш детально розглянуто метод опорних векторів – проаналізовано функції ядра класифікатора і запропоновану для використання модифікацію дистанційної функції.

Запропоновано власну модель профілювання поведінки користувачів за допомогою метода опорних векторів, в основі якого лежить запропонована функція класифікатора ядра. Модель не має недоліків, що притаманні сучасним системам аналізу поведінки користувачів:

- Система не потребує великої кількості розрахункових потужностей.
- Система збирає інформацію лише про вповноважених користувачів, а всі не потрібні дані відсікаються на етапі передобробки, тому профілювання проходить швидше.
- У запропонованій моделі, згідно дослідів проведених у другому розділі, використовується найкращий метод профілювання поведінки.
- В даній моделі будуються профілі користувачів, через що системи захисту здатні виявляти не лише відомі сценарії атак, а й нові.

А також запропоновано та реалізовано програмне рішення, для вирішення задачі побудови проділей поведінки користувачів, в основі яких лежить запропонована модель.

ВИСНОВКИ

В результаті виконання роботи було повністю досягнуто поставленої мети - підвищення рівня захищеності внутрішньої мережі інфраструктури, шляхом розробки власної моделі профілювання поведінки користувачів за допомогою метода опорних векторів, в основі якого лежить запропонована функція ядра класифікатора. Було розроблено власну модель профілювання поведінки користувачів, математично описано метод, за допомогою якого відбувається профілювання поведінки, а також представлено механізм, на основі якого було реалізовано дану модель.

Використання запропонованої моделі, у внутрішніх мережах інфраструктури різноманітних компаній та організацій дозволить більш детально аналізувати загрози в середині мережі і завчасно виявляти порушників. Механізм, за допомогою якого можливо реалізувати дану модель, є досить простим у налаштуванні та роботі і не вимагає великого рівня знань від адміністратора.

Дана модель не має аналогій, адже працюю згідно розроблених вході роботи принципів, а в основі процесу профілювання лежить запропонована функція ядра, що вперше використовувалась для вирішення даних типів задач.

ПЕРЕЛІК ПОСИЛАНЬ

1. INSIDER THREAT 2018 [Електронний ресурс]. – 2018. – Режим доступу до ресурсу: https://crowdresearchpartners.com/wp-content/uploads/2017/07/Insider-Threat-Report-2018.pdf.
2. Get the 2018 Gartner SIEM Magic Quadrant [Електронний ресурс]. – 2018. – Режим доступу до ресурсу: https://logrhythm.com/gartner-magic-quadrant-siem-report-2018/.
3. IBM QRadar SIEM [Електронний ресурс] – Режим доступу до ресурсу: https://www.ibm.com/ru-ru/marketplace/ibm-qradar-siem.
4. ArcSight Enterprise Security Manager (ESM) [Електронний ресурс] – Режим доступу до ресурсу: https://www.microfocus.com/ru-ru/products/siem-security-information-event-management/features.
5. Splunk® Enterprise Security [Електронний ресурс] – Режим доступу до ресурсу: https://www.splunk.com/en_us/software/enterprise-security.html.
6. McAfee Enterprise Security Manager [Електронний ресурс] – Режим доступу до ресурсу: https://www.mcafee.com/enterprise/ru-ru/products/enterprise-security-manager.html.
7. Passeri P. 2018: A Year of Cyber Attacks [Електронний ресурс] / Paolo Passeri // HACKMAGEDDON. – 2019. – Режим доступу до ресурсу: https://www.hackmageddon.com/2019/01/15/2018-a-year-of-cyber-attacks/.
8. Потери банков от киберпреступности [Електронний ресурс] // Tadviser. – 2019. – Режим доступу до ресурсу: https://is.gd/yfIxoV.
9. Ilascu I. Russian Banks Under Phishing Attack [Електронний ресурс] / Ionut Ilascu // BleepingComputer. – 2018. – Режим доступу до ресурсу: https://www.bleepingcomputer.com/news/security/russian-banks-under-phishing-attack/.
10. GORELIK M. COBALT GROUP 2.0 [Електронний ресурс] / MICHAEL GORELIK // Morphisec. – 2018. – Режим доступу до ресурсу: http://blog.morphisec.com/cobalt-gang-2.0.

11. Lazarus [Електронний ресурс]. – 2017. – Режим доступу до ресурсу: https://xakep.ru/2017/10/18/lazarus-hacked-feib/
12. 74% банков не готовы [Електронний ресурс]. – 2018. – Режим доступу до ресурсу: https://xakep.ru/2019/02/19/banks-not-ready/.
13. Li H. Research and Implementation of an Anomaly Detection Model Based on Clustering Analysis [Текст]/ Han Li., 2010.
14. Онищук Р. М. ОПИС АЛГОРИТМУ ДЛЯ ПОБУДОВИ НАЇВНОГО БАЄСОВСЬКОГО КЛАСИФІКАТОРА [Текст]/ Онищук Р. М.. – 3 с.
15. Vapnik V. N. A Training Algorithm for Optimal Margin Classifiers[Текст] / Vapnik Vladimir N., 2017. – 9 с.
16. Метрики в задачах машинного обучения [Електронний ресурс] – Режим доступу до ресурсу: https://habr.com/ru/company/ods/blog/328372/.
17. User databases [Електронний ресурс] // 2017 – Режим доступу до ресурсу: http://kdd.ics.uci.edu/databases/
18. Machine Learning & Artificial Intelligence [Електронний ресурс] // 2019 – Режим доступу до ресурсу: https://aws.amazon.com/marketplace/solutions/machinelearning?nc2=h_ql_mp.
19. Паттерны поведения в психологии [Електронний ресурс] – Режим доступу до ресурсу: https://experimental-psychic.ru/povedencheskie_patterny/.
20. Hofmann T. Kernel Methods in Machine Learning[Текст]/ Thomas Hofmann, 2008. – (3). – С. 758–823.