УДК 004.056 DOI: 10.31673/2412-4338.2021.044153

Гайдур Г.І., Гахов С.О. Дмітрієв В.Є., Бондаренко Н.В.
Державний університет телекомунікацій, Київ

ВИЯВЛЕННЯ АНОМАЛІЙ ТРАФІКУ В ІНФОРМАЦІЙНИХ СИСТЕМАХ ОРГАНІЗАЦІЙ З ВИКОРИСТАННЯМ МЕТОДІВ MACHINE LEARNING НА ОСНОВІ АЛГОРИТМІВ ПРОГНОЗУВАННЯ КАТЕГОРІЙНИХ ПОЛІВ

Анотація: *У статті досліджено проблему виявлення аномалій в мережевому трафіку інформаційних систем організацій. Виявлення аномалій в мережевому трафіку дозволить визначати скриту шкідливу активність даних, отриманих на основі протоколів, які збирають статистичні дані мережевого трафіку інформаційної системи. Це в свою чергу дозволить зменшити навантаження та налаштувати атрибути за якими буде здійснюватися моніторинг та аналіз мережевого трафіку. Авторами запропонована архітектура виявлення аномалій мережевого трафіку, яка розбита на функціональні рівні. Для збору статистичних даних проаналізовано протоколи, які дозволяють отримувати статистичні дані, а саме протокол Net Flow/IPFIX, який надає вичерпну інформацію на основі заголовків пакетів. Для обробки та аналізу отриманих даних авторами розроблено модель виявлення аномалій в трафіку інформаційної системи. Модель виявлення аномалій використовує статистичні дані для подальшої їх обробки, а також можливість зберігання даних в репозиторії. Всі отримані дані проходять фільтрацію щодо виявлення шкідливих процесів, передаються та зберігаються в репозиторії бази атак з можливістю створення попереджень та ідентифікації атаки.. Для зазначеної моделі запропоновано використання Machine Learning на основі методів прогнозування категорійних полів. В роботі було використано датасет з даними фаєрвола, в якому міститься інформація про кількість та розмір переданих та отриманих пакетів пакетів, дані щодо використання шкідливого програмного забезпечення. Застосовуючи метод було проведено експериментальне дослідження даних щодо прогнозування наявності в них шкідливого програмного забезпечення. Було досліджено метод прогнозування категорійних полів з використанням алгоритмів класифікації Logistic Regression, SVM, Random Forest Classifier та інші. На основі отриманих даних побудовано матрицю плутанини, яка дозволяє оцінити похибку роботи алгоритмів.*

Ключові слова*: інформаційна система, аномалія, атака, метод, модель, алгоритм, машинне навчання.*

Haidur G.I., Gakhov S.O. Dmitriiev V.E., Bondarenko N.V.
State University of Telecommunications, Kyiv

DETECTION OF TRAFFIC ANOMALIES IN THE INFORMATION SYSTEMS OF ORGANIZATIONS USING MACHINE LEARNING METHODS ON THE BASE OF ALGORITHMS FOR FORECASTING CATEGORY FIELDS

Abstract: *The article examines the problem of detecting anomalies in the network traffic of information systems of organizations. Detection of anomalies in network traffic will allow to determine the hidden malicious activity of the data obtained on the basis of protocols that collect statistical data of the network traffic of the information system. This, in turn, will allow you to reduce the load and configure the attributes that will be used to monitor and analyze network traffic. The authors proposed the network traffic anomaly detection architecture, which is divided into functional levels. Protocols were analyzed to collect statistics, namely the Net Flow/IPFIX protocol, which provides comprehensive information based on packet headers. To process and analyze the received data, the authors developed a model for detecting anomalies in the traffic of the information system. The anomaly detection model uses statistical data for their further processing, as well as the possibility of storing data in a repository. All received data is filtered to detect malicious processes, transferred and stored in the repository of the attack database with the possibility of creating warnings and identifying the attack. For the specified model,*

the use of Machine Learning based on methods of predicting categorical fields is proposed. The work used a dataset with firewall data, which contains information on the number and size of transmitted and received packets of packets, data on the use of malicious software. Using the method, an experimental study of the data was conducted to predict the presence of malicious software in them. The method of forecasting categorical fields using Logistic Regression, SVM, Random Forest Classifier and other classification algorithms was investigated. Based on the obtained data, a confusion matrix was built, which allows to estimate the error of the algorithms.

Keywords: *information system, anomaly, attack, method, model, algorithm, machine learning.*

Гайдур Г.И., Гахов С. А., Дмитриев В. Е., Бондаренко Н.В.

Государственный университет телекоммуникаций, Киев

ВЫЯВЛЕНИЕ АНОМАЛИЙ ТРАФИКА В ИНФОРМАЦИОННЫХ СИСТЕМАХ ОРГАНИЗАЦИЙ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ MACHINE LEARNING НА ОСНОВЕ АЛГОРИТМОВ ПРОГНОЗИРОВАНИЯ КАТЕГОРИЙНЫХ ПОЛЕЙ

Аннотация: *В статье исследована проблема выявления аномалий в сетевом трафике информационных систем организаций. Выявление аномалий в сетевом трафике позволит определять скрытую вредную активность данных, полученных на основе протоколов, собирающих статистические данные сетевого трафика информационной системы. Это в свою очередь позволит уменьшить нагрузку и настроить атрибуты, по которым будет осуществляться мониторинг и анализ сетевого трафика. Авторами предложена архитектура обнаружения аномалий сетевого трафика, разбитая на функциональные уровни. Для сбора статистических данных проанализированы протоколы, позволяющие получать статистические данные, а именно, протокол Net Flow/IPFIX, предоставляющий исчерпывающую информацию на основе заголовков пакетов. Для обработки и анализа полученных данных авторами разработана модель обнаружения аномалий в трафике информационной системы. Модель обнаружения аномалий использует статистические данные для дальнейшей обработки, а также возможность хранения данных в репозитории. Все полученные данные проходят фильтрацию по выявлению вредных процессов, передаются и хранятся в репозитории базы атак с возможностью создания предупреждений и идентификации атаки. Для данной модели предложено использование Machine Learning на основе методов прогнозирования категорийных полей. В работе был использован датасет с данными файервола, в котором содержится информация о количестве и размере переданных и полученных пакетов, данные по использованию вредоносного программного обеспечения. Используя метод, было проведено экспериментальное исследование данных по прогнозированию наличия в них вредоносного программного обеспечения. Был исследован метод прогнозирования категорийных полей с использованием алгоритмов классификации Logistic Regression, SVM, Random Forest Classifier и другие. На основе полученных данных построена матрица путаницы, позволяющая оценить погрешность работы алгоритмов.*

Ключевые слова: *информационная система, аномалия, атака, метод, модель, алгоритм, машинное обучение.*

1. Вступ

Забезпечення кібербезпеки інформаційних систем (ІС) є однією з найважливіших проблем для ефективного виконання бізнес-процесів організацій. Адже їх робота залежить від комп'ютерів та обладнання, яке використовується в роботі інформаційних систем. Швидкий розвиток технологій, які забезпечують роботу інформаційних систем викликають зацікавленість та постійну увагу атакуючих зловмисників, які намагаються будь-якими способами отримати гроші, скомпрометувати або нанести просто збиток організаціям. Крім того, інформаційні системи стають все більш складними і взаємопов'язаними, що в свою чергу ускладнює забезпечити відсутність помилок, непередбачених шпаринок, які відкривають доступ атакуючим зловмисникам до інформаційних систем.

Для захисту корпоративного сегменту організацій частіше використовується (рис.1):

– захист кінцевих точок EDR;
– захист периметру (Firewall, NGFirewall).

При організації таких заходів, як правило недостатня увага приділяється видимості мережі (network visibility and security).

Рис 1. Технології захисту корпоративного сегменту інформаційної системи

Захист периметру інформаційної системи полягає у захисті територіально-розподілених інформаційних систем як великого, так і маленького розміру. Це досягається шляхом використання в організації міжмережевих екранів (пристроїв захисту периметра), які в залежності від технології обробки інформації можуть бути в різній комплектації і забезпечувати різний функціонал, такий як розмежування та контроль доступу, ідентифікація та аутентифікація користувачів, трансляція IP-адрес (NAT), застосовувати VPN, виконувати аналіз трафіку та ряд інших можливостей.

Захист кінцевих точок інформаційної системи, до яких відносяться персональні комп'ютери або сервери, полягає у видимості шкідливої діяльності на кінцевій точці, що дозволяє фахівцям з кібербезпеки контролювати кінцеві точки для стримування та пом'якшення атак. Механізми виявлення шкідливої діяльності допомагає виявляти атаки на пристрої кінцевих точок і надає швидкий доступ до інформації, яка може допомогти розслідувати атаку. Дії реагування можуть автоматично відповідати на атаки, виконуючи дії на рівні пристрою, такі як карантин кінцевої точки або блокування шкідливих процесів.

Видимість мережі має велике значення для кібербезпеки. Це пов'язано з проблемами, які виникають через велику кількість об'ємів даних, появі «сліпих зон» мережі. Згідно з опитування Vanson Bourne приблизно 67% організацій говорять про те, що саме сліпі зони є великою проблемою при забезпеченні захисту своїх даних [1, 2, 3].

Поширення невідомих шкідливих програм, ставлять під загрозу внутрішні системи, руйнівні DDoS-атаки, APT та загрози, що обходять традиційні системи кібербезпеки, змінили ландшафт IT-безпеки. Будувати захист периметру і покладатися на рішення, що базуються на сигнатурах, вже недостатньо.

Відповіддю на цю проблему, за рекомендацією Gartner [18], є попереджувальне виявлення та пом'якшення аномалій мережі. Рішення NBAD постійно спостерігають за мережевим трафіком, аналізуючи обмін даними для пошуку аномалій та виявлення підозрілої поведінки. Це дозволяє реагувати на ще невідомі загрози кібербезпеки, які не виявляються іншими технологіями.

Тому дана тема потребує особливого розгляду, а саме використання сучасних методів виявлення аномалій мережевого трафіку в інформаційних системах організацій.

2. Аналіз літературних даних і постановка проблеми

Видимість мережі дозволяє IT-фахівцям та фахівцям кібербезпеки бути в курсі всього всередині мережі та пересуватись по ній за допомогою інструментів контролю мережі. Інструменти моніторингу мережі використовуються для спостереженням за мережевим трафіком, відслідковуванням за додатками, продуктивністю мережею, керуванням мережевими ресурсами. Чим більша видимість мережі організації, тим більше контролю над мережевими даними інформаційної системи організації. Такий підхід дозволяє виявляти аномалії мережевого трафіку на ранніх стадіях, шляхом отримання даних з використанням протоколів SNMP, Net Flow/IPFIX, які збирають дані з мережевих пристроїв. Ці дані є вихідними даними для розробки методів виявлення аномалій мережевого трафіку притаманних різному виду атак.

Як зазначено в [18], моніторинг мережі є вимогою для багатьох підприємств та гарною практикою для всіх. Аналіз мережевої поведінки (NBA) виходить за рамки пошуку відомих поганих сигнатур атак і дозволяє зрозуміти, що відбувається у мережі. NBA не замінює системи, що базуються на підписах, вона їх доповнює, щоб дати фахівцям з кібербезпеки більш повне уявлення про мережу.

В [3] зазначено що уразливості та атаки, що постійно розвиваються є загрозою кібербезпеки сучасного кіберпростору. Застосування методів виявлення аномалій в мережевому трафіку дозволить підвищити ефективність засобів кібербезпеки.

В [4,5,6] зазначається що кібератаки постійно модифікуються з розвитком засобів захисту від кібератак. Виявлення кібератак класичними засобами не завжди є ефективними. Саме з використанням аналізу аномалій мережевого трафіку можна завчасно вжити заходи, щодо виявлення та запобігання вторгнень в інформаційній системі організації.

В [7] досліджено теоретико-методологічні і практичні аспекти методів аномальних станів, які дозволяють створити методологію подуви систем виявлення вторгнень на основі аналізу мережевої поведінки.

Застосування методів аналізу поведінки мережного трафіку, аномалії в мережі можна буде виявляти якнайшвидше, наприклад, атака з впровадженням ботів, атака з використанням міжсайтових сценаріїв (XSS), атака з обходом каталогу та інші типи атак. Методи виявлення аномалій необхідні для підвищення адаптивності і масштабованості через збільшення обсягів даних про трафік. Традиційні методи, які використовують системи виявлення вторгнень не завжди спрацьовують. Наприклад, зловмисникам легко обійти визначені правила виявлення, і нові невідомі атаки не можуть бути виявлені за допомогою правил, що базуються на існуючих атаках. Таким чином, методи, що ґрунтуються на правилах, часто страждають від високого рівня хибних спрацьовувань. По суті, виявлення аномалій у мережевому трафіку це завдання класифікації даних. Він спрямований на те, щоб відрізняти дані про атаки від звичайної поведінки. Крім традиційних методів виявлення на основі правил [8, 9], у задачі виявлення аномального трафіку широко використовуються зазначені методи, які засновані на статистичній теорії [10, 11], теорії інформації [12, 13] та машинному навчанні [14, 15]. Модель виявлення на основі машинного навчання є багатообіцяючим методом інтелектуального виявлення аномалій у великомасштабному мережному середовищі з високою пропускною здатністю.

Тому основну увагу даного дослідження буде зосереджено на створені моделі виявлення аномалій мережевого трафіку на основі впровадження методів машинного навчання.

3. Мета і задачі дослідження

Метою дослідження є розробка моделі процесу виявлення аномалій у мережевому трафіку інформаційної системи.

Для досягнення мети було вирішено наступні завдання, які розбито на етапи:

1. Розроблена архітектура моніторингу та збору даних мережевого трафіку. Проведено вибір статистичних даних.

2. Запропоновано модель виявлення атак на основі аномальної поведінки.

3. Розроблено модель виявлення аномалій на основі алгоритмів Machine Learning.

4. Здійснено моделювання згідно запропонованої моделі на основі алгоритмів прогнозування категорійних полів.

4. Архітектура збору даних мережевого трафіку інформаційної системи

Згідно [12, 13] архітектуру виявлення аномалій мережевого трафіку можна представити у вигляді відповідних рівнів, кожен з яких виконує свої функції (рис. 2).

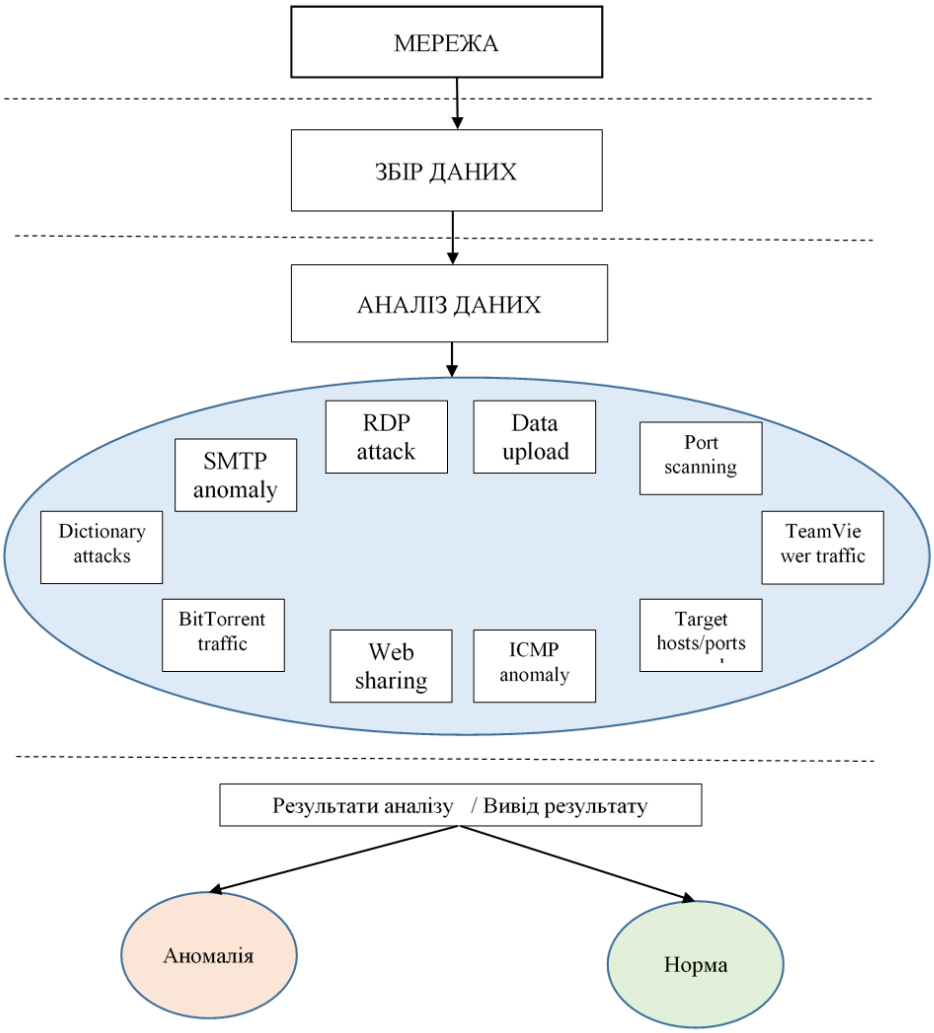


Рис 2. Архітектура виявлення аномалій мережевого трафіку

Збір даних в мережі відбувається на основі протоколів моніторингу мережевого трафіку може базуватись на таких технологіях [10,12]:

1. SNMP - це традиційний і простий метод для ресурсів IT-інфраструктури, що спочатку призначений для управління мережею. Він збирає дані з мережевих пристроїв, надаючи інформацію про їх доступність і стан (використання ЦП і ОЗП, пропускну здатність, що споживається мережевим пристроєм і т.д.)

2. Net Flow/IPFIX є пасивною безагентною технологією, призначеною для моніторингу мережі з кількома операційними програмами та програмами безпеки. Протокол Net Flow/IPFIX надає вичерпну інформацію про те, хто з ким спілкується, коли, як довго і як часто (IP-адреси,

обсяги даних, час, порти, протоколи та інші технічні характеристики зв'язку TCP/IP на третьому та четвертому мережевих рівнях). Моніторинг мережного трафіку за допомогою NetFlow генерує статистику як за базовими передачами даних, так і за вмістом повідомлення. Статистика NetFlow надається мережевими елементами (маршрутизаторами, комутаторами) або спеціалізованими автономними апаратними датчиками. Зонди прозоро підключаються до контрольованої мережі як пасивні пристрої, створюючи точний та докладний потік статистики з копії мережного трафіку. Цей підхід використовується для подолання різних обмежень продуктивності та функцій моніторингу NetFlow на основі маршрутизатора.

З вище наведених технологій найбільш використовуваним є метод Net Flow/IPFIX, який аналізує тільки заголовки пакетів (рис.3). Корисне навантаження не відслідковується і не зберігається. Співвідношення об’єму даних та статистики 500:1, що дозволяє скоротити час на аналіз та збір статистики.

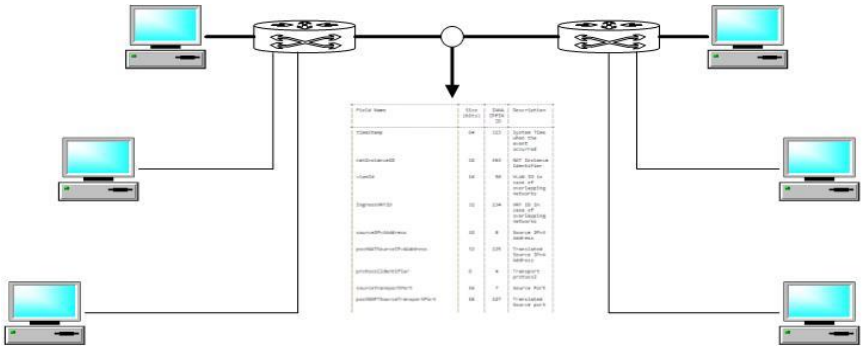


Рис 3. Збір даних для аналізу

Усі пакети, що належать до певного потоку, мають набір загальних характеристик. Кожна властивість визначається як результат застосування функції до значень (рис. 3). Статистичні дані, які збираються до бази даних записуються наступним чином.

```
protocol = UInt8Field(description=_("IP protocol type"))
src_tos = UInt8Field(description=_("Type of Service byte"))
tcp_flags = UInt8Field(description=_("TCP flags cumulative byte"))
l4_src_port = UInt8Field(description=_("TCP/UDP src port number"))
ipv4_src_addr = IPv4Field(description=_("IPv4 source address"))
src_mask = UInt8Field(description=_("Number of mask bits in src adr"))
ipv4_dst_port = UInt8Field(description=_("TCP/UDP dst port number"))
ipv4_dst_addr = IPv4Field(description=_("IPv4 destination address"))
dst_mask = UInt8Field(description=_("Number of mask bits in dst adr"))
output_snmp = UInt16Field(description=_("Output interface index"))
ipv4_next_hop = IPv4Fielddescription=_("IPv4 adr of nexthop router"))
src_as = UInt32Field(description=_("Src BGP AS number"))
        .
        .
dst_as = UInt32Field(description=_("Dst BGP AS number"))
bgp_ipv4_next_hop = UInt32Field(description=_("Nexthop router's IP in BGP domain")).
```

5. Модель аналізу даних щодо виявлення аномалій на базі Machine Learning

Виявлення аномального трафіку допоможе визначити обставини та мотиви, шляхом вилучення необхідної інформації з отриманих даних [9,11,13]. Отриманні дані щодо проходження

атаки необхідно розуміти які данні задіяні в її проходженні і необхідні при застосуванні методів виявлення аномалій. База даних містить різні дані, але для певного виду атаки, для аналізу трафіка необхідно застосувати певні данні, які можуть мати сліди атаки і ознаки за якими їх буде визначено. У зв'язку з тим, що класичні методи не завжди в повній мірі справляються з поставленими задачами виявлення аномалій розроблено запропоновану модель виявлення аномального трафіку на рис.4.

Ефективність виявлення аномалій в трафіку інформаційної системи можна підвищити за рахунок впровадження нових методів. До таких методів відносяться методи Machine Learning. Machine Learning використовуються для вирішення різних задач, таких як прогнозування значення поля, прогнозування майбутніх значень, виявлення закономірностей даних і виявлення аномалій нових даних. Набір інструментів машинного навчання (MLTK) дозволяє користувачам створювати, перевіряти, керувати та впроваджувати моделі машинного навчання з використанням керованого інтерфейсу користувача[13-15,17,18].

Для проведення експерименту було застосовано програмне середовище Splunk_ML_Toolkit [18], яке використовує команди SPL, надає налаштовану візуалізацію, приклади для вивчення різних концепцій машинного навчання Machine Learning. Splunk_ML_Toolkit включає наскрізні приклади з наборами даних, з можливістю візуалізацію та команди SPL до обраних даних.(рис. 5)

Експериментальна модель базується на даних датасету, які отримані з базового набору бібліотеки Splunk_ML_Toolkit *Firewall_trafic.csv*. Обраний датасет (http://127.0.0.1:8000/en-US/app/Splunk_ML_Toolkit/smart_prediction?experimentId=%2FservicesNS%2Fgaydur%2FSplunk_ML_Toolkit%2Fmltk%2Fexperiments%2Fff7987f35e6f45f7bbfd7e861f8b74ce) складається з таких даних:

bytes_received
bytes_sent
dest_port
dst_ip
has_known_vulnerability
packets_received
packets_sent
receive_time
serial_number
session_id
src_ip src_port

Одним з класичним методів використання методів Machine Learning - є метод прогнозування категорійних полів, який відноситься до типу навчання класифікації. Даний тип класифікації вивчає тенденцію приналежності даних до тієї чи іншої категорії на основі зв'язаних даних [14, 17, 18].

Прогнозування категорійних полів використовує наступні алгоритми класифікації:

- Logistic Regression;
- SVM;
- Random Forest Classifier;
- Gaussian NB;
- Bernoulli NB;
- Decision Tree Classifier.

Розглянемо приклад прогнозування використання злоякісного програмного забезпечення для даних які отримано з використанням протоколу Net Flow/IPFIIX, з яких було отримано дані щодо переданих байтів, отриманих байтів, порт призначення, пакети отримано, пакети

відправлено. Застосуємо кожен алгоритм в програмному середовищі Splunk ML toolkit згідно представленої моделі згідно рис. 5 [14].

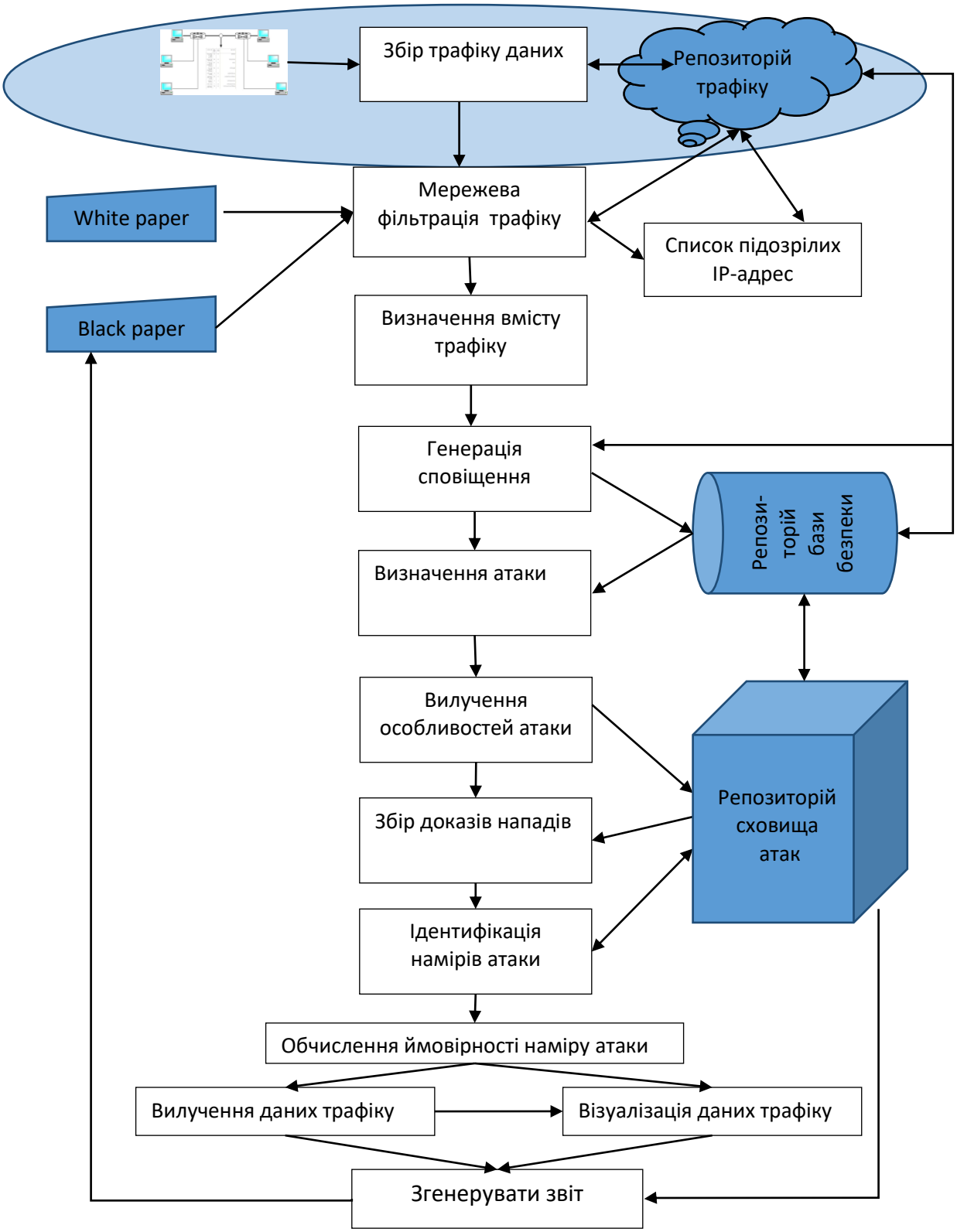


Рис 4. Модель виявлення аномалій в трафіку інформаційної системи

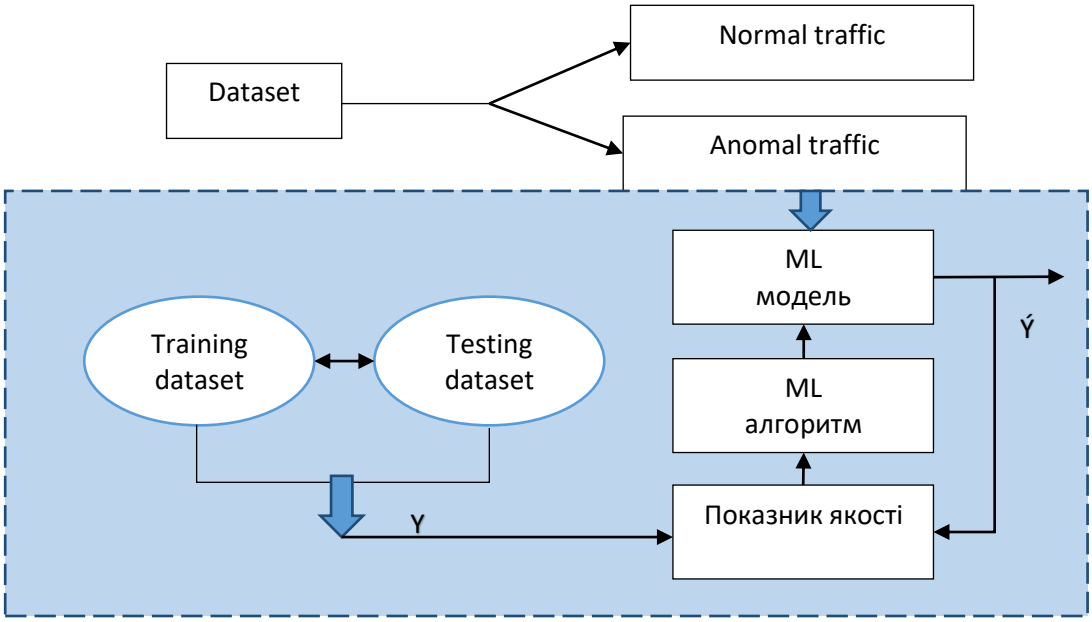


Рис. 5. Модель роботи алгоритмів Machine Learning

6. Результати експерименту

Отримані данні були оброблені та розділені на Testing dataset та Testing dataset у співвідношенні 70/30.

Застосування вищезгаданих алгоритмів дозволить оцінити роботу моделі за наступними критеріями (табл.1):

Precision/ Точність –це статистика, яка відображає процент часу коли модель показує що передбачений клас є правильний.

Recall - це статистика, яка відображає процент часу коли модель прогнозує правильний клас.

Accuracy - це статистика, яка відображає модель що передбачає загальний процент правильних прогнозів.

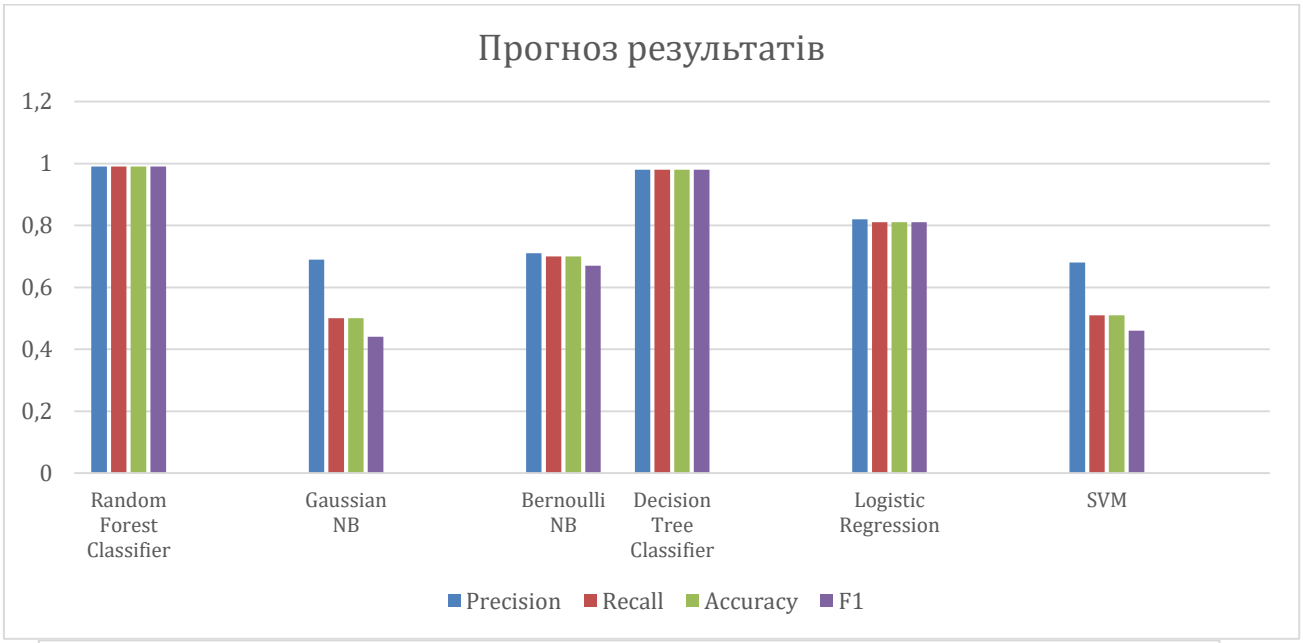
F1- це статистика є середньозваженим значенням точності і повноти, засноване на шкалі від нуля до одиниці. Що ближче статистика до одиниці, то краще підходить модель.

Таблиця 1

Оцінка прогнозів алгоритмів

Алгоритм	Precision	Recall	Accuracy	F1
Random Forest Classifier	0.99	0.99	0.99	0.99
Gaussian NB	0.69	0.50	0.50	0.44
Bernoulli NB	0.71	0.70	0.70	0.67
Decision Tree Classifier	0.98	0.98	0.98	0.98
Logistic Regression	0.82	0.81	0.81	0.81
SVM	0.68	0.51	0.51	0.46

Результат експериментального дослідження показав, що для заданих критеріїв прогнозування класифікації полів кращі результати показали алгоритм Random Forest Classifier та Decision Tree Classifier (рис. 6). Це свідчить про те, що дані алгоритми краще працюють з обраним набором даних для виявлення аномальної поведінки мережевого трафіку щодо використання шкідливого програмного забезпечення.



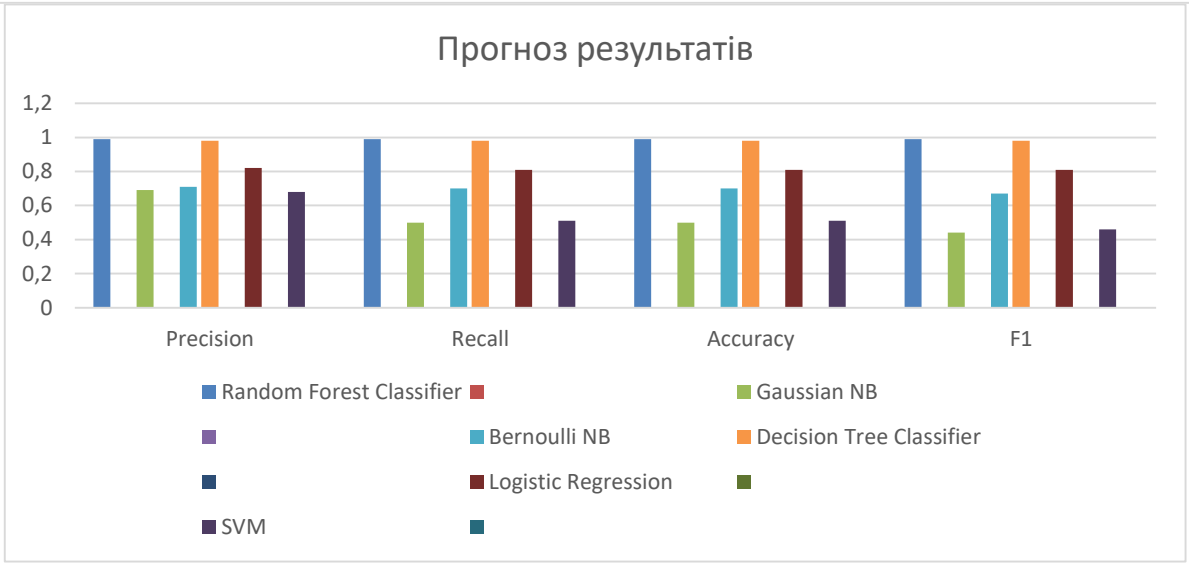


Рис 6. Результати роботи алгоритмів а) прогноз за алгоритмами;
б) прогноз за критеріями

Проведемо оцінку продуктивності моделі для обраних алгоритмів прогнозування категорійних полів на основі Machine Learning. Визначення продуктивності моделі на наборі тестових даних основана на матриці плутанини.

Матриця плутанини візуалізує точність класифікатора, порівнюючи фактичні та прогнозовані класи. Матриця двійкової плутанини складається з квадратів: Predicted actual False та Predicted actual True (табл.2), де:

• TP: True Positive: прогнозовані значення, що правильно прогнозуються як фактичні позитивні;

• FP: Передбачені значення неправильно передбачають фактичний позитивний результат. тобто. негативні значення прогнозуються як позитивні;

• FN: False Negative: позитивні значення прогнозуються як негативні;

• TN: True Negative: прогнозовані значення, які правильно прогнозуються як фактичні негативні.

Таблиця 2

Таблиця плутанини

Predicted actual False	TN	FP
Predicted actual True	FN	TP

Розрахунок тесту точності з матриці плутанини (табл.3):

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Таблиця 3

Матриця плутанини

Алгоритм	Predicted actual	Predicted False	Predicted True
Random Forest Classifier	False	5896 (98.8%)	74 (1.2%)
	True	83 (0.9%)	9073 (99.1%)
Gaussian NB	False	5565 (94.2%)	343 (5.8%)
	True	7245 (79%)	1924 (21%)
Bernoulli NB	False	2062 (35.5%)	3753 (64.5%)
	True	689 (7.6%)	8414 (92.4%)
Decision Tree Classifier	False	5828 (98.3%)	101 (1.7%)
	True	137 (1.5%)	8992 (98.5%)
Logistic Regression	False	4770 (82.1%)	1038 (17.9%)
	True	1845 (20.2%)	7282 (79.8%)
SVM	False	5591 (93.2%)	405 (6.8%)
	True	6898 (76.7%)	2095 (23.3%)

Таким чином матриця плутанини показує як обрана модель класифікації плутається, коли вона робить прогноз. Отримана матриця дозволяє зробити представлення про похибки, які допущені класифікатором і тим самим отримати типи похибок, які допускаються.

Експертним методом доведено, що за отриманими прогнозами очікування класифікації категорійних полів для заданих вхідних даних ефективним є алгоритм Random Forest Classifier та Decision Tree Classifier. Для інших алгоритмів похибка досить велика, що приведе до хибних спрацювань в системі виявлення аномалій. Тож для використання інших алгоритмів необхідно доопрацювання вхідних даних мережевого трафіку та математичного апарату класифікації.

8. Висновки

Виявлення аномалій в мережевому трафіку інформаційних систем організацій, це ефективні заходи щодо виявлення атак, проходження та виявлення яких класичними методами стають не ефективними. Існуючі методи не відповідають швидкій реакції на зміни, які відбуваються при кібератаках. Запропонована архітектура моніторингу та збору мережевого трафіку дозволяє проводити аналіз отриманих даних для виявлення атак на основі статистичних даних. Ці дані отримуються з використанням протоколу Net Flow/IPFIX, який використовує дані тільки заголовків пакетів. Це дозволяє зменшити навантаження на мережу. Однак для аналізу статистичних даних згідно моделі виявлення аномалій доцільно використовувати Machine Learning. Тому в даній роботі було проведено дослідження застосування методу прогнозування категорійних полів, на основі розробленої моделі роботи алгоритмів Machine Learning, а саме алгоритмів методу прогнозування категорійних полів.

В результаті дослідження було експериментальним шляхом отримано дані щодо точності прогнозів обраних алгоритмів та проведено оцінку продуктивності мережі з використанням матриці плутанини. Подальші дослідження можуть бути присвячені підвищенню ефективності алгоритмів Machine Learning для виявлення аномалій в мережевого трафіку відповідно до обраного типу атаки.

Список використаної літератури

1 Detecting Abnormal Cyber Behavior Before a Cyberattack. March 5, 2021. Online: https://www.nist.gov/blogs/manufacturing-innovation-blog/detecting-abnormal-cyber-behavior-cyberattack (viewed on July, 27, 2021).

2. Гайдур Г.І., Гахов С.О., Марченко В.В. Метод побудови динамічної моделі логічного об'єкта інформаційної системи та визначення закону його функціонування. Radioelectronic and Computer Systems, 2022, no. 1(101). С. 129-14. doi: 10.32620/reks.2022.1.10.

3. Qian Ma, Cong Sun, Baojiang Cui, A Novel Model for Anomaly Detection in Network Traffic Based on Support Vector Machine and Clustering Security and Communication Networks. Security and Communication Networks. Volume 2021. doi: 10.1155/2021/2170788.

4. Казмірчук С.В., Корченко А.О., Паращук Т.І. Аналіз систем виявлення вторгнень. Захист інформації. Том 20 № 4 (2018), 2018. С. 259-276.

5. O. Lawal, "Analysis and Evaluation of Network Based Intrusion Detectionand Prevention System in an Enterprise Network Using Snort Freeware", African Journal of Computing & ICT, Ibadan, Vol. 6, no. 2, 2013, pp. 169-184.

6. S. Cooper, 11 Top Intrusion Detection Tools for 2021. [Electronic resource]. Online: https://www. comparitech.com/net-admin/network-intrusion-detection-tools. (viewed on July, 27, 2021).

7. Анна Корченко. Методи ідентифікації аномальних станів для систем виявлення вторгнень. Монографія. Київ. ЦП «Компринт», 2019, 361с.

8. RFC 7011. Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of Flow Information, September 2013.

9. M. V. Mahoney and P. K. Chan, "Learning rules for anomaly detection of hostile network traffic," in Proceedings of the Third IEEE International Conference on Data Mining, pp. 601–604, IEEE, Leipzig, Germany, July 2003.

10. Гайдур Г.І., Гахов С.О. Теоретичний підхід до вирішення проблеми виявлення шкідливих процесів на основі аналізу станів логічного об'єкта інформаційної системи. Телекомунікаційні та інформаційні технології. № 1 (70), 2021. С.79-87.

11. E. Eskin, Anomaly Detection over Noisy Data Using Learned Probability Distributions, Citeseer, Princeton, New Jersey, USA, 2000.

12. W. Lee and D. Xiang, "Information-theoretic measures for anomaly detection," in Proceedings of the 2001 IEEE Symposium on Security and Privacy, S&P 2001, pp. 130–143, IEEE, Philadelphia, PA, USA, November 2000.

13.M. A. Ambusaidi, Z. Tan, X. He, P. Nanda, L. F. Lu, and A. Jamdagni, "Intrusion detection method based on nonlinear correlation measure," International Journal of Internet Protocol Technology, vol. 8, no. 2-3, 2014, pp. 77–86.

14. Splunk® Machine Learning Toolkithttps://docs.splunk.com/Documentation/MLApp (viewed on September, 18, 2021).

15. M. Ahmed, A. Naser Mahmood, and J. Hu, "A survey of network anomaly detection techniques," Journal of Network and Computer Applications, vol. 60, 2016, pp. 19–31.

16. Clarence Chio, David Freeman, Machine Learning and Security, O'Reilly Media, Inc. 118. 2018. ISBN: 9781491979907

17. Alexey Nefedov, Support Vector Machines: A Simple Tutorial, Creative Commons Attribution, 32, 2016.

18. Network Behavior Analysis: Moving Beyond Signatures https://www.gartner.com/en/documents/1405498 (viewed on October, 2, 2021).

References

1 Detecting Abnormal Cyber Behavior Before a Cyberattack. March 5, 2021. Online: https://www.nist.gov/blogs/manufacturing-innovation-blog/detecting-abnormal-cyber-behavior-cyberattack (viewed on July, 27, 2021).

2. Haydur H.I., Gakhov S.O., Marchenko V.V. The method of building a dynamic model of a logical object of the information system and determining the law of its functioning. Radioelectronic and Computer Systems, no. 1(101). pp. 129-14, 2022. doi: 10.32620/reks.2022.1.10.

3. Qian Ma, Cong Sun, Baojiang Cui, A Novel Model for Anomaly Detection in Network Traffic Based on Support Vector Machine and Clustering Security and Communication Networks. Security and Communication Networks. Volume 2021. doi: 10.1155/2021/2170788.

4. Kazmirchuk S.V., Korchenko A.O., Paraschuk T.I. Analysis of intrusion detection systems. Protection of information. Volume. 20 № 4 (2018), pp. 259-276. 2018.

5. O. Lawal, "Analysis and Evaluation of Network Based Intrusion Detectionand Prevention System in an Enterprise Network Using Snort Freeware", African Journal of Computing & ICT, Ibadan, Vol. 6, no. 2, pp. 169-184, 2013.

6. S. Cooper, 11 Top Intrusion Detection Tools for 2021. [Electronic resource]. Online: https://www. comparitech.com/net-admin/network-intrusion-detection-tools/ (viewed on July, 27, 2021).

7. Anna Korchenko, Methods of identifying abnormal states for intrusion detection systems. Monograph. Kyiv. TSP "Komprynt", 361 p., 2019.

8. RFC 7011. Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of Flow Information, September 2013.

9. M. V. Mahoney and P. K. Chan, “Learning rules for anomaly detection of hostile network traffic,” in Proceedings of the Third IEEE International Conference on Data Mining, pp. 601–604, IEEE, Leipzig, Germany, July 2003.

10. Haydur H.I., Hakhov S.O. A theoretical approach to solving the problem of detecting malicious processes based on the analysis of the states of the logical object of the information system. Telecommunications and information technologies. №1 (70), pp.79-87. 2021.

11. E. Eskin, Anomaly Detection over Noisy Data Using Learned Probability Distributions, Citeseer, Princeton, New Jersey, USA, 2000.

12. W. Lee and D. Xiang, “Information-theoretic measures for anomaly detection,” in Proceedings of the 2001 IEEE Symposium on Security and Privacy, S&P 2001, pp. 130–143, IEEE, Philadelphia, PA, USA, November 2000.

13.M. A. Ambusaidi, Z. Tan, X. He, P. Nanda, L. F. Lu, and A. Jamdagni, “Intrusion detection method based on nonlinear correlation measure,” International Journal of Internet Protocol Technology, vol. 8, no. 2-3, pp. 77–86, 2014.

14. Splunk® Machine Learning Toolkithttps://docs.splunk.com/Documentation/MLApp (viewed on September, 18, 2021).

15. M. Ahmed, A. Naser Mahmood, and J. Hu, “A survey of network anomaly detection techniques,” Journal of Network and Computer Applications, vol. 60, pp. 19–31, 2016.

16. Clarence Chio, David Freeman, Machine Learning and Security, O'Reilly Media, Inc. 118. 2018. ISBN: 9781491979907

17. Alexey Nefedov, Support Vector Machines: A Simple Tutorial, Creative Commons Attribution, 32, 2016.

18. Network Behavior Analysis: Moving Beyond Signatures https://www.gartner.com/en/documents/1405498. (viewed on October, 2, 2021).