DOI 10.15276/imms.v12.no1-2.29
УДК 004.056.53: 004.77

Informatics and Mathematical Methods in Simulation
Vol. 12 (2022), No. 1-2, pp. 29-39

ВИЯВЛЕННЯ ВТОРГНЕНЬ В МЕРЕЖУ ІНТЕРНЕТУ РЕЧЕЙ ШЛЯХОМ АНАЛІЗУ НАЯВНОСТІ АНОМАЛІЙ В МЕРЕЖЕВОМУ ПОТОЦІ ТРАФІКУ AAA ПРОТОКОЛІВ

Л.Ю. Гальчинський

Національний Технічний Університет України «КПІ» імені Ігоря Сікорського
просп. Перемоги, 37, Київ, 02056 Україна; e-mail: hleonid@gmail.com

Одним з найпотужніших трендів інноваційних змін у сучасному світі зараз є Інтернет речей IoT. І хоча цей тренд по-суті тільки на початковому етапі, він вже грає помітну роль в сучасному світі, у сучасній цивілізації. Можна впевнено прогнозувати подальше розширення сфер його застосування. Проте поява цього феномена супроводжується, також, появою пов'язаних з цим проблем. В першу чергу це стосується кіберзагроз, які можуть принести значну шкоду як самим мережам Інтернету речей, так і людям, що користуються цими мережами. Відповідно виникає необхідність у засобах боротьби з цими загрозами. Це складна проблема, яка, в першу чергу, вимагає розробку методів та технологій виявлення такого роду загроз. При цьому ще не склалися чіткі поняття про релевантні рішення для кіберзахисту різних об'єктів застосувань IoT, особливо там, де фінансові можливості обмежені. Тому актуальним є дослідження щодо знаходження рішень, які можуть бути прийнятними для порівняно малобюджетних об'єктів. Метою роботи є отримання методу виявлення вторгнень в мережу IoT, шляхом виявлення аномалій в мережевому потоці трафіку AAA протоколів. Ця проблема ускладнюється тим, що знайдене рішення має відповідати основним політикам безпеки. Проведене нами дослідження показало, що таке рішення принципово можливе шляхом використання можливостей AAA протоколу RADIUS та використання машинного навчання. Методи машинного навчання були перевірені на датасеті UNSW_NB15. Для проведення машинного навчання була проведена попередня обробка з метою нормалізації та стандартизації даних. Комп'ютерні експерименти дозволили перевірити придатність виявлення набору кібератак для різних моделей машинного навчання. Було показано, що сформульований підхід має потенціал для використання у практичних розробках виявлення вторгнень в мережі IoT.

Ключові слова: Інтернет речей, IoT, аномалії у мережевому потоці, AAA протоколи, метод аналізу ієрархій, машинне навчання, датасет

Вступ

Інтернет речей IoT став помітним трендом у сучасному світі, керованому технологіями. Інтернет речей поширюється на різні сфери — від розумних переносних пристроїв до розумних міст, від сфери побуту до промислових застосувань. За оцінками Gartner на 2020 рік Інтернет речей (IoT) мав складати до 26 мільярдів встановлених одиниць, а постачальники продуктів і послуг Інтернету речей мали отримати додатковий дохід, що перевищує 300 мільярдів доларів [1].

Однак цей процес породжує також і складні проблеми, зокрема безпекові. Різноманітні пристрої мають доступ один до одного через віртуальну приватну мережу. Ці комунікаційні процеси дуже вразливі до атак [2].

Основні методи атак на IoT мережі включають атаки типу «відмова в обслуговуванні» (DoS), підробка даних та моніторинг мережі [3].

Зараз методи виявлення вторгнень поділяються на виявлення вторгнень на основі хоста та виявлення вторгнень на основі мережі.

Існує значний корпус досліджень, які висвітлюють питання виявлення вторгнень в мережу на основі виявлення аномалій в мережевому потоці. В цих дослідженнях пропонуються різноманітні моделі виявлення та аналізу аномалій, такі як глибинна

нейронна мережа [3], мережа Маркова [4], метод зворотного поширення помилки [5], об'єднання інформації на основі графу атаки та аналіз основних компонентів [6].

Проте на даний час за даними існуючих досліджень проблема виявлення аномалій мережевого трафіку IoT пристроїв все ще далека від остаточного вирішення. В першу чергу це обумовлено різноманітністю мережевого трафіку різних пристроїв Інтернету речей, гетерогенністю пристроїв в мережах IoT та мережевого середовища, слабкою захищеністю пристроїв тощо. Важливим фактором також є можливості суб'єктів, що використовують мережі IoT, оскільки впровадження систем безпеки може бути досить витратним і дорога система кіберзахисту може бути суб'єкту не по кишені.

В даній роботі основний фокус уваги зосереджено на ефективності та простоті впровадження запропонованого рішення виявлення вторгнень у вже існуючі мережі пристроїв Інтернету речей на основі аналізу даних обліку AAA протоколів в реальному часі.

Аналіз досліджень і публікацій

Проведено багато досліджень по всьому світу щодо безпеки різноманітних IoT інфраструктур, в основному зосереджені на виявленні шкідливого коду та мережевих атак [7].

Для досягнення швидкого виявлення атак зловмисного коду автори в [8] запропонували безпечну та захищену конфіденційну схему агрегації, що ґрунтується на адитивному гомоморфному шифруванні та операціях перешифрування проксі в криптосистемі Paillier.

У роботі [9] автори використовували дизассемблер і метод статистичного аналізу для боротьби з виявленням шкідливого коду. Наразі, згідно статистичних даних, головною точкою входу хакерів для здійснення атаки на мережу IoT, особливо на IIoT, є розумний лічильник.

Механізм виявлення вторгнення відстежує події, що відбуваються в розумному лічильнику, і аналізує їх. Після атаки або виявлення потенційної загрози безпеці механізм виявлення вторгнення видає сигнал тривоги, щоб система та менеджери застосували відповідні механізми реагування.

Зі збільшенням об'єму трафіку IoT та шумів, метод виявлення аномалій, що ґрунтується на традиційному машинному навчанні, стикається з проблемами низької точності та низької надійності вилучення ключових ознак, що знижує продуктивність виявлення атак мережевих атак. Тому зараз активно розвивається метод виявлення аномалій, що ґрунтується на глибокому навчанні [10,11].

Ряд дослідників звернули увагу на організаційні аспекти впровадження систем кіберзахисту для мереж IoT. Зокрема в роботі [12] запропоновано систему виявлення аномалій та ідентифікації пристроїв розумних будинків із застосуванням колективної комунікації. Суть нового підходу в організації безпеки розумних будинків, яке передбачає їх об'єднання у кластери, що дає можливість не тільки об'єднати ресурси власників для впровадження системи захисту, але ще й отримати більш високий рівень достовірності виявлення вторгнень. Якщо це можливо, то це безумовно слушна ідея.

Мета і задачі дослідження

Метою роботи є отримання методу виявлення вторгнень в мережу IoT, на основі виявлення аномалій в мережевому потоці для суб'єкта, який, не маючи значних ресурсів, бажає отримати систему захисту мережі IoT, що задовольняла б вимогам більшості політик безпеки.

Шляхів вирішення проблеми захисту від вторгнень декілька.

1. Самостійна розробка. Це досить довгий і витратний шлях, що передбачає створення команди аналітиків, програмістів та IT- спеціалістів.
2. Послуги спеціалізованих організацій, які надають всі необхідні послуги разом. На перший погляд це виглядає більш ефективним, У такому разі умовна компанія лише налаштовує в своїй мережі переадресацію пакетів в оточення спеціалізованої компанії, де

спеціалізована компанія їх аналізує та надсилає звіт про можливі вторгнення. Проте тут можуть стати на заваді політики безпеки, які можуть забороняти третім сторонам мати повне бачення всього, що відбувається у внутрішній мережі компанії.

3. Придбання IDS або IPS. Таке рішення однак потребує значних витрат на впровадження. Навіть, якщо використовувати рішення з відкритим кодом, такі як Snort, Suricata або Bro IDS.

Мета цього дослідження полягає, у знаходженні можливостей рішення, яке не буде потребувати від умовної компанії значних бюджетних та часових витрат, а також не буде суперечити більшості політик безпеки.

Основна частина

У сучасних корпоративних мережах для розмежування доступу до мережі переважно використовують AAA (authentication, authorization, accounting) сервери [13]. Враховуючи цю обставину, в даній роботі пропонується наступне рішення: вже існуючі AAA сервери налаштувати для перенаправлення запитів обліку на вже розгорнуті та налаштовані AAA сервери. Ці AAA сервери роблять попередню обробку даних обліку, стандартизовану для всіх інфраструктур, та завантажують оброблені дані в центр обробки даних. В центрі обробки отримані дані аналізуються у реальному часі методами машинного навчання, та у випадку виявлення аномалій надсилається відповідне повідомлення.

Фактично, запропоноване рішення не буде потребувати практично жодних додаткових бюджетних витрат, а тільки незначні витрати часу для налаштування переадресації даних обліку на AAA сервері. У той же час дані обліку вже мають стандартизований формат для будь-якого пристрою, відповідно попередня обробка даних буде однаковою для будь якої інфраструктури.

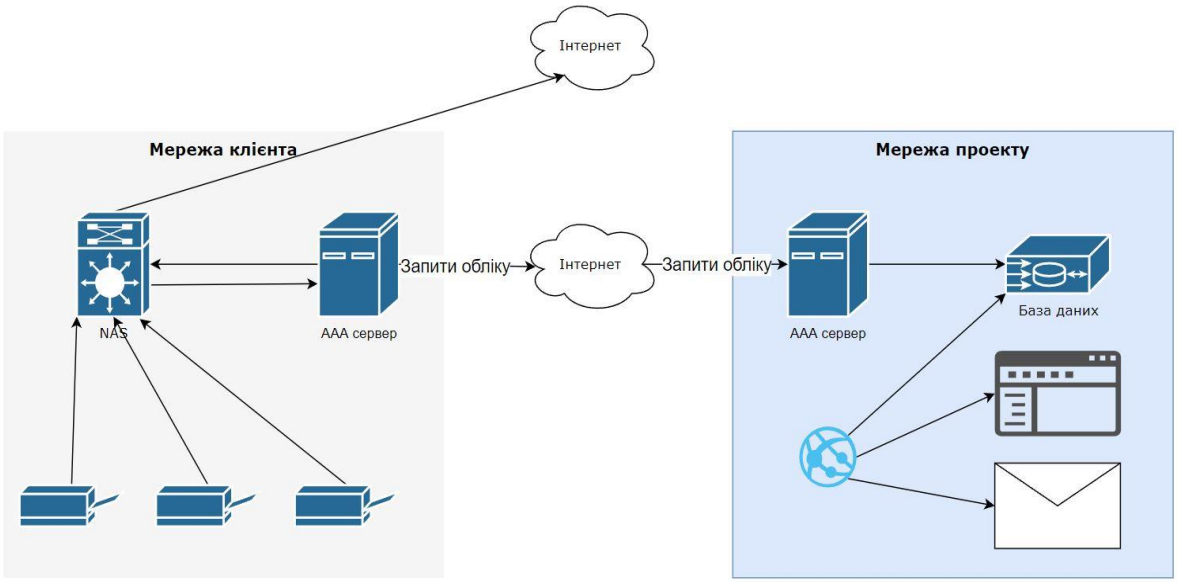

Рис.1. Загальна архітектура запропонованого рішення

AAA сервер підключений до Active Directory Domain Services компанії і коли користувач підключається до мережі зі своїми обліковими даними, мережева система зберігання даних(Network attached storage) NAS перенаправляє цей запит на AAA сервер, а сервер в свою чергу перевіряє чи є такий користувач в директорії і відповідно дозволяє або забороняє доступ до мережі. IoT-пристрої також підключаються до мережі через AAA сервери, однак оскільки далеко не завжди для є них можливість ввести логін та пароль, вони підключаються через MAC автентифікацію. Автентифікація пристроїв відбувається на основі їх фізичних MAC-адрес.

Ключовим моментом тут є те, що ААА сервери призначені для збору даних про використання мережі кожним окремим пристроєм. Однак в даній роботі пропонується використовувати дані обліку для виявлення аномалій.

Подальша конкретизація запропонованого рішення потребує вибору ААА протоколу з множини можливих та вибір методу визначення аномалій. Оскільки ці етапи не взаємопов'язані їх можна вирішувати послідовно.

У сучасних мережах використовують два основних рішення для ААА: Remote Authentication Dial-In User Service (RADIUS) та Cisco's Terminal Access Controller TACACS+ протоколи. Існує ще й третій ААА-протокол, відомий як DIAMETER, який переважно використовується мобільними операторами. Далі коротко охарактеризуємо ці рішення. Функціональні можливості ААА протоколів приведені в табл.1

Таблиця 1

Функціональні можливості ААА протоколів

Протокол / Функція	**RADIUS**	**TACACS+**	**Diameter**
Автентифікація	Автентифікація здійснюється шляхом дешифрування пакета запиту доступу NAS, автентифікації джерела NAS та перевірки параметрів запиту доступу до файлу користувача. Потім сервер повертає одну з трьох відповідей автентифікації: access-accept; access -reject; access-challenge.	Має три типи пакетів: Start, Continue, Reply. Клієнт починає автентифікацію з пакета «Start», який описує тип автентифікації, яку потрібно виконати. Для простих типів автентифікації, таких як PAP, пакет також може містити ідентифікатор користувача та пароль. Сервер відповідає типом пакету «Reply».	Клієнт (тобто NAS) надсилає запит на автентифікацію серверу, що містить команду AA-Request (AAR), ідентифікатор сеансу (session-ID), адресу та ім'я хоста клієнта, за якими слідують ім'я та пароль користувача та значення стану.
Авторизація	Авторизація не є окремою функцією в протоколі RADIUS, вона є просто частиною відповіді автентифікації. Коли сервер RADIUS перевіряє запит на доступ, він повертає клієнту NAS усі атрибути підключення, зазначені у файлі користувача.	Авторизація в TACACS+ здійснюється за допомогою пар атрибут / значення запиту (Request) та відповіді (Response) для встановлення прав доступу та привілеїв, зворотного виклику, фільтрації вхідних і вихідних пакетів та ін.	Запити на авторизацію мають виконуватися протягом існуючого сеансу; їх не можна використовувати для ініціювання сеансів, але їх можна пересилати за допомогою проксі-сервера Diameter.
Облікові можливості	Облік RADIUS збирає дані для статистичних цілей, моніторингу пристроїв в мережі, а також використовується для точного виставлення рахунків користувачам.	На додачу до стандартних облікових даних, які підтримує RADIUS, TACACS має можливість запису подій, яка може записувати зміни на системному рівні прав доступу або привілеїв.	Перевищує облікові можливості RADIUS і TACACS+, за рахунок опцій моніторингу подій, періодичні звіти, передачу записів у реальному часі та інші
Додаткові можливості	Дозволяють виконувати роль проксі-сервера для запитів клієнтів, пересилаючи їх на сервери в інших доменах автентифікації. Пересилання може ґрунтуватися на ряді критеріїв, включаючи іменований або номерний домен. Це особливо корисно, коли єдиний модемний пул є спільним для відділів або організацій.	Розширені функцій автентифікації та авторизації TACACS в чому вони значно покращені відносно функцій протоколу RADIUS завдяки двом спеціальним можливостями: рекурсивним пошуком і викликом (callout).	Підтримку декілька конфігурацій проксі, включаючи дві моделі RADIUS і дві додаткові моделі брокерів.

Порівняння ААА-протоколів

Клієнтами ААА-серверів є пристрої доступу до мережі (NAS). Вони передають інформацію про користувача на сервер ААА і діють відповідно до відповідей, які повертає сервер. Сервери отримують запити на підключення користувача, автентифікують користувача і повертають NAS інформацію про конфігурацію, необхідну для надання послуг користувачеві. Ця інформація може включати транспортні параметри, параметри протоколу, додаткові вимоги автентифікації (зворотній виклик, SecureID), директиви авторизації (дозволені послуги, застосування фільтрів) та вимоги до обліку. Ці функціональні якості притаманні усім трьом з перелічених ААА-протоколів. Проте існують і суттєві відмінності. Далі охарактеризуємо їх базові функції: Авторизація; Автентифікація; Облік; Додаткові можливості.

Вибір ААА протоколу для запропонованого рішення

Кожен з трьох розглянутих ААА протоколів має свої переваги та недоліки, тому їх досить складно порівнювати. Який з них обрати для реалізації запропонованого рішення? Наприклад, переваги протоколу Diameter в розширюваності та чотири варіанти проксування запитів, не є такими значущими, коли умовна компанія просто хоче налаштувати безпечний доступ до мережі. Коли адміністратор мережі компанії хоче налаштувати безпечний мережевий доступ до ресурсів, то його мало хвилює посилені можливості розширюваності та переадресації. Він бажає, щоб працювала автентифікація і авторизація, тут і зараз, і бажано без глибокого занурення в конфігураційні файли, що власне і потрібно для мережі IoT. Тому треба обрати критерії, на основі яких можна об'єктивно обрати ААА протокол, який би найкраще відповідав би даній задачі. На наш погляд такими критеріями можуть слугувати наступні характеристики: можливості переадресації запитів; детальність обліку; швидкість; підтримка виробниками мережевого обладнання; розповсюдженість. Даний набір критеріїв дозволяє оцінювати різні аспекти застосування ААА протоколів в мережах, зокрема і в мережах IoT. Відтак, якщо кожний з критеріїв оцінити кількісно, то ми зможемо об'єктивно оцінити кожний з протоколів і зробити обґрунтований вибір. Однак, при цьому ми маємо справу з багатокритеріальною задачею. І тому маємо застосувати релевантний метод для її розв'язку. Для обґрунтування вибору протоколу був застосований метод аналізу ієрархій(MAI). Метод аналізу ієрархій являє собою метод прийняття рішень ієрархічної композиції задачі та експертного рейтингування. Після того, як ієрархія побудована, особи, що приймають рішення, систематично оцінюють різні її елементи, порівнюючи їх один з одним попарно, з огляду на їх вплив на елемент над ними в ієрархії. Суть MAI полягає в поєднанні експертних оцінок та строгим формальним алгоритмом визначення кращого вибору з можливих. В ролі експертів виступали мережеві адміністратори компаній, в яких використовувались описані вище ААА протоколи.

Побудова ієрархії.

На рис. 2. представлена побудована MAI ієрархія для вирішення задачі вибору кращого ААА протоколу для цілей даної роботи.

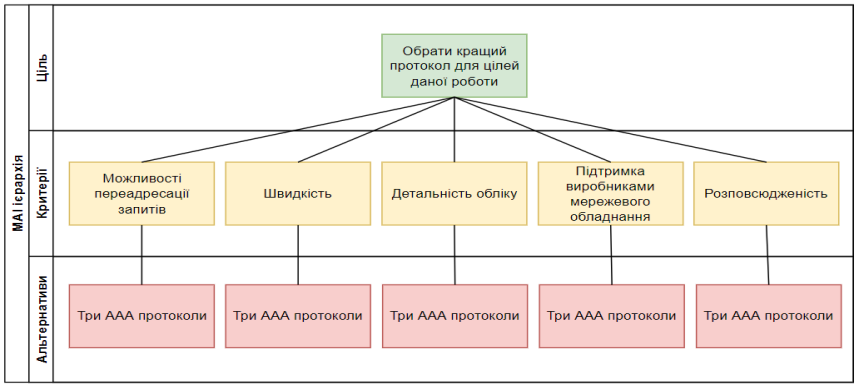

Рис. 2. MAI ієрархія для задачі вибору ААА протоколу

Далі було здійснено порівняння критеріїв по шкалі відношень Сааті щодо того, наскільки вони на думку експертів важливі.

На основі суджень експертів, представлених в табл. 2, було визначено пріоритети для критеріїв, за якими будуть порівнюватися кожен трьох протоколів

Таблиця 2

Результати попарних порівнянь критеріїв експертами

Критерій – Критерій	Можливості пере-адресації запитів	Детальність обліку	Швидкість	Підтримка виробниками мережевого обладнання	Розповсю-дженість
	1	2	3	4	5
1	1	2	5	1/3	1/5
2	1/2	1	4	1/5	1/7
3	1/5	1/4	1	1/7	1/7
4	3	5	7	1	2
5	5	7	7	1/2	1

Далі було здійснено попарне порівняння всіх трьох AAA протоколів за визначеними критеріями. На основі даних, отриманих на кроках попарного порівняння, було сформовано глобальні пріоритети для кожної з альтернатив. Їх загальна сума становить 1, тобто спочатку була проведена процедура нормування. Кожна альтернатива має глобальний пріоритет, що відповідає її «відповідності» всім судженням експертів щодо розглянутих критеріїв. Глобальні пріоритети всіх альтернатив представлені в Табл.3.

Таблиця 3

Глобальні пріоритети всіх альтернатив

Критерій Прлтокол	Можливості переадресації запитів	Детальність обліку	Швид-кість	Підтримка виробниками мережевого обладнання	Розповсю-дженість	Глобальний пріорітет
RADIUS	0.037	0.008	0.023	0.305	0.258	**0.631**
TACACS+	0.010	0.018	0.004	0.061	0.077	0.170
Diameter	0.084	0.057	0.010	0.027	0.022	0.200

RADIUS протокол з глобальним пріоритетом 0.631 краще за інші протоколи підходить для цілей даної роботи.

Експериментальні дослідження

З метою оцінювання ефективності виявлення аномалій на основі використання AAA протоколів було проведено серію випробувань. Спочатку необхідно провести дослідження придатністі пакетів обліку RADIUS-протоколу для для тренування моделей машинного навчання по виявленню аномалій в трафіку повідомлень. Для тестування було використано датасет UNSW_NB15-dataset [13] з набором шкідливого ПЗ та зразки їх трафіку та нешкідливого ПЗ. Далі також були протестовані різні моделі машинного навчання з метою оцінки якості розпізнавання аномалій.

Набори даних NIDS можна концептуалізувати як реляційні дані, які можна вважати табличними даними, тобто наборами записів даних (також визначених як вектори, події, зразки або спостереження), кожен з яких складається зі колонок (тобто атрибутів/областів) з різними типами даних, такими як цілі, з плаваючою точкою, бінарні або категоріальні дані. У випадку вхідних багатоваріантних даних, те, чи є їхні характеристики однаковими чи різними типами даних, визначає застосовність методів виявлення аномалій для їх обробки. Формат пакету RADIUS-протоколу[14] показано на Рис.3.

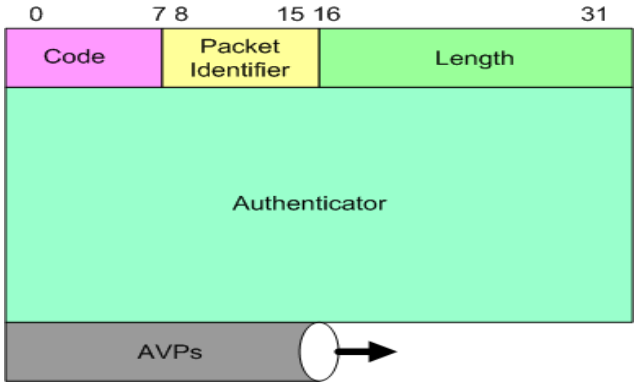

Рис. 3. Формат пакету RADIUS

Атрибути RADIUS містять конкретні відомості про аутентифікацію, авторизацію, облік і конфігурацію для запиту та відповіді. Атрибути обмінюються між NAS і сервером. Кожен атрибут пов'язаний з набором властивостей, які визначають, як його інтерпретувати. Найважливішою властивістю є тип даних, який оголошує тип даних, які ідентифікує атрибут (рядок символів, ціле число, IP-адреса або вихідні двійкові дані). Інформація, яка передається разом із запитом, упакована в набір пар атрибут/значення (AV-пари, або AVP), які складаються з номерів атрибутів та пов'язаних із ними даних. Таким чином, RFC 2865 (Стандартний трек) визначає аутентифікацію та авторизацію, а облікова частина визначена в інформаційних RFC 2866 і 2869. Є ще багато атрибутів.

Не всі поля в цьому наборі даних відповідають полям пакетів обліку RADIUS-протоколу. В наборі даних UNSW_NB15 наявні поля, які неможливо буде отримати зі стандартного трафіку обліку RADIUS. Відповідно для навчання моделі такі поля не будуть використовуватись. Необхідно врахувати, що цей датасет фактично являє собою набір перехоплених пакетів, а облік RADIUS відображає статистичні дані про використання мережі певним пристроєм за певний час. Тому для забезпечення можливості коректного тренування моделей машинного навчання треба провести попередню обробку датасету UNSW_NB15. Попередня обробка включає такі операції, як очищення даних, дедуплікація, агрегація та нормалізація даних. В кінцевому рахунку ми отримуємо нормалізовані і стандартизовані дані представлені на рис. 4.

	dur	spkts	dpkts	sbytes	dbytes	rate	sttl	dttl	sload	dload	-
0	-0.208678	-0.130765	-0.165331	-0.046480	-0.098409	-0.002151	0.722026	-0.751628	0.590935	-0.27285	-
1	-0.208679	-0.130765	-0.165331	-0.039194	-0.098409	0.210460	0.722026	-0.751628	4.363255	-0.27285	-
2	-0.208679	-0.130765	-0.165331	-0.043188	-0.098409	0.678204	0.722026	-0.751628	4.220037	-0.27285	-
3	-0.208679	-0.130765	-0.165331	-0.044155	-0.098409	0.470318	0.722026	-0.751628	2.850314	-0.27285	-
4	-0.208678	-0.130765	-0.165331	-0.037100	-0.098409	0.054546	0.722026	-0.751628	4.198501	-0.27285	-

Рис. 4. Нормалізовані дані з датасету UNSW_NB15

Для проведення машинного навчання необхідно підібрати набір параметрів датасету який би дозволив досягнути компроміс між можливостями формату обліку протоколу RADIUS та достатньо високим рівнем виявлення аномалій в мережі IoT. В першу чергу це можна зробити за рахунок виключення суттєво скорельованих показників. Пороговою величиною такого вилучення в даному дослідженні була значення кореляції 0,5 за Пірсоном. Наприклад, розрахунок кореляції показав, що між spkts та sloss, dpkts (кількість пакетів відправлених до джерела) та dbytes (від призначення до вихідних байтів транзакції), tcprtt та ackdat (час встановлення TCP-з'єднання, час між пакетами SYN_ACK та ACK) перевищують 0,9, і існує тривала позитивна кореляція. З іншого боку кореляція між показниками spkts і state, dbytes і tcprtt менше 0,1.

Крім того, методика машинного навчання говорить що кращий результат буде тоді, коли ентропія, тобто перемішаність ознак буде якомога меншою. Кількісним показником перемішаності служить домішка Джині(DS).

$$DS = 1 - \sum_{i=1}^{n} p_i^2 \quad (1)$$

де p_i - частоти представників різних ознак.

Домішка Джині говорить нам про «забрудненість» множини вибору. Знижуючи домішку Джіні, ми можемо з упевненістю зробити висновок, що чистота буде більше, а отже, вищий шанс на однорідність множини. Чим менше значення домішки Джині для ознаки, тим менше забрудненість об'єкта в наборі даних і тим буде краще навчальний ефект для ознаки.

Процедура відбору ознак була проведена у два етапи. На першому етапі були відібрані всі ознаки з пороговим значенням коефіцієнта кореляції 0,5. На другому етапі були залишені саме ті з них, які мали відповідне менше значення домішки Джині.

В результаті процедури відбору з набору даних UNSW_NB15 в даній роботі було обрано наступні ознаки для подальшого навчання моделі:

- dur - тривалість з'єднання;
- sload - кількість вихідних бітів в секунду (вхідне навантаження);
- dload - кількість вихідних бітів в секунду (вихідне навантаження);
- dpkts - кількість пакетів від пункту призначення до джерела;
- spkts - кількість пакетів від джерела до призначення.

Характерно, що всі ці атрибути визначаються авторами дата сету як основні.

Експерименти, представлені у цій роботі, проведено на Dell Inspiron 15 3000, завантаженому з операційною системою Windows 10 з процесором: AMD A6-6310 1800 МГц 1,80 ГГц. Моделі ML будуються, навчаються, оцінюються та тестуються на базі бібліотек Scikit-Learn,pandas Python. Для формування навчальної та тестової вибірок було було налаштовано як навчальний набір та набір для тестування, а саме, UNSW_NB15_training-set.csv та UNSW_NB15_testing-set.csv відповідно. Кількість записів у навчальному наборі становить 175 341 запис, а у тестовому наборі – 82 332 записів різних типів, атакуючих та звичайних.

Об'єктами для виявлення були наступні різновиди атак:
Fuzzers; Analysis; Backdoors; DoS; Exploits; Generic; Reconnaissance; Shellcode; Worms

Для ідентифікації атак були обрані наступні алгоритми:
Дерево рішень; Random Forest; Багатошаровий перцептрон; Gradient Boosting;
Classifier Support; Vector Classifier

Для верифікації точності моделей використовувся метод перехресної перевірки. Під час навчання моделі набір даних розбивався на рівні апроксимовані сегменти. Потім кожен сегмент використовувався для перевірки точності моделі, а інші сегменти використовувалися для навчання. Процедура повторюється поки всі сегменти набору даних не будуть використані для перевірки точності.

Для об'єктивної оцінки ефективності алгоритмів машинного навчання в даній роботі була використана класична тріада з трьох показників: кількість правильних відповідей (Precision); повнота (recall); F-міра – поєднує повноту та влучність.Precision можна інтерпретувати як частку об'єктів, названих класифікатором позитивними і при цьому дійсно позитивними, а recall показує, яку частку об'єктів позитивного класу з усіх об'єктів позитивного класу знайшов алгоритм.

Результат та обговорення

Були проведені комп'ютерні розрахунки для кожної з моделей машинного навчання для розпізнавання кожної з визначених атак. В таблиці надані результати оцінки

ефективності алгоритмів навчання за значенням F-міри, як агрегованого критерія якості. F-міра (в загальному випадку) є середнє гармонійне precision і recall.

Таблиця 4

Результати перевірки точності

Алгоритм машинного навчання – Назва атаки	Дерево рішень	Random Forest	Багатошаровий перцептрон	Gradient boosting classifier	Support vector classifier
Analysis	0,19	0,19	0,1	0,8	0,1
Backdoor	0,14	0,15	0,11	0,10	0,1
DoS	0,33	0,27	0,2	0,13	0,1
Exploits	0,69	0,72	0,72	0,73	0,69
Fuzzers	0,61	0,65	0,58	0,59	0,42
Generic	0,99	0,99	0,99	0,99	0,98
Normal	0,92	0,93	0,92	0,91	0,87
Reconnaissance	0,82	0,83	0,82	0,83	0,58
Shellcode	0,6	0,67	0,61	0,59	0,0
Worms	0,49	0,23	0,19	0,52	0,0

Для побудованих моделей був застосований метод перехресної перевірки, а також був визначений середній час класифікації. Всі отримані результати представлені в таблиці.

Таблиця 5

Порівняння моделей

Назва алгоритму	F-міра (%)	FPR (%)	Точність (перехресна перевірка) (%)	Час навчання (с)	Час класифікації (мкс)
Дерево рішень	80.9	21.1	80,53	**27,6**	**0,99**
Random Forest	82,6	19,2	82,48	143,6	23,38
Багатошаровий перцептрон	**83,7**	**18,3**	**83,24**	1082,8	2,54
Gradient Boosting Classifier	82	19,9	81,8	3323,1	25,9
Support Vector Classifier	76,8	25,7	76,54	715,1	0,89

Аналіз табл. 5 показує, що два методи навчання: Random Forest та багатошаровий перцептрон має певну перевагу у порівнянні з іншими методами. Причому багатошаровий перцептрон має кращі результати по точності. Однак він поступається в швидкості навчання дереву рішень.

Отримані у цьому дослідженні результати дозволяють стверджувати про можливість виявлення вторгнень в мережі IoT на основі виявлення аномалій шляхом аналізу трафіка за AAA протоколом та дозволяють окреслити практичне рішення для захисту мережі IoT. Очевидно, в подальшому необхідно буде дещо покращити точність виявлення аномалій.

Програмний продукт на основі мікросервісної архітектури буде поставлятися користувачам як SaaS. Ядром цього мікросервісу є авторська програма[15], яка відповідає за обробку пакетів і взаємодію з мережею.

Висновки

У роботі проведені дослідження пошуку підходу виявлення кібератак на мережі IoT на основі аналізу аномалій в трафіку AAA протоколів. В процесі дослідження були:

- проведено аналіз можливостей існуючих протоколів дозволив сформулювати критерії якості для задач виявлення аномалій;
- поставлена і вирішена задача вибору найбільш прийнятного AAA протоколу методом аналізу ієрархій;
- попередня обробка даних датасету UNSW_NB15 з метою використання даних для машинного навчання шляхом проведення двохетапної процедури відбору ознаки, що дало можливість нормалізувати та стандартизувати дані для подальшого машинного навчання;
- проведено комп'ютерні експерименти, в якому було перевірено можливості різних методів машинного навчання для виявлення 9 різновидів кібератак, зафіксованих в дата сеті;
- запропоновано програмне рішення поставленої проблими у вигляді мікросервісної архітектури.

Найкращі результати по критерію точності показали методи: Багатошаровий персептрон та Random Forest. Точність з врахуванням перехресної перевірки складає 83.24%, що говорить про потенційну можливість реалізації підходу виявлення вторгнень в мережу IoT шляхом виявлення аномалій в трафіку за протоколом RADIUS.

Зазначимо, що проведені дослідження не можна вважати завершеними, але дають підставу стверджувати про перспективу подальших досліджень у цьому напрямку.

Список літератури

1. Gartner Says 6.4 billion Connected "Things" Will Be in Use in 2016, Up 30 Percent From 2015". Gartner. 2015. 10 November.

2. MAC Authentication Bypass (MAB). URL: https://networklessons.com/ cisco/ccie-routing-switching-written/mac-authenticationbypass-mab.

3. Nasim B. M., Jelena M., Vojislav B. M., Hamzeh K. A framework for intrusion detection system in advanced metering infrastructure. 2014. P. 195–205.

4. Deep learning. URL: https://en.wikipedia.org/wiki/Deep_learning.

5. Markov random field. URL: https://en.wikipedia.org/wiki/Markov_random_field.

6. Backpropagation. URL: https://en.wikipedia.org/wiki/Backpropagation.

7. Park Y., Nicol D.M., Zhu H. Prevention of Malware Propagation in AMI. Proceedings of the IEEE International Conference on Smart Grid Communications. 2013. P. 474–479.

8. Neetesh S., Bong J. C., Santiago G. Secure and Privacy-Preserving Concentration of Metering Data in AMI Networks. Proceedings of the 2017 IEEE International Conference on Communications (ICC). 2017.

9. Euijin C., Younghee P., Huzefa S., Identifying malicious metering data in advanced metering infrastructure. Proceedings of the 2014 IEEE 8th International Symposium on Service Oriented System Engineering, 2014. P. 490–495

10. DoS and DDoS vulnerability of IoT: A review. URL: https:// www.researchgate.net/publication/39862422_DoS_and_DDoS_vulnerability_of_IoT_A_review.

11. Fernandes G., Rodrigues J. J. P. C., Carvalho L. F., Al-Muhtadi J. F, Proenca M. L.

A Comprehensive Survey oon Network Anomaly Detection. Telecommunication Systems. 2019. V. 70. No. 3. P. 447–489.

12. Нічепорук, А., Нічепорук, А., Савенко, О., & Казанцев, А. An Intelligent System For Detecting Anomalies and Identifying Smart Home Devices Based on the Collective Communication. Electrotechnic and Computer Systems. 2021. No. 34(110). P.50-61.

13. Nour M. Designing an Online and Reliable Statistical Anomaly Detection Framework for Dealing with Large High-Speed Network Traffic. Diss. University of New South Wales, Canberra, Australia.

14. RADIUS. URL: https://en.wikipedia.org/wiki/RADIUS.

15. Нікітін Є.Є., Гальчинський Л.Ю. Комп'ютерна програма «Radius сервер на платформі .NET». Свідоцтво про реєстрацію авторського права на твір №109412. від 12 листопада 2021.

DETECTION OF INTRUSIONS INTO THE INTERNET OF THINGS NETWORK BY ANALYZING THE PRESENCE OF ANOMALIES IN THE NETWORK TRAFFIC FLOW OF AAA PROTOCOLS

L.Yu. Galchynsky

National Technical University of Ukraine "KPI" named after Igor Sikorsky
ave. Peremohy, 37, 02056 Ukraine; e-mail: hleonid@gmail.com

Today, one of the most powerful trends in innovative change in the modern world is the Internet of Things IoT. And although this trend is essentially only at an early stage, it already plays a significant role in the modern world, in modern civilization. It is safe to predict that further expansion of its scope. However, the emergence of this phenomenon is also accompanied by the emergence of related problems. This is especially true of cyber threats, which can cause significant harm to both the Internet of Things itself and the people who use those networks. Accordingly, there is a need for means to combat these threats. This is a rather complex problem, which primarily requires the development of methods and technologies for detecting such threats. At the same time, there are no clear concepts of relevant solutions for cyber security of various IoT applications, especially where financial resources are limited. Therefore, research is currently relevant to find solutions that may be acceptable for relatively low-budget facilities. The aim of the work is to obtain a method for detecting intrusions into the IoT network by detecting anomalies in the network traffic flow of AAA protocols. This problem is complicated by the fact that the solution found must comply with basic security policies. The study conducted in this paper showed that such a solution is possible in principle by using the capabilities of the AAA RADIUS protocol and the use of machine learning. Machine learning methods were tested on the UNSW_NB15 dataset. For machine learning, pre-processing was performed to normalize and standardize the data. Computer experiments have tested the suitability of detecting a set of cyberattacks for different models of machine learning. It was shown that the formulated approach has the potential to be used in practical developments to detect intrusions in the IoT network.

Keywords: Internet of Things, malicious intrusion into the network, anomalies in the network flow, AAA protocols, method of analysis of hierarchies, NIDS data sets, machine learning methods, UNSW_NB15.