ISSN 2219-9365

УДК 004.092
DOI: 10.31891/2219-9365-2021-67-1-7

ШАГІН В. Ю., НІЧЕПОРУК А. А., КАШТАЛЬЯН А. С.
Хмельницький національний університет

ЦЕНТРАЛІЗОВАНА РОЗПОДІЛЕНА СИСТЕМА ВИЯВЛЕННЯ АТАК У КОРПОРАТИВНИХ КОМП'ЮТЕРНИХ МЕРЕЖАХ НА ОСНОВІ МУЛЬТИФРАКТАЛЬНОГО АНАЛІЗУ

У роботі запропоновано архітектуру та компоненти розподіленої системи виявлення мережевих атак, в якій поєднано вимоги централізованості, розподіленості, самоорганізованості та на її основі здійснено розробку централізованої розподіленої системи визначення мережевих атак в корпоративних комп'ютерних мережах на основі мультифрактального аналізу. Проведені експериментальні дослідження з реалізованою централізованою розподіленою системою визначення мережевих атак в комп'ютерних мережах підтвердили ефективність функціонування в комп'ютерній мережі.
Ключові слова:Розподілена система, Виявлення атаки, Мультифрактальний аналіз, Комп'ютерна мережа.

SHAGIN V., NICHEPORUK A., KASHTALIAN A.
Khmelnytskyi National University

CENTRALIZED DISTRIBUTED ATTACK DETECTION SYSTEM IN CORPORATE COMPUTER NETWORKS BASED ON MULTIFRACTAL ANALYSIS

The architecture and components of a distributed network attack detection system are proposed in the paper, which combines the requirements of centralization, distribution, self-organization and on its basis develops a centralized distributed network attack detection system in corporate computer networks based on multifractal analysis. Experimental studies with a centralized distributed system for detecting network attacks in computer networks have confirmed the effectiveness of functioning in a computer network.

The proposed centralized system is based on the use of multifractals. Multifractals are complex fractals that occur, as a rule, in nature. In fact, the multifractal approach means that some object under study can be divided into parts that have their own similarity characteristics that are different from others. Network traffic is self-similar at some intervals. Therefore, the method of maxima of wavelet transform modules will be used for its analysis, which allows to determine the features of the signal.

To conduct experimental research, a distributed system was implemented and software was deployed to detect attacks on local computer networks. The study found that the total data processing time is about 8 seconds, while the amount of data is almost 4.9 million lines, in the format of a text document, this data is more than 700 megabytes. Thus, this indicates the high speed of the algorithm and the efficiency of system resources.
Keywords: Distributed System, Attack Detection, Multifractal Analysis, Computer Network.

Вступ. Зі зростанням об'ємів інформації, що передаються через мережу зростає і число апаратних та програмних засобів для втручання в процес передачі даних. Щодня все більше користувачів в інтернеті стають жертвами дій зловмисного програмного забезпечення. Особливу небезпеку становлять атаки та корпоративні системи, що можуть спричинити фінансові збитки організаціям чи безпосередньо їх працівникам.

Для мінімізації кількості успішних мережевих атак в корпоративних комп'ютерних мережах може бути використана математична статистика. Оскільки в корпоративних комп'ютерних мережах, як правило, присутня велика кількість комп'ютерів необхідні ефективні методи виявлення та реагування на активність в мережі. Існує твердження, що немає абсолютно захищеної системи, проте можливе суттєве зниження шансу на несанкціонований доступ до даних чи комп'ютерної системи загалом. Для протидії активності зловмисників необхідний комплексний підхід до розробки методів виявлення вторгнень та систем, в які інтегруються ці методи.

Концепція розподілених систем. На сьогоднішній день питання надійності комп'ютерної мережі є надзвичайно важливим. Тому для бізнесу широко використовуються розподілені мережі, які можуть бути реалізовані за допомогою програмного продукту. Концепт розподіленої мережі можна трактувати по різному. В широкому розумінні це робочі станції, що обмінюються інформацією та спільно її обробляють, при цьому кожна робоча станція може мати різну обчислювальну потужність та ємність сховища даних.

Модель розподіленої мережі передбачає поділ завдань між клієнтам чи серверами, що залежить від того, чи підходить конкретна машина для обробки даних [1–5]. За цим типом архітектури частина додатку виконується клієнтом, а інша частина сервером. З клієнт-серверним типом побудови програмного забезпечення користувач працює лише з обмеженою кількістю даних, включаючи інтерфейс користувача, користувацький ввід та запити до бази даних. Контроль доступу до бази даних, отримання чи обробку даних користувачем здійснює безпосередньо сервер. При цьому важливо перевіряти чи справді клієнт має дозвіл на оперування інформацією. Для цього проводять процедуру двох голосувань. Під час першого голосування кожен вузол отримує довіру, розподіляючи інформацію про відкликання сусідів розподіленої мережі.

Сусідні вузли створюють довірене оточення. Якщо вузол обмінюється інформацією з іншими в достатній мірі – він може і далі спілкуватись з рештою в мережі, в іншому випадку він відхиляється. Якщо вузол загрожує мережі та проявляє підозрілу поведінку – відбувається друга процедура відкликання голосування. В цьому випадку, якщо оточуючі вузли вказують на підозрілу активність свого сусіда – блокуються весь обмін даними з цим вузлом по всій мережі.

Основними властивостями розподіленої системи за теоремою CAP [6] є цілісність, доступність та стійкість до відмов (Consistency, Availability, Partition tolerance). Проте будь-яка розподілена система може мати не більше двох властивостей.

Цілісність або узгодженість системи за теоремою CAP означає її лінійність. Кожна наступна операція повинна мати інформацію про результат попередньої. Доступність системи означає безперебійну успішну обробку запитів на всіх активних вузлах. Якщо частина вузлів не відповідає чи система відповідає не на всі запити – за теоремою CAP це непостійна доступність. Стійкість до відмов в першу чергу забезпечує стабільну роботу всієї системи. Наприклад, якщо кластер із декількох серверів втратив з'єднання з половиною своїх вузлів – робота системи продовжується, хоч і було втрачено доступність. В іншому випадку частини кластера працюють незалежно один від одного та відповідають на запити користувачів. В такому випадку тільки після відновлення зв'язку між всіма ланками розподіленої системи буде зрозуміло, що це була єдина система.

Згідно теореми CAP можна виділити три варіанти проєктування розподіленої системи: AC – доступність та цілісність; CP – цілісність та стійкість; AP – доступність та стійкість.

AC системи мають суттєвий недолік – відсутність стійкості до відмов мережі. Використання цього типу систем вимагає чіткого усвідомлення всіх ризиків. В іншому випадку необхідно обирати між цілісністю та доступністю системи.

Таким чином, з огляду на широку розповсюдженість централізованої архітектури, важливим завданням є проектування системи, яка дозволить здійснити виявлення атак та підвищити загальний рівень безпеки всієї мережі.

Централізована система виявлення атак в корпоративних комп'ютерних мережах на основі мультифрактального аналізу. В основі запропонованої централізовано системи закладено використання мультифракталів. Мультифракталами називають складні фрактали, що зустрічаються, як правило, в природі. Фактично мультифрактальний підхід означає, що деякий досліджуваний об'єкт можна розділити на частини, що мають власні характеристики подібності, відмінні від інших. Мережевий трафік є самоподібним на деяких часових проміжках. Тому для його аналізу буде використано метод максимумів модулів вейвлет-перетворення, що дозволяє визначити особливості сигналу.

Для аналізу параметрів мультифрактального спектру використаємо наступний алгоритм:

Декомпозиція вихідного сигналу $f(t)$ на коефіцієнти батьківським вейвлетом $\psi(t)$:

$$W_f(u,j) = \left(f(t), \psi_{u,s}(t)\right) = 2^{-j/2} \int_{-\infty}^{\infty} \frac{t-u}{2^j} dt$$

де u – параметр масштабу,

j – просторова координата чи момент часу.

В масиві коефіцієнтів знаходимо позиції локальних максимумів $\{u_p(j)\}_{p \in Z}$ та знаходимо їх абсолютне значення та формуємо масив максимумів

$$|W_f(u_p, j)|;$$

Визначення функції розбиття:

$$S(q,j) = \sum_p |W_f(u_p,j)|^q;$$

Для кожного q ∈ R обчислюємо показник масштабу:

$$\tau(q,j) = lim \inf_{j \to 0} \frac{\ln S(q,j)}{\ln 2^j};$$

Обчислюємо мультифрактальний спектр за допомогою перетворення Лежандра:

$$f_L(\alpha) = \min_{q \in R}[q(\alpha + 1/2) - \tau(q)];$$

Для кожного проміжку j обчислюємо мультифрактальні розмірності порядку q:

$$D_{q,j} = \frac{1}{q-1}[q(\alpha(q,j) - f(\alpha(q),j)].$$

Для ілюстрації та аналізу мережевого трафіку обрано його інтенсивність, тобто кількість відправлених та прийнятих пакетів за одиницю часу.

Принцип виявлення вторгнень запропонованої системи наступний. Нехай X – часовий ряд звичайного трафіку, Y – часовий ряд шкідливого трафіку, Z – часовий ряд аномалій. Звідси Y=X+Z. Незалежно від наявності властивостей самоподібності в часовому ряді аномалій – Y все ще буде самоподібним процесом, якщо X стаціонарний самоподібний процес. Проте ступінь самоподібності може змінюватись. Нехай $s_X s_Y s_Z$ функції автокореляції для X, Y та Z відповідно. Тоді під час атаки акцентуємо увагу на $\|s_Y - s_X\|$, при цьому $s_Y = s_X + s_Z$. Для кожного $H \in (0.5, 1)$ існує лише одна функція автокореляції з самоподібністю. Тому розглядається $\|H_Y - H_X\|$, де H_Y та H_X – середнє значення показників Херста X та Y відповідно. Коефіцієнт Херста вводиться для підвищення точності оцінки самоподібності системи. Недоліком підходу є необхідність перезапуску визначення порогу самоподібності трафіку для кожного масштабу. Тому сигнал про зміну самоподібності буде подано незалежно від того, чи існує він для іншого масштабу. Після визначення мережевої атаки трафік розбивається на декілька частин. Інтенсивність атаки можна визначити за допомогою аналізу показника Херста та швидкості його зміни, тобто різницю між показниками Херста до факту атаки та після [7].

Визначення точки зміни самоподібності трафіку базується на тому, що ентропія послідовності зі змінною граничною точкою самоподібності більша ніж ентропія послідовності з фіксованою точкою.

На рис. 1 представлено схему оцінки безпеки мережевого трафіку. Алгоритм складається з п’яти етапів:

1. Збір трафіку.
2. Статистичний аналіз.
3. Оцінка показника Херста.
4. Визначення аномалій.
5. Оцінка безпеки.

Для зменшення впливу на функціонування мережі трафік дублюється на сервер, що збирає мережеву інформацію кожного з підключених клієнтів.

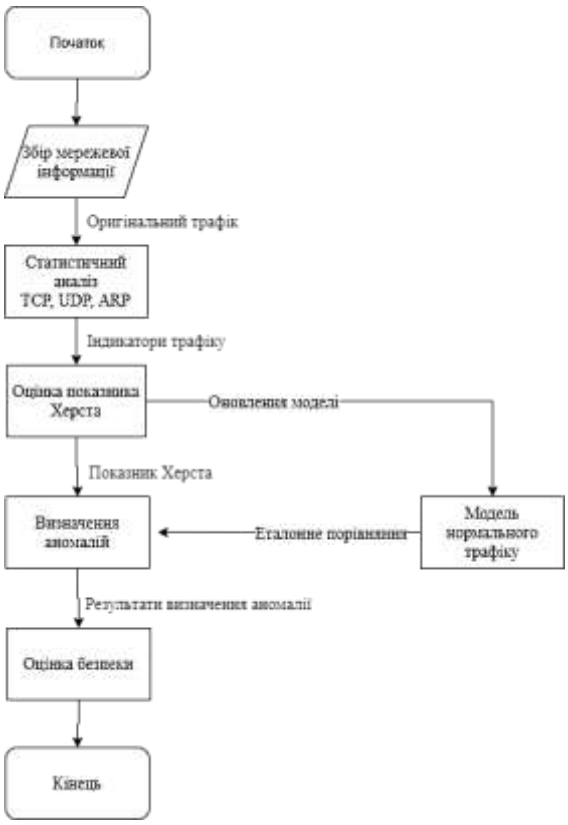


Рис. 1. Визначення оцінки безпеки мережевого трафіку

Експериментальні дослідження. Для розробки централізованої системи виявлення атак було використано середовище Microsoft Visual Studio Community 2019. Досліджувана централізована система складалась з одного сервера та двох клієнтських станцій.

Для реєстрації мережевих пакетів було обрано бібліотеку з відкритим вихідним кодом SharpPcap.

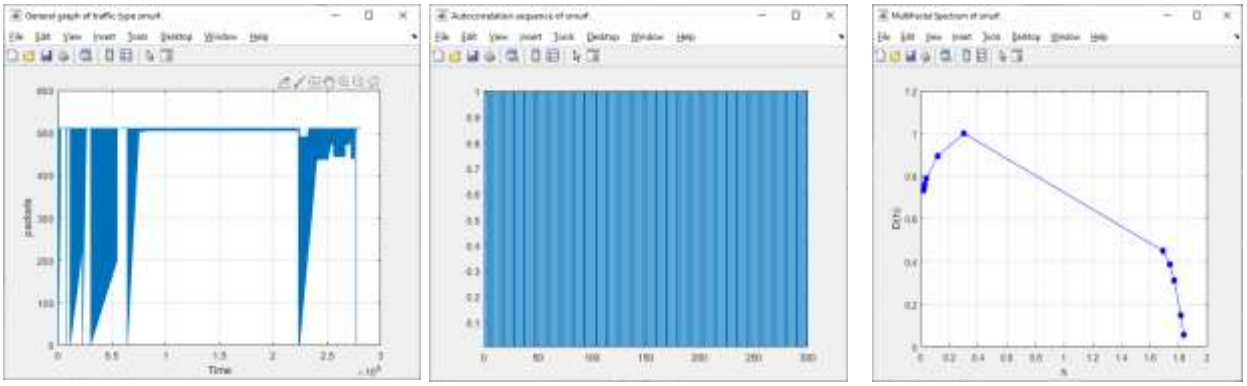
Бібліотека SharpPcap підтримує наступні мережеві протоколи: Ethernet, LinuxSLL, Ip (IPv4 and IPv6), Tcp, Udp, ARP, ICMPv4 и ICMPv6, IGMPv2, PPPoE, PTP, Link Layer Discovery Protocol (LLDP), Wake-On-LAN (WOL). Основними можливостями цього пакету є визначення активних мережевих пристроїв, формування статистики, зчитування мережевої інформації з активних мережевих пристроїв та читання з файлів, фільтрація пакетів, збереження мережевої інформації в файл. Форматом даних є Pcap та pcap-ng, за умови використання npcap чи libpcap версії 1.1.0 та вище.

В результаті дослідження було реалізовано наступні модулі: «WebAPI», «Common», «Database», «EntityFramework», «Repository» та «Services». Клієнтський додаток виступає в ролі фонової служби та не має графічного інтерфейсу чи ключів запуску для повної автономності. Тому використовуючи методи бібліотеки SharpPcap додаток самостійно знаходить активний мережевий інтерфейс та відкриває підключення до нього.

Для дослідження методу було використано пакет прикладних програм аналізу та програмування Matlab. Візуалізація мультифрактального спектру відбувається вбудованими засобами програмного забезпечення.

Розроблене програмне забезпечення використовує набір даних KDD Cup [8], який конвертовано в таблицю засобами Matlab. Для виведення графіків трафіку використано значення кількості пакетів. В даному наборі даних кожен рядок можна розглядати як статистику активності за одиницю часу. Перед створенням графіків програма фільтрує статистичні дані за кожним наявним типом мережевих атак. Далі формуються графіки мережевого трафіку для кожного виду мережевих атак та окремо графік нормального трафіку.

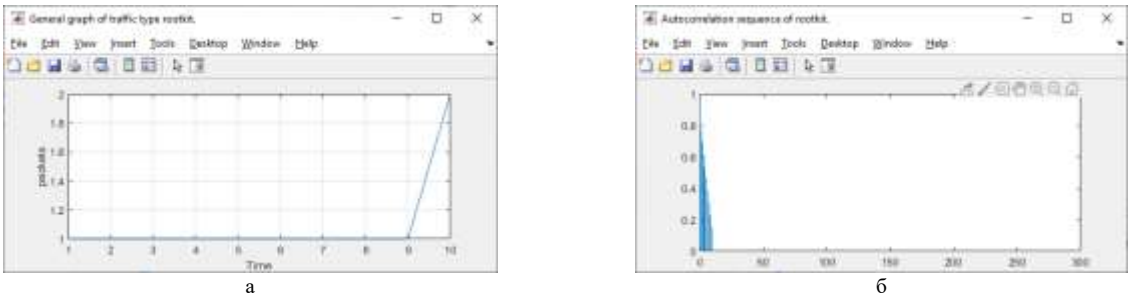
Результатом виконання представленого вище коду є графік мережевого трафіку, графік автокореляційної послідовності та мультифрактальний спектр (рис. 2).



а б в

Рис. 2. Результати проведеного експерименту:
а) графік активності мережевої атаки smurf;
б) автокореляційна послідовність мережевої атаки smurf;
в) мультифрактальний спектр мережевої атаки smurf

Слід відзначити, що запропонований метод має обмеження: при недостатній кількості значень в автокореляційній послідовності неможливо сформувати мультифрактальний спектр. На це впливає кількість випадків визначеного типу атаки на часовій шкалі. Яскравим прикладом з набору даних є активність мережевої атаки rootkit, що зображено на рисунку та її автокореляційна послідовність (рис. 3).



а б

Рис. 3. Результати системи при недостатній кількості значень в автокореляційній послідовності на прикладі мережевої атаки RootKit:
а) графік активності мережевої атаки; б) автокореляційна послідовність мережевої атаки

Проте, разом з тим запропонована система, при достатній кількості даних, дозволяє здійснити не тільки факт виявлення атаки, а й розрізнити її тип (рис. 4).

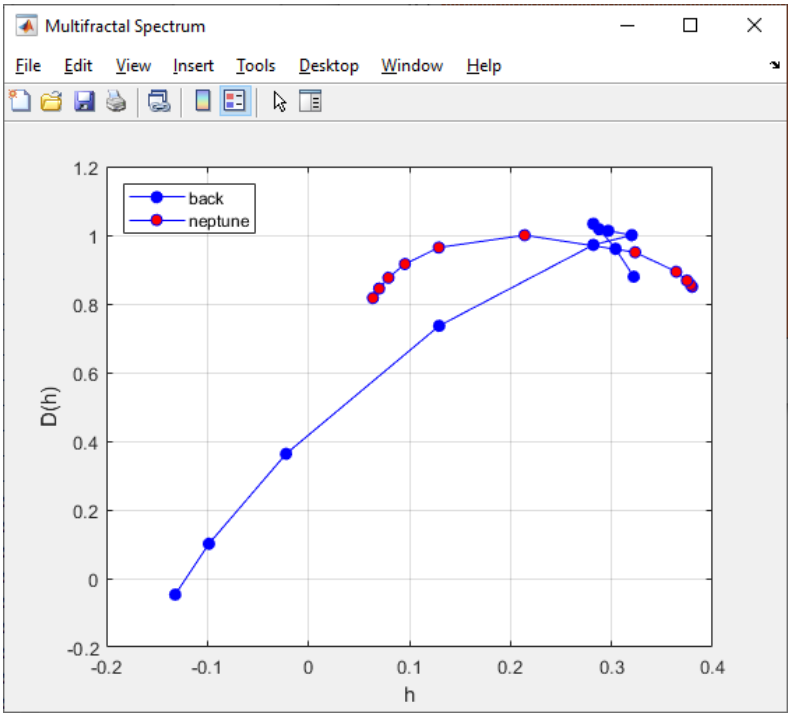
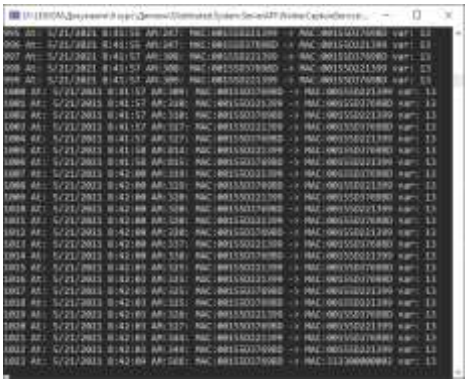
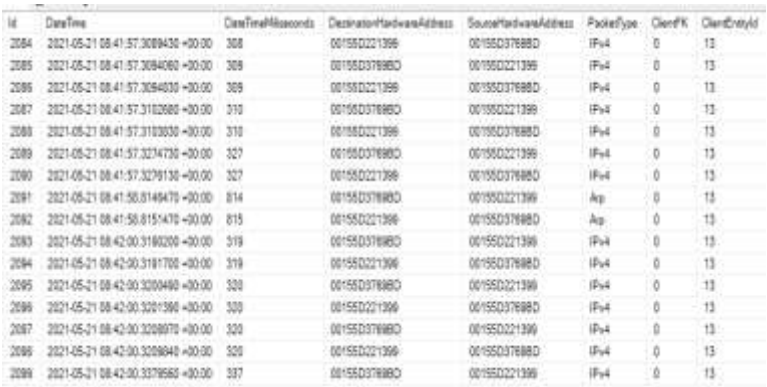


Рис. 4. Мультифрактальний спектр двох різновидів DOS атак

Виконання клієнтського програмного забезпечення не потребує багато ресурсів та невидиме для користувача, оскільки відбувається в якості фонової служби. Для налагодження програми було використано консоль, де відбувається виведення активності вузла (рис. 5).





Id	DateTime	DateTimeMilliseconds	DestinationHardwareAddress	SourceHardwareAddress	PacketType	ClientPK	ClientEntryId
2084	2021-05-21 08:41:57.3089430 +00:00	308	00155D221396	00155D37688D	IPv4	0	13
2085	2021-05-21 08:41:57.3094060 +00:00	309	00155D37688D	00155D221396	IPv4	0	13
2086	2021-05-21 08:41:57.3094630 +00:00	309	00155D221396	00155D37688D	IPv4	0	13
2087	2021-05-21 08:41:57.3102680 +00:00	310	00155D37688D	00155D221396	IPv4	0	13
2088	2021-05-21 08:41:57.3103030 +00:00	310	00155D221396	00155D37688D	IPv4	0	13
2089	2021-05-21 08:41:57.3274730 +00:00	327	00155D37688D	00155D221396	IPv4	0	13
2090	2021-05-21 08:41:57.3276130 +00:00	327	00155D221396	00155D37688D	IPv4	0	13
2091	2021-05-21 08:41:58.8146470 +00:00	814	00155D37688D	00155D221396	Arp	0	13
2092	2021-05-21 08:41:58.8151470 +00:00	815	00155D221396	00155D37688D	Arp	0	13
2093	2021-05-21 08:42:00.3180200 +00:00	318	00155D37688D	00155D221396	IPv4	0	13
2094	2021-05-21 08:42:00.3181700 +00:00	318	00155D221396	00155D37688D	IPv4	0	13
2095	2021-05-21 08:42:00.3200460 +00:00	320	00155D37688D	00155D221396	IPv4	0	13
2096	2021-05-21 08:42:00.3201390 +00:00	320	00155D221396	00155D37688D	IPv4	0	13
2097	2021-05-21 08:42:00.3208970 +00:00	320	00155D37688D	00155D221396	IPv4	0	13
2098	2021-05-21 08:42:00.3209840 +00:00	320	00155D221396	00155D37688D	IPv4	0	13
2099	2021-05-21 08:42:00.3379560 +00:00	337	00155D37688D	00155D221396	IPv4	0	13

а б

Рис. 5. Результати роботи розробленого програмного забезпечення:
а) консоль клієнтського додатку, б) фрагмент таблиці отриманих мережевих пакетів

В процесі реалізації запропонованої системи було визначено також швидкість її роботи. Так загальний час обробки даних складає близько 8 секунд, тоді як об'єм даних складає майже 4,9 мільйони рядків, в форматі текстового документу ці дані займають більше 700 мегабайт. Таким чином, це свідчить про високу швидкість роботи алгоритму та ефективність використання системних ресурсів.

Висновки. В результаті проведеного дослідження розроблено архітектуру та компоненти розподіленої системи виявлення мережевих атак, в якій поєднано вимоги централізованості, розподіленості, самоорганізованості та на її основі здійснено розробку централізованої розподіленої системи визначення мережевих атак в корпоративних комп'ютерних мережах на основі мультифрактального аналізу. Проведені експериментальні дослідження з реалізованою централізованою розподіленою системою визначення мережевих атак в комп'ютерних мережах підтвердили ефективність функціонування в комп'ютерній мережі.

References

1. C. Sarkar, A. U. Nambi S. N., R. V. Prasad, A. Rahim, R. Neisse and G. Baldini, "DIAT: A Scalable Distributed Architecture for IoT," in IEEE Internet of Things Journal, vol. 2, no. 3, pp. 230-239, June 2015, doi: 10.1109/JIOT.2014.2387155.

2. F. Jammes et al., "Technologies for SOA-based distributed large scale process monitoring and control systems," IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society, 2012, pp. 5799-5804, doi: 10.1109/IECON.2012.6389589.

3. S. Helen, IRM: Integrated file replication and consistency maintainence in P2P Systems, IEEE Trans. on Parallel and Distributed Systems, Vol. 21, No. 1, January 2010, pp. 100-113.

4. A. Sperotto, G. Schaffrath, R. Sadre, C. Morariu, A. Pras and B. Stiller, "An overview of ip flow-based intrusion detection", IEEE communications surveys & tutorials, vol. 12, no. 3, pp. 343-356, 2010.

5. A. Gupta, O. J. Pandey, M. Shukla, A. Dadhich, S. Mathur and A. Ingle, "Computational intelligence based intrusion detection systems for wireless communication and pervasive computing networks," 2013 IEEE International Conference on Computational Intelligence and Computing Research, 2013, pp. 1-7, doi: 10.1109/ICCIC.2013.6724156.

6. S. S. Y. Shim, "Guest Editor's Introduction: The CAP Theorem's Growing Impact," in Computer, vol. 45, no. 2, pp. 21-22, Feb. 2012, doi: 10.1109/MC.2012.54.

7. P. Dymora, M. Mazurek, Anomaly detection in iot communication network based on spectral analysis and hurst exponent, Applied Sciences 9, 5319 (2019). doi: 10.3390/app9245319

8. KDD Cup 1999 Dataset, URL: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html