ВОЛОКОННО-ОПТИЧНІ ТЕХНОЛОГІЇ В ІНФОРМАЦІЙНИХ (INTERNET, INTRANET ТОЩО) ТА ЕНЕРГЕТИЧНИХ МЕРЕЖАХ

УДК 004.056.5:004.056.8

І.Є. РОМАНЕЦЬ

ОНТОЛОГІЧНИЙ ПІДХІД В СИСТЕМІ ЗАБЕЗПЕЧЕННЯ БЕЗПЕКИ ВИКОРИСТАННЯ IP-ТЕЛЕФОНІЇ

Західноукраїнський національний університет,
46009, Львівська, 11, м. Тернопіль, Україна, e-mail: i.romanets@wunu.edu.ua

Анотація. У сучасному інформаційному світі використанням VoIP стає привабливим варіантом для спілкування користувачів. Враховуючи тенденцію зниження оплати за базові широкосмугові послуги та швидке зростання швидкості Інтернету, використання VoIP має лише набирати популярності. Однак у міру збільшення використання VoIP зростають і потенційні загрози для звичайних користувачі. У роботі розглянуто особливості організації корпоративної системи VoIP телефонії, виділено основні проблеми в системі захисту інформації в VoIP телефонії, та окреслено шляхи їх вирішення. Особливо актуальним є розвиток методів аналізу мовлення та відповідної обробки природної мови, який дозволяє створювати більш точні та ефективні системи виявлення аномального трафіку та потенційно небезпечних комунікацій. З постійним розвитком технологій штучного інтелекту цікавим стає напрямок використання інтелектуальних засобів для аналізу контенту в системі VoIP. Запропоновано метод виявлення аномалій в трафіку IP-телефонії на основі групування VoIP повідомлень на основі контекстно-частотного аналізу. Запропоновано метод автоматизованого наповнення онтології тематичних повідомлень в корпоративній системі IP-телефонії, який ґрунтується на формалізовано представленні повідомлень за допомогою деревовидних структур та на описі операцій взаємодії засобами алгебри кортежів. Здійснено програмну реалізацію перетворення голосових повідомлень в текстові представлення з використанням бібліотеки SpeechRecognition для перетворення голосу в текст у мові програмування Python. Проведено експериментальні дослідження запропонованих підходів, імплементовано програмну підсистему виявлення аномальних повідомлень на основі онтологічного підходу в діючу корпоративну IP-телефонію.

Abstract. In today's information world, the use of VoIP has become an attractive option for user communication. With the downward trend in paying for basic broadband services and the rapid increase in internet speeds, the use of VoIP should only continue to grow in popularity. However, as the use of VoIP increases, so do the potential threats to ordinary users. This article examines the peculiarities of organizing corporate VoIP telephony systems, highlights the main problems in information protection systems in VoIP telephony, and outlines ways to solve them. The development of methods for speech analysis and corresponding processing of natural language, which allows for creating more accurate and effective systems for detecting anomalous traffic and potentially dangerous communications, is especially relevant. With the continuous development of artificial intelligence technologies, the direction of using intelligent means for content analysis in the VoIP system is becoming interesting. A method for detecting anomalies in IP-telephony traffic, based on grouping VoIP messages through context-frequency analysis, is proposed. Additionally, a method for automated filling of the ontology of thematic messages in corporate IP-telephony systems is proposed, based on the formalized presentation of messages using tree-like structures and the description of interaction operations through tuple algebra. Furthermore, a software implementation for converting voice messages into text representations using the SpeechRecognition library for voice-to-text conversion in the Python programming language was created. Experimental studies of the proposed approaches were conducted, and a software subsystem for detecting anomalous messages based on the ontological approach was implemented in the current corporate IP-telephony system.

Keywords: IP-telephony systems, VoIP, ontology, information protection, SpeechRecognition.

DOI: 10.31649/1681-7893-2024-47-1-240-252

ВОЛОКОННО-ОПТИЧНІ ТЕХНОЛОГІЇ В ІНФОРМАЦІЙНИХ (INTERNET, INTRANET ТОЩО) ТА ЕНЕРГЕТИЧНИХ МЕРЕЖАХ

ВСТУП

VoIP – це технологія, яка дозволяє здійснювати передачу медіа-даних у реальному часів через широкосмугове підключення до Інтернету замість типових аналогових телефонних ліній. Зазвичай VoIP дає змогу дзвонити іншим людям, які також отримують дзвінки через використання протоколів TCP/IP. Взаємозв'язані сервіси VoIP також дозволяють здійснювати та отримувати дзвінки з традиційних номерів стаціонарних телефонів, але за окрему плату. Для деяких послуг VoIP потрібен комп'ютер або виділений телефон VoIP, інші реалізовуються шляхом використання стаціонарного телефону для здійснення VoIP дзвінків через спеціальний адаптер [1-3].

У сучасному інформаційному світі використанням VoIP стає привабливим варіантом для спілкування користувачів. Враховуючи тенденцію зниження оплати за базові широкосмугові послуги та швидке зростання швидкості Інтернету, використання VoIP має лише набирати популярності. Однак у міру збільшення використання VoIP зростають і потенційні загрози для звичайних користувачі. Хоча уразливості VoIP зазвичай подібні до тих, з якими стикаються користувачі в Інтернеті, з'явилися нові загрози, такі як шахрайство та атаки, які є унікальні для IP-телефонії [4,5].

Користувачі Інтернету знайомі з неприємністю зловживання електронною поштою у вигляді спроб спаму та фішингу. VoIP відкриває ще один шлях для цих неприємностей, які можуть призводити до спаму через Інтернет-телефонію (SPIT), спуфінгу та крадіжки особистих даних. Крім того, конфіденційність самих розмов VoIP опинилася під вразливістю залежно від типу послуги або конфігурації VoIP.

Зі збільшенням використання VoIP зростатимуть і надокучливі маркетингові стратегії, які з ним пов'язані. Новим різновидом гібриду цих двох концепцій є SPIT, або спам через інтернет-телефонію. Спам електронною поштою, надсилання комерційних повідомлень через VoIP – це швидко та дешево. На відміну від традиційного телемаркетингу, VoIP пропонує потенціал для великої кількості небажаних дзвінків через широкий спектр інструментів, які вже доступні зловмисникам в Інтернеті. Продавці телемаркетингу можуть легко надсилати велику кількість повідомлень клієнтам VoIP. На відміну від традиційних спам-повідомлень електронної пошти, які у середньому займають лише 10–20 Кбайт, небажана голосова пошта VoIP може вимагати вже мегабайти пам'яті для зберігання інформації [6].

Технічно для зловмисника можливо маскуватися під іншого абонента VoIP. Наприклад, зловмисник може вставити фальшивий ідентифікатор абонента у звичайний виклик VoIP, щоб одержувач вважав, що дзвінок надійшов із відомого та надійного джерела (наприклад, банку). Тоді інший користувач, який обдурений електронною ідентифікацією абонента, може невиправдано довіряти людині на іншому кінці. У такому разі одержувач може шляхом обману розкрити особисті дані, такі як інформація щодо номерів рахунків, номери банківських карток або іншу конфіденційну інформацію. Ця схема, по суті, є версією традиційного VoIP фішингу, коли користувач переходить за посиланнями в небажаному електронному листі та обманом змушений надати особисті дані на фальшивому веб-сайті.

Багато критиків VoIP ставлять під сумнів його конфіденційність. Проблема полягає в тому, що іноді дані VoIP передаються в незашифрованому вигляді через Інтернет. Тому технічно є можливість комусь перехопити дані VoIP і спробувати реконструювати розмову. Хоча це надзвичайно важко зробити, проте деякі програмні засоби розроблені для збирання бітів і фрагментів даних VoIP для здійснення спроб реконструювати розмове.

Виходячи з цього, розробка методів захисту VoIP залишається дуже актуальним напрямком наукових досліджень, особливо в контексті постійного розвитку інтернет технологій та відповідним зростання кіберзагроз.

АНАЛІЗ ЛІТЕРАТУРНИХ ДАНИХ ТА ПОСТАНОВКА ЗАДАЧІ ДОСЛІДЖЕННЯ

Використання VoIP на сьогодні дуже швидко розвивається і набуває все більшого значення у бізнесі та особистому користуванні. Це пов'язано насамперед із зміною тенденції та постійним розвитком VoIP технологій. На рисунку 1 представлено аналіз основних напрямків, які безпосередньо впливають на розвиток та використання сервісів VoIP.

Розглянемо ці напрямки більш детально. Зростання популярності хмарних сервісів VoIP - хмарні VoIP-системи стають більш популярними серед бізнесу та окремих користувачів через їхню гнучкість, швидкість впровадження та економічність. Завдяки постійному розвитку мережевих технологій та протоколів передачі даних, які використовуються в VoIP, якість голосу та зв'язку постійно покращується, що безпосередньо впливає на якість обслуговування. Сучасні VoIP-системи надають широкий спектр додаткових функцій, таких як конференц-зв'язок, зміна голосу, відеодзвінки, інтеграція з CRM-системами [7,8].

Зростання мобільності відбувається завдяки використанню мобільних додатків та можливість робити виклики через Інтернет з будь-якого місця. Це надає можливість мобільним користувачам

доступу до VoIP-зв'язку незалежно від їхнього місця перебування. Інтеграція з іншими технологіями - VoIP стає складовою частиною інтегрованих комунікаційних рішень, які включають в себе електронну пошту, чат, відеоконференції та інші засоби спілкування. Особливо необхідно відзначити такий напрям, як забезпечення безпеки. Розвиток технологій шифрування, аутентифікації та контролю доступу допомагає забезпечити безпеку VoIP-зв'язку та захистити дані користувачів від несанкціонованого доступу та перехоплення. Підсумовуючи, необхідно відзначити, що розвиток VoIP відбувається швидкими темпами, відкриваючи нові можливості для спілкування та покращення бізнес-комунікацій, а напрямок забезпечення безпеки VoIP є особливо актуальним в сучасних умовах [9,10].

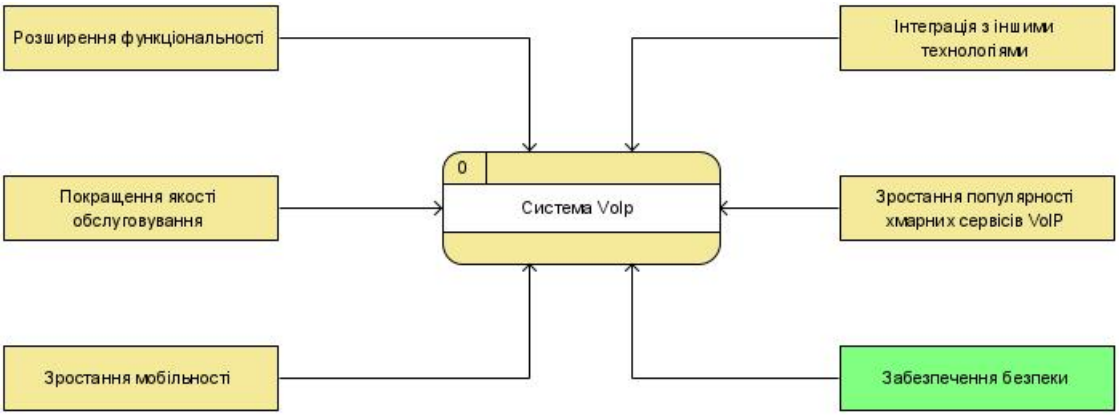

Рисунок 1 – Особливості розширення та використання VoIP технологій

Оскільки забезпечення безпеки функціонування VoIP є одним із ключових аспектів, які безпосередньо впливають на якість IP-телефонії та розширення сфер її використання, то проаналізуємо більш детально відомі методи та засоби, які використовуються для захист інформації у VoIP (рисунок 2).

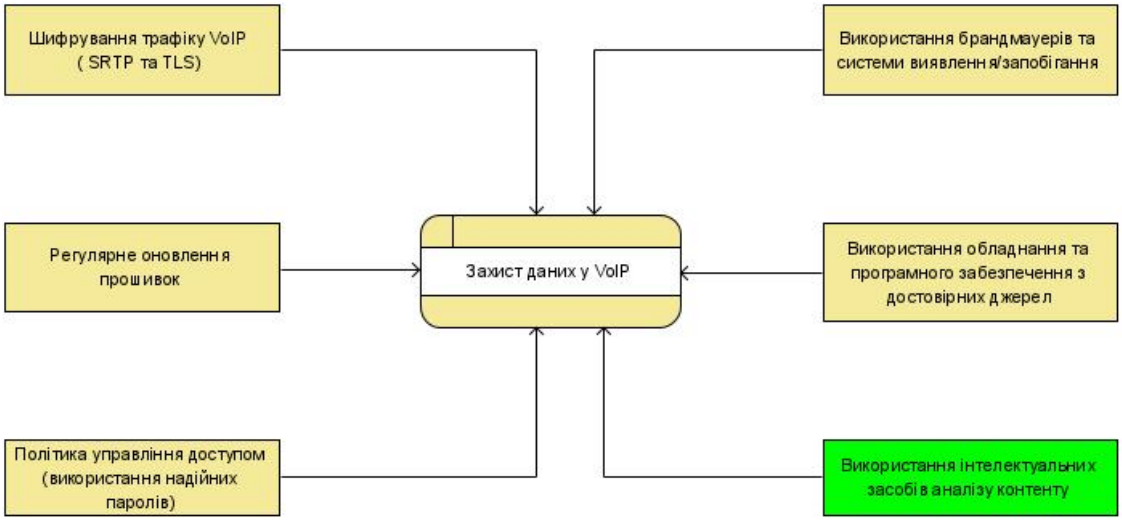

Рисунок 2 – Підходи до захисту інформації в системі VoIP

Системи VoIP вразливі до різних загроз безпеки, включаючи підслуховування, фішинг і атаки на відмову в обслуговуванні (DoS). Відомими підходами до зміцнення безпеки в IP-телефонії є наступні [11,12]:

- трафіку VoIP за допомогою різних протоколів, таких як безпечний транспортний протокол реального часу (SRTP) і безпечний протокол транспортного рівня (TLS);
- підходи до ефективної політики управління доступом, що включає використання надійних паролей та контроль доступу, щоб запобігти несанкціонованому проникненню;
- регулярне оновлення прошивок та програмного забезпечення для виправлення вразливостей безпеки;
- шифрування використання брандмауерів та систем виявлення/запобігання вторгненням для захисту від кіберзагроз;

- використання обладнання та програмного забезпечення VoIP від перевірених постачальників, які дотримуються відповідних стандартів та протоколів;
- використання інтелектуальних засобів аналізу контенту.

З постійним розвитком технологій штучного інтелекту, особливо цікавим є напрямок використання інтелектуальних засобів для аналізу контенту в системі VoIP. Розглянемо основні сучасні підходи, які можна використовувати для аналізу VoIP контенту (рисунок 1.3).

Використання алгоритмів машинного навчання та обробки природної мови (NLP) для сканування тексту VoIP-повідомлень на ключові слова, фрази або семантичні зв'язки дозволяють знаходити інформацію, яка носить зловмисний характер. Використання методів та спеціалізованого програмного забезпечення для голосового аналізу допомагає визначати специфічні звуки, мовні ознаки або відтінки, що характеризують пропаганду або загрози. Дослідження метаданих VoIP-повідомлень, таких як IP-адреси, час і довжина дзвінка, для виявлення аномалій або звичайних патернів, що можуть вказувати на можливу злочинну діяльність. Моніторинг і фільтрація контенту у поєднанні з використанням спеціалізованого програмного забезпечення для моніторингу та фільтрації VoIP-повідомлень на основі заздалегідь визначених критеріїв. Використання інформації з аналізу для формування блеклистів IP-адрес, номерів телефонів чи інших ідентифікаторів, а також для виявлення повторюваних патернів комунікацій, які можуть вказувати на злочинну діяльність.

Важливо відзначити, що ефективність аналізу VoIP-повідомлень (як голосових так і текстових) на вміст злочинної інформації може залежати від комбінації цих методів та технологій, а також від постійного вдосконалення і адаптації до нових загроз.

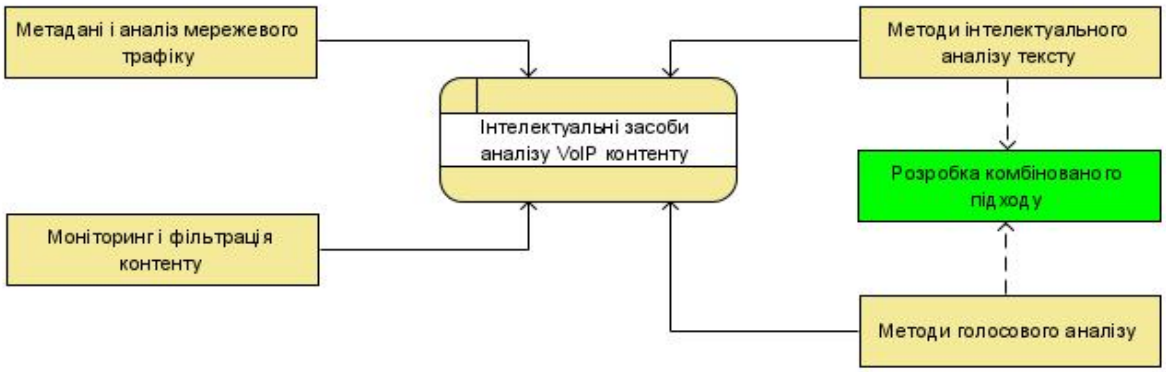

Рисунок 3 – Підходи до інтелектуального аналізу контенту в системі VoIP

Особливо актуальним є розвиток методів аналізу мовлення та відповідної обробки природної мови, який дозволяє створювати більш точні та ефективні системи виявлення аномального трафіку та потенційно небезпечних комунікацій. Метою цієї статті є розробка методів та засобів опрацювання даних в IP-телефонії на основі знання-орієнтованого підходу з використанням онтологій [13,14] для покращення безпеки використання VoIP.

Створення онтології функціонування VoIP вимагає співпраці між експертами в предметній області, інженерами та дослідниками, які знайомі з технологіями, стандартами та найкращими підходами до організації VoIP. Така онтологія сприятиме полегшенню взаємодії, семантичній інтеграції та допомагатиме в системному проектуванні, управлінні, виправлені несправностей, а також сприятиме покращенню системи захисту інформації в VoIP.

Онтологія даної предметної області формується на основі структурованого представлення концептів, сутностей, зв’язків у сфері технології VoIP. Вона дозволить отримати формалізовану та семантично насичену структуру для розуміння та представлення знань про протоколи, компоненти та їх взаємодію.

Онтологія VoIP повинна включати [16-18]:

- поняття та сутності: стек протоколів VoIP, представлення багаторівневої архітектури протоколів VoIP, включаючи такі протоколи, як SIP (протокол ініціації сеансу), RTP (транспортний протокол реального часу), SDP (протокол опису сеансу) та інші; компоненти VoIP: об’єкти, залучені до зв’язку VoIP, такі як користувацькі додатки (мобільні телефони з додатками, IP-телефони), шлюзи VoIP, проксі-сервери, реєстратори та медіа-сервери; опис життєвого циклу VoIP виклику: представлення різних етапів VoIP виклику, включаючи встановлення виклику, узгодження, розрив виклику та систему управління викликом;

- відношення: комунікаційні зв'язки: зв'язки між об'єктами VoIP, такі як зв'язок користувач-користувач, зв'язок користувач-сервер, зв'язок сервер-сервер і зв'язок медіа-потоку; відношення залежностей між компонентами та протоколами VoIP, такі як залежність SIP від базових транспортних протоколів, таких як UDP або TCP, і залежність RTP від RTCP для зворотного зв'язку та управління;
- властивості та атрибути: властивості протоколу VoIP, такі як формати повідомлень, заголовки, методи, параметри та коди стану, визначені SIP, RTP та іншими відповідними протоколами; атрибути компонентів: атрибути компонентів VoIP, наприклад IP-адреси, порти, кодеки, підтримувані функції, можливості та конфігурації; події VoIP: події, пов'язані зі зв'язком VoIP, такі як ініціювання виклику, завершення виклику, утримання/відновлення виклику, переадресація виклику, переадресація виклику та події медіапотоку; дії: дії, які виконують об'єкти VoIP у відповідь на події, такі як надсилання SIP-повідомлень, узгодження параметрів медіа, встановлення та розрив медіа-сеансів і обробка сигналів управління викликами;
- ієрархічна таксономія: класифікація об'єктів і протоколів VoIP в ієрархічні таксономії на основі їхніх ролей, функцій і зв'язків; категоризація технологій VoIP на основі моделей розгортання (локальні, хмарні), мережевих архітектур (одноранговi, клієнт-сервер) і сценаріїв використання (корпоративні VoIP, мобільний VoIP);
- семантичні обмеження: обмеження на зв'язки та взаємодію між об'єктами VoIP, що забезпечують узгодженість в онтології; правила протоколу: правила, що регулюють поведінку та використання протоколів VoIP, включаючи правила обробки повідомлень, встановлення сеансу та узгодження медіа для окремих протоколів.

ПОБУДОВА ОНТОЛОГІЇ ПОВІДОМЛЕНЬ VOIP

При побудові онтології опису VoIP повідомлень в системі IP-телефонії необхідно здійснити формалізацію основних понять у формі окремих концептів, та описати зв'язки між цими поняттями. Розробка онтології для повідомлень VoIP передбачає створення структурованого представлення концептів, сутностей та зв'язків у межах відповідного домену [18,19]. На рисунку 4 представлено онтологічний граф для формалізованого опису VoIP повідомлень.

На першому етапі здійснюємо ідентифікацію домену, яка включає визначення області та меж описуваної онтології. У нашому випадку це буде зв'язок за протоколом передачі голосових повідомлень через Інтернет (VoIP), що охоплює концепції, пов'язані з передачею голосових даних через IP-мережі [20].

На другому етапі визначаємо основні поняття та сутності, які пов'язані з VoIP повідомленнями:

- протоколи VoIP (SIP, RTP, RTCP);
- компоненти VoIP (клієнт VoIP, сервер VoIP, шлюз VoIP);
- типи повідомлень VoIP (пакети INVITE, ACK, BYE, RTP);
- параметри VoIP (IP-адреса клієнта, IP-адреса призначення, номери портів, тип корисного навантаження);
- повідомлення (голосові та текстові), текстові повідомлення характеризуються основної частиною, автором, переліком ключових слів, оцінкою, та комбінаційним представленням.

На третьому етапі формуємо опис ієрархічних зв'язків та будуємо таксономію класів, наприклад: протоколи VoIP (SIP,RTP,RTCP), компоненти VoIP (клієнт, сервер, шлюз). Будуємо відповідні відношення між класами, тобто визначаємо відношення між різними поняттями. Цей процес може включати такі зв'язки «є» (успадкування), «частина» (композиція), «має» (асоціація) тощо. Наприклад:

- клієнт VoIP має сеанс SIP;
- пакет RTP є типом пакета даних VoIP;
- сервер VoIP є «частиною» інфраструктури VoIP.

На четвертому етапі визначаємо властивості та атрибути, пов'язані з кожним поняттям. Вони представляють характеристики або ознаки, які їх описують. Наприклад, SIP-повідомлення має наступні атрибути:

- тип запиту (ACK, BYE);
- код відповіді (200 OK, 404 Not Found);
- заголовки SIP (Call-ID);
- порядковий номер;
- мітка часу.

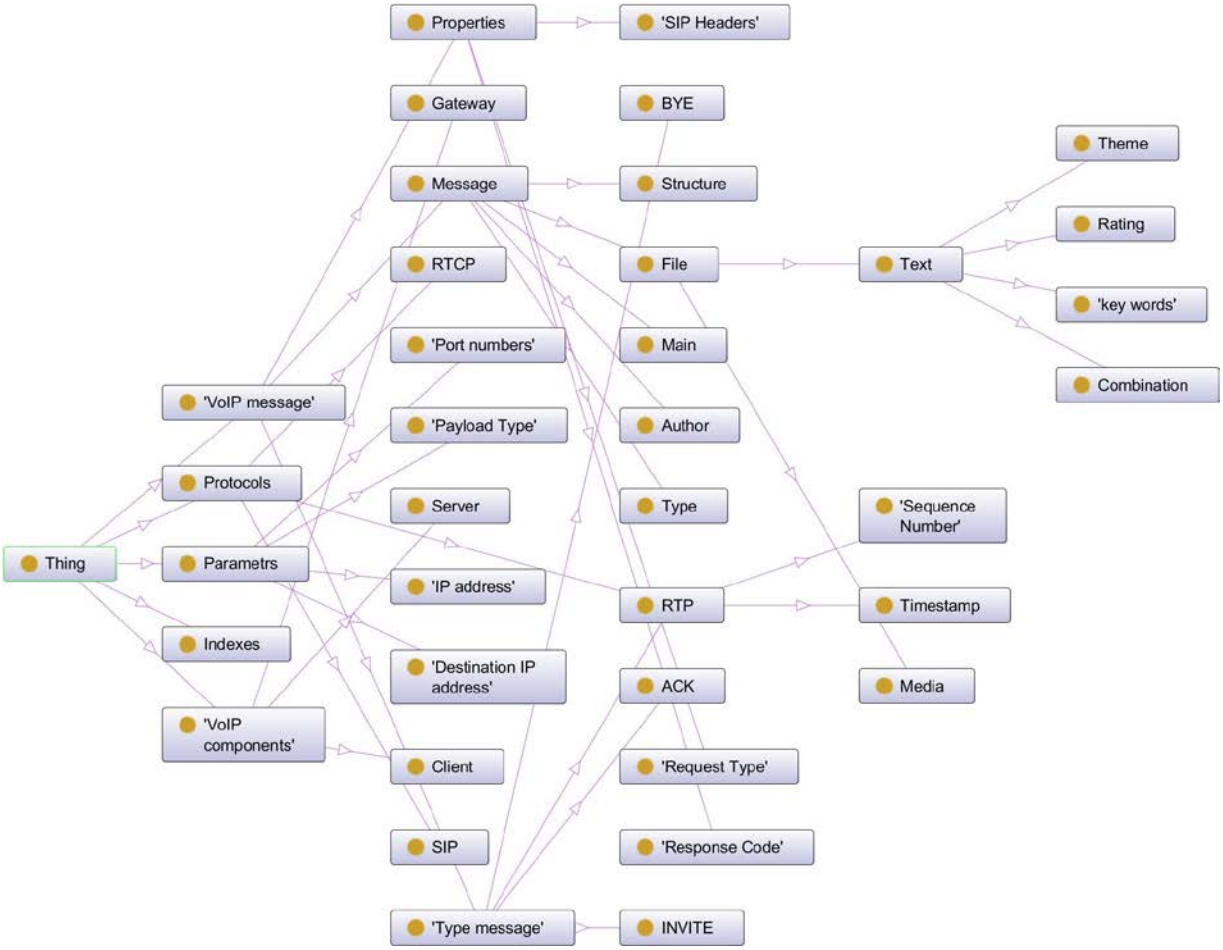

Рисунок 4 – Представлення формалізованої предметної області «VoIP повідомлення» онтологічного графу

На п’ятому етапі формуємо систему правил та обмеження для забезпечення узгодженості онтології. Це може включати обмеження відповідності правил домену предметної області, обмеження типів даних.

Заключний етапом є перевірка онтології на реальних сценаріях VoIP, яка включає перевірку адекватності відношення до домену і відповідності до вимог практичного використання.

МЕТОД ВИЯВЛЕННЯ АНОМАЛІЙ В ТРАФІКУ IP-ТЕЛЕФОНІЇ НА ОСНОВІ ГРУПУВАННЯ VOIP ПОВІДОМЛЕНЬ

Для аналізу VoIP контенту необхідно формалізувати метод побудови базового повідомлення в IP-телефонії. Для цього необхідно спочатку здійснити перетворення VoIP медіа повідомлення в текстове представлення і таким чином сформувати базу даних вже текстових повідомлень.

Для перетворення голосового повідомлення в текстове використаємо бібліотеки SpeechRecognition для перетворення голосу в текст у мові програмування Python. Відповідна програмна реалізація такого перетворення представлена на рисунку 5.

Тематику таких повідомлень задаємо загальним текстовим ідентифікатором IMP предметної області та відповідним специфікатором SM. За допомогою перетворення голосового повідомлення у текстове представлення та процедури символьної конкатенації $VPS = IMP \,\&\, SM$ отримаємо множину VP текстових повідомлень, які представлені наборами відповідних компонентів конкретних понять

$$VP(VPS, P) = \{VP_i\}_{i=1}^{P}, \quad . \tag{1}$$

де P — потужність множини VP ; VP_i — елемент множини, що визначає набір понять i - того повідомлення.

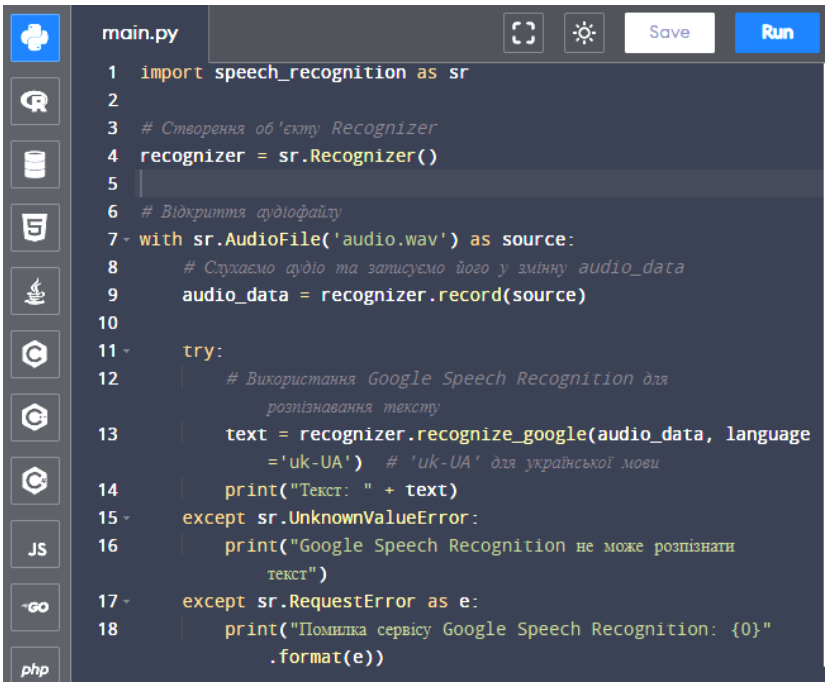

Рисунок 5 – Фрагмент лістингу програмного коду для перетворення VoIP медіа повідомлення в текстове представлення

На наступному кроці для аналізу понять повідомлень використовуємо лише ті поняття, які відносяться до аналізованої предметної області, тобто належать множині VPK, а не здійснюємо повну перевірку усієї множини понять. Критерієм виконання такої умови є наявність ідентифікатора предметної області в темі повідомлення:

$$VPK = \{VP_{i*}^{*} | VP_{i*} \in VP, VP_{i*}.theme \cap IMP \neq \emptyset\}_{i=1}^{P*} \quad . \tag{2}$$

Із набору текстових тверджень необхідно вибрати інформацію, яка визначає тематику повідомлень. У першу чергу інформація про тематику повідомлення визначається у наборі тверджень, з яких починається повідомлення.

Якщо початок повідомлення не дозволяє визначити його тематику, то інформацію про тематику вибираємо з впорядкованої множини елементів на основі частотного аналізу понять. Інформація такого типу буде впорядкованим списком $LKB(VP_i^*)$ виділених ключових понять :

$$LKB(VP_i^*) =< C_{ik}(VP_i^*) >_{k=1}^{CPi} \quad . \tag{3}$$

Така множина може також містити також і випадкову інформацію. Однак елементи множини понять, які будуть повторюватися, дають нам інформацію про структуру і тематику такого повідомлення. Тому на основі впорядкованої множини та множини ключових понять сформуємо базову BCM та узагальнену GCM множини пар поняття - частота появи поняття, які визначаємо наступним чином:

$$BCM = \{(CS_l, NCS_l) | CS_l\}, \tag{4}$$

$$GCM = \{(CS_m, NCS_m) | CS_m \in \bigcup_i LKB(VP_i^*)\} \tag{5}$$

Для знаходження понять, які будуть релевантних до заданої предметної області, впорядкуємо елементи множини GCM у порядку спадання відповідних частот елементів, а деяку частину понять включаємо відразу в базову концептуальну множину CSC

$$\text{CSC} = \left\{(CS_l, NCS_l) | FCL_l = \frac{NCS_l}{P^*} > FF_0\right\}, \tag{6}$$

де величина F_0, належить інтервалу [0.200; 0.500], а конкретне значення цієї величини вибираємо із врахуванням особливостей та специфіки предметної області. Далі аналізуємо множину понять, які мають низьку частоту, такий аналіз доцільний, якщо нам не вдалося наповнити множину понять на основі відношення (6). Розглянемо деяку визначену і впорядковану вибірку SCMF, яка також включає відповідні частоти понять у повідомлені

$$\text{SCMF} = \{NCS_l | (CS_l, NCS_l) \in BCM\}. \tag{7}$$

Частоти з найвищими значеннями перевіряємо за критерієм 4σ на характеристику аномальності. Концепти, що відповідають критерію аномальності, включаємо в сформовану множину CSC.

При формуванні множини SK ключових концептів, що визначають тематику повідомлення, то здійснюємо впорядкування цієї множини. Рангові номери присвоюємо лише ключовим концептам. Так CRC_{ik} позначає ранг k-го концепту в і-тому повідомлені. При аналізі повідомлень необхідно також врахувати вплив тривалості голосового повідомлення (довжини перетвореного в текст повідомлення) на предметну аудиторію, оскільки є багато повідомлень, які є короткотривалими. Для того, щоб врахувати таку особливість використаємо деяку вагову функцію $wcs(i)$.

Оскільки із спаданням тривалості голосових повідомлень їх важливість також буде плавно спадати, і чим коротше повідомлення, тим швидше відбуватиметься це спадання. Для моделювання такого процесу можна використати кубічний сплайн. Для її однозначного визначення необхідно накласти хоча б 4 умови. Зокрема найтриваліше повідомлення буде мати важливість для цього аргументу рівну 1, якщо похідна дорівнює 0. Вводимо систему ваг $CSG(i) = \frac{csg(i)}{\sum_i csg(i)}$., яка матиме нормований характер та обчислюємо усереднений ранг k-го концепту наступним співвідношенням:

$$RA_k = \sum_i R_{ik} CSG(i) \quad (8)$$

Ранжування концептів здійснюємо на основі усереднення рангів. Узгодженість рангів перевіряємо із використанням коефіцієнта конкордації [12]:

$$CSW = \frac{12 \sum_{k=1}^{K} (\sum_{i=1}^{I} R_{ik} CSG(i) - \bar{R})}{I^2 (K^3 - K)}, \quad (9)$$

де $\bar{R} = \frac{1}{K} \sum_{k=1}^{K} \sum_{i=1}^{I} R_{ik} CSG(i)$. Якщо виконується наступна умова

$$I(K-1)CSW > \chi^2_{K-1,\alpha} \ , \quad (10)$$

то таке ранжування є значущим [12].

Якщо ранжування повного списку концептів не є значущим, то відкидаємо один елемент із списку в порядку зростання відповідних ваг. Відкидання проводимо до отримання значущого ранжування, починаючи із концепту, який матиме найменшу вагу.

Запропонований метод дозволяє згрупувати повідомлення відповідно до класифікованої структури ключових понять. Таке групування доцільно використовувати для виявлення аномальних повідомлень в спеціалізованій корпоративній мережі з розгорнутою системою IP-телефонії.

МЕТОД АВТОМАТИЗОВАНОГО НАПОВНЕННЯ ОНТОЛОГІЇ ТЕМАТИЧНИХ ПОВІДОМЛЕНЬ В КОРПОРАТИВНІЙ СИСТЕМІ IP-ТЕЛЕФОНІЇ

Як описано в [5] онтологія описується деревоподібною системою концептів досліджуваної предметної області. Під концептами предметної області розуміються як і базові поняття предметної області, так і узагальнюючі поняття, які часто є спільними у багатьох суміжних предметних областях. Онтологію для VoIP можна змоделювати деякою деревовидною структурою наступного виду

$$OC\_Voip =< IdConcept, ParentConcept, Base >, \quad (11)$$

де $IdConcept$ ідентифікатор концепту, $ParentConcept$ ідентифікатор батьківського концепту та атрибут $Base$, який використовується для розмежування базові поняття від узагальнюючих понять. Базове поняття відрізняється від узагальнюючого тим, що узагальнюючі поняття не мають посилання на батьківський елемент, тобто $ParentConcept = 0$. Кожне поняття може описуватися деякими лінгвістичними представленнями у вигляді OC_Prs словосполучень. Онтологійні поняття описуються конкретними словоформами OC_Form , а також OC_Base основами цих понять. Словоформи можуть використовуватися для опису понять для користувачів, а основи понять – для автоматичного визначення еквівалентності представлень відповідно до вибраної мови. Атрибути таких понять групуємо в структури:

$$OC\_Base =< IdLanguage, IdBase, WordBase >, \quad (12)$$

$$OC\_Form =< IdLanguage, IdForm, IdBase, WordForm >, \quad (13)$$

$$OC\_Prs =< IdConcept, IdLanguage, IdPhrase, IdBase, IdForm, IdParentBase >, \quad (14)$$

де $IdLanguage$ ідентифікатор мови, $IdBase$ - ідентифікатор визначеної основи слова, $WordBase$ визначена основа слова, $IdForm$ ідентифікатор визначеної форми слова, $WordForm$ визначена словоформа, $IdConcept$ ідентифікатор аналізованого концепту, $IdPhrase$ ідентифікатор аналізованої фрази, $IdParentBase$ ідентифікатор основи поняття, що належить до батьківської категорії.

Для наповнення онтології VoIP можна використовувати:

- аналіз предметної області VoIP, включаючи протоколи, компоненти систем, типи повідомлень, характеристики зв'язку тощо;
- термінологічні словники, які використовуються у VoIP. Це включає технічні терміни, абревіатури, назви протоколів, тощо;
- стандарти та специфікації VoIP, такі як SIP (Session Initiation Protocol), RTP (Real-time Transport Protocol), SDP (Session Description Protocol) тощо. Ці документи надають інформацію про

протоколи, повідомлення, функції та інші аспекти VoIP;

- інформацію, що опублікована у відкритому доступі, така як повідомлення в тематичних групах в соціальних мережах, форуми, вікі, які можуть містити корисні відомості про проблеми, вирішення, нові можливості розвитки в галузі VoIP;
- співпрацю з експертами у галузі VoIP, такими як інженери, адміністратори мереж, дослідники, які можуть допомогти у розумінні складних питань;
- програмне забезпечення VoIP, такі як Asterisk, FreeSWITCH, Kamailio, які містять документацію, код та інші ресурси, які можна використовувати для розуміння архітектури та принципів роботи VoIP.
- для формування бази тематичних медіа-повідомлень, можна також використовувати спеціальні записи з медіа ресурсів (Youtube, Tiktok, тощо).

Для пошуку та виборки структурованої інформації, яка відноситься до аналізованої предметної області, формуємо множину $KeyConSet$, яка включає ключові поняття, які характеризують її визначення та поведінку. Для аналізу структури VoIp повідомлень визначимо деяку допоміжну множину $AdvSet$:

$$AdvSet =< IdMessAn, IdListElem, IdElement, IdBase, IdForm, IdParentBase >, \quad (15)$$

де $IdMessAn$ — ідентифікатор аналізованого повідомлення, $IdListElem$ — ідентифікатор набору всіх понять, $IdElement$ — ідентифікатор окремого елемента набору.

Аналіз повторюваних понять дозволяє визначити найбільш значущі поняття, і дозволяє відокремити їх від несуттєвих понять. Для аналізу та контролю важливих понять використаємо деяку структуру $BaseFr$ частот основ:

$$BaseFr =< IdBase, BsFreq, IdLastMess, ConBack > \quad (16)$$

де $IdBase$ — ідентифікатор основи, $BsFreq$ — частота появи основи в різних повідомленнях, $IdLastMess$ - ідентифікатор останнього повідомлення, де була основа поняття, $ConBack$ — відмітка фоновості поняття, що приймає $NULL$ значення за замовчуванням при деякій невизначеності.

При аналізі набору текстових повідомлень, які згруповані за відповідним критерієм встановлюємо деякий ідентифікатор такого набору:

$$CurrMessId := max\left(\pi_{IdLastMess}(BaseFr)\right) + 1. \quad (17)$$

Аналізовані елементи сформованої множини розбиваємо на набір елементарних понять із використанням роздільників, списку розбиваються на елементарні поняття $ElemCon$ із використанням роздільників.

До частин елементарного поняття необхідно застосувати процедуру $BaseConcept$ для формування їх основ. Ця процедура реалізовується шляхом відсікання закінчень від слів. При умові, що

$$BaseConcept(ElemCon_{Mess,List}) \subset BaseConcept(KeyConSet), \quad (18)$$

всі елементарні поняття з аналізованої множини з відповідними атрибутами додаємо в визначену структуру $AdvSet$. Словоформи вибираємо з елементів списку $ElemCon_{Mess,List}$ і розпізнаємо за допомогою відношення OC_Form або поповнюємо його.

Нехай $Word_k(ElemCon_{Mess,List})$ —k – те слово, яке відокремлене із елемента списку $ElemCon_{Mess,List}$. Якщо основа слова вже існує в аналізованій множині, то цей ідентифікатор $IdBaseWord$визначаємо з відношення OC_Base:

$$IdBaseWord = \pi_{IdBase}\left(\sigma_{WordBase=BaseConcept(Word_k(ElemCon_{Mess,List}))}(OC\_Base)\right). \quad (19)$$

Якщо основа слова вже знаходиться у відношенні $BaseFr$ і номер аналізованого повідомлення не співпадає із номером зарахованого на попередньому кроці, то частотний індекс $BsFreq$ збільшуємо на 1, а номер поточного повідомлення заноситься в поле $IdLastMess$:

$$\left(Count\left(\pi_{IdBase}\left(\sigma_{IdBase=IdBaseWord\ AND\ IdLastMess<>CurrMessId}(BaseFr)\right)\right) \neq 0\right) \Rightarrow$$
$$\Rightarrow (BaseFr.BsFreq := BaseFr.BsFreq + 1) \wedge (BaseFr.IdLastMess := CurrMessId). \quad (20)$$

Такий підхід забезпечує врахування частот використання основ у різних повідомленнях. Коли основа у відношенні $BaseFr$ не знайдена, її враховуємо з частотним індексом, який рівний 1 та додаємо у відношення, а також враховуємо номеру поточного повідомлення:

$$\left(Count\left(\pi_{IdBase}\left(\sigma_{IdBase=IdBaseWord}(BaseFr)\right)\right) = 0\right) \Rightarrow$$
$$\Rightarrow (BaseFr.IdBase := IdBaseWord) \wedge (BaseFr.BsFreq := 1) \wedge (BaseFr.IdLastMess := CurrMessId). \quad (21)$$

Якщо основа $Word_k(ElemCon_{Mess,List})$ основи не визначена, то вона додається у список основ, а слово додається в множину словоформ, а відношення часто визначаємо на основі (21).

Для додавання поняття в онтологію експерту із знанням предметної області у IP-телефонії пропонуються тільки основи з частотою, яка більша за визначене мінімальне значення $BaseFr_0 \geq 2$, яке

вибирається користувачем. Експерт із запропонованого списку основ здійснює вибір і включення в онтологію, коли дану умову задовольняє не менше $BaseFr_0$ основ

$$Count\left(\pi_{IdBase}\left(\sigma_{BaseFr>=BaseFr_0}(BaseFr)\right)\right) > BaseFr_0. \quad (22)$$

В процесі прийняття відповідного рішення, основи відображаються лише з контексту аналізованих повідомлень. Це дозволяє формувати поняття з кількох слів і не повторювати основи, які не обрані спеціалістом для включення в онтологію. Основу для контексту $ContextBase$ вибираємо виходячи із критерію максимальної частоти $MaxBsFreq$:

$$MaxBsFreq = max\left(\pi_{BsFreq}(\sigma_{ConBack=Null})\right), \quad (23)$$

$$ContextBase = rand\left(\pi_{IdBase}\left(\sigma_{BsFreq=MaxBsFreq}\right)\right). \quad (24)$$

Формування контексту визначеної основи здійснюється із використанням двовимірного масиву $ArrayCont$,який містить словоформні ідентифікатори. Перший індекс характеризує номер фрази в контексті, а інший – ідентифікатор номера слова у фразі. $NumbPhrCont$ – показник, який характеризує кількість фраз контексту:

$$NumbPhrCont = count\left(\pi_{IdElement}\left(\sigma_{IdBase=CBase}(AdvSet)\right)\right). \quad (25)$$

Ідентифікатори усіх елементів з аналізованої структури заносимо в сформований масив $AdvPhrId$, який носить допоміжний характер:

$$AdvPhrId = \pi_{IdElement}\left(\sigma_{IdBase=CBase}(AdvSet)\right) \quad (26)$$

розмірності $NumbPhrCont$. І-та стрічка даного контекстного масиву визначається за наступним правилом:

$$AdvContext[i] = \left(\pi_{IdForm}\left(\sigma_{IdElement=AdvPhrN[i]\wedge IdParentBase=NULL}(AdvSet)\right) AS\ HeadLine\right) \cup$$
$$\cup\left(\pi_{IdForm}\left(\sigma_{IdElement=AdvPhrN[i]\wedge(IdParentBase=HeadLine.IdBase\ OR\ IdParentBase=HeadLine.IdBase)}(AdvSet)\right) AS\ HeadLine_1\right) \quad (27)$$

Сформований набір словоформ представляється спеціалістові в галузі VoIP та відповідної досліджуваної предметної області для формування елементів, які будуть поповнювати онтологію. Основи, які потрапляють в онтологію отримують фоновий показник $ConBack := 2$, щоб уникнути процедур повторного аналізу. Основи, які не використовувалися жодного разу, не можуть формувати контекстні основи, і для них показник $ConBack := 1$. Вибрані елементи, які використовуються для онтологійного наповнення, можуть формувати онтологійну ієрархію, яка дозволяє формалізувати експертні знання у формі онтології.

ПРОГРАМНА РЕАЛІЗАЦІЯ СИСТЕМИ ВИЯВЛЕННЯ АНОМАЛЬНИХ ПОВІДОМЛЕНЬ В VOIP

Для реалізації підходу до автоматизованого наповнення онтології тематичних повідомлень в корпоративній системі VoIP використано мову програмування Python та спеціалізовану бібліотеку Pandas для обробки та аналізу даних, що надає зручні та ефективні структури даних та інструменти для роботи з ними. Для збереження сформованих онтологічних структур використовується СУБД MySQL.

На рисунку 6 наведено інтерфейс аналізу ключових понять в окремих повідомленнях, які сформовані в процесі експлуатації засобів корпоративної IP-телефонії.

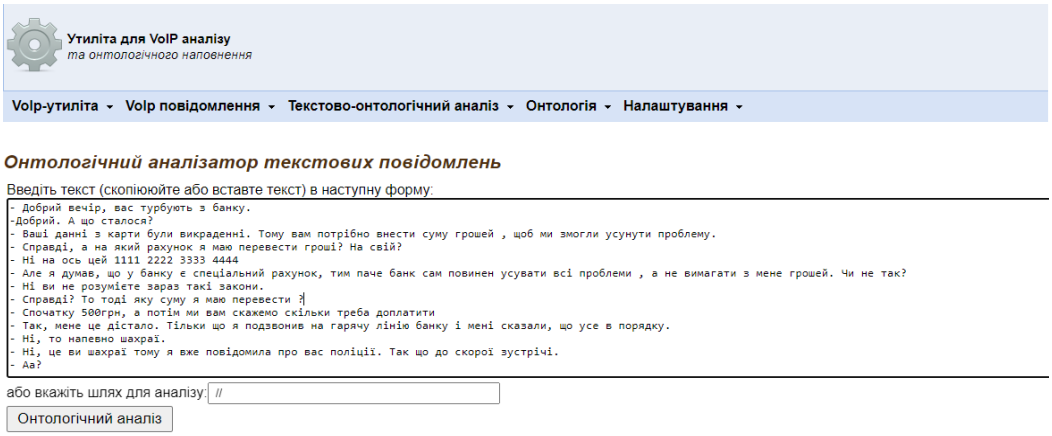

Рисунок 6 – Інтерфейс підсистеми аналізу ключових понять у VoIP повідомленнях

Після виконання процедури попереднього онтологічного аналізу, яка описано в попередньому розділі статті і програмно реалізована в системі, отримаємо результат, який представлено на рисунку 7. Сервіс дозволяє знайти ключові поняття в даному повідомлені, визначити необхідні коефіцієнти та відповідно дозволяє сформувати інформацію або для занесення в онтологічне представлення аналізованої області, або сформує схему для порівняння з формалізованими еталонними повідомленнями для виявлення аномальності такого повідомлення.

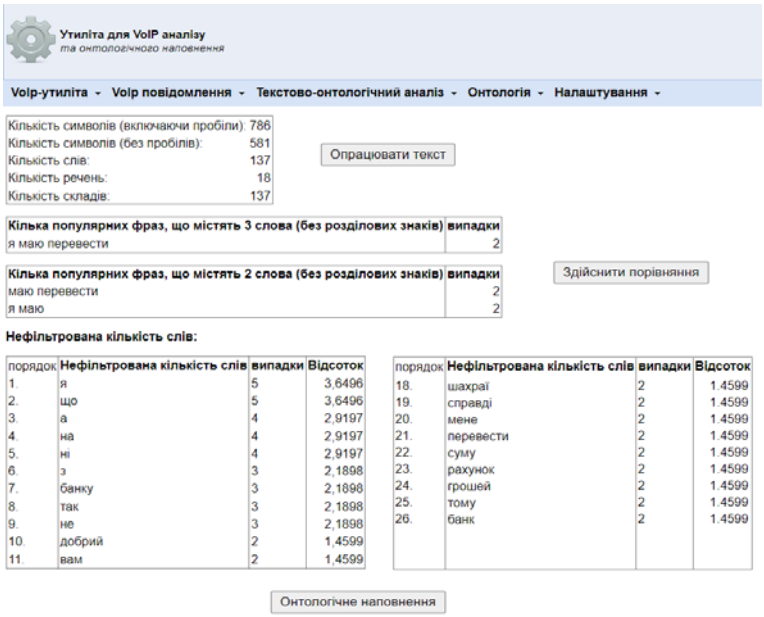

Рисунок 7 – Інтерфейс підсистеми аналізу ключових понять у VoIP повідомленнях та формування необхідних вагових коефіцієнтів для прийняття рішень

В процесі апробації запропонованих методів було здійснено інтеграцію даної інтелектуалізованої підсистеми в діючу корпоративну VoIP систему Західноукраїнського національного університету. На рисунку 8 представлено сторінку, яка окрім підсистеми аналізу ключових понять у VoIP включає також модуль для перетворення голосових повідомлень в текстові представлення і адміністративний модуль, що використовується для прийняття рішень відносно користувачів системи та відповідне управління виявленими аномальними повідомленнями.

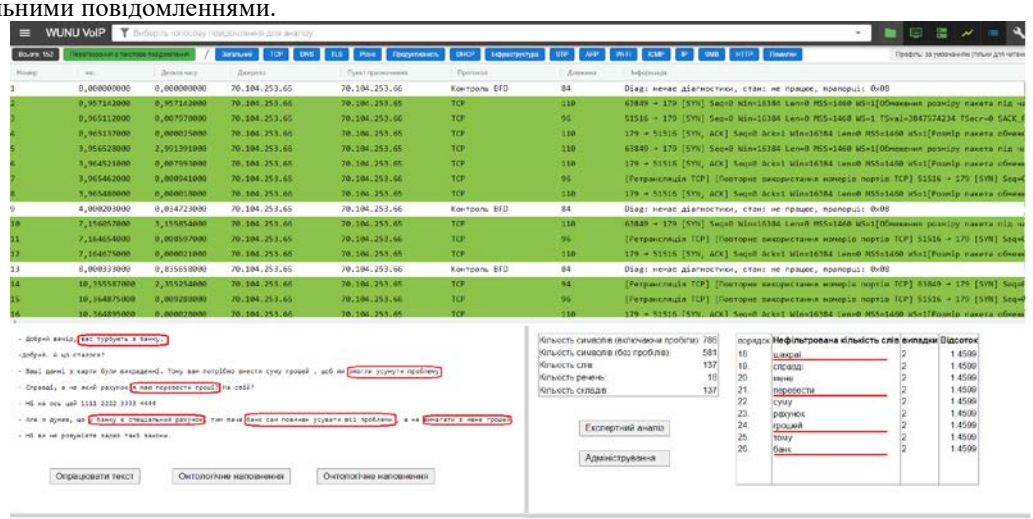

Рисунок 8 – Інтерфейс інтеграції підсистеми аналізу VoIP повідомленнях в систему управління корпоративною IP-телефонією

Дана інтеграція може також використовуватися і для виявлення голосових повідомлень, які здійснюють користувачі при дзвінках на платні сервіси, що також є дуже актуальним питанням в системі практичного використання IP-телефонії.

ВИСНОВКИ

У роботі розглянуто особливості організації корпоративної системи VoIP телефонії, виділено основні проблеми в системі захисту інформації в VoIP телефонії, та окреслено шляхи їх вирішення. Особливо актуальним є розвиток методів аналізу мовлення та відповідної обробки природної мови, який дозволяє створювати більш точні та ефективні системи виявлення аномального трафіку та потенційно небезпечних комунікацій.

З постійним розвитком технологій штучного інтелекту цікавим стає напрямок використання інтелектуальних засобів для аналізу контенту в системі VoIP. Запропоновано онтологію опису VoIP повідомлень в системі IP-телефонії, здійснено формалізацію основних понять у формі окремих концептів, та описано зв'язки між цими поняттями. Запропоновано метод виявлення аномалій в трафіку ip-телефонії на основі групування voip повідомлень на основі контекстно-частотного аналізу.

Запропоновано метод автоматизованого наповнення онтології тематичних повідомлень в корпоративній системі IP-телефонії, який грунтується на формалізовано представленні повідомлень за допомогою деревовидних структур та на описі операцій взаємодії засобами алгебри кортежів.

Здіснено програмну реалізацію перетворення голосових повідомлень в текстові представлення з використанням бібліотеки SpeechRecognition для перетворення голосу в текст у мові програмування Python. Проведено експериментальні дослідження запропонованих підходів, імплементовано програмну підсистему виявлення аномальних повідомлень на сонові онтологічного підходу в діючу корпоративну IP-телефонію.

Подальші дослідження у цьому напрямку будуть спрямовані на розширення бази онтологічних представлень у системі функціонування корпоративної IP-телефонії, апробації підсистеми онтологічного групування тематичних повідомлень та вдосконалені захисту корпоративної VoIP та реалізації підсистеми виявлення зловживань при дзвінках на платні сервіси.

СПИСОК ЛІТЕРАТУРИ / REFERENCES

1. Waleed Nazih, Alnowaiser Khaled, etc. "Detecting SPIT Attacks in VoIP Networks Using Convolutional Autoencoders: A Deep Learning Approach" Applied Sciences 13, no. 12: 6974, 2023. https://doi.org/10.3390/app13126974
2. Tuanhua L., "Interactive Behavior Analysis Based on Social Network," 2021 International Conference of Social Computing and Digital Economy (ICSCDE), 2021, pp. 188-191.
3. Pascual, S.; Bonafonte, A.; Serra, J. SEGAN: Speech Enhancement Generative Adversarial Network. In Proceedings of the 18th Annual Conference of the International Speech Communication Association (INTERSPEECH 2017), Stockholm, Sweden, 20–24 August 2017; pp. 3642–3646.
4. Hernes, M., Nguyen, N.T., Maleszka, M., Bytniewski, A. The automatic summarization of text documents in the cognitive integrated management information system (2015) Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, FedCSIS 2015, art. No 8, pp. 1387-1396.
5. Dyvak, M., Papa, O., Melnyk, A., Pukas, A., Porplytsya, N., Rot, A. Interval model of the efficiency of the functioning of information web resources for services on ecological expertise (2020) Mathematics, 8 (12), art. no. 2116, pp. 1-12.
6. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, JMLR Workshop and Conference Proceedings, Sardinia, Italy, 13–15 May 2010; pp. 249–256
7. Alvares, C.; Dinesh, D.; Alvi, S.; Gautam, T.; Hasib, M.; Raza, A. Dataset of attacks on a live enterprise VoIP network for machine learning based intrusion detection and prevention systems. Comput. Netw. 2021, 197, 108283
8. Pereira, D.; Oliveira, R. Detection of Abnormal SIP Signaling Patterns: A Deep Learning Comparison. Computers 2022, 11, 27.
9. Ruff, L.; Kauffmann, J.R.; Vandermeulen, R.A.; Montavon, G.; Samek, W.; Kloft, M.; Dietterich, T.G.; Müller, K.R. A unifying review of deep and shallow anomaly detection. Proc. IEEE 2021, 109, 756–795.
10. Mohamed, Amira A., Amira Eltokhy, and Abdelhalim A. Zekry. 2023. "Enhanced Multiple Speakers' Separation and Identification for VOIP Applications Using Deep Learning" Applied Sciences 13, no. 7: 4261. https://doi.org/10.3390/app13074261

11. Kafke, John, and Thiago Viana. 2022. "Call Me Maybe: Using Dynamic Protocol Switching to Mitigate Denial-of-Service Attacks on VoIP Systems" Network 2, no. 4: 545-567. https://doi.org/10.3390/network2040032
12. Ormazabal, G.; Sarvesh, N.; Eilon, Y.; Henning, S. Secure sip: A scalable prevention mechanism for dos attacks on sip based voip systems. In Proceedings of the International Conference on Principles, Systems and Applications of IP Telecommunications, Berlin/Heidelberg, Germany, 1–2 July 2008; pp. 107–132.
13. Cauteruccio, F.; Cinelli, L.; Corradini, E.; Terracina, G.; Ursino, D.; Virgili, L.; Savaglio, C.; Liotta, A.; Fortino, G. A framework for anomaly detection and classification in Multiple IoT scenarios. Future Gener. Comput. Syst. 2021, 114, 322–335.
14. Pasichnyk R. and Sachenko A., "Semantic WEB-Search Developing by Problem-Oriented Ontology Means," 2007 4th IEEE Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, Dortmund, Germany, 2007, pp. 445-448, doi: 10.1109/IDAACS.2007.4488457.
15. Dyvak Mykola, Melnyk Andriy, Rot Artur, Hernes Marcin, Pukas Andriy, "Ontology of mathematical modelling based on interval data", Complexity, vol. 2022, Article ID 8062969, 24 p., 2022
16. Tas, I.M.; Unsalver, B.G.; Baktir, S. A Novel SIP Based Distributed Reflection Denial-of-Service Attack and an Effective Defense Mechanism. IEEE Access 2020, 8, 112574–112584.
17. Kovbasistyi, A. Melnyk, M. Dyvak, V. Brych and I. Spivak, "Method for detection of non-relevant and wrong information based on content analysis of web resources," 2017 XIIIth International Conference on 18.
18. Pasichnyk N.R. The method of forming ontological content based on the analysis of noisy, poorly structured information on specialized websites, Inductive modeling of complex systems. Collection of scientific works, Iss. No. 4, 2012, pp.158-168.
19. Pasichnyk N.R. The method of forming ontological content based on the analysis of information on specialized websites, Bulletin of the Khmelnytskyi National University: Technical Sciences, 2012. — Vol. No. 5. — P. 241-244.
20. Korel, Lukáš, Uladzislau Yorsh, Alexander S. Behr, Norbert Kockmann, and Martin Holeňa. 2023. "Text-to-Ontology Mapping via Natural Language Processing with Application to Search for Relevant Ontologies in Catalysis" *Computers* 12, no. 1: 14. https://doi.org/10.3390/computers12010014
21. Memariani, A.; Glauer, M.; Neuhaus, F.; Mossakowski, T.; Hatings, J. Automated and Explainable Ontology Extension Based on Deep Learning: A Case Study in the Chemical Domain. In Proceedings of the 3rd International Workshop on Data Meets Applied Ontologies, Hersonissos, Greece, 29 May–2 June 2021; pp. 1–16.

Надійшла до редакції: 17.04.2024

РОМАНЕЦЬ ІГОР ЄВГЕНОВИЧ – старший викладач, Західноукраїнський національний університет, 46009, Львівська, 11, м. Тернопіль, Україна, ***e-mail: i.romanets@wunu.edu.ua***

I.Eu. ROMANETS

ONTOLOGICAL APPROACH IN THE USING SECURITY SYSTEM IP TELEPHONY

West Ukrainian National University