УДК 004.056

Сухарев Нікіта
Куперштейн Леонід

ЗАСІБ ДЛЯ РОЗВІДКИ З ВІДКРИТИХ ДЖЕРЕЛ НА ОСНОВІ ШІ-АГЕНТІВ

Вінницький національний технічний університет

Анотація

Описано сучасні методи розвідки з відкритих джерел та обмеження традиційних підходів. Запропоновано мультиагентну архітектуру на базі фреймворку CrewAI із п'ятьма агентами на основі великої мовної моделі.

Ключові слова: мультиагентна система, розвідка з відкритих джерел, велика мовна модель, crewAI, кібербезпека.

Abstract

Modern open-source intelligence methods and limitations of traditional approaches are described. A multi-agent architecture based on the CrewAI framework with five agents based on a large language model is proposed.

Keywords: multi-agent system, open-source intelligence, large language model, CrewAI, cybersecurity

Вступ

У сучасних інформаційних системах обсяги публічно доступних даних стрімко зростають: соціальні мережі, публічні реєстри, корпоративні веб-сайти та спеціалізовані форуми генерують великі масиви інформації, що мають ключове значення для виявлення та аналізу кіберзагроз. Завдяки розвідці з відкритих джерел фахівці з інформаційної безпеки можуть безконтактно отримувати дані про технічні інфраструктури, організаційну структуру та потенційні вразливості об'єктів захисту.

Традиційні методи розвідки з відкритих джерел, що базуються на послідовному опитуванні різних інструментів та ручній обробці результатів, виявляють обмеження при роботі з великими обсягами даних і потребують значного часу на фільтрацію та структурування інформації. Зокрема, для критичних інформаційно-комунікаційних систем затримки у зборі й аналізі даних можуть призводити до зниження оперативності реагування на інциденти та підвищення ризиків компрометації.

Для подолання цих викликів у роботі запропоновано мультиагентну архітектуру на базі фреймворку CrewAI, яка розподіляє завдання між п'ятьма спеціалізованими агентами, кожен із яких відповідає за окремий аспект процесу розвідки з відкритих джерел, від координації запитів до узагальнення результатів. Такий підхід забезпечує гнучкість, масштабованість і прискорення обробки даних шляхом паралельного виконання типових операцій.

Результати дослідження

У сучасних умовах кібербезпеки розвідка з відкритих джерел або Open Source Intelligence (OSINT) стає невід'ємним інструментом для виявлення потенційних загроз. Цей підхід охоплює збір і аналіз даних із веб-сторінок, соціальних мереж, публічних реєстрів і спеціалізованих форумів, дозволяючи аналітикам безконтактно оцінювати стан інформаційної інфраструктури та виявляти індикатори компрометації . Крім того, OSINT дає змогу оперативно реагувати на інциденти, збираючи актуальні відомості про нові вразливості й тактики атак [1].

Розвиток OSINT супроводжується інтеграцією суміжних напрямів, а саме, TECHINT, SOCMINT, HUMINT і GEOINT, кожен із яких доповнює загальну картину дослідження. TECHINT зосереджується на технічних аспектах об'єкта, SOCMINT аналізує поведінкові патерни користувачів у соціальних мережах, HUMINT враховує людський фактор, а GEOINT забезпечує геопросторовий контекст дослідження. Такий комплексний підхід підвищує глибинність аналізу й дозволяє врахувати різні аспекти загроз [2].

Аналіз понад тридцяти класичних інструментів OSINT засвідчив їхню обмежену інтеграцію та необхідність ручної обробки результатів. Більшість утиліт призначені для однофункціонального використання і вимагають послідовного запуску та об'єднання даних вручну, що значно знижує

швидкість розвідки та збільшує трудозатрати аналітиків. При цьому відсутність єдиного механізму семантичного аналізу ускладнює фільтрацію релевантної інформації.

Великі мовні моделі (LLM), зокрема GPT-4o, демонструють високу здатність до семантичного аналізу й класифікації неструктурованих текстових даних. Завдяки вбудованим знанням та здатності розуміти контекст LLM можуть автоматизувати підготовчі етапи OSINT, від попередньої фільтрації шуму до узагальнення ключових фактів. Поєднання LLM із мультиагентною архітектурою відкриває можливості для побудови систем, які одночасно виконують декілька завдань паралельно, що суттєво підвищує масштабованість розвідки [3].

У запропонованій архітектурі на базі фреймворку CrewAI [4] використано YAML-сценарії для керованої взаємодії між агентами. Система включає координуючий агент, який ініціалізує процес і розподіляє завдання, три профільних агенти, які досліджують різні джерела: домени, організації, людей, а також звітуючий агент. Така модель забезпечує модульність і гнучкість, а саме, додавання нових агентів або зміна порядку їх виклику здійснюється через редагування YAML-сценарію, без необхідності модифікації системного коду.

Загальний алгоритм роботи програмного засобу для розвідки з відкритих джерел представлений на рисунку 1. Алгоритм ілюструє основні компоненти системи, порядок їхньої взаємодії, а також напрямки передачі даних між агентами.

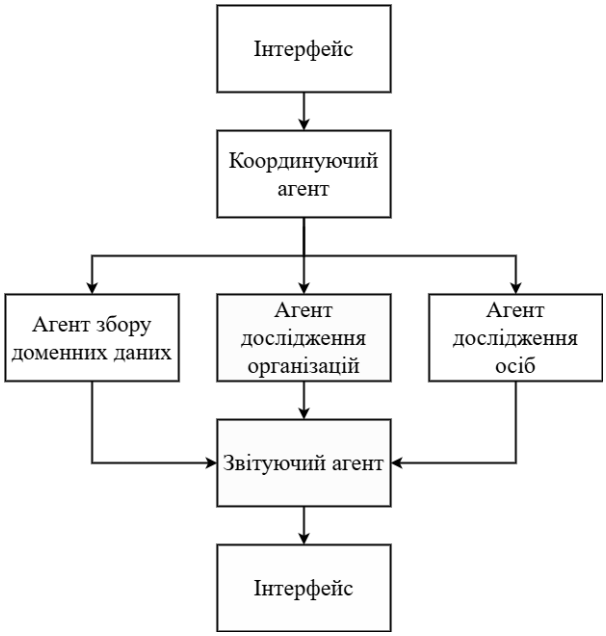

Рисунок 1 – Загальний алгоритм роботи програмного засобу для розвідки з відкритих джерел

Кожен агент мультиагентної системи є автономним модулем із чітко визначеною роллю та описом своєї функції, передісторією, задачами та очікуваним виходом у вигляді структурованого звіту, а за обробку даних відповідає вбудована LLM (наприклад, GPT-4o). Така схема дозволяє швидко додавати або переналаштовувати агентів через конфігурацію без змін у коді.

Ключовими перевагами засобу є висока масштабованість завдяки паралельній обробці, гнучкість у налаштуванні сценаріїв через і модульність, яка дозволяє оперативно впроваджувати нові аналітичні компоненти. Ці результати підтверджують доцільність застосування мультиагентних систем із LLM для автоматизації OSINT і закладають основу для розширення функціоналу у напрямі соціальної та геопросторової розвідки.

Висновки

Запропонована мультиагентна архітектура на базі CrewAI із п’ятьма ШІ-агентами, кожен із яких описаний через роль, опис, передісторію, задачу, очікуваний результат та інтегровану велику мовну модель, демонструє високу модульність, гнучкість і масштабованість процесів розвідки з відкритих джерел. Використання YAML-сценаріїв дозволяє змінювати послідовність та функціонал агентів без модифікації коду, що спрощує підтримку та подальший розвиток системи в умовах змінних вимог.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Розвідка з відкритих джерел (Open-source intelligence – OSINT) [Електронний ресурс]. – Режим доступу: https://www.maxzosim.com/rozvidka-z-vidkritikh-dzherel-osint/
2. Intelligence Collection Methods: A Comprehensive Overview [Електронний ресурс]. – Режим доступу: https://kector.com/intelligence-collection-methods/
3. What Is a Large Language Model (LLM)? [Електронний ресурс]. – Режим доступу: https://medium.com/@maximemalige/what-is-a-large-language-model-llm-7c4ce0d1c5b1
4. Multi AI Agent Systems with crewAI [Електронний ресурс]. – Режим доступу: https://www.deeplearning.ai/short-courses/multi-ai-agent-systems-with-crewai/

Куперштейн Леонід Михайлович − доцент кафедри захисту інформації, Вінницький національний технічний університет, м. Вінниця, Україна, email: kupershtein@vntu.edu.ua
Сухарев Нікіта Олександрович − студент групи 1БКС-21б, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, email: snikitasuharev@gmail.com

Kupershtein Leonid − associated professor of information protection department, Vinnytsia National Technical University, Vinnytsia, email: kupershtein@vntu.edu.ua
Sukharev Nikita − student of 1БКС-21б group, Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, Ukraine, email: snikitasuharev@gmail.com