НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
ІНСТИТУТ ПРОБЛЕМ РЕЄСТРАЦІЇ ІНФОРМАЦІЇ НАН УКРАЇНИ

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА БЕЗПЕКА

МАТЕРІАЛИ XXI МІЖНАРОДНОЇ НАУКОВО-ПРАКТИЧНОЇ КОНФЕРЕНЦІЇ

ВИПУСК 21

Київ – 2021

Рекомендовано до друку Вченою радою
Інституту проблем реєстрації інформації НАН України
(протокол № 12 від 28 грудня 2021 р.)

Інформаційні технології та безпека. Матеріали XXI Міжнародної науково-практичної конференції ІТБ-2021. – Київ: Інжиніринг. – 174 с. ISBN: 978-966-2344-84-4

До збірника увійшли матеріали доповідей, представлених на XXI Міжнародній науково-практичній конференції «Інформаційні технології та безпека» (ІТБ-2021, 9 грудня 2021 року, м. Київ, Україна).

У збірнику представлені матеріали, присвячені питанням безпеки та живучості критичних інфраструктур, комп'ютерного моделювання складних систем, технологій аналітики великих обсягів даних (Big Data), аналітичних систем на основі відкритих джерел інформації (OSINT), моделювання, аналізу та прогнозування процесів мережевої взаємодії, методів і засобів підтримки прийняття рішень.

Для фахівців в області інформаційних технологій, інформаційної і кібернетичної безпеки, а також для аспірантів і студентів старших курсів вищої школи відповідних спеціальностей.

ISBN 978-966-2344-84-4

РОЗПОДІЛЕНІ ІНТЕЛЕКТУАЛЬНІ АГЕНТИ ДОБУВАННЯ КОНТЕНТУ ІЗ СОЦІАЛЬНИХ МЕРЕЖ

А.М. Соболєв[1], Д.В. Ланде[1,2]
[1] Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", Київ, Україна
[2] Інститут проблем реєстрації інформації НАН України, Київ, Україна

dwlande@gmail.com

Запропоновано розподілені агенти добування контенту із соціальних мереж, які складається з декількох серверів, що проводять збір інформації і розміщені в різних дата центах, які контролюються єдиною інтелектуальною системою, яка забезпечує належний рівень відмовостійкості та достовірності отриманої інформації.

Останнім часом соціальні мережі стали зручним та ефективним засобом комунікації. Вони надають величезну свободу дій в інформаційному просторі, який є відкритим та доступним для всіх бажаючих. При ефективному їх використанні вони стають потужним інструментом для проведення OSINT (Open-Source Intelligence) та водночас дають можливість отримання стратегічно важливої, але загальнодоступної інформації у питаннях національної безпеки та критичної інформації, що дозволяє проводити оцінку настрою суспільства у визначеному інформаційному полі.

Популярні соціальні мережі зберігають та контролюють величезний об’єм даних про повсякденне життя та соціальні взаємодії людей і тому вони прискіпливо перевіряти кожен запит, що приходить до них та не дозволяють стороннім сервісам без їх згоди копіювати таку інформацію собі. У разі коли ці сервіси помічають нестандартну поведінку від клієнта, що робить запит до їх даних вони негайно заблоковують йому доступ та надалі прискіпливо перевіряють запити, що приходять з IP Address заблокованого клієнта. Ці дії провокують до використання певної технології у таких клієнтів, щоб забезпечити стандартний рівень поведінки, що притаманний звичайному користувачу їх сервісів.

Для забезпечення можливості одночасного процесу добування інформації з соціальних мереж без використання сторонніх платних сервісів та для контролю і управління такою системою з єдиного місця, запропоновано впровадити команди агентів, що дозволяють завантажувати дані та обмінюватись такою інформацією між собою та забезпечує цілісність отриманих даних і розподіляє навантаження між собою.

Основою для контролю, управління, синхронізації навантаження та сумісної роботи агентів використано документо-орієнтовану систему керування базами даних MongoDB, що дозволяє зберігати інформаціє про запуски агентів, список їх джерел та алгоритм їх поведінки. NoSQL база

даних MongoDB дозволяє з легкістю впроваджуватись у найбільш популярні операційні системи, невибаглива до ресурсів та витримує велике навантаження

Оскільки системи управління доступом в популярних соціальних мережах націлені на виявлення нестандартної поведінки користувачів, що тим самим призвело то створення ефективного алгоритму доступу агента до таких мереж, який базується на кількості часу проведеного в цих мережах, об'ємі інформації, що збирається за один запит та кількості інформація, що надається таким агентом. Також слід відмітити, що основною увагою таких сервісів є поведінка клієнтів в нічний час (із регіону з якого відбувається запит), оскільки в такий період кількість запитів від клієнтів має мінімізуватись а в іншому випадку такий клієнт вже є об'єктом для дослідження.

Для розподіленої взаємодії використано 3 сервери, що територіально розміщені в різних дата центрах та знаходяться на великих відстанях між собою:

1. Нідерланди;
2. Україна;
3. Сполучені Штати Америки.

На даних серверах Рис. 1 розгорнуто MongoDB кластер, інтелектуальну систему управління та команди агентів для добування інформації. Протоколом для керування та взаємодії між агентами використовується протокол HTTPS, оскільки він являється найбільш популярним в глобальній мережі Інтернет, дозволяє швидко оптимізувати команди для мережевих агентів та має належний рівень безпеки.

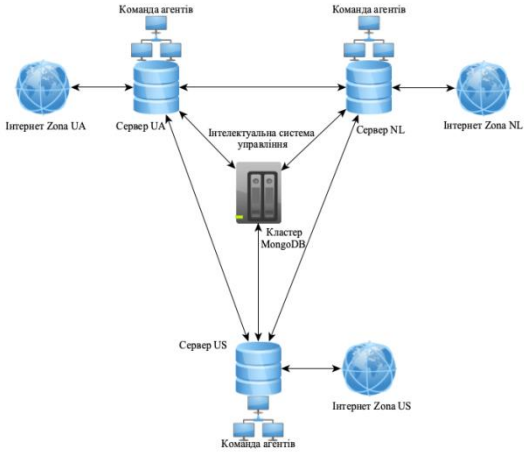

Рис. 1 – Схема розподіленого добування інформації на основі 3 серверів

Також дані команди агентів використовують в основі архітектуру RESTful сервісу, що дозволяє ефективно проводити налаштування та їх контроль роботи. Також, в основі їх взаємодії використовуються системні повідомлення типу Heartbeat, що дозволяють ефективно досліджувати життєвий цикл агентів і у разі відмови будь-якого з них швидко виявити проблему без втрати отриманих ним даних.

Запропоновані команди агентів ∇ являють собою кластер високої доступності, у якому при відмові одного агента його функції переймає інший доступний агент. Таким чином, процес добування інформації із соціальних мереж продовжує працювати без зупинки завдяки інтелектуальній системі контролю цих агентів. Щоб побудувати відмовостійку структуру, потрібно мінімум два фізичні сервери із системами зберігання даних, тому для забезпечення відмовостійкості у двух серверів використовується допоміжний третій сервер, який дозволяє ефективно використовувати ресурси та рівномірно розподіляти навантаження між двума серверами. Також, інтелектувальна система контролю за мережевими агентами працює за таким принципом: якщо один з агентів виходить з ладу, в роботу автоматично включається інший при цьому йде оповіщення про збій агента. Загальна логіка кластера агентів створюється лише на рівні програмних протоколів і дає можливість:

1. Керувати всіма мережевими агентами за допомогою одного інтелектуального модуля;
2. Додавати та вдосконалити програмні та апаратні ресурси, без зупинки системи та масштабних архітектурних перетворень;
3. Забезпечувати безперебійну роботу системи при виході з ладу одного або двух агентів;
4. Синхронізувати дані між кластерами агентів;
5. Ефективно розподіляти запити на кластери агентів.

Один із серверів, що входять до складу кластера, є центральним сервером кластера. Центральний сервер, крім обслуговування клієнтських з'єднань, управляє роботою всього кластера і зберігає при цьому реєстр кластера.

Під час встановлення з'єднання агент звертається до центрального сервера кластера. Центральний сервер, на основі аналізу статистики завантаженості агентів, спрямовує його до конкретного робочого процесу, який йому необхідно виконати.

Отже головним завданням кластера агентів є виключення простою системи та надання звітів, скільки агенти зібрали самі, а скільки взяли в інших агентів. В ідеалі будь-який інцидент, пов'язаний із зовнішнім втручанням або внутрішнім збоєм у роботі ресурсу, повинен дозволяти продовжувати роботу системи.

Висновки

На основі проведеного тестування та оцінки роботи даних мережевих агентів, що добувають контент із соціальних мереж та складаються з 3 серверів, можна зробити висновок, про ефективність запропонованої взаємодіє оскільки це надало системі належний рівень відмовостійкості та достовірності отриманої інформації і забезпечило рівномірність завантаження кластерів агентів. Також представлена система виконує задачі безпеки сервісу моніторингу, що дозволяє забезпечити цілісність отриманих даних, обходячи обмеження у зборі інформації, доступність даних для моніторингу і повноту отриманої інформації. У разі, якщо для якось країни інформація буде змінена, агенти при взаємодії між собою відобразлять ці зміни і збережуть всі копії отриманих даних.

ЗМІСТ

Національна академія наук України
Інститут проблем реєстрації інформації НАН України

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА БЕЗПЕКА

Матеріали XXI Міжнародної науково-практичної конференції

Випуск 21

Підп. до друку 28.12.2021. Формат 60x84 1/16. Папір офс. Гарнітура Times.
Спосіб друку – ризографія. Ум. друк. арк. 10,5. Обл.-вид. арк. 24,36. Наклад 100 пр.
Зам. № 15-200.

ТОВ "Інжиніринг" 978-966-2344-84-4