УДК 519.2:004.9

Д.В. Рудченко

Харківський національний університет радіоелектроніки, Харків

ВИЯВЛЕННЯ СПІЛЬНОТ ТА ЇХ ЛІДЕРІВ В СОЦІАЛЬНИХ МЕРЕЖАХ ДЛЯ ЗАБЕЗПЕЧЕННЯ БЕЗПЕКИ

У роботі приведено приклад вирішення маркетингових завдань та питань кібербезпеки відносно аналізу інформації, отриманої у соціальних мережах. Зазначено основні загрози, які виникають в соціальних мережах, наведені деякі методи захисту від вказаних загроз. Обґрунтовано, що важливе значення займає аналіз даних соціальних мереж. Описано розвідувальний цикл для збору і аналізу даних та основні методи аналізу даних. На основі аналізу тестової вибірки з соціальної мережі Вконтакті визначено лідерів серед користувачів, а так само серед спільнот.

Ключові слова: *аналіз соціальних мереж, пошук лідера, пошук спільноти, загрози, кібербезпека.*

Вступ

Соціальні мережі як один з найважливіших класів мереж давно привертає до себе увагу. Широке поширення інтернету породило вибухове зростання доступних даних для аналізу, у тому числі мережевих. Аналіз подібних даних може допомогти при виявленні лідера [1]. Залежно від масштабу доступних даних виявлення може здійснюватися як в межах однієї соціальної групи, так і на глобальному рівні. Очевидним застосуванням є рішення маркетингових питань або пошук лідера терористичної спільноти. Необхідність аналізу соціальних мереж з кожним роком зростає, адже інтернет став середовищем, що впливає на поведінку багатьох мільйонів користувачів, як дорослих, так і підлітків і дітей, на їх дії, знання і думки, установки і ціннісні орієнтири, поведінкові навички. Таке різноманіття методів використання та користувачів несе багато загроз.

Метою роботи є збір та проведення аналізу даних соціальної мережі, побудова взаємозв'язків між користувачами, по їх друзям і групам, а також визначення центральних елементів серед цих зв'язків.

Загрози в соціальних мережах

У наші дні в соціальних мережах існує цілий ряд різних загроз. Наведемо основні загрози кібербезпеки організації чи особистості з соціальних мереж.

1. Спам в соціальних мережах. Спам є однією з найбільш класичних атак. Зловмисниками генеруються фіктивні облікові записи, і тисячі запитів друзів автоматично відправляються в надії, що хтось їх прийме. Після прийняття запитів жертвами атакуючий може почати відправляти спам-повідомлення. Крім того, у зловмисників часто є можливість використовувати кілька облікових записів паралельно, поки кожна з них не блокується денним лімітом.

2. Загрози соціальної інженерії. Найпопулярніша тактика для кіберзлочинців. Соціальні мережі дозволяють зловмисникам знаходити конфіденційну інформацію, яка може бути використана для майнової та моральної шкоди.

3. Розміщення приманок в соціальних мережах. Ідея проста: використовувати ключові слова і посилання таким чином, щоб спам-повідомлення отримували кращий список облікових записів. Зловмисники шукають нові повідомлення, що містять гарячі ключові слова. Потім шахрай приймає це повідомлення, замінює вихідний скорочений URL своїм власним посиланням, яке вказує на шкідливий сайт, і повторно читає повідомлення.

4. Уособлення друзів. Повідомлення, здається, виходить від облікового запису одного, люди схильні довіряти їй. Ці повідомлення про оновлення часто містять посилання на інші шкідливі сайти, щоб отримати більше паролів до облікових записів.

5. Крадіжка паролів і фішинг. Після успішної крадіжки аутентифікаційних даних стане можливим відправляти рекламу, деяку інформацію від імені інших або мотивувати одержувачів на будь-які негативні дії, зокрема, на передачу посилань на шкідливий сайт або запускати шкідливий код і виконувати інші (часто нелегальні) дії.

Методи моніторингу та аналізу соціальних мереж є більш ефективними з методів захисту в довгостроковій перспективі, але вимагають участі фахівців з різних галузей науки.

Оскільки віртуальні соціальні групи мають здатність до реорганізації, основним завданням моніторингу та аналізу віртуальних спільнот, що представляють загрозу для національної безпеки інформації, є не їх знищення, а управління і контроль їх діяльністю за допомогою різноманітних методів.

Розвідувальний цикл

У сучасному суспільстві Інтернет і соціальні мережі відіграють дуже важливу роль, одна з яких є вплив звичайних людей, користувачів соціальних мереж на міжнародну політику та соціальні процеси. Для забезпечення кібербезпеки в розвідувальних, контррозвідувальних та правоохоронних заходах служби безпеки країн використовують системи моніторингу та аналізу даних Інтернет, та соціальних мереж. При аналізі соціальних мереж використовують розвідувальний цикл:

1. Постановка завдання – необхідність збору всієї інформації про людину, групи і т.ін.

2. Планування – визначення джерел, план пошуку і аналізу даних.

3. Збір даних – формування запитів до джерел і збереження результатів для поточного і наступного порівняння та аналізу.

4. Структурування та обробка даних – витяг метаданих з документів.

5. Аналіз даних – обробка первинних даних, обчислення базових показників, виявлення статистичних та структурних закономірностей.

6. Підготовка звіту та презентація результатів – візуальне представлення результатів аналізу та відповідний опис.

Веб-інтерфейси соціальних мереж є джерелами даних реального часу і призначені для перегляду і взаємодії зі сторінками соціальної мережі в веббраузері або для використання даних користувачів спеціалізованими додатками. Оскільки сценарії використання інтерфейсів соціальних мереж не передбачають автоматичного збору даних безлічі користувачів з метою побудови соціального графа, то виникає ряд проблем:

– приватність даних – часто доступ до даних користувачів дозволений тільки для зареєстрованих і авторизованих учасників мережі, що вимагає підтримки емуляції користувальницької сесії за допомогою спеціальних облікових записів (акаунтів);

– слабка структурованість даних – у багатьох випадках програмні інтерфейси (API) соціальних мереж мають обмежений функціонал, що вимагає підтримки отримання за допомогою призначеного для користувача веб-інтерфейсу статичних копій HTML-сторінок, коректної обробки їх динамічної частини (включаючи виконання асинхронних запитів до сервера соціальної мережі), вилучення потрібних даних за допомогою алгоритму і / або шаблону і побудови їх структурованого уявлення, зручного для подальшої автоматичної обробки;

– обмеження доступу і блокування – з метою запобігання несанкціонованому автоматичному збору даних і обмеженню навантаження на інфраструктуру сервісу соціальної мережі власники сервісів часто вводять явні чи приховані обмеження на допустиму кількість запитів від одного користувача акаунта і / або IP-адреси в одиницю часу, що вимагає врахування кількості запитів, що посилаються, а також підтримки динамічної ротації використовуваних для збору даних користувача для аккаунтів і IP-адрес;

– розмірність даних обумовлює необхідність в паралельному методі збору даних, а також в методах отримання репрезентативної вибірки користувачів соціальної мережі (семплірування).

Виявлення спільнот в мережі

Спільноти в мережі характеризуються наявністю великої кількості зв'язків між їх учасниками і значно меншою кількістю зв'язків з іншими членами мережі. Спільнота може відповідати групам вебсторінок, які мають схожі теми [1], групи пов'язаних осіб в соціальних мережах [2] і т.ін. Найпростішим випадком спільноти є така, де кожен учасник пов'язаний з кожним, а інші члени мережі не спілкуються з членами спільноти (кліка). Виявлення спільнот (явних і неявних) є важливим завданням аналізу мереж, що включає в себе класифікацію членів спільноти, і, як результат, ідентифікацію однорідних груп, груп лідерів або експертів [3–6].

Виявлення спільноти в багатьох випадках є задачею кластеризації. Підходи до розподілу цільових груп шляхом виявлення спільнот дозволяють побудувати математичні моделі, а потім використовувати моделі інформаційного впливу та управління [7]. У той же час аналіз мереж досліджує структуру відносин між учасниками в різних областях, і виявляє неявні зв'язки між ними з використанням теорії графів [8]. Більш детальний огляд методів виявлення спільноти можна знайти в [9].

Виявлення лідерів в спільнотах

Пошук лідерів в співтоваристві є важливим завданням ACC, оскільки в дослідженні і моделюванні інформаційного впливу важливо мати дані про характер взаємодій членів спільноти, зв'язку між ними і законами розподілу інформаційних потоків. Згідно [10], деякий учасник є лідером, якщо після вчинення певної дії значне число інших повторює одну і ту ж дію в заданий інтервал часу. Завдання виявлення лідерів широко поширені в багатьох областях. Наприклад, в [11] «гіпотеза впливових членів» розглядається в зв'язку з маркетинговими завданнями; вибір багатьох осіб для пропозиції будь-якого продукту або інновацій [12]; поширення і максимізація впливу в конкурентних соціальних мережах і залучення послідовників, вірусний маркетинг [10]; поширення соціального впливу [13; 14] та ін.

Аналіз даних

Аналіз інформації соціальних мереж має широкий спектр варіацій. Дана робота охоплює лише деякі з них. В ході аналізу соціальних мереж засто-

совувалися пари (людина-частота появи, група-частота появи), сортування (для відсіювання найменш важливих елементів), багатопоточність (для прискорення обробки великих обсягів даних) і деякі інші прийоми.

У зв'язку з постійною необхідністю отримання великих наборів даних із соціальних мереж, була розроблена програма для збору даних, яка підтримує завантаження даних про друзів, групи і записи зі стіни користувача з соціальній мережі Вконтакті. Крім того, програма підтримує багатопотокове скачування. Для оцінки продуктивності програми були проведені експерименти, в яких скачували профілі користувачів соціальної мережі Вконтакті. Були досягнуті показники у 2000 акаунтів на годину.

Збір даних здійснювався з соціальної мережі ВКонтакті. Для збору використовувалися методи VK API для розробників додатків (https://vk.com/dev/methods). Збір інформації охоплює всіх користувачів, але не всі спільноти. При зборі даних з профілів користувачів, а також скачуванні графа дружби, краулер попередньо отримує список ідентифікаторів всіх користувачів з каталогу (https://vk.com/catalog.php).

Для збору профілів користувачів використовуються методи API users.get і groups.getById. Методи приймають на вхід списки ідентифікаторів користувачів або спільнот і повертають списки їх профілів у форматі JSON. Для збору інформації про друзів і групи використовуються методи API friends.get і groups.get, а для інформації про стіни – wall.get. Методи приймають ідентифікатор одного користувача і повертають списки ідентифікаторів його друзів, груп та повну інформацію про кожний пост на його стіні (текст поста, кількість лайків, репостів, дата публікації, номер поста на стіні, ідентифікатор людини чи спільноти, якщо даний пост є Фортеця, а також багато іншої інформації). Всі розроблені методи збору даних використовують версію API 5.63.

Для аналізу даних соціальних мереж було розроблено чотири режими роботи (рис. 1–4):

1. Виявлення 20-ти найбільш популярних користувачів (знаходження більшого числа зв'язків користувача серед дерева соціальної групи).

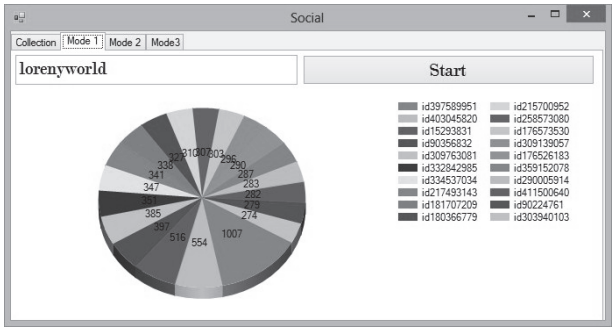

Рис. 1. Результати роботи першого режиму

2. Знаходження 20-ти найбільш популярних груп серед вищепредставлених користувачів (групи, на які підписано найбільшу кількість користувачів з усього спектра зібраних).

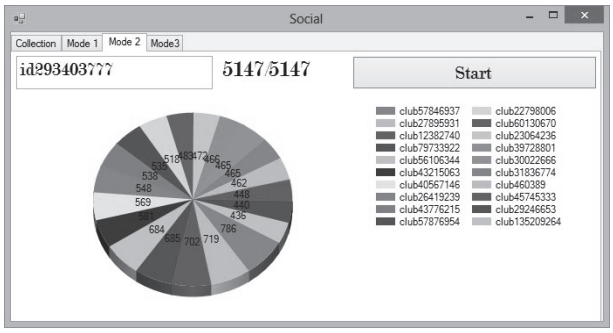

Рис. 2. Результати роботи другого режиму

3. Відображення найбільш популярних груп у 3-х користувачів (групи, на які підписані більш 1 людини з трьох представлених).

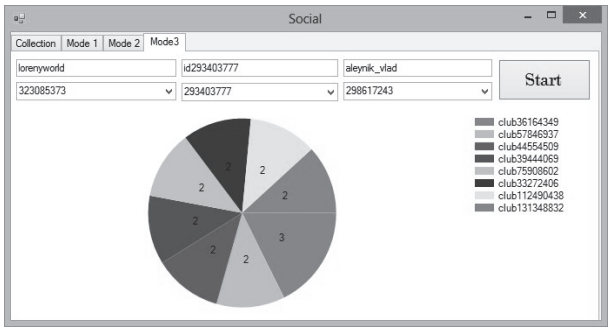

Рис. 3. Результати роботи третього режиму

4. Побудова графового дерева для конкретного користувача.

Соціальні мережі мають деревоподібну структуру, яку можна представити у вигляді графів. Коренева вершина являє собою акаунт користувача, з якого було розпочато збір даних. На першому рівні графових вершин знаходяться акаунти друзів даного користувача. На другому рівні знаходяться акаунти друзів їх друзів. Дана структура є основною для всіх соціальних мереж.

Рис. 4. Результати роботи четвертого режиму

Висновки

В даній роботі обгрунтовано важливість аналізу даних соціальних мереж. Описано розвідувальний цикл для збору і аналізу даних. Також описані основні методи аналізу даних, такі як виявлення спільнот в мережі, виявлення лідерів в спільнотах і інші. Для проведення аналізу даних з соціальної мережі Вконтакті було створено програмне забезпечення, яке підтримує завантаження даних про друзів, групи и записи зі стіни користувача. було проведено аналіз соціальної мережі Вконтакті. Збір даних тестової вибірки користувачів соціальної мережі включає в себе: айді, їх друзів та спільноти, на які вони підписані. На основі зібраних даних був проведений пошук лідера шляхом вибору найбільш популярного акаунта користувача серед друзів та друзів їх друзів. Також був проведений пошук лідера серед спільнот шляхом виявлення, на яку спільноту підписана найбільша кількість користувачів с тестової вибірки. Також в роботі було побудовано графове дерево користувача соціальної мережі, його друзів та друзів його друзів. Таким чином результати даної роботи можна застосовувати при аналізі соціальних мереж для вирішення маркетингових питань, питань кібербезпеки та інших.

Список літератури

1. Flake G.W. Self-organization and identification of Web communities / G.W .Flake, S. Lawrence, C.L. Giles, F.M. Coetzee // Computer. – 2002. – Iss. 3. – P. 66-70.

2. Girvan M. Community structure in social and biological networks / M. Girvan, M.E. Newman // Proceedings of the National Academy of Sciences of the United States of America. – 2002. – Iss. 12. – P. 7821-7826.

3. Бузун Н. Выявление пересекающихся сообществ в социальных сетях / Н. Бузун, А. Коршунов. – Москва, 2012. – 18 с.

4. Coscia M. A classification for community discovery methods in complex networks / M. Coscia, F. Giannotti, D. Pedreschi // Statistical Analysis and Data Mining. – 2011. – P. 512-546.

5. Коломейченко М.И. Алгоритм обнаружения сообществ в социальных сетях / М.И. Коломейченко, А.А. Чеповский, А.М. Чеповский // Фундаментальная и прикладная математика. – 2014. – Вып. 19(1). – С. 21-32.

6. Mona Jalal. A Survey on Community Mining in Social Networks [Electronic resourse] / Mona Jalal, AnHay Doan // Electronic data. – [Github, 2017]. Mode of access: http://monajalal.github.io/assets/pdf/CS784_report.pdf . Accessed 07 March 2017.

7. Губанов Д.А. Социальные сети: модели информационного воздействия, управления и противоборства: монография / Д.А. Губанов, Д.А. Новиков, А.Г. Чартишвили. – М., 2010. – 228 с.

8. Ehrlich K. Inside Social Network Analysis / K. Ehrlich, I. Carboni // IBM Watson Research Center. New York, USA, Technical Report, 2005. – P. 5-10.

9. Fortunato S. Community detection in graphs / S. Fortunato // Physics Reports. – 2010. – Vol. 486, Iss. 3–5. – P. 75-174.

10. Goyal A. Discovering leaders from community actions. / A. Goyal, F. Bonchi, Laks V.S. Lakshmanan // Proceedings of the 17th ACM Conference on Information and Knowledge Management, Napa Valley, California, USA, 2008. – P. 499-508.

11. Watts D.J. Influentials, Networks, and Public Opinion Formation / D.J. Watts, P.S. Dodds // Journal of consumer research. – 2007. – Iss. 4. – P. 441-458.

12. Kempe D. Maximizing the spread of influence through a social network / D. Kempe, J. Kleinberg, É. Tardos // Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, Washington, USA, 2003. – P. 137-146.

13. Dodds P.S. A generalized model of social and biological contagion. / P.S. Dodds, D.J. Watts // Journal of theoretical biology. – 2005. – Iss. 4. – P. 587-604.

14. Слабченко О.О. Методы и алгоритмы выявления сообществ потенциальных абитуриентов и их лидеров в социальных сетях. / О.О. Слабченко, В.Н. Сидоренко, Р.А. Пономарчук // Бюллетень Национального Кременчугского университета, 2013. – Вып. 1 (78). – С. 53-61.

Надійшла до редколегії 12.05.2017

Рецензент: д-р техн. наук проф. Л.О. Кіріченко, Харківський національний університет радіоелектроніки, Харків.

ВЫЯВЛЕНИЕ СООБЩЕСТВ И ИХ ЛИДЕРОВ В СОЦИАЛЬНЫХ СЕТЯХ ДЛЯ ОБЕСПЕЧЕНИЯ БЕЗОПАСНОСТИ

Д.В. Рудченко

В работе приведен пример решения маркетинговых задач и вопросов кибербезопасности относительно анализа информации полученной в социальных сетях. Указаны основные угрозы, возникающие в социальных сетях, приведены некоторые методы защиты от указанных угроз. Обосновано, что важное значение занимает анализ данных социальных сетей. Описаны разведывательный цикл для сбора и анализа данных и основные методы анализа данных. На основе анализа тестовой выборки из социальной сети Вконтакте определено лидеров среди пользователей, а так же среди сообществ.

Ключевые слова: *анализ социальных сетей, поиск лидера, поиск сообщества, угрозы, кибербезопасность.*

COMMUNITIES AND THEIR LEADERS DETECTION IN SOCIAL NETWORKS FOR SECURITY

D. Rudchenko

An example of solving marketing tasks and questions of cybersecurity regarding the analysis of information received in social networks is given in the article. The main threats that arise in social networks are indicated, some methods of protection from these threats are given. It is justified that the analysis of social networking data is important. A reconnaissance cycle is described for the collection and analysis of data and basic methods for analyzing data. Based on the analysis of the test sample from the social network Vkontakte identified leaders among users, as well as among the communities.

Keywords: *social network analysis, leader search, community search, threats, cybersecurity.*