УДК 004.67 (043.2)

Польгуль Т.Д., Яровий А.А.
Вінницький національний технічний університет

ВИЗНАЧЕННЯ ШАХРАЙСЬКИХ ОПЕРАЦІЙ ПРИ ВСТАНОВЛЕННІ МОБІЛЬНИХ ДОДАТКІВ З ВИКОРИСТАННЯМ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

Проблема визначення шахрайських інсталяцій є актуальною, оскільки розробники мобільних додатків витрачають великі кошти на компанії, які у свою чергу зобов'язані здійснити вказану кількість інсталяцій програмного додатку. Проте, багато з цих компаній застосовують шахрайські способи. У кінцевому результаті замовник констатує, що компанія здійснила замовлену кількість встановлень програмного додатку, а в кінцевому результаті, через невеликий проміжок часу заявленої кількості користувачів немає.

Метою даного дослідження є розробка моделі визначення шахрайських інсталяцій мобільних додатків з використанням інтелектуального аналізу даних.

Серед найбільш поширених шахрайських способів встановлення програмних додатків можна виділити: кліковий спам (Click Spamming), мобільне викрадення (Mobile Hijacking), ферми дій (Action Farms).

Відповідно, до сучасних та найбільш очевидних контрзаходів, які вже стали своєрідним стандартом у цій галузі, можна віднести: IP-фільтрацію, блокування видавця, зовнішню фільтрацію натиснень, виявлення стрибків при кліках або запитах інсталяції. Також, існують методи, які визначають співвідношення населення по геолокації, використовують аналіз дельти часу між подіями (наприклад, такі відомі фірми як Adjust та Kochava), аналізують показники продуктивності для визначення шахрайства.

Зважаючи на вищевказані шахрайські способи інсталяції мобільних додатків та аналізуючи дані власного мобільного додатку, можна зробити висновок, що події, які відбуваються шахрайським способом, мають спільні ознаки. Використовуючи методи кластеризації, залучених шахрайським способом користувачів можна віднести до одного кластеру, при цьому правильно визначивши ознаки, за якими здійснюється кластеризація.

Проаналізувавши відомі методи кластеризації та класифікації, у даній роботі запропоновано математичну модель визначення подібних користувачів. Розроблена модель базується на модифікованому методі колаборативної фільтрації і розв'язує багатокритеріальну задачу визначення подібних користувачів. Для вирішення задачі кластеризації та класифікації вхідних даних використовується комбінована метрика схожості, яка формується на основі коефіцієнта косинусної схожості між двома векторами (1) та коефіцієнта Танімото (2).

Подібність користувачів визначається в модулі визначення схожості користувачів інтелектуальної системи за допомогою коефіцієнта кореляції Пірсона.

$$k = \cos(a,\ b) = \frac{(a \cdot b)}{|a| \cdot |b|} \tag{1}$$

де a, b – вектори, елементами яких є множини частот появи окремих дій у заданому наборі інформації.

$$k = T(A,\ B) = \frac{N_c}{N_a + N_b - N_c} \tag{2}$$

де Na – кількість елементів в певному наборі даних користувача A, Nb – кількість елементів у наборі даних користувача B, Nc – кількість елементів в їх перетині.

Отже, у даній роботі проаналізовано шахрайські способи встановлення мобільних додатків, запропоновано модель класифікації користувачів на основі модифікованого методу колаборативної фільтрації з метою визначення користувачів, створених при встановленні мобільних додатків шахрайськими способами та при органічних встановленнях додатку.