

МЕТОДИ ВИЯВЛЕННЯ АНОМАЛІЙ ПОВЕДІНКИ КОРИСТУВАЧА В ІНФОРМАЦІЙНИХ СИСТЕМАХ

В. О. Горбенко^{1, а}, В. М. Ткач¹

¹ Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»,
Фізико-технічний інститут

Анотація

Своєчасне виявлення аномальної поведінки користувачів у інформаційних системах має велику важливість у сучасних інформаційних системах, адже при виявленні аномалії у власника системи є можливість зреагувати на таку поведінку до моменту, коли системі буде завдано шкоди. В роботі проаналізовано можливі джерела інформації та методи виявлення аномальної поведінки користувачів інформаційних систем в мережі інтернет. Оглянуті методи було оцінено з аспекту їх можливості застосування до різних систем з мінімізацією вклада користувача. Отримані результати показують, що найефективнішим методом виявлення аномалій, котрий може бути перенесений на різні системи та мінімізує роботу користувача є системи засновані на машинному навчанні.

Ключові слова: інформаційна система, аналіз поведінки, користувач, аномалія

Вступ

З розповсюдженням інформаційних систем та переміщенням багатьох систем обслуговування та надання послуг до мережі інтернет, значного розповсюдження здобули й атаки на такі сервіси. Зі всього широкого спектру атак на сервіси ми можемо виділити одну, спільну для всіх атак характеристику: нестандартну/аномальну поведінку користувачів або клієнтів системи. Так, наприклад, при роботі з системою онлайн банкінгу аномальним буде спроба зняття коштів з рахунку користувача у системах іншої країни чи деяка кількість спроб списання суми з рахунку. Для переважної більшості систем без відповідного захисту аномальна поведінка користувачів може виявлятися у частих запитах до певного ресурсу, що може за результатом привести до відмови до обслуговування, яка буде заважати звичайним користувачам системи та може використовувати ресурси системи не за основними цілями (наприклад виконувати код програм-майнерів). Відмова від обслуговування може призвести до тимчасових незручностей у користуванні системою та навіть до втрати спільноти користувачів системи. Прикладами для такого є нещодавня атака на сервіс github.com де за допомогою раніше виявленої в'язливості у серверах memcached було виконано DoS атаку за рахунок значного збільшення трафіку з багатьох клієнтів [1]. Боротьба з користувачами-зловмисниками також набула розвитку та на сьогоднішній день являє собою одну з галузей кібербезпеки. Серед багатьох сфер захисту та роботи у кібербезпеці виділяють превентивні дії та аналіз поведінки користувачів. Виявлення аномальної поведінки та аномальних дій користувачів передувє запобіганню кібератак на інформаційні системи

і є невід'ємною частиною процесу захисту самої системи, оскільки без методів та процесів виявлення та сповіщення не було б можливості відреагувати на зміни поведінки користувачів та системи до самого моменту атаки. Також серед систем аналізу поведінки користувачів існують наступні різновиди систем за направленістю моніторингу:

- Аналіз поведінки користувача (User and Entity Behavior Analytics) – це процес у кібербезпеці, направлений на виявлення загроз, направлених атак. Аналіз направлений на пошук та виділення шаблонів поведінки користувачів, використання алгоритмів та статистичного аналізу для виявлення аномалій, що сигналізують про потенційні загрози.
- Управління потоками інформації про безпеку та події (Security information and event management). Системи та технології SIEM направленої на аналіз подій та загроз в реальному часі. На практиці SIEM представляється додатками, приладами, послугами для агрегації даних з компонентів системи, сповіщення про перевищення порогових значень метрик, засобами візуалізації показників.

У роботі буде оглянуто методи автоматичного виявлення аномалій на основі даних про події в системі та дії користувача.

1. Джерела інформації про поведінку користувачів

Системи аналізу поведінки користувача аналізують дані, отримані з великої кількості джерел задля забезпечення себе максимальною кількістю інформації. Більшість існуючих систем тим чи іншим чином сканують системні повідомлення, відслідковують по-

^аvladimir4152@gmail.com

дії в системі та активність використання мережевого з'єднання. Для захисту веб-сервісу кількість отриманої про користувача інформації сильно обмежується через недоступність самого користувача та його знаходження у невідвласній мережі. У такому випадку можливо використати тільки інформацію з HTTP повідомлень [2, 3], тобто адресу на яку користувач відправляє запит, параметри запиту, зміст тіла запиту, зміст заголовків запиту.

2. Методи виявлення

2.1. Виділення ключових метрик та обмеження їх значень

Типовим способом регулювання дій користувача є обмеження кількості певних дій, що може виконати користувач. При відповідному налаштуванні, цей підхід дає можливість користувачу використовувати необхідні ресурси та не зтікати з обмеженнями, а системі – знати про активність користувача та перевищення нормального значення кількості певної дії і відповідно реагувати на це. Такий метод є простим у реалізації, проте він не має гнучкості та потребує детального налаштування, важко адаптується до змін в системі чи навантаженні. Налаштування системи потребує наявності кваліфікованого спеціалісту та наявності статистичних даних про моделі користування системою.

2.2. Кореляційний аналіз

Кореляційний аналіз – метод, що дозволяє виявити залежність між кількома випадковими величинами. Припустимо, проводиться незалежне вимірювання різних параметрів у одного типу об'єктів. З цих даних можна отримати якісно нову інформацію – про взаємозв'язок цих параметрів. Незважаючи на те, що величини носять випадковий характер, в загальному випадку спостерігається деяка залежність – величини корелюють. У випадку з аналізом поведінки користувача можемо провести аналіз кореляції різних показників активності. Ці величини не випадкові та регламентовані багатьма факторами. Наприклад, робочим розкладом та наявністю вихідного дня – для користувача корпоративної інформаційної системи чи користувача сервісу перегляду відео

2.3. Алгоритми машинного навчання

- K-Means
 - Простий у реалізації
 - Дозволяє кластеризувати дані та визначити найближчу групу
 - Використовує багато ресурсів пам'яті
 - Сильна залежність від функції порівняння (відстані)
- Нейронна мережа [4]
 - Відсутня необхідність перепрограмування під кожний конкретний випадок
 - Потребує навчання
 - Значний час роботи
 - Не прослідковуються принципи та причини відповіді
- Дерево рішень
 - Потрібна спеціальна підготовка даних
 - Можливо оброблювати нечислові дані
 - Процес прийняття рішення можливо відслідкувати
 - Малий час виконання
 - Не адаптується до систем, потребує перебудови
 - Використовує значні ресурси системи
- Алгоритм нечіткої кластеризації Fuzzy C-means (FCM)
 - Дозволяє класифікувати точку до декількох груп
 - Необхідно визначити кількість кластерів
 - Потрібно визначити порогове значення належності точки кластеру
 - Чутливість к початковому визначенні кількості кластерів

3. Отримані результати

В результаті проведеного дослідження методів виявлення користувачів було отримано інформацію про доцільність використання наявних алгоритмів та виявлено, що алгоритми машинного навчання та класифікації є найбільш перспективними, оскільки надають системі виявлення аномалій можливості адаптуватися до нових умов та нових систем без значного переналаштування. Також виявлено, що подібні алгоритми потребують навчання на даних системи, але це не є критичним, оскільки навчання та процес виявлення аномалій можуть бути одночасні, так як джерелом даних для системи є інформація з запитів до сервера, а для навчання можливо використати дані з існуючих записів про запити до системи.

Висновки

Для аналізу поведінки користувачів у інформаційних системах варто будувати моделі на основі класифікаторів та машинного навчання, оскільки такі алгоритми є достатньо гнучкими задля адаптації до нових систем та можуть бути відносно просто змінені власником системи при зміні характеристик системи чи типового користувача.

Перелік використаних джерел

1. Kottler Sam. February 28th DDoS Incident Report. — 2018. — Access mode: <https://githubengineering.com/ddos-incident-report/>.
2. Li J. Research of Analysis of User Behavior Based on Web Log. — 2013. — June. — P. 601-604.
3. User behavior analysis based on user interest by web log mining / X. Luo, J. Wang, Q. Shen et al. — 2017. — Nov. — P. 1-5.
4. Intrusion detection with autoencoder based deep learning machine / O. Kaynar, A. G. Yüksek, Y. Görmez, Y. E. Işık. — 2017. — May. — P. 1-4.