УДК 004

Кожухівський А. Д., професор
Ганусяк С. І., аспірант
Державний університет інформаційно-комунікаційних технологій

ПРОГНОЗУВАННЯ БОТНЕТ-АКТИВНОСТІ В ІНФОРМАЦІЙНІЙ СИСТЕМІ ПІДПРИЄМСТВА ЗА ДОПОМОГОЮ РЕГРЕСІЙНИХ МОДЕЛЕЙ

Ботнети – це розподілені мережі скомпрометованих пристроїв, які потрапили під контроль зловмисників і використовуються ними для виконання різноманітних шкідливих дій. До складу таких мереж можуть входити персональні комп'ютери, сервери, мобільні пристрої, а в умовах стрімкого розвитку Інтернету речей – навіть розумні телевізори, побутова техніка та системи відеоспостереження. Усі ці пристрої об'єднуються під централізованим або децентралізованим управлінням, що дозволяє зловмисникам координувати масовані кібератаки – зокрема, розповсюдження шкідливого програмного забезпечення, виконання DDoS-атак, розсилку спаму, крадіжку персональних даних або збір облікової інформації [3].

В умовах активної цифровізації та автоматизації виробничих процесів підприємств різного профілю зростає ризик того, що ботнет-активність спричинить значні економічні збитки. Зокрема, вона може призвести до втрати критично важливої або конфіденційної інформації, зупинки роботи окремих інформаційних систем, порушення логістичних процесів і навіть репутаційних втрат компанії. Особливо актуальною ця загроза є для великих корпорацій, банківських установ, промислових підприємств та органів державної влади.

З огляду на складність виявлення ботнетів у реальному часі та їхню здатність до швидкої зміни шаблонів поведінки, особливої актуальності набуває завдання прогнозування ботнет-активності. Ефективне прогнозування надає можливість виявляти загрозу ще до її реалізації – зменшуючи потенційні ризики для інформаційної безпеки [2].

Одним із найбільш перспективних підходів до вирішення цього завдання є застосування методів машинного навчання, зокрема регресійного аналізу, що дозволяє працювати з часовими рядами та знаходити аномальні закономірності в мережевому трафіку, які можуть вказувати на приховану ботнет-активність.

Регресійний аналіз – це метод статистичного моделювання, що полягає у визначенні залежності між змінною-ціллю (у цьому випадку – рівнем ботнет-активності) та незалежними змінними (ознаками, які впливають на неї) [1]. Серед таких ознак можуть бути:

- загальний обсяг вхідного і вихідного трафіку;
- кількість встановлених з'єднань за одиницю часу;
- частота звернень до нестандартних портів;
- тип використаного мережевого протоколу (TCP, UDP, ICMP тощо);
- географічне розташування джерела трафіку;
- повторюваність підозрілих шаблонів активності [2].

Регресійні моделі дозволяють не лише прогнозувати розвиток ситуації, але й виявляти приховані взаємозв'язки між параметрами, які людина може не побачити під час звичайного аналізу.

У сфері кібербезпеки ці моделі широко використовуються для:

- моделювання поведінки користувачів та пристроїв у мережі для виявлення аномалій;
- аналізу трафіку з метою прогнозування пікових навантажень;
- виявлення нетипових змін у поведінці системи, які можуть свідчити про приховану активність ботнетів;
- розрахунку ймовірності виникнення інцидентів у майбутньому [4].

Найчастіше використовуються наступні типи регресій:

- лінійна регресія – проста, але ефективна при стабільних і передбачуваних даних;
- поліноміальна регресія – підходить для моделювання складних, нелінійних взаємозв'язків;
- регресії з регуляризацією (Ridge, Lasso) – корисні для роботи з великими наборами ознак, де існує ризик перенавчання та мультиколінеарності [5].

Методика дослідження:

1. Збір даних. На першому етапі проводиться збір логів мережевого трафіку підприємства. Вони можуть бути отримані з систем захисту, міжмережевих екранів, роутерів або SIEM-систем. До даних зазвичай входять:

- IP-адреси джерела і призначення;
- часові мітки;
- обсяги переданих пакетів;
- використані порти та протоколи;
- кількість з'єднань та їхня тривалість;
- напрямок трафіку (вхідний або вихідний).

2. Попередня обробка. Отримані дані нормалізуються, видаляються пропуски, проводиться агрегація за часовими вікнами (наприклад, кожні 5 хвилин або 1 годину). Формуються часові ряди, а також проводиться маркування даних – визначення, яка активність є

типовою, а яка потенційно пов'язана з ботнет-діяльністю (на основі експертної оцінки або знань про інциденти в минулому).

3. Побудова моделей. У дослідженні використовуються:

- Лінійна регресія – як базовий інструмент для аналізу залежностей.
- Поліноміальна регресія – для моделювання випадків із різкими змінами навантаження.
- Ridge/Lasso-регресія – для побудови стійких моделей за умов високої варіативності.

4. Оцінка ефективності моделей. Для перевірки якості прогнозування використовуються стандартні метрики:

- MAE (середня абсолютна похибка) – показує середню різницю між прогнозом і фактичним значенням.
- RMSE (корінь середньоквадратичної похибки) – дає більше ваги великим відхиленням.
- R^2 (коефіцієнт детермінації) – демонструє, яка частка варіацій пояснюється моделлю [3].

Також проводиться тестування моделей на нових, раніше не бачених даних, щоб оцінити їхню здатність адаптуватися до змін у поведінці ботнетів.

Результати експериментів показали, що:

- лінійна регресія добре справляється в умовах стабільного трафіку, але має обмеження в умовах складних змін;
- поліноміальна регресія другого або третього ступеня краще моделює стрибкоподібні навантаження;
- регуляризовані моделі є найбільш гнучкими та стійкими до шуму й надлишкових змінних [3].

Загалом, інтеграція регресійного аналізу в системи кіберзахисту підприємств дозволяє вчасно виявляти відхилення в мережевій поведінці, що потенційно пов'язані з ботнет-мережами. Це, своєю чергою, підвищує рівень кіберстійкості та дозволяє діяти на випередження.

У майбутньому доцільним є розширення даного підходу шляхом використання глибоких нейронних мереж, зокрема LSTM (Long Short-Term Memory) або GRU (Gated Recurrent Unit). Ці моделі краще адаптовані для роботи з часовими рядами та дозволяють враховувати довгострокові залежності, контекст попередніх атак, сезонність трафіку та інші складні характеристики поведінки мереж.

Список використаних джерел:

1. Грайворонський М. В., Новіков О. М. Безпека інформаційно-комунікаційних систем. К.: Видавнича група BHV, 2009. 608с

2. Жилін А. В. Технології захисту інформації в інформаційно-телекомунікаційних системах: навч. посіб. ІСЗЗІ КПІ ім. Ігоря Сікорського. Київ: КПІ ім. Ігоря Сікорського, Вид-во «Політехніка», 2021. 213 с.

3. Захист від ботнетів. URL: https://www.eset.com/ua/support/information/entsyklopediya-zahroz/zakhyst-vid-botnetiv/?srsltid=AfmBOop5nxAw5V0fXFktltYE3qK47Cr-yXSHhIlTsDN42MTLrIt_Olsi

4. Остапов С. Е. Технології захисту інформації: навчальний посібник. Х.: Вид. ХНЕУ, 2013. 476 с

5. Ткаченко М. П., Кучеренко Є. А., Журавська І. М. Автоматизована система виявлення бот-атак на основі аналізу користувацьких профілів у соціальних мережах. Інформаційні технології та інженерія: тези доп. Всеукр. наук.-практ. конф. Миколаїв, 8–11 лютого 2022 р. Миколаїв: Чорном. нац. ун-т ім. Петра Могили, 2022. С. 28–31.