КІБЕРНЕТИКА

УДК 004.05

О. ЛЕТИЧЕВСЬКИЙ
Інститут кібернетики ім. В.М. Глушкова НАН України, Київ, Україна,
e-mail: *oleksandr.letychevskyi@litsoft.com.ua.*

Б. ПАНЧУК
Інститут кібернетики ім. В.М. Глушкова НАН України, Київ, Україна,
e-mail: *bogdanscloud@gmail.com.*

ПРОБЛЕМА ТОЧНОСТІ В СИСТЕМАХ ПРОТИДІЇ КІБЕРАТАКАМ ТА ВЕРИФІКАЦІЯ НЕЙРОННИХ МЕРЕЖ НА ПРИКЛАДІ ЗАДАЧІ ВИЯВЛЕННЯ БОТНЕТІВ

Анотація. Розглянуто проблему точності виявлення вторгнень у програмних системах на основі нейронних мереж глибокого навчання. Представлено приклад системи виявлення ботнетів — зловмисного програмного забезпечення, яке є джерелом потенційних атак, зокрема атаки типу «Distributed Denial of Service» або «відмова в сервісі» (DDoS). Систему створено, як модель класифікації, що виявляє поведінку ботнетів на заражених ресурсах. Проведено низку експериментів на відкритому наборі даних Канадського інституту кібербезпеки (Canadian Institute for Cybersecurity). Для підвищення точності класифікації застосовано спосіб розширення набору даних за допомогою генерації прикладів змагальних атак. Представлено метод верифікації надійності нейронної мережі з використанням автоматичного доведення властивості стійкості моделі на основі SMT-розв'язувачів. Для підвищення точності виявлення атак розглянуто також нейросимвольний підхід, який поєднує алгебраїчні методи з моделями класифікації.

Ключові слова: кібербезпека, ботнет, алгебраїчне моделювання, нейронна мережа глибокого навчання, змагальні атаки, верифікація.

ВСТУП

Виявлення вторгнень у мережеві та комп'ютерні середовища є однією з головних задач сучасної кібербезпеки. Для її розв'язання активно використовують технологію штучного інтелекту (ШІ), а саме нейронні мережі глибокого навчання.

Нейронну мережу-класифікатор навчають на основі наборів даних, які містять інформацію про мережеві дані або трафік та задіяні у ньому мережеві протоколи. Дані розділяють на такі, що представляють нормальне функціонування спостережуваної системи, і такі, що містять елементи атаки. Сучасні системи виявлення здатні не лише виявляти аномальну поведінку системи, а й класифікувати конкретний тип атаки. Проте під час створення цих систем виникають певні проблеми, які є результатом статистичної природи методів штучного інтелекту.

Насамперед це точність класифікації нейронної мережі, що цілком залежить від навчального набору даних, який є обмеженим і не може охопити абсолютно всі стани нормальної роботи системи. З іншого боку, якщо навчальна вибірка містить забагато однотипних даних, то може виникнути явище перенавчання. У цьому разі модель занадто сильно підлаштовується під повторювані дані у вибірці, втрачаючи здатність до генералізації. Також суттєвою проблемою є незбалансованість даних у навчальних наборах. Класифікація тих різно-