УДК 004.056.55

ЗАДАЧА РОЗПІЗНАВАННЯ ОБРАЗІВ В ІНТЕЛЕКТУАЛЬНИХ СИСТЕМАХ ВИЯВЛЕННЯ КІБЕРАТАК

Петренко Т.А., ст. викладач кафедри кібербезпеки та математичного моделювання,
Коротка Г.М., студ.гр. КБ-161
Чернігівський національний технологічний університет

Існує багато задач, які вирішують системи штучного інтелекту: задачі розпізнавання образів, розуміння тексту, створення експертних систем, доведення теорем, моделювання процесів і явищ, діагностики та постановки діагнозу. Проте, однією з основних є задача розпізнавання образів.

Розпізнавання образів (об'єктів, сигналів, явищ чи процесів) – задача, яку людині доводиться вирішувати практично кожну секунду. Для вирішення цієї задачі людина застосовує великі ресурси свого мозку, включаючи одночасно біля 10 – 12 млрд. нейронів. Саме це дає можливість людям миттєво впізнавати один одного, з великою швидкістю читати печатні та рукописні тексти, безпомилково водити автомобілі у складному вуличному трафіку, здійснювати відбраківку деталей на конвеєрі, розгадувати коди та інше.

Розпізнавання являє собою задачу перетворення вхідної інформації, в якості якої доречно розглядати деякі параметри, ознаки розпізнаваних образів, в вихідну, що представляє собою висновок про те, до якого класу належить розпізнаваний образ. Розпізнавання використовуються в різноманітних сферах життя людини: у медицині, металургії, банківській справі, інформаційних технологіях, біометрії, охоронній системі, тощо. У кожній галузі розпізнавання виконує певну роль. Такі системи набули широкого розповсюдження і в кібербезпеці під час розпізнавання загроз, аномалій та кібератак в інформаційних системах на основі аналізу даних що зчитуються з певних датчиків інформаційної системи (службового програмного забезпечення, операційних систем, апаратного забезпечення комп'ютерів та мереж, тощо). Крім того, методи розпізнавання застосування і в робототехніці, так як роботи повинні безпосередньо сприймати зовнішній світ, і, відповідно, мати пристрої машинного зору.

Застосування методів розпізнавання в інформаційних системах в першу чергу пов'язано з завданнями захисту інформації. Створюються системи розпізнавання кіберзагроз, які ідентифікують кібератаки по притаманним їм ознакам. Для створення інтелектуальної системи розпізнавання кібератак необхідно вирішити 8 основних задач, пов'язаних з розробкою інтелектуального захисту інформації.[1] Саме тому, метою цієї роботи є розгляд основних завдання, що виникають в процесі проектування і побудови інтелектуальних систем розпізнавання кібератак.

Завдання 1 полягає в тому, щоб визначити повний перелік ознак (параметрів), що характеризують кібератаки, аномалії та загрози для розпізнавання яких розробляється дана система. В якості первинних ознак можна використовувати параметри які зчитуються з певних програм, наприклад, кількість пакетів та час за який вони надходять до системи або експериментальні дані одержані в ході реалізації тестів на проникнення у комп'ютерну систему. В якості вторинних ознак для розпізнавання аномалій, загроз та кібератак можна використати різноманітні статистичні характеристики, наприклад дані моніторингу.

Завдання 2 полягає в проведенні первісної класифікації розпізнаваних об'єктів або явищ, в складанні апріорного алфавіту класів. Кібератаки поділяються на такі класи: DOS, R2L, U2R, PROBE. Denial of Service (DOS) – кібератаки відмови систем від обслуговування, яка характеризується генеруванням великого об'єму трафіку, що призводить до перевантаження і блокування серверу; Remote to User (R2L) – кібератаки, що характеризуються одержання доступу нелегітимним (незареєстрованим) користувачем несанкціонованого віддаленого доступу до інформації управління; User to Root (U2R) – кібератаки, що передбачають несанкціоноване розширення повноважень нелегітимних (незареєстрованих) користувачів до рівня локального суперкористувача (адміністратора); Probing (PROBE) – кібератаки сканування портів з метою одержання конфіденційної інформації.

Завдання 3 складається в розробці апріорного словника ознак. До словника ознак можна віднести: базові ознаки – ця група містить ознаки, які можна отримати з заголовка мережевого пакету; мережеві ознаки – ця група включає у себе ознаки, які можна порахувати по відношенню до часового вікна у 2 секунди; контентні ознаки – ця група містить ознаки, які можна отримати з вмісту пакетів, такі як спроба авторизації чи спроба створення файлу.

Завдання 4 складається в описі всіх класів апріорного алфавіту класів на мові ознак, включених в апріорний словник ознак. При створенні системи розпізнавання кібератак будуть враховуватися параметри мережевого трафіка бази даних KDD 1999 Kup Data. 41 параметр – максимальна кількість інформації, яку можна було б отримати з аналізу пакетів.

Завдання 5 полягає в розбитті апріорного простору ознак кібератак на області, відповідні класам апріорного алфавіту класів. На підставі вхідних параметрів оточення формуються множина можливих атак та відповідна їм множина можливих параметрів, згідно значень яких можна виявити аномальний стан,

породжений відповідним елементом з множини. Для виявлення аномального стану кожному типу атаки множини ставиться у відповідність підмножина параметрів з множини параметрів, згідно з яким можна виявити підозрілу активність в системі. Таким чином формується множина пар - “атака→параметри”, в якій кожній атаці буде відповідати набір параметрів.

Завдання 6 полягає у виборі алгоритмів розпізнавання, що забезпечують віднесення розпізнаваної кібератаки до того чи іншого класу або їх деякої сукупності. Алгоритми розпізнавання ґрунтуються на порівнянні тієї чи іншої міри близькості або міри схожості розпізнається об'єкта з кожним класом. В алгоритмах розпізнавання, що базуються на використанні детермінованих ознак, в якості міри близькості використовується середньоквадратичне відстань між даним об'єктом зі і сукупністю об'єктів, які являють собою клас. В алгоритмах розпізнавання, що базуються на використанні імовірнісних ознак, в якості міри близькості використовується ризик, пов'язаний з рішенням про приналежність розпізнаваного об'єкта до класу. В алгоритмах розпізнавання, що базуються на використанні логічних ознак, не використовується поняття «міра близькості». Коли побудовано опис класів на мові логічних ознак у вигляді відповідних співвідношень, при підстановці в ці співвідношення значень ознак, що характеризують розпізнаваний об'єкт, автоматично виникає відповідь: до якого класу або класів цей об'єкт може бути віднесений і до яких він не відноситься.[2]

Метод інтелектуального розпізнавання кіберзагроз з використанням логічних ознак дає можливість отримати результат навіть в ситуації, коли немає даних по функціях, що описують розподіл значень ознак кібератаки. В рамках методу запропоновані логічні процедури розпізнавання загроз.[2] Парадигмою побудови логічних процедур розпізнавання загроз є відшукання інформативних фрагментів описів об'єктів. Ці фрагменти при створенні, конкретних проектних рішень для системи розпізнавання, дозволять однозначно робити висновок про наявність (або відсутність) атаки (аномалії, загрози) в рамках класу. Вхідні дані - ознаки аномалій, атак і кіберзагроз.

Завдання 7 полягає у визначенні робочого алфавіту класів і робочого словника ознак системи розпізнавання. Інформативними вважаються ознаки, які відображають певні закономірності в описах об'єктів, використовуваних для навчання. У системі розпізнавання кіберзагроз інформативними вважаються такі фрагменти, які зустрічаються в описах об'єктів одного типу загроз, але не зустрічаються в описах об'єктів інших класів кіберзагроз.

Наприклад, при розпізнаванні DOS-атак системою розпізнавання будуть враховуватися наступні параметри:

Параметр	Описання
duration	тривалість з’єднання
protocol_type	тип протокола (tcp, udp, та ін.)
service	мережева служба отримувача(http, telnet)
flag	стан з’єднання
src_bytes	число байтів переданих від джерела до отримувача
dst_bytes	число байтів переданих від отримувача до джерела
land	1 якщо з’єднання по ідентичним портам, 0 в інших випадках
wrong_fragment	кількість невірних пакетів
urgent	кількість пакетів з флагом urg

Завдання 8 полягає у виборі показників ефективності системи розпізнавання та алгоритмів оцінки їх значень. Для кожного класу кількість ознак варіюється від 3 до 9. Інформативність ознаки змінюється в діапазоні від -1 до +1. Для оцінки ефективності процедур розпізнавання можна використовувати метод ковзного контролю [3]. Під час тестування розробленого методу інтелектуального розпізнавання загроз, в якості вхідних даних для навчання та тестування використовувалася база даних KDD Cup Data.

Таким чином, ми проаналізували основні завдання, які необхідно вирішити в процесі проектування і побудови інтелектуальних систем розпізнавання кіберзагроз. Без їх вирішення не можлива побудова ефективно-функціонуючих систем інтелектуального розпізнавання образів в галузі кібербезпеки.

Список використаних джерел

1. Горелик А.Л., Современное состояние проблемы распознавания: Некоторые аспекты / А.Л. Горелик, И.Б. Гуревич, В.А. Скрипкин. – М.: Радио и связь, 1985. – 160с.
2. Петренко Т.А., Лахно В.А., Григорян Г.С. Розробка адаптивної системи розпізнавання кіберзагроз / Безпека українського суспільства в концепції вступу в постіндустріальне суспільство ЄС: Наукові доповіді та тези учасників науково-практичної конференції (м. Київ, 16 грудня 2015 р.), К., 2015. С. 66–76.
3. Мірошник, М. А. Розробка методів оцінки ефективності захисту інформації в розподілених комп'ютерних системах [Текст] / М. А. Мірошник // Інформаційно-керуючі системи на залізничному транспорті: науково-технічний журнал. – 2015. – № 4 (113). – С. 39–43.