https://www.scilook.eu/index.php/slif/article/view/slif11-055
DOI: 10.30888/2415-7538.2018-11-01-055

DATA MINING AND MACHINE LEARNING TECHNIQUES TO DETECT INTRUSION INTO CYBERSECURITY OF ROBOTIC SYSTEMS

DATA MINING ТА МАШИННІ ТЕХНІКИ НАВЧАННЯ ДЛЯ ВИЯВЛЕННЯ ВТОРГНЕННЯ В КІБЕРБЕЗПЕКУ РОБОТОТЕХНІЧНИХ СИСТЕМ

Byrlakov V.M / Бурлаков В.М.,
c.t.s., as.prof / к.т.н., доцент,
Petrukhno I. R / Петрухно І.Р,
student / студент
Національний технічний університет України «КПІ ім. Ігоря Сікорського»
Україна, м. Київ,
National Technical University of Ukraine «Igor Sikorsky KPI»,
Ukraine. Kiev.

Анотація. *У статті запропонований метод паралельних обчислень на базі програмного засобу Hadoop для виявлення вторгнення в кібер безпеку робототехнічних систем.*

Ключові слова: *Кібербезпека, робототехнічні системи, Data Mining, Hadoop, машинні техніки навчання, виявлення вторгнень.*

Вступ

Практично по визначенню всі роботи оснащені здатністю сприймати, обробляти і записувати навколишній світ. Щоб забезпечити кращу продуктивність, вони постійно збирають інформацію. У цих умовах, якщо ці роботи скомпрометовані, виникає двовимірна проблема безпеки: по-перше, проблеми безпеки щодо віртуального боку робота (дані, повідомлення і т. Д.), А по-друге, проблеми, пов'язані з фізичною стороною, яка стосується як робота так і цілісності його системи. Сучасний стан являє собою термін «Cyber-physical security», який охоплює віртуальні та фізичні проблеми. Кіберфізичні атаки стикаються з кількома проблемами, з якими доводиться стикатися. З одного боку, проблеми безпеки охоплюють свідомість, пов'язану з фізичною недоторканністю особистості. Люди зазвичай турбуються про проблеми, які роботи можуть заподіяти людям або їх речам. Наприклад, є кілька комерційно доступних хірургічних роботів, як роботи Da Vinci, підключених до мереж зв'язку, що дозволяє дистанційним операціями з боку фахівців, що станеться, якщо ці роботи або їх комунікація будуть захоплені? Були претензії до зламаним військовим роботам, але навіть сервісні роботи в домашніх умовах створюють проблеми з безпекою, вони можуть завдати серйозної шкоди будинкам (підпал, наїзд на машини і т. д). З іншого боку, проблеми конфіденційності, пов'язаних з роботами, поширюються в багатьох областях. Існує широкий спектр сервісних роботів, які призначені для будинків та торгових приміщень. Вони можуть використовуватися в якості мобільних телеконференціонних платформ, вітальних помічників, віртуальних домашніх тварин, іграшок і т. д. Якби ці роботи були зламані, вони могли б надати багато особистої інформації про користувачів, що взаємодіють з роботом. Ця інформація може бути отримана із загальних даних (вік, розмір і т. Д.), Особисті фотографії, призначена для користувача інформація, економічна і т.

Д., Що відкриває новий етап кібербезпеки роботизованих систем.

Основна частина

Існують різні способи атаки робота та різних рівнів тяжкості вторгнень:

- "Stealth Attacks" може бути реалізовані різними способами. У цьому типі атаки хакери в основному намагаються змінити показання датчиків робота, щоб викликати помилку. Це може бути досягнуто шляхом зміни середовища або втручання в роботу датчиків. Деякі рішення могли бути запропоновані для виявлення подібних атак, наприклад, з використанням сукупної суми (CUSUM) для виявлення помилок в показаннях датчиків діапазону, які можуть спричинити зіткнення у мобільному роботі.
- "replay attack". Якщо хакери здатні перехопити зв'язок системи, вони можуть відтворювати захоплені пакети, навіть якщо вони зашифровані. Якщо протокол зв'язку не підготовлений до такого типу атаки, система розглядає ці повторно відтворені пакети як легітимні та помилково приймає рішення. Цей тип атаки також використовується як попередник, щоб виявити внутрішність системи, шукаючи нові слабкі сторони. Є й інші види атак на роботів, не пов'язаних з датчиками. Вони можуть бути орієнтовані на когнітивні елементи системи управління. Наприклад, у медичного робота, якщо система отримує неправдиву інформацію про стан хворого, робот може прийняти неправильні рішення, які потенційно можуть спричинити серйозні збитки. Цей тип «введення помилкових даних» також може бути використаний в інших типах роботів, наприклад, надання мобільних роботів помилкових карт для приведення зіткнення або фальшивої інформації робототехніці, що допомагає ввести користувачів в оману. З точки зору конфіденційності, "підслуховування" є однією з найбільш побоюваних потоків в комп'ютерних системах. Те саме стосується робототехнічних систем. Якщо робототехнічні системи обмінюються інформацією з іншими зовнішніми системами, це повідомлення може бути скомпрометованим та отримати приватну інформацію про користувачів.
- (DoS) - це ще один класичний тип атаки. DoS-атаки в робототехніці зазвичай означають, що робот перестає працювати, тому роботам не постраждали пошкодження, а роботи не пошкоджують людей чи їх оточення. Збитки виникають через відсутність сервісу, наданого роботом. Серйозність атаки залежить від критичності наданої послуги. Найгірший випадок виникає, коли робот не просто зупинено, а викрадено. Це відоме як кібербезпека як "Віддалений доступ". У цій ситуації роботів створюють проблеми з безпекою, а не лише конфіденційність. Класифікація суворості нападів є складною проблемою. Це дуже важко передбачити наслідки нападів. Навіть невелика втрата даних може мати катастрофічний вплив на репутацію компанії. Для шантажу магната може бути використаний єдиний приватний образ, який піддавався підслухуванню з домашнього робота. DoS-атаки, які тільки перешкоджають роботі виконувати свою роботу, можуть виглядати менш небезпечно, але вони дійсно важливі, наприклад, в теле-хірургічних робототехнічних системах.

Методології Data Mining

Зібрані історичні / журнальні дані в мережі вивчаються за алгоритмами класифікації для побудови прогнозної моделі для ідентифікації хакерів та нападників. У цьому розділі описуються найпопулярніші класифікатори, які також називаються учнями під наглядом, а саме імовірнісний алгоритм Nave Bayes (NB), дерева на основі C4.5 (J48) та інстанції на базі IB1 (на інстанції)

1. Наївні Байєс (NB): цей алгоритм класифікації використовує теорему Байєса, і особливості набору навчальних матеріалів, як показано в таблиці 1, вважаються незалежними від даного класу міток для побудови прогнозованої моделі. Цей класифікатор спирається на дискримінантну функцію, як показано в рівнянні.

$$f_i(X) = \prod_{j=1}^{N} P(x_j \mid c_i) P(c_i)$$

Нехай набір даних буде D з функціями X = (x1, x2, ... xN) і класу C з j лейблами Cj, j = 1, 2, ... N. Цей алгоритм обчислює умовні ймовірнісні значення P (xj | ci) та попередні імовірності P (ci) на заданому наборі тренінгів для побудови прогнозованої моделі. P (ci) обчислюються шляхом підрахунку даних, що містяться в класі Ci, розподіляє отриманий рахунок на підставі кількості навчальних даних. Той же спосіб слід застосовувати для обчислення імовірностей за допомогою спостережуваної частоти розподілу функцій в xj у межах навчального набору, який позначений. Подана вірогідність обчислюється на кожному класі для прогнозування невідомих відмічених даних

2. C4.5 (J48): Цей алгоритм використовує дерево рішень для побудови прогнозованої моделі. Дерево рішень будується за допомогою численних методів. Всі ці методи перетворюють даний набір даних у деревоподібну структуру. Вузли дерева являють собою функції, а ребра представляють зв'язок між функціями за значенням функцій, найнижчий рівень якого представляє відмітку класу. Рекурсивно значення функцій обчислюється за допомогою засобу отримання інформації або ентропії для перетворення набору даних навчання в деревоподібну структуру. Низька ентропія і висока інформація, що посилюють значення функції, вибирається як повторюваний вузол як головний вузол для розбиття набору даних і перетворення набору даних в структуру дерева. Структура дерева, як правило, використовується для прогнозування незамічених даних у прогнозі

3. IB1: Цей алгоритм використовує принцип найближчого сусіда для побудови прогнозної моделі. У цьому підході, відстань між навчальним екземпляром та даним іспитом випробувань розраховується за допомогою евклідової відстані. Якщо для декількох екземплярів найменша відстань до тесту, використовується перший знайдений екземпляр. Найближчий сусід є одним з найбільш значущих алгоритмів навчання; вона може бути пристосована до вирішення більш широких проблем. Нехай набір даних D має X-екземпляри (X1, X2, X3 ... Xn) та функцію F (F1, F2, F3, ... Fm) з міткою класу Cj, де j = 1,2 ... K. Цей алгоритм вказує значення відстані сусідніх екземплярів для прогнозування незамічених даних X з міткою класу. Метод

евклідової відстані використовується для обчислення ваги сусідів екземплярів X. Таким чином, незамічені дані передбачаються голосуванням за вагою для обчислення найближчих сусідів того класу для прогнозування невідомих даних. Цей метод не реле на попередній імовірності, як NB алгоритми. Визначається висока вартість обчислень, коли кількість екземплярів є більшою, оскільки вимірювання відстані є обчислювально дорогим. Отже, алгоритм вибору властивостей використовується для зменшення розмірності набору даних навчання для ефективної обчислювальної вартості цього алгоритму

Опис методу паралельних обчислень з використанням програмного засобу Hadoop

Метод паралельної обробки розподіляє обробку великої кількості даних про події безпеки між різними вузлами комп'ютерного кластеру. Вузли обробляють дані паралельно, що значно покращує час відгуку системи. У підприємстві є декілька джерел, які генерують дані про події безпеки. Ці джерела включають, але не обмежуються ними, мережні пристрої (наприклад, комутатори та маршрутизатори), діяльність бази даних, дані додатків та дії користувачів. Дані про події безпеки створюються на дуже високій швидкості. Автономний комп'ютер, який послідовно обробляє такий великий обсяг даних про події безпеки, займе багато часу, щоб виявити атаку, яка не допустима в таких критичних ситуаціях безпеки. омпонент збору даних збирає дані подій безпеки з різних джерел залежно від типу аналітики безпеки та вимог безпеки підприємства. Колектор даних пересилає зібрані дані на компонент зберігання даних, в якому зберігаються дані. Дані можуть зберігатись кількома способами, такими як розподілена файлова система Hadoop (HDFS), HBase та система управління реляційною базою даних (RDBMS). Щоб увімкнути паралельну обробку, збережені дані потрібно розділити на блоки фіксованого розміру (наприклад, 64 МБ або 128 Мб). Після розбиття дані обробляються в компоненті аналізу даних через кілька вузлів, що працюють паралельно, відповідно до принципів розподіленої структури, таких як Hadoop або Spark. Результат аналізу надається користувачеві через компонент візуалізації. Обмеження. Тактика паралельної обробки передбачає, що аналітична система безпеки, яка включає цю тактику, вже інтегрована з кластером вузлів, здатних обробляти дані паралельно. Ще один важливий фактор, який потребує догляду, полягає у розбитті логічного запису на два блоки при розподілі даних на блоки. У такій ситуації важливо зберігати достатньо інформації про тип файлу даних, щоб запис можна було реконструювати. Метод паралельної обробки залежить від методу динамічного навантаження та моніторингу передачі даних для балансування навантаження серед вузлів та керування потік даних у вузли відповідно. MapReduce - паралельна структура обробки, яка широко використовується в розподіленій установці. Ця структура складається з двох етапів - Map and Reduce. Пари ключових значень з вузлів mapper подається у вузли редуктора після заданого інтервалу часу (наприклад, 5-хвилинну агрегацію або 1-годинну агрегацію). Коли редуктор генерує результати, тригер виконує порогове значення. Цей заздалегідь визначений інтервал часу вводить затримку. Наприклад, атака може бути запущена при t = 5 сек заздалегідь

визначеного інтервалу часу, і тригер буде виконуватися при t = 5 хв. За ці 5 хв цілком можливо, що вже було завдано значної шкоди.

Висновки

Аналіз даних з використанням з використанням методу паралельних обчислень на базі програмного засобу Hadoop - це чудовий спосіб для користувачів аналізувати дані, не турбуючись про те, щоб вони були під загрозою. Коли мова йде про великі дані, ви можете зберігати, обробляти та аналізувати їх без зайвого місця в мережі.

Саме тому більшість мереж, як правило, використовують Hadoop як надійне джерело захисту великих даних. Він може не лише зберігати ваші великі дані, але й вдосконалені принципи кібербезпеки полегшують їх збереження..

Література

1. M. Nikhil Kumar, K.V.S. Koushik, K. John Sundar (2018). Data Mining and Machine Learning Techniques for Cyber Security Intrusion Detection

2. D.ASIR ANTONY GNANA SINGH, E.JEBAMALAR LEAVLINE (2013) Data mining in network security - techniques & tools: a research perspective

3. Priya James (2018) Protecting Big Data with Hadoop: A Cyber Security Protection Guide

References:

1. M. Nikhil Kumar, K.V.S. Koushik, K. John Sundar (2018). Data Mining and Machine Learning Techniques for Cyber Security Intrusion Detection

2. D.ASIR ANTONY GNANA SINGH, E.JEBAMALAR LEAVLINE (2013) Data mining in network security - techniques & tools: a research perspective

3. Priya James (2018) Protecting Big Data with Hadoop: A Cyber Security Protection Guide

Abstract: *The article proposes a method of parallel computing on the basis of the Hadoop software to detect the invasion of cyber security of robotic systems. Describe existing method data mining technique, describes the existing vulnerabilities of robotic systems and the types of threats.*

Key words: *Cybersecurity, Robotic Systems, Data Mining, Hadoop, Machine Learning Techniques, Intrusion Detection.*