УДК 004.056
DOI: 10.31673/2409-7292.2019.010615

С.В. Довбешко, С.В. Толюпа, Я.В. Шестак

ЗАСТОСУВАННЯ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ДЛЯ ПОБУДОВИ СИСТЕМ ВИЯВЛЕННЯ АТАК

В статті розглянуто методи інтелектуального аналізу даних, основне завдання яких полягає у виявленні в даних неструктурованої інформації та подання її в наочному вигляді. Проаналізовано сферу застосування даних методів та зроблено висновки стосовно перспектив їх використання в системах виявлення вторгнень та атак. Безліч параметрів для виявлення мережевих атак становить значний обсяг даних, що визначає можливість їх обробки саме методами інтелектуального аналізу даних. На сьогодні системи виявлення вторгнень і атак зазвичай являють собою програмні або апаратно-програмні рішення, які автоматизують процес контролю подій, що відбуваються в інформаційній системі або мережі, а також самостійно аналізують ці події в пошуках ознак проблем безпеки. Оскільки кількість різних типів і способів організації несанкціонованих проникнень в чужі мережі за останні роки значно збільшилася, системи виявлення вторгнень і атак побудованих на методах інтелектуального аналізу даних стали необхідним компонентом інфраструктури безпеки інформаційних систем й мереж.

Ключові слова: атака, вторгнення, нечітка логіка, сигнатура, трафік, інформаційна система, кластеризація, опорні вектори, база даних, нейромережі, генетичні алгоритми.

Розвиток обчислювальних засобів та інформаційних технологій призводить до автоматизації різних процесів практично у всіх сферах життя суспільства: збільшуються обчислювальні потужності комп'ютерних засобів, удосконалюються технології мережевої взаємодії, змінюються формати і вимоги до побудови інформаційних систем. Слідом за розвитком інформаційних технологій з не меншою швидкістю з'являються нові загрози інформаційній безпеці, тому проблема захисту інформації залишається ключовим напрямком наукових досліджень.

За останні роки було створено низку шкідливих засобів, що використовують принципово нові методи і підходи, які дозволяють традиційних засобів захисту виявляти і адекватно реагувати на такі загрози. Прикладом таких засобів є поліморфні віруси, що не дозволяють виявити себе за допомогою сигнатурних антивірусів або можуть бути використані постійної великий обчислювального навантаження засобів захисту, або руткіт, що використовують апаратну віртуалізацію, повністю контролюють будь-які дії антивірусів і навіть простих антируткітів. Ці загрози стосуються як окремо взятих призначених для користувача або серверних комп'ютерів, так і мережевої безпеки. Для виявлення деяких видів сучасних мережевих атак необхідно зберігати великий обсяг сигнатур і використовувати безліч додаткових обчислень для контролю трафіку.

Останніми роками спостерігається тенденція до об'єднання обчислювальних ресурсів в розподілені обчислювальні мережі. Принципи обробки даних в розподілених обчислювальних мережах мають суттєві відмінності від роботи простих електронно-обчислювальних машин, що стосується і різних аспектів захисту інформації. Мережеві атаки є одним з основних видів порушення інформаційної безпеки в розподілених обчислювальних мережах. Якісний розвиток даного виду загроз викликає необхідність постійного вдосконалення засобів захисту, пропонуючи принципово нові методи виявлення мережевих атак.

Одним з ключових засобів захисту обчислювальних систем є системи виявлення вторгнень (СОВ, Intrusion Detection System). Система виявлення вторгнень - це програма або програмно-апаратний засіб, призначений для виявлення фактів несанкціонованого доступу в обчислювальну систему або мережу. Системи виявлення вторгнень використовуються для виявлення різних видів шкідливої активності: мережевих атак проти безлічі сервісів; атак, спрямованих на підвищення призначених для користувача привілеїв, неавторизованого доступу до важливих системних і призначеним для користувача файлів, а також дій шкідливого програмного забезпечення (комп'ютерних вірусів, троянів і черв'яків).

Традиційні СОВ працюють за схожою з більшістю антивірусних засобів сигнатурної схемою і стикаються зі схожими проблемами, що й інші засоби захисту.

З розвитком інформаційних технологій особливо актуальною стала проблема обробки великих даних. В цьому випадку недостатньо простого статистичного аналізу, що викликає перехід до більш складного інтелектуального аналізу даних (ІАД). Основне завдання методів інтелектуального аналізу даних полягає у виявленні в даних неструктурованої інформації та подання її в наочному вигляді. Безліч параметрів для виявлення мережевих атак становить значний обсяг даних, що визначає можливість їх обробки саме цими методами ІАД.

Основним засобом захисту інформаційно-телекомунікаційних систем та мереж (ІТСМ) від інформаційноруйнівних впливів (втручань) у вигляді кібернетичних вторгнень (КВ) є системи виявлення та/або запобігання вторгненням (СВВ/СЗВ/СВА), основна задача яких зводиться до оперативної їх ідентифікації (встановлення відповідності між об'єктом і його ідентифікатором (унікальним атрибутом) та в ідеальному випадку ініціювання ефективного захисного сценарію щодо припинення факту порушення конфіденційності, доступності та цілісності інформаційних ресурсів, сервісів. Практика застосування СВВ сформувала два напрямки протидії КВ: виявлення зловживань (Misuse detection) та виявлення аномалій (Anomaly detection). Перший підхід орієнтований на виявлення лише класифікованих (відомих) вторгнень на основі підходів синтаксичного порівняння відповідності структурних (сигнатур/патернів), інваріантних та кореляційних ознак виконуваного процесу (системи) з існуючою базою відомих шаблонів. Головними недоліками такого підходу є неможливість виявлення нових модифікацій КВ чи кібернетичних атак нульового дня (0-day) та неможливість автоматичного вводу нових шаблонів, що свідчить про їх достатньо малу ефективність. Другий підхід, навпаки, зводиться до задачі виявлення невідомих КВ на основі знаходження набору ознак, який не відповідає очікуваній поведінці об'єкта (користувача/системи) – шаблони характеристик, які не задовольняють визначеному поняттю нормальної поведінки фіксуються як аномалії [1].

Всі розробники систем виявлення атак і організації, які використовують СВА повинні розуміти й вивчати їх класифікацію, щоб вибрати кращі рішення для систем захисту інформації. При дослідженні різних аспектів таксономії і застосуванні різних варіантів ми зможемо досягти більш високого рівня безпеки інформаційних систем [2].

На сьогодні системи виявлення вторгнень і атак зазвичай являють собою програмні або апаратно-програмні рішення, які автоматизують процес контролю подій, що відбуваються в інформаційній системі або мережі, а також самостійно аналізують ці події в пошуках ознак проблем безпеки. Оскільки кількість різних типів і способів організації несанкціонованих проникнень в чужі мережі за останні роки значно збільшилася, системи виявлення атак (СВА) стали необхідним компонентом інфраструктури безпеки більшості організацій.

В [3] представлений матеріал являє собою сучасний погляд на таксономію систем виявлення атак з коротким поясненням та обґрунтуванням кожної ознаки в систематиці. Щоб зробити дану класифікацію всеосяжною і повною окрім звичних ознак, таких як: середовище моніторингу, метод виявлення, архітектура, характер відповіді, принцип роботи та час реакції, були включені наступні характеристики: джерело аудиту, технологія побудови, парадигма виявлення та режим збору даних [7] (рис. 1.).

При розробці і проведенні досліджень систем виявлення вторгнень однією з ключових завдань є вибір масивів даних, на яких буде проводитися тестування. Великі компанії-розробники в першу чергу орієнтуються на власні бази даних, спеціалізовані під конкретні завдання і область застосування.

На сьогоднішній день можна виділити дві найбільш поширені тренувальні бази даних з відомими атаками - DARPA і KDD.

Тренувальна база даних DARPA (Defense Advanced Research Project Agency) була сформована в рамках досліджень лабораторії Лінкольна Массачусетського технологічного інституту (MIT Lincoln Laboratory) в рамках дослідження можливостей різних систем

виявлення вторгнень. Під час цього дослідження використовувалися дані мережевого трафіку і відомості від файлової системи для можливості ідентифікації змодельованих вторгнень, проведених фахівцями під час запису мережевих дампов. Тренувальні дані містять як реальний потік мережевого трафіку, так і спеціально змодельований фоновий трафік. Всі атаки були спрямовані на реальні обчислювальні системи.

Після проведення дослідження роботи різних систем виявлення вторгнень, DARPA надало збережені тренувальні дані у вільному доступі. В даний час ці тренувальні бази доступні всім дослідникам, тому значна частина публікацій у науковій літературі, пов'язаних з пропозицією нових методів і підходів з виявлення мережевих атак або аномалій, спираються на ці тестові дані. Використання даної бази даних дозволяє дослідникам порівняти основні характеристики якості виявлення: ймовірності помилок пропуску (false negative) і помилкового спрацьовування (false positive).

Загальна кількість типів атак, включених в тестові дані DARPA, склало 32 атаки. З точки зору атакуючого ці атаки можна розділити на чотири категорії:

- атаки відмови в обслуговуванні (Denial of Service, DoS);
- атаки переходу від віддаленого використання до локального (Remote to Local);
- атаки отримання користувачами прав суперкористувача (User to Root);
- атаки сканування або проб (Probing/surveillance).

Інформація про атаки DARPA зберігається у вигляді текстового опису, в якому вказується час початку атаки, тривалість, адреса жертви, назва атаки, категорія атаки та інші параметри.

На відміну від тренувальних даних DARPA, база даних KDD містить не дампи мережевого трафіку, а оброблені відомості у вигляді масивів з 42 ключових значень. Дана база успішно застосовується багатьма дослідниками для аналізу застосовності різних математичних методів в завданню виявлення мережевих атак, в основному через можливість використання масивів даних з більшості програмних засобів без виконання додаткової обробки.

Склад 42 параметрів, що розглядаються в базі даних KDD, був обгрунтований декількома науковими працями [7], присвяченими виявлення аномалій в мережевому трафіку. Однак при дослідженні можливостей по виявленню конкретних мережевих атак виявляється недостатньо аналізувати тільки представлені параметри, але також необхідно розглядати корисне навантаження мережевих пакетів - вищі рівні стека протоколів TCP / IP. Крім позначених тренувальних баз даних існує безліч більш вузько спеціалізованих, але вони не набули такого широкого поширення в науковому середовищі.

Міжмережеві екрани організовують фільтрацію вхідного і вихідного інформаційного потоку, відсікаючи все заборонені дані і пропускаючи дозволені. Міжмережеві екрани працюють на різних рівнях стека протоколів TCP/IP. Реалізований контроль доступу з зовнішньої мережі до внутрішньої сильно залежить від рівня аналізованих протоколів. Ступінь захисту підвищується зі збільшенням рівня, але разом з нею підвищується обчислювальне навантаження проведеного аналізу і складність адміністрування [3].

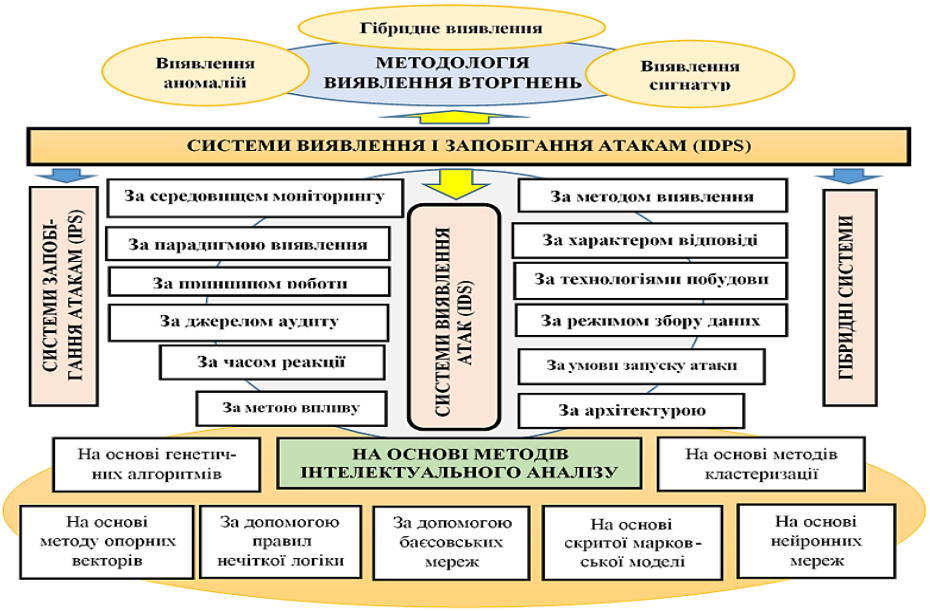


Рис. 1. Класифікаційні ознаки систем виявлення і запобігання атак.

З основних недоліків міжмережевих екранів можна виділити неможливість детального контролю активності авторизованих користувачів. На думку експертів переважна більшість порушень інформаційної безпеки відбувається в зв'язку з діяльністю легітимних користувачів, що призводить до необхідності контролю призначених для користувача дій на більш високому рівні.

У своїй роботі системи виявлення вторгнень керуються не тільки мережевим трафіком і безліччю правил, а й аудитом системи, різними журналами, показниками роботи операційної системи і т.д. Також існують системи запобігання вторгнень, що дозволяють не тільки виявити факт реалізації вторгнення в систему, але і мінімізувати наслідки, розірвавши мережеве з'єднання, заблокувавши підозрілу активність користувача або навіть адміністратора.

Рис. 2. Загальна структура системи вивлення вторгнень.

Найбільш ефективним способом запобігання несанкціонованому використанню інформаційних систем і мережевих ресурсів є підтримка багаторівневого захисту, коли спільно використовуються міжмережеві екрани, системи виявлення вторгнень, системи аудиту, політика безпеки і інші засоби захисту.

Найбільш загальна структура системи виявлення вторгнень, розроблена групою дослідників CIDF (Common Intrusion Detection Framework) [4], представлена на рис. 2.

Блок збору даних (сенсор, Event-box) - аналізує дані для обробки та прийняття рішення аналізатором. У даних можуть міститися імена контрольованих параметрів, їх особливості та

значення. Сенсор може виконувати перетворення даних для перетворення в необхідний формат або для скорочення обсягу даних, що передаються.

Блок аналізу (Analyzer-box) - приймає рішення про наявність або відсутність ознак атаки або аномалії на підставі даних від сенсорів. В рамках аналізу даний блок може виконувати функції фільтрації, нормалізації, перетворення і кореляції даних. При виявленні атаки блок аналізатора може додати до вихідних даних опис виявленої атаки. Блок аналізатора може мати багаторівневу систему.

Блок бази даних (сховище даних, Database-box) - містить множини вирішальних правил і семантичний опис атак, а також накопичувальну інформацію від сенсорів. Дані можуть перебувати в текстових файлах, базі даних, і т.д.

Блок корекції (Response-box) - інформує адміністратора про зафіксовану атаку, а в випадку системи запобігання вторгнень формує активну реакцію. Системи запобігання вторгнень відстежують активність в режимі реального часу і швидко реалізовують дії щодо запобігання атак. Можливі заходи - блокування потоків трафіку в мережі, скидання з'єднань, видача сигналів оператору. Також системи запобігання вторгнень можуть виконувати дефрагментацію пакетів, упорядкування пакетів TCP для захисту від пакетів з зміненими номерами послідовності і підтвердження.

Системи виявлення мережевих атак збирають інформацію з пакетів мережевого трафіку, системних журналів і показників функціонування системи. Традиційні системи виявлення мережевих атак будуються на сигнатурному підході: за допомогою набору правил або сигнатур, що формуються експертами і розміщені в базу вирішальних правил, описуються всі можливі сценарії і особливості атак. У цього підходу існує безліч відомих недоліків. За допомогою аналізу сигнатур неможливо виявити нові види атак, тому що база вирішальних правил не містить інформації про відповідну атаці. Процес аналізу сигнатур для розподілених атак є вкрай складним завданням. Крім того, бази вирішальних правил популярних систем виявлення вторгнень практично є загальнодоступними, тому порушник може протестувати можливості приховування атаки.

Перераховані проблеми підходу пошуку сигнатур змушують фахівців шукати альтернативні шляхи для організації захисту від мережевих атак. Одним з популярних напрямків досліджень є застосування різних методів ІАД в системах виявлення мережевих атак [5]. ІАД (інтелектуальний аналіз даних, глибинний аналіз даних) - сукупність методів виявлення в даних раніше невідомих, нетривіальних, практично корисних і доступних інтерпретації знань, необхідних для прийняття рішень в різних сферах людської діяльності. В основі даних методів лежить припущення, що вся легітимна активність в системі може бути представлена у вигляді математичної моделі. Застосовувані для виявлення мережевих атак методи ІАД переслідують одну з наступних цілей: виявлення порушень; виявлення аномалій.

Перші моделюють атаки і застосовують засоби класифікації, другі моделюють нормальну поведінку і виконують пошук винятків.

При використанні методів ІАД для виявлення мережевих атак можна виділити наступні проблеми: дані, аналізовані системами виявлення, мають високу розмірність і обсяг; вимога обробки даних в режимі реального часу; велика кількість шумів і невідповідностей в даних, що обробляються що викликають неадекватну реакцію методів інтелектуального аналізу даних.

Проаналізуємо СВА основаних на ***методах ІАД***. Одним з таких методів є виявлення атаки за допомогою ***скритої марковської моделі***. Скрита марковська модель представляє собою статистичну модель [6], де система моделюється як процес Маркова з невідомими параметрами. Задача методу полягає в оцінці скритих параметрів, що базуються на параметрах, які спостерігаються. Послідовності подій, зібрані з нормальних операційних систем, використовуються в якості навчальної вибірки для оцінки параметрів прихованої марковської моделі. Після навчання скритої марковської моделі ймовірнісні оцінки

використовуються в якості порогових значень для ідентифікації мережевих аномалій в тестових даних.

Виявлення атак за ***допомогою байєсовських мереж***. Байєсовська мережа являє собою модель, яка кодує імовірнісні взаємозв'язки між змінними. Основний метод застосування баєсовських мереж передбачає незалежність серед атрибутів. Кілька варіантів застосування байєсовських мереж були запропоновані для виявлення мережевих аномалій [7]. Більшість методів направлено на формування умовних залежностей між атрибутами з використанням складних мереж Байеса [8, 9]. Байєсовські методи часто використовуються в процедурі класифікації і локалізації помилкових спрацьовувань. Для виявлення вторгнень або прогнозування поведінки порушника байєсовські мережі можуть бути ефективними в деяких випадках, але в загальному випадку точність цього методу залежить від припущень, пов'язаних з поведінкою моделі цільової системи. Таким чином, будь-яке значне відхилення від припущень призведе до зменшення точності виявлення.

Виявлення атак за допомогою ***методів кластеризації***. Методи кластеризації групують дані в кластери на підставі схожості об'єктів. Більшість методів кластеризації починається з вибору центральної точки для кожного кластера [10], а множина елементів розподіляється по кластерам. Після цього центри коригуються, а елементи перерозподіляються. Кластеризація дозволяє вивчити і виявити аномалії, не вимагаючи множини класів або типів аномалій, тобто для виявлення аномалій за допомогою методів кластеризації не виникає потреби в навчальній множині. Кластеризація досить широко застосовується для виявлення мережевих аномалій [11, 12].

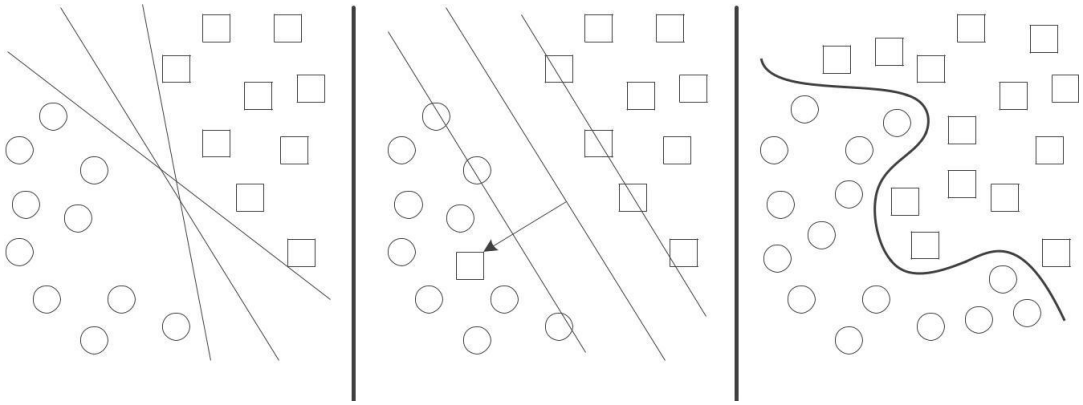
Виявлення невідомих мережевих атак найчастіше будується саме на методах кластеризації. Однорідні групи зі схожими характеристиками або кластери формуються шляхом розбиття набору елементів без будь-яких позначок. В системі вкрай важливо правильно визначити кластери, щоб максимально віддалити їх від викидів. Кінцева мета даних методів полягає у визначенні ступеня відхилення викидів від кластерів. За допомогою простого порівняння з пороговим значенням викиди з високим ступенем відхилення від кластерів позначаються як аномалії.

Особливу увагу заслуговує метод опорних векторів (Support Vector Machine, SVM), який представляє собою набір схожих алгоритмів категорії «навчання з учителем», застосовуваних у задачах класифікації та регресійного аналізу. Даний метод належить до сімейства лінійних класифікаторів. Характерною особливістю методу опорних векторів є постійне скорочення емпіричної помилки класифікації і збільшення зазору між класами. Тому даний метод часто називають методом класифікатора з максимальним зазором.

Метод відшукує елементи, що знаходяться на кордонах між двома класами, які і називаються опорними векторами.

На рис. 3 показані різні випадки, що виникають при застосуванні SVM для двовимірних даних:

- приклади поділяють площин (а);
- розділяє площині зі штрафом (б);
- лінійна неподільність (в).



а б в

Рис. 3. Метод опорних векторів

Метод опорних векторів здійснює пошук лінійної функції, яка дозволяє віднести елементи набору даних до одного з двох класів. Завдання бінарної класифікації може бути сформульована як пошук лінійної функції f (x), яка приймає значення менше нуля для елементів одного класу і більше нуля для елементів іншого.

Розподілена гіперплощина має наступний вигляд:

$$f(x) = w * x - b = 0 \qquad (1)$$

де **w** - вектор, перпендикулярний до розподіленої гіперплощини,

параметр **b** визначає відстань гіперплощини від початку координат.

Гіперплощина, паралельні оптимальної гіперплощини і найближчі до опорним векторах двох класів, можуть бути описані наступними рівняннями:

$$\begin{cases} wx - b = 1 \\ wx - b = -1 \end{cases} \qquad (2)$$

Якщо навчальна множина даних лінійно нероздільні, то можна вибрати гіперплощини так, щоб в смугу між ними не потрапляла жодна точка навчальної вибірки і потім максимізувати відстань між гіперплоскостями. Ширина смуги в цьому випадку дорівнює $\frac{2}{\|w\|}$, тому слід мінімізувати $\|w\|$. Для виключення всіх точок з смуги, повинна виконуватися умова:

$$c_i(wx_i - b) \geq 1{,}1 \ \leq i \ \leq n, \qquad (3)$$

де c_i - мітка класу, що приймає значення -1 і +1,

x_i - вектор робочої вибірки з міткою класу c_i .

Це завдання квадратичної оптимізації еквівалентна задачі пошуку сідлової точки функції Лагранжа [7]:

$$\begin{cases} -L(\lambda) = \sum_{i=1}^{n} \lambda_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j c_i c_j (x_i x_j) \rightarrow min_\lambda \\ \lambda_i \geq 0{,}1 \leq i \leq n \\ \sum_{i=1}^{n} \lambda_i c_i = 0 \end{cases} \qquad (4)$$

де L - функція Лагранжа,

λ_i - множники Лагранжа.

Щоб узагальнити SVM на випадок лінійної нероздільності (рис. 3(в)), вводиться константа C - внутрішній параметр методу, що дозволяє регулювати відношення між максимізацією ширини розділової смуги і мінімізацією сумарної помилки [8].

Даний метод є одним з найбільш популярних методів класифікації. Метод будує оптимальну гіперплощину в просторі характеристик: $w \times x - b = 0$, що розділяє нормальні і аномальні елементи [14]. У підсумку завдання можна звести до квадратичного програмування: $min \parallel \omega \parallel \frac{2}{H}/2 + c\sum_{i=1}^{N}\xi_i$, при $c_i\ (wx_i\text{-}b) \geq 1\text{-}\xi_i$, $1 \leq i \leq N$, де ξ_i - величина помилки на об'єктах x_i.

Параметр C є компромісом між точністю опису моделі, яка визначається величиною помилки $\sum_{i=1}^{N}\xi_i$ і можливостями моделі до узагальнення тобто значенням межі $1/\parallel \omega \parallel \frac{2}{H}$.

Основною проблемою застосування методу опорних векторів в завданні бінарної класифікації є складність пошуку лінійної кордону між двома класами. У разі якщо таку кордон побудувати не вдається, одне з рішень - це збільшення розмірності (перенесення даних в інший простір, більш високої розмірності), де існує можливість побудови площині, що розділяє безліч елементів на два класи.

Виявлення атак за ***допомогою правил нечіткої логіки***. Нечіткі системи виявлення мережевих вторгнень використовують множину нечітких правил для визначення ймовірності конкретних або загальних мережевих атак. Нечітка множина може бути сформована для опису трафіку в конкретній мережі. В роботі [17] описується метод для побудови класифікаторів, що використовують нечіткі асоціативні правила, які застосовуються для виявлення вторгнення в мережу. Нечіткі набори правил асоціації використовуються для опису нормальних і аномальних класів. Належність запису певному класу визначається за допомогою відповідної метрики. Нечіткі асоціативні правила формуються на основі звичайних навчальних вибірок. Тестований зразок класифікується як нормальний, якщо згенерований сукупністю правил показник буде вище певного порогового значення. Зразки з більш низьким показником вважаються аномальними.

Звичайно протидіяти вторгненням і атакам основуючись тільки на одному з методів ІАД малоефективно, тому необхідно підійти до цього питання комплексно і побудувати інтелектуальну систему протидії вторгненням (рис. 3). При побудові такої інтелектуальної (експертної) системи пропонується вибрати нечітку модель. Це пов'язано з тим, що значна частина інформації про причини і джерела атак може бути отримана тільки експертним шляхом або у вигляді евристичних описів процесів. Для визначення джерел атак система безпеки має бути представлена моделлю тієї інформаційної мережі на яку вона орієнтується. Данна модель ділить завдання переміщення інформації між комп'ютерами через середовище мережі на кількість рівнів менш великих і легше вирішуваних підзадач. Кожна з цих підзадач вирішується за допомогою одного рівня мережі. Тому первинне завдання після фахівця безпеки може бути представлене декомпозицією завдань безпеки по окремих рівнів мережі.

Представимо окремий рівень системи безпеки у вигляді нелінійного об'єкту з безліччю вхідних змінних $\{x_i\}, i = \overline{1,n}$ і однієї вихідної змінної y:

$$y = f_y(x_1, x_2, ..., x_n) \qquad (5)$$

Як вхідні змінні вибираються ознаки джерел атак. Вихідна змінна y є показником ступеня можливості стану рівня мережі.

У моделі використовуються наступні допущення і обмеження:

- вхідні змінні $\{x_i\}$ в межах одного рівня незалежні;
- на кожному з рівнів мережі ізолюються окремі мережеві функції.

Комплексна інтелектуальна система підтримки прийняття рішень для визначення вторгнень містить набір функціональних компонент, що дозволяють максимально

автоматизувати і прискорити вироблення дій, що управляють, при зміні ситуації в системі безпеки (рис. 3).

Сучасний підхід до побудови систем виявлення атак на інформаційні системи сповнений недоліків і вразливостей, що дозволяють, на жаль, шкідливим впливам успішно долати системи захисту інформації. Перехід від пошуку сигнатур атак до виявлення передумов виникнення загроз інформаційної безпеки має сприяти тому, щоб докорінно змінити дану ситуацію, скоротивши дистанцію відставання в розвитку систем захисту від систем їх подолання. Крім того, такий перехід має сприяти підвищенню ефективності управління інформаційною безпекою і, нарешті, більш конкретних прикладів застосування нормативних і керівних документів, що вже стали стандартами.

Результати різних методів інтелектуального аналізу даних для виявлення вторгнень представлені в таблиці 1.

Метод інтелектуального аналізу даних	Показник розпізнавання (%)	Помилкове спрацьовування (%)
Метод k-найближчих сусідів	92	1
Метод опорних векторів (SVM)	95,5	1
SVM+ k-найближчих сусідів	96,3	0,84
SVM+нечітка логіка	97	0,73
SVM+ k-найближчих сусідів+ нечітка логіка	99,56	0,44

Проведене імітаційне моделювання та застосування інтелектуальної системи підтвердило правильність вибору множини методів інтелектуального аналізу даних в якості побудови систем виявлення вторгнень. Так метод опорних векторів дозволив ідентифікувати більшість атак з результатом 96-98%. Метод головних компонент скоротив обсяг інформації, необхідної для класифікації мережевих пакетів, і підвищив швидкість формування модулів виявлення, але виявив проблему перенавчання. Методи кластеризації дозволили сформувати безліч модулів виявлення, виділивши типові фрагменти атак в окремі модулі виявлення і розбивши комплексні атаки на окремі модулі. Застосування нечіткої логіки підвищило результати роботи системи і дозволило класифікувати вектори, що мають різні мітки в навчальній вибірці.

Список літератури:

1. Субач І.Ю., Фесьоха В.В. Модель виявлення аномалій в інформаційно – телекомунікаційних мережах органів військового управління на основі нечітких множин та нечіткого логічного виводу. Збірник наукових праць ВІТІ № 3 – 2017.

2. І.М. Павлов, С.В. Толюпа, В.І. Ніщенко Аналіз таксономії систем виявлення атак у контексті сучасного рівня розвитку інформаційних систем. Сучасний захист інформації №4, 2014,с. 44-52.

3. Толюпа С.В., Штаненко С.С., Берестовенко Г. Класифікаційні ознаки систем виявлення атак та напрямки їх побудови. Збірник наукових праць Військового інституту телекомунікацій та інформатизації імені Героїв Крут Випуск № 3. 2018р. с. 56-66.

4. Зоріна Т.І. Системи виявлення і запобігання атак в комп'ютерних мережах. Вісник східноукраїнського національного університету імені Володимира Даля № 15 (204) ч.1 2013. – с. 48-54.

5. 4-104. Valdes, A. Adaptive model-based monitoring for cyber attack detection / A. Valdes, K. Skinner // In: Proc. of the Recent Advances in Intrusion Detection (Toulouse, France, 2000) — 2000. — P. 80–92.

6. 5-3. Барсегян, А. А. Технологии анализа данных : Data Mining, Visual Mining, Text Mining, OLAP / А. А. Барсегян, М. С. Куприянов, В. В. Степаненко, И. И. Холод. — СПб.: БХВ-Петербург, 2007. — 384 с.

7. 6-60. Ghahramani, Z. An Introduction to hidden Markov models and Bayesian networks / Z. Ghahramani // International Journal of Pattern Recognition and Artificial Intelligence — 2001. — Vol. 15. — P. 9–42.

8. Barbara, D. Detecting novel network intrusions using Bayes estimators / D. Barbara, J. Couto, S. Jajodia, N. Wu. // In: Proc. of the 1st SIAM International Conference on Data Mining. — 2001. — 17 p.

9. Kruegel, C. Bayesian event classification for intrusion detection / C. Kruegel, D. Mutz, W. Robertson, F. Valeur // In: Proc. of the 19th Annual Computer Security Applications Conference — 2003. — P. 14–23.

10. Valdes, A. Adaptive model-based monitoring for cyber attack detection / A. Valdes, K. Skinner // In: Proc. of the Recent Advances in Intrusion Detection (Toulouse, France, 2000) — 2000. — P. 80–92.

11. Portnoy, L. Intrusion detection with unlabeled data using clustering / L. Portnoy, E. Eskin, S. J. Stolfo // In: Proc. of ACM Workshop on Data Mining Applied to Security. — 2001. — P. 1–14.

12. Sequeira, K. ADMIT: Anomaly-based data mining for intrusions / K. Sequeira, M. Zaki // In: Proc. of the Eighth ACM SIGKDD Int'nl Conference on Knowledge Discovery and Data Mining (New York, NY, USA, 2002), ACM. — 2002. P. 386–395.

13. Yang, H. Clustering and classification based anomaly detection / H. Yang, F. Xie, Y. Lu // Fuzzy Systems and Knowledge Discovery — 2006. — Vol. 4223. — P. 1082–1091.

14. Bhattacharyya, D. K. Network Anomaly Detection. A Machine Learning Perspective / D. K. Bhattacharyya, J. K. Kalita. — CRC Press, 2014. — 364 p.

15. Васильев, В. И. Применение нейронных сетей при обнаружении атак на компьютеры в сети Internet (на примере атаки SYNFLOOD) / В. И. Васильев, А. Ф. Хафизов // Нейрокомпьютеры: разработка и применение. — 2001. — №4-5. — С. 108-114.

16. Bankovic Z. Improving network security using genetic algorithm approach / Z. Bankovic, D. Stepanovich, S. Bojanic, O. Nieto-Taladris // Computers and Electrical Engineering. — 2007. — Vol. 33. — No. 5-6. — P. 438-451.

17. Tajbakhsh, A. Intrusion detection using fuzzy association rules / A. Tajbakhsh, M. Rahmati, A. Mirzaei // Applied Soft Computing — 2009 — Vol. 9. — No. 2. — P. 462.