УДК 004.056.5

ПРОБЛЕМИ ВИКОРИСТАННЯ БАЗИ ОЗНАК МЕРЕЖЕВИХ АТАК KDD-99 В ІНТЕЛЕКТУАЛЬНИХ СИСТЕМАХ ВИЯВЛЕННЯ ВТОРГНЕНЬ

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

На сьогоднішній день комп'ютерні мережі відіграють важливу роль у повсякденному житті людини, адже вони використовуються майже у всіх сферах її діяльності. Яке б не було підприємство, комерційне, державне, муніципальне чи бюджетне, але у всіх випадках його функціонування так чи інакше забезпечується комп'ютерною мережею. Тож зрозуміло, що забезпечення безпеки є необхідним, оскільки наслідки недостатньої захищеності найрізноманітніші – крадіжка, знищення або поширення конфіденційної інформації (комерційної таємниці), персональних даних, підміна інформації, блокування доступу до неї, обмеження функціональності або повна зупинка діяльності комп'ютерної мережі. А останнє навіть призводить до фактичної зупинки бізнес-процесів, що може завдати великих збитків. Одним із способів вирішення такої проблеми є побудова інтелектуальних систем виявлення вторгнень. Але їх результативність можлива лише за наявності ефективного набору даних. База KDD99 – є прикладом такого набору даних.

База даних KDD99 - це база, що містить стандартний набір даних, який включає в себе широкий спектр вторгнень, імітованих в умовах військової мережі.

Загалом така база містить близько 5000000 записів про мережеві з'єднання. Кожний запис представляє собою образ мережевевого з'єднання та включає 41 параметр мережевого трафіка і позначається як "атака" або "не атака" [1].

№ з/п	Параметр	Опис
1.	*duration*	Тривалість (у секундах) з'єднання
2.	*protocol_type*	Тип протоколу (TCP, UDP, etc.)
3.	*service*	Атакований сервіс
4.	*src_bytes*	Кількість байтів від джерела до призначення
5.	*dst_bytes*	Кількість байтів відповіді клієнту
6.	*flag*	Прапорці з'єднання
7.	*land*	1, якщо з'єднання від/до того самого хоста/порта
8.	*wrong_fragment*	Кількість „хибних" фрагментів
9.	*urgent*	Кількість термінових пакетів
10.	*hot*	Кількість „гарячих" індикаторів
11.	*num_failed_logins*	Кількість невдалих спроб реєстрації
12.	*logged_in*	1, якщо успішний вхід в систему; 0 неуспішне
13.	*num_compromised*	Кількість „компроментуючих" умов
14.	*root_shell*	1, якщо root shell отриманий; інакше 0
15.	*su_attempted*	1, якщо виконувалась „su root" ; інакше 0
16.	*num_root*	Кількість „root" доступів
17.	*num_file_creations*	Кількість операцій створення файлів
18.	*num_shells*	Кількість запитів на надання оболонки
19.	*num_access_files*	Кількість операцій на доступ до контролю файлів
20.	*num_outbound_cmds*	Кількість вихідних команд для FTP сесії
21.	*is_hot_login*	1, якщо логін належав до „гарячого" списку
22.	*is_guest_login*	1, якщо „гостьовий" вхід
23.	*count*	Кількість з'єднань на хост в поточній сесії за останні 2 с
24.	*serror_rate*	% з'єднань що мали „SYN" помилки
25.	*rerror_rate*	% з'єднань що мали „REJ" помилки
26.	*same_srv_rate*	% з'єднань що мали однаковий сервіс
27.	*diff_srv_rate*	% з'єднань на різні сервіси
28.	*srv_count*	Кількість з'єднань на такий самий сервіс за останні 2 с
29.	*srv_serror_rate*	% з'єднання з помилкою в „SYN" пакеті
30.	*srv_rerror_rate*	% з'єднання, що мають „REJ" помилки
31.	*srv_diff_host_rate*	% з'єднання від інших хостів
32.	*dst_host_count*	Кількість з`єднань до локального хоста, встановлених віддаленою стороною
33.	*dst_host_srv_count*	Кількість з`єднань до локального хоста, встановлених віддаленою стороною та використовуючих одну службу
34.	*dst_host_same_srv_rate*	% з`єднань до локального хоста, встановлених віддаленою стороною та використовуючих одну службу
35.	*dst_host_diff_srv_rate*	% з`єднань до локального хоста, встановлених віддаленою стороною та використовуючих різні служби
36.	*dst_host_same_src_port_rate*	% з`єднань до даного хоста при поточному номері порту джерела
37.	*dst_host_srv_diff_host_rate*	% з`єднань до служби різних хостів
38.	*dst_host_serror_rate*	% з`єднань з помилкою типу SYN для даного хост-приймача
39.	*dst_host_srv_serror_rate*	% з`єднань з помилкою типу SYN для даної служби приймача
40.	*dst_host_rerror_rate*	% з`єднань з помилкою типу REJ для даного хост-приймача
41.	*dst_host_srv_rerror_rate*	% з`єднань з помилкою типу REJ для даної служби приймача

Також в базі представлені 22 типи атак. При цьому атаки поділяються на 4 основні категорії: DoS, U2R, R2L і Probe.

DoS атаки - це мережеві атаки, спрямовані на виникнення ситуації, коли на системі, що є атакованою, відбувається відмова в обслуговуванні. Дані атаки характеризуються генерацією великого обсягу трафіку, що призводить до перевантаження і блокування сервера. Виділяють шість DoS атак: back, land, neptune, pod, smurf, teardrop.

U2R атаки передбачають отримання зареєстрованим користувачем привілеїв локального суперкористувача (мережевого адміністратора). Виділяють чотири типи U2R атак: buffer_overflow, loadmodule, perl, rootkit.

R2L атаки характеризуються отриманням доступу незареєстрованого користувача до комп'ютера з боку віддаленого комп'ютера. Виділяють вісім типів R2L атак: ftp_write, guess_passwd, imap, multihop, phf, spy, warezclient, warezmaster.

Probe атаки полягають в скануванні мережевих портів з метою отримання конфіденційної інформації. Виділяють чотири типи Probe атак: ipsweep, nmap, portsweep, satan [2].

Зовнішній вигляд бази KDD99 – текстовий файл у якому у вигляді матриць представлено набір параметрів певного типу атаки або нормального з'єднання.

```
0,tcp,http,SF,215,45076,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,1,1,0.
00,0.00,0.00,0.00,1.00,0.00,0.00,0,0,0.00,0.00,0.00,0.00,0.00,
0.00,0.00,0.00,normal.
0,tcp,http,SF,162,4528,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,2,2,0.0
0,0.00,0.00,0.00,1.00,0.00,0.00,1,1,1.00,0.00,1.00,0.00,0.00,0
.00,0.00,0.00,normal.
```

Рисунок 1 – Зовнішній вигляд бази NSL-KDD

Переваги:

1. Не включає надлишкових записів в набір ознак.
2. У запропонованих тестових наборах не містить дублікатів записів.
3. Кількість відібраних записів з кожної складної групи є обратно пропорційною відсотковій кількості записів у вихідному наборі даних KDD. Як результат, класифікаційні показники різних методів машинного навчання змінюються в більш широкому діапазоні, що робить більш ефективним точне оцінювання різних методів навчання.

Недоліки:

1. Досить велика кількість параметрів, що знижує час виявлення вторгнень. Тому на практиці використовують не всі.

Список використаних джерел

1. KDD Cup 1999 Data [Електронний ресурс] – Режим доступу до ресурсу: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.
2. Марков Р. А., Бухтояров В. В., Попов А. М., Бухтоярова Н. А. Дослідження нейромережевих технологій для виявлення інцидентів інформаційної безпеки // Молодий вчений. - 2015. - №23. - С. 55-60. - URL https://moluch.ru/archive/103/23866/
3. DERIVED FEATURES [Електронний ресурс] – Режим доступу до ресурсу: http://kdd.ics.uci.edu/databases/kddcup99/task.html

---

УДК 004.457

UBA/UEBA - СИСТЕМИ

Гринько В.В., студ. гр. КБ-171, **Мехед Д.Б.,** к.п.н., доцент кафедри
кібербезпеки та математичного моделювання
Чернігівський національний технологічний університет

Сьогодні дуже важливим чинником належного функціонування підприємств та компаній є високий рівень інформаційної безпеки. Інформаційна безпека означає безпеку всієї інформаційної середовища: це значить, що під захистом повинні знаходитися не тільки самі дані, але і їх носії, а також вся інфраструктура. Рішення для забезпечення ІБ повинні охоплювати технічні, адміністративні, правові аспекти, а також поведінку користувача, щоб не допустити витоків і розголошення комерційної таємниці. Цільових атак стало більше, вони стали більш витонченішими і більш продуманими, зловмисники стали хитрішими та розумнішими, а кількість інформаційних систем збільшилась. В такому світі контролювати і реагувати на інциденти інформаційної безпеки стає все складніше і дорожче. Тому перед індустрією інформаційної безпеки стоїть велика кількість завдань по автоматизації процесів реагування на інциденти і загрози та їх виявлення. Одне з цих завдань вирішують системи UBA/UEBA..

UBA – система є одним з основних інструментом захисту ІБ. UBA - система, що дозволяє на основі даних про користувачів з допомогою алгоритмів машинного навчання і аналізу будувати моделі поведінки користувачів і визначати відхилення від цих моделей, тобто ця система використовує технології