УДК 519.7

КЛАСИФІКАЦІЯ ЗАГРОЗ КОМП'ЮТЕРНІЙ ІНФОРМАЦІЇ З ВИКОРИСТАННЯМ КАРТ КОХОНЕНА

С. Копитко

Львівський інститут банківської справи Університет банківської справи НБУ

Розглядається проблема класифікації множини загроз комп'ютерної інформації. Запропоновано нейромережеву модель класифікації інформаційних загроз з використанням самоорганізуючих карт Кохонена. На підставі описів реальних хакерських атак на комп'ютерні системи, включаючи банківські, з використанням запропонованої моделі отримано характеристики типових загроз комп'ютерній інформації. За типову загрозу вибрано ядро відповідного класу інформаційних загроз.
Ключові слова: захист інформації, система захисту інформації, класифікація загроз, модель класифікації загроз, карти Кохонена.

Постановка проблеми. При розробці і впровадженні інформаційних систем однією з основних задач є розмежування доступу співробітників підприємства до інформаційних і інших ресурсів обчислювальної системи. Такі тенденції притаманні практично для всіх рівнів ієрархії сучасних інформаційних технологій, починаючи з архітектурного рівня в цілому (Internet і Intranet), включаючи мережні технології (наприклад, IP v.4.0 і IP Sec), і закінчуючи рівнем загальносистемних засобів (ОС, СУБД) і їх застосувань.

Масштаби використання інформаційних технологій стали такі, що разом з проблемами продуктивності, надійності і стійкості функціонування інформаційних систем гостро постає проблема захисту циркулюючої в них інформації від несанкціонованого доступу. На сьогоднішній день існує достатньо велика статистика загроз ОС, направлених на подолання вбудованих в ОС механізмів захисту, які дозволяють змінити налаштування механізмів безпеки, обійти розмежування доступу і т.д.

Моніторинг фактів несанкціонованого доступу до інформації показує, що більшість розповсюджених систем достатньо вразливі з точки зору безпеки, і це не дивлячись на тенденцію до підвищення рівня їх захищеності. Наприклад, за даними Computer Emergency Readiness Team Coordination Center(CERT/CC) з 1998 по 2002рр. кількість інцидентів, пов'язаних з питаннями безпеки, зросла на 2099%, середній річний приріст їх склав 116%, причому спостерігається тенденція прискорення росту в останній час[1]: протягом 2003р. зафіксовано 13759 випадків порівняно з 82094 у 2002р. Також у [1] як стійку тенденцію відзначено суттєве збільшення кількості вразливих місць сучасних програмних продуктів(майже чотирикратний приріст у 2003р. порівняно з 2002-м.). За різними оцінкам, від 70 до 80 % втрат від злочинів в сфері IT припадають на внутрішні атаки. При цьому топ-менеджери IT-компаній приділяють велику увагу мережевій безпеці своїх систем, проте часто недооцінюють ту шкоду, яку може нанести бізнесу їх власний працівник.

Отже, проблема захисту комп'ютерної інформації стає надзвичайно важливою на сучасному етапі інформатизації суспільства. Традиційно для її вирішення при впровадженні ІС проектують систему захисту інформації, яка мала б реалізувати передбачені заходи протидії несанкціонованому доступові до комп'ютерної інформації. Характерною особливістю методів проектування систем захисту інформації є орієнтація на деяку множину загроз, типову для конкретної предметної області [2]. Важливими процедурами таких методів виступають ідентифікація загроз комп'ютерній інформації та формування їх базової множини для предметної області ІС, класифікація загроз на підставі вибраної сукупності ознак, формування досьє типових загроз комп'ютерної інформації. Оскільки для оптимізації системи захисту використовуються саме типові загрози, то названі процедури фактично визначають якість спроектованої системи захиту комп'ютерної інформації для ІС.

Аналіз систем класифікації загроз інформації. На сьогодні існує значна кількість різних підходів щодо визначення класифікаційних схем загроз інформації. Це зумовлено насамперед великою кількістю різних типів ІС, додаткових пристроїв та прикладних програм.

Анін Б.Ю. у [3, с.18] описує три методи зламу комп'ютерних систем. Він виходить з того, що в загальному випадку програмне забезпечення будь-якої універсальної комп'ютерної системи складається з трьох основних компонентів: операційної системи, програмного забезпечення мережі і системи керування базами даних (СКБД). Тому всі спроби зламу захисту комп'ютерних систем можна розділити на три групи: 1) атаки на рівні операційної системи; 2) атаки на рівні програмного забезпечення мережі; 3) атаки на рівні систем управління базами даних.

Варто звернути увагу на іншу класифікацію, а саме класифікацію загроз за способом їх здійснення. Ця класифікація необхідна для формування заходів додаткового захисту інформації.

Відповідно до даної класифікації загрози комп'ютерної безпеки поділяються на явні і приховані. Під явними розуміємо такі загрози, які зрозумілі однозначно передбачені. Вони не вимагають для протидії їм яких-небудь

додаткових відомостей про статистику загроз і неочевидних припущень про можливі атаки зловмисника. Явні загрози пов'язані з некоректною реалізацією і налаштуванням системи захисту. До таких можуть бути віднесені: 1)некоректність реалізації механізму захисту; 2)неповнота покриття каналів доступу до інформації механізмами захисту; 3)суперечність можливих налаштувань механізмів захисту.

Під прихованими розуміємо такі загрози, які не очевидні, і вимагають для протидії їм додаткових припущень про можливі атаки зловмисника. Приховані загрози пов'язані з нерегламентованими діями користувача перш за все за допомогою запуску власних програм, а також з використовуванням зловмисником помилок і закладок в системному і прикладному ПЗ. При цьому прихована загроза може бути охарактеризована двома властивостями: 1) характеристикою об'єкту загрози (наприклад, обліковий запис користувача); 2) загальною характеристикою атаки зловмисника (наприклад, модифікація облікового запису із застосуванням власне запущеної програми).

З урахуванням сказаного можна зробити наступний найважливіший висновок: будь-який механізм захисту повинен проектуватися з урахуванням як явних, так і прихованих (у тому числі і невідомих) загроз інформаційної безпеки, оскільки тільки в цьому випадку можна говорити про можливість реалізації механізмом захисних властивостей.

Розглянемо ще один різновид класифікаційної схеми загроз інформації, який має на меті чітко розмежувати напрямки атаки. Тобто множина загроз класифікується за ознакою "тип атаки" [4], де виділяють локальні і віддалені атаки та атаки на потік даних.

Локальною атакою називатимемо випадок, коли зловмисник виявився безпосередньо перед клавіатурою (дисководом, CD-ROM і т. п.) даного комп'ютера.

Віддалена атака — це варіант атаки, коли зловмисник не бачить (і можливо ніколи не побачить) ту робочу станцію (або сервер), з якою він працює. При цьому сам комп'ютер, що атакується, можливо, не проявляє ніякої активності в мережі.

Атака на потік даних — інцидент, коли комп'ютер, що атакується, активно відправляє, приймає або обмінюється даними з іншими комп'ютерами мережі, локальної або глобальної, а місцем застосування атакуючої дії є сегмент мережі або вузол мережі між цими системами.

За своїм характером такі атаки можуть бути пасивними, коли зловмисник копіює собі дані для подальшого аналізу, і активними, коли зловмисник вносить зміни в дані або повністю підміняє їх.

Постановка завдання. Проаналізувавши вище названі типи класифікацій загроз інформації, можна побачити їхню неоднозначність, недостатній рівень конкретизації, хоч кожна з них є теоретично обґрунтованою і застосовується при проектуваннях систем захисту інформації. Це пояснюється насамперед широким спектром застосувань інформаційних технологій, що унеможливлює створення єдиної класифікаційної схеми. Тому постає задача виділення типових загроз інформації на підставі множини характеристик існуючих конкретних загроз для заданої предметної області. Тобто із заданої множини потрібно виділити групи загроз, які є близькими між собою за певними характеристиками. Як відомо [3,с.50-52], такі проблеми зводяться до задач класифікації. Суть задачі класифікації полягає у розбиті об'єктів заданої множини на класи з використанням векторів параметрів (характеристик) об'єктів. Об'єкти одного класу вважаються еквівалентними з точки зору критерію розбиття. Часто класи наперед невідомі, а формуються динамічно. Класи залежні від вхідної множини об'єктів, тому доповнення її новим об'єктом може вимагати коригування сформованої системи класів.

Складність реалізації сформульованої мети зростає з огляду на досить велику кількість характеристик, які бажано врахувати. Тому для вирішення даної проблеми застосуємо один з методів нейромережевої класифікації, а саме карти Кохонена.

Економіко-математична модель класифікації загроз комп'ютерної інформації. Нехай M - множина конкретних загроз інформації, кожна з яких характеризується k ознаками (характеристиками). Позначимо для $p \in M$ через $\overline{x}^p = \left(x_1^p, x_2^p, ..., x_k^p\right)^T$ вектор стовбець, що визначає параметри p -ої загрози із множини M, де $x_j^p \ \left(1 \le j \le k\right)$ - значення j -ої характеристики p -ої загрози. Очевидно, що $\overline{x}^p \in E^k$, де E^k - евклідовий k - вимірний простір.

Введемо множину класів загроз $C^1, C^2, ..., C^R = \left\{C^r\right\}$ у просторі класів C, до того ж виконується включення $C^1 \cup C^2 \cup ... \cup C^R \subset C$.

(1)

Простір класів C може не збігатися з простором E^k і мати іншу розмірність. Визначимо ядра класів загроз $\left\{\overline{c}^r\right\} = \{\overline{c}^1, \overline{c}^2, ..., \overline{c}^R\} \ \left(1 \le r \le R\right)$ у просторі класів C як загрози, що є типовим для свого класу. Тобто ядро класу $\overline{c}^r$ - це така загроза, яка найбільше за своїми характеристиками наближається до узагальнених властивостей загроз даного класу C^r.

Для оцінки наближення загрози до ядра класу потрібно ввести міру близькості. Позначимо через $d(\bar{x}^p, \bar{c}^r)$ скалярну функцію від загрози і ядра класу, значення якої тим менша, чим більше загроза схожа на ядро класу. Тобто, $d(\bar{x}^p, \bar{c}^r)$ - міра близькості конкретної загрози $\bar{x}^p \in E^k$ вхідної множини M до ядра $\bar{c}^r$ класу $C^R \subset C$. Найчастіше використовують дві міри:

$$d(\bar{x}^p, \bar{c}^r) = \sum_{j=1}^{k} (x_j^p - c_j^r)^2$$

– евклідову міру, коли ; (2)

$$d(\bar{x}^p, \bar{c}^r) = \sum_{j=1}^{k} |x_j^p - c_j^r|.$$

– nfr pdfye vihe “city block”? lt

(3)

У формулах (2), (3) c_j^r - значення узагальненої j -ої характеристики загроз для ядра класу C^r.

Задавши число класів R і маючи множину загроз M, можна сформулювати таку задачу класифікації: знайти R ядер класів $\{\bar{c}^r\}$ та розбити загрози $\{\bar{x}^p\}$ на класи $\{\bar{C}^r\}$, тобто побудувати функцію $m(p)$ так, щоб мінімізувати суму мір близькості на множині M:

$$\min_M D(M) = \min_{m(p)} \sum_{p \in M} d(\bar{x}^p, \bar{c}^{m(p)}),$$

(4)

де функція $m(p)$ визначає номер класу загроз на підставі індексу (номера) загрози у вхідній множині M.

Нейромережева модель класифікації загроз комп’ютерній інформації. Ідея застосування сітки Кохонена до пошуку розв’язку задачі (4) коротко формулюється в наступний спосіб. Вибираємо як вхідні дані вектор параметрів $\bar{x}^p$ єдиної загрози $p \in M$. Результатом роботи сітки Кохонена буде код класу загроз, до якого належить загроза, характеристики якої подавали на вхід сітки. У нейромережах прийнято кодувати виходи номером каналу. Тому сітка буде мати R виходів (за кількістю класів) і чим більше буде значення виходу з номером r_0, тим більша впевненість сітки в тому, що загроза належить класу C^{r_0}. Кожний вихід сітки Кохонена можна трактувати як ймовірність того, що загроза належить даному класу. Так як сума виходів сітки рівна 1, то всі виходи утворюють повну групу (використаємо рівність (1)) і кожна загроза із M попадає в один клас.

Зведення моделі (4) до вигляду, потрібного для застосування сітки Кохонена, здійснюємо шляхом послідовного виконання таких етапів[3, с.52-54].

На першому етапі вибираємо евклідову міру близькості (2). Тоді ядро класу загроз, що лімітує суму мір близькості загроз цього класу, співпадає з центром тяжіння:

$$\bar{c}^{r_0} = \frac{1}{N(r_0)} \sum_{p \in M(r_0)} \bar{x}^p$$

(5)

де $N(r_0)$ - кількість загроз $\bar{x}^p$ в класі r_0, а $M(r_0)$ - сукупність загроз із М, що належать до класу C^{r_0}, і $M(r_0) = \{p \mid m(p) = r_0\}$.

На другому етапі з урахуванням вибраної міри конкретизуємо вираз під знаком $\min$ у формулі (4). У координатній формі маємо:

$$D = \sum_{p \in M} \sum_{j=1}^{k} (x_j^p - c_j^{m(p)})^2 = \sum_{p \in M} \left[< \bar{x}^p, \bar{x}^p > -2 < \bar{x}^p, \bar{c}^{m(p)} > + < \bar{c}^{m(p)}, \bar{c}^{m(p)} > \right],$$

де через $< \bar{a}, \bar{b} >$ позначено скалярний добуток векторів $\bar{a}, \bar{b} \in E^k$. Оскільки у попередньому виразі два доданки, а саме $\sum_{p \in M} < \bar{c}^{m(p)}, \bar{c}^{m(p)} > = const$, $\sum_{p \in M} < \bar{x}^p, \bar{x}^p > = const$, то задача пошуку мінімуму D із (4) еквівалентна задачі пошуку максимуму виразу:

$$\max_M \sum_{p \in M} \sum_{j=1}^{k} x_j^p \cdot c_j^{m(p)}$$

(6)

На третьому етапі зводимо задачу мінімізації (4) до задачі максимізації (6). Згідно (4), розбиття на класи загроз має здійснюватись так, щоб сумарна міра близькості для всієї вхідної множини M загроз з характеристиками $\{\bar{x}^p\}$ була мінімальною. Отже, задача класифікації загроз інформації (4) з евклідовою мірою близькості (2) зводиться до пошуку такої функції розбиття $m(p)$, яка на заданій множині M загроз для вказаної кількості R класів максимізує значення виразу (6).

Алгоритмічні аспекти реалізації нейромережевої моделі класифікації загроз комп'ютерної інформації. Алгоритм класифікації множини загроз інформації, який максимізує вираз (6) легко реалізується у формі нейронної мережі. Для цього потрібно R суматорів, що знаходять всі D^{rp}, та інтерпретатор, який має визначити суматор з максимальним виходом. Значення D^{rp} може бути розраховане формальним нейроном. Для цього потрібно вибрати x_j^p в якості вхідного сигналу, а компоненти ядер класів c_j^r як вагові коефіцієнти ознаки. Тоді кожний формальний нейрон з числом входів, рівним кількості компонент у вхідному векторі $\bar{x}^p$, буде давати на виході одну із сум $D^{r,p}$.

Для визначення класу загроз, до якого належить вхідна, серед всіх нейронів даного шару потрібно буде вибрати один з максимальним виходом, що є задачею інтерпретатора. Інтерпретатор – це або програма, що вибирає нейрон з максимальним виходом, або шар нейронів з оберненими зв'язками. На звичайних ЕОМ програмний інтерпретатор ефективніший, аніж шар нейронів з оберненими зв'язками. Тоді структура сітки Кохонена має вигляд:

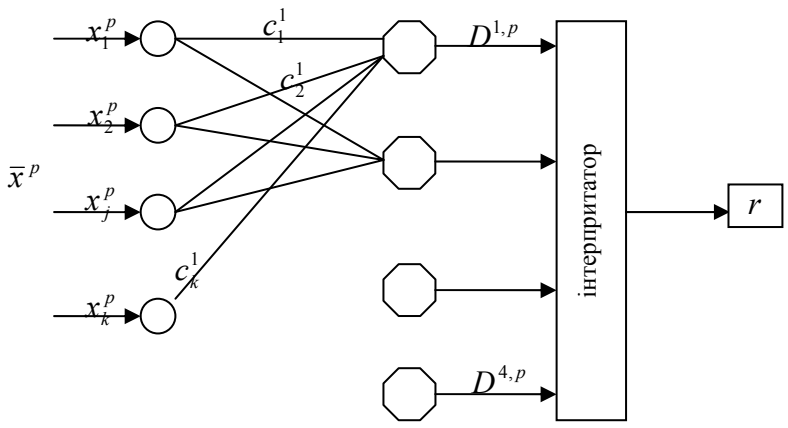

Рис. 1. Сітка Кохонена для класифікації загроз (R=4).

Нейрони шару Кохонена генерують сигнали $D^{r,p}$. Інтерпретатор вибирає максимальний сигнал шару Кохонена і видає номер класу r, відповідний номеру входу, за яким він отримав максимальний сигнал. Це відповідає номеру класу загрози, яка була подана на вхід у формі вектора характеристик $\bar{x}^p (p \in M)$. Ядра $\bar{c}^r$ виступають як вагові коефіцієнти нейронів. Кожний нейрон Кохонена запам'ятовує одне ядро класу, і відповідає за ідентифікацію загроз у своєму класі. Тобто величина виходу нейрону тим більша, чим ближче загроза до ядра даного класу.

Загальна кількість класів загроз збігається з кількістю нейронів Кохонена. Нейрони Кохонена мають лінійну функцію активізації. Вхідні вектори сітки найчастіше формуються так:

$$\frac{\bar{x}^p}{\left\|\bar{x}^p\right\|} \to \bar{x}^p \quad або \quad \frac{\bar{x}^p}{\sum\limits_{p\in M}\left\|\bar{x}^p\right\|^2} \to \bar{x}^p,$$

де $\|\cdot\|$ - норма вектора.

Апробація моделі. Як вхідні дані для моделі використані основні характеристики загроз інформації, отриманні з аналізу 16-и атак, які були здійснені хакерами на підприємства фінансової сфери [4]. Це джерело інформації містить детальний опис самої атаки, спосіб її виявлення та процедури усунення загрози. Кожна загроза інтерпретувалась множиною 13-и основних параметрів. Цільовою ознакою вибрано втрати від несанкціоно-

ваного доступу до інформації із застосуванням відповідної загрози, а до основних факторів віднесли тип організації, об'єкт загрози, вид сервера, тип мережі, мету загрози, тип ОС, кількість рівнів захисту, період доби коли проводилася атака, частота атак, тривалість атаки, час її виявлення, кількість етапів атаки.

Побудова сітки Кохонена на підставі вхідних даних [5] проводилася з використанням програмного засобу Kohonen Map, який реалізований як додаток до MS Excel. В результаті опрацювання програмою параметрів зазначеної множини загроз було виділено чотири кластери загроз інформації, загальна характеристика яких включена в табл.1.

Таблиця 1

Кластери загроз інформації

номер кластера	назва	Вартість компонент системи захиту	Експертна оцінка втрат, млн. $
1	A1	6520	310
2	A2	1200	220
3	B1	6380	150
4	B2	5240	190

Джерело:[власні розрахунки]

Деталізація основних характеристик ядер виділених кластерів включена в таблицю2. Набір таких характеристик приймемо як опис типової загрози інформації.

Таблиця 2

Характеристики типових загроз комп'ютерній інформації.

№ клас-тера	Об'єкт загрози	Мета загрози	ОС	Рівні захисту	Додаткові утиліти
A1	сервери та Web-вузли компаній	злом Web-сервера та модифікація інформації.	Windows NT, Windows 2000, Macintosh та Solaris.	одно- та дворів-нева система захисту	зовнішній та внут-рішній брандмауери, утиліти сканування портів (SNORT) та комплекти захисту FileGuard.
A2	Web-сервери та SQL-сервери	отримання конфіденційної інформації.	Windows NT	дворів-нева система захисту	зовнішній та внут-рішній брандмауер і утиліти SNORT.
B1	мережі VNP та маршрути-затори компаній.	використання мережі для подальших атак та отримання безплатного інтернету	Windows NT та Windows 2000.	дворів-нева система захисту	зовнішній та внутрішній брандмауер
B2	Web-сервер, поштовий сервер і DNS-сервер	отримання конфіденційної інформації та доведення недосконалості систем захисту.	Windows 2000 та Windows NT	дворів-нева система захисту	відсутні

Джерело:4, результати застосування Kohonen Map

Висновки. Як показав проведений аналіз розробка систем захисту комп'ютерної інформації суттєво залежить від характеристик множини загроз, від яких мають бути захищені дані. Тому проблема формування адекватного опису інформаційних загроз стає надзвичайно актуальною. Традиційним методом формування такого опису є класифікація сукупності загроз на класи з метою виділення характерстик типових загроз. Існуючі системи класифікації в значній мірі залежні від вподобань дослідника і це впливає на адекватність результатів їх застосування.

У контексті мети дослідження запропоновано нейромережеву модель класифікації загроз комп'ютерній інформації, яка дозволяє адекватніше визначити характеристики типових інформаційних загроз. Цю модель можна реалізувати з допомогою самоорганізуючих карт Кохонена, що показано у статті.

Запропонована модель апробована на сукупності описів реальних хакерських атак на комп'ютерні системи. З використанням досить простого інструментального засобу Kohonen Map для MS EXCEL 97 було отримано опис чотирьох типових загроз, що дозволяє сформувати досьє типових загроз для подальшого використання з метою оцінки якості проектованої системи захисту інформації.

Очевидно, що якість процедури класифікації інформаційних загроз залежить від обсягу їх вхідної множини. Тому в подальшому передбачається поповнення бази даних загроз новими описами реальних атак.

1. Нупур Дєвіс, Уоттс Хамфри, Сємюєл Редвайн, Герлинда Цибульски, Гєри Макгро. Реферат отёта участников форума National Cybersecurity Summit.//Открытые системы.-2004.-№8.-http://www.osp.ru/os/2004/08/045.htm.
2. Захарова М.В. Метод синтезу механізмів зихисту інформації в спеціалізованих автоматизованих системах за умовою погроз // Проблеми і перспективи розвитку банківської системи України: Збірник наукових праць. Т. 18. – Суми: У АБС НБУ, 2006. – с 327-331.
3. Анин Б.Ю. "Защита компьютерной информации".-Спб.: БХВ – Санкт-Петербург, 2000. – 384с.:ил.
4. Конев И.Р., Беляев А.В. Информационная безопасность предприятия. - Спб.: БХВ – Санкт-Петербург, 2003. – 752с.:ил.
5. Защита от хакеров // CD: Эксперт: криптографическая защита данных. 2004р.

CLASSIFICATION OF COMPUTER INFORMATION THREATS USING KOHONEN MAPS

S. Kopytko

Lviv Banking Institute of Banking University of National Bank of Ukraine

The problem of classifying the ensemble of computer information threats is examined. A neuron network model of information threats classification using self-organizing Kohonen maps is suggested. The characteristics of typical computer information threats were defined on the basis of descriptions of real hacker attack on computer systems, including banking systems. A typical threat is defined by a nucleus of a relevant class of information threat.

Keywords: information security, information security, threats classification, threats classification model, Kohonen maps.