УДК 512.972:519.6

з 983.202-01+3818.6 + В151.5

[1]**Мінаєв Ю. М.** д-р техн. наук
[2]**Філімонова О. Ю.** канд. техн. наук

МОДЕЛЮВАННЯ ТРАФІКА КОМП'ЮТЕРНИХ МЕРЕЖ В БАЗИСІ ТЕНЗОРНИХ ОРТОГОНАЛЬНИХ ІНВАРІАНТІВ

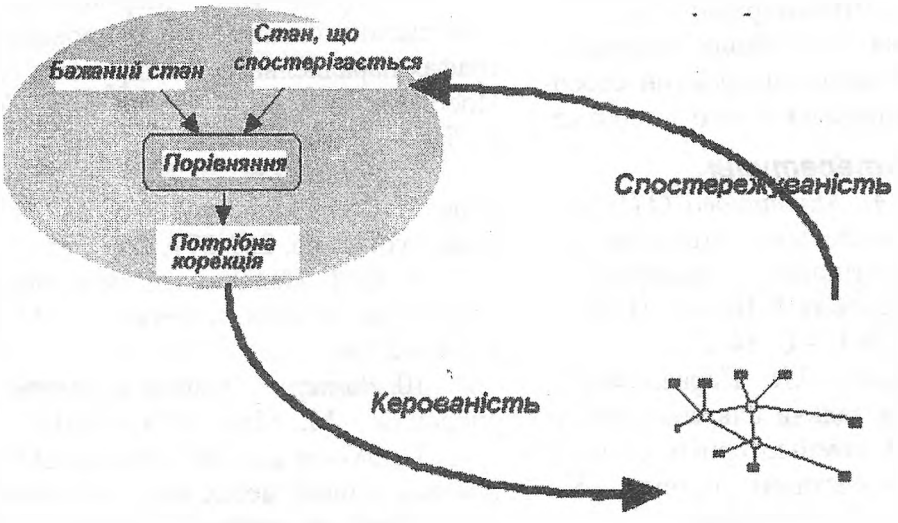
[1]Інститут комп'ютерних технологій Національного авіаційного університету
[2]Київський національний університет будівництва та архітектури

Розглядаються питання моделювання трафіка комп'ютерних мереж шляхом уявлення трафіка у вигляді тензора другого рангу. Показано, що застосування ортогональних інваріантів тензора дозволяє практично однозначно ідентифікувати стан комп'ютерної мережі, зокрема, виявити наявність аномальних станів.

Постановка проблеми (мережевий трафік та особливості його застосування до ідентифікації атак на комп'ютерні мережі)

Відомо, що трафік комп'ютерних мереж (КМ) вміщує 9 компонент, серед яких, зокрема, характеристики адресного простору, характеристики пакету, що передається, та особливості протоколу передачі. Для того, щоб ідентифікувати трафік як такий, що представляє атаку, треба мати певну множину патернів – зразків трафіка, які одназначно характеризують трафік. На рис. 1 наведено технологію цієї процедури.

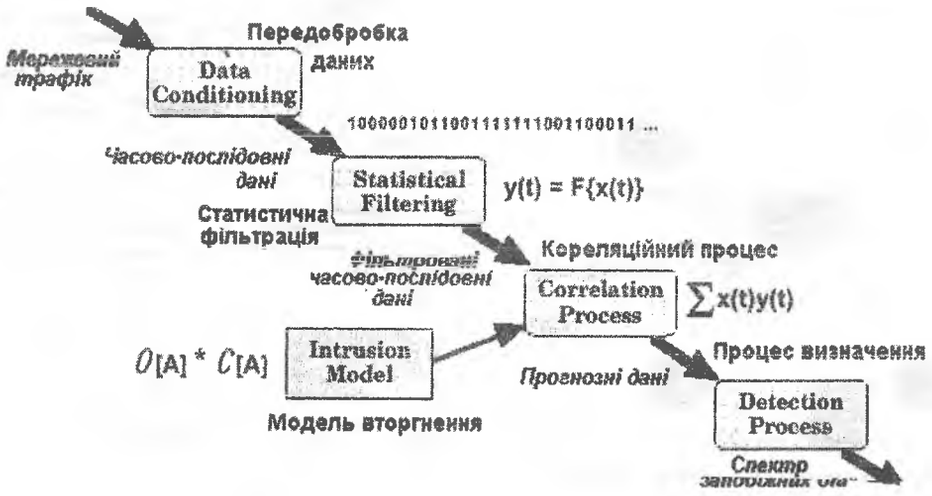
Одним з потужних засобів розв'язання цієї проблеми (ідентифікації атак на підставі аналізу трафіка) є використання нейронних мереж, які працюють як класифікатори, розпізнаючи два стани – атака або її відсутність. Переважна більшість робіт, які присвячені цій проблемі, використовують нейромережу «прямо», тобто подаючи на її входи параметри трафіка без будь-якої попередньої обробки. Дослідження, виконані авторами, довели, що ефективність ідентифікації атаки може бути суттєво підвищена, якщо, по-перше, уявити трафік як тензор, по-друге, виконати попередню обробку тензора трафіка і знайти інваріанти тензора, зокрема такі, які утворюють ортогональний базис. Нижче буде показано, що вони характенизують наявність атаки практично однозначно. На рис. 2 наведено приклад такої постановки задачі.



A Focused External Attacker Utilizes a Control Loop

Виділене зовнішнє джерело атаки (атакувальник) використовує петлю управління

Рис. 1. Загальна схема використання параметрів стану КМ для ідентифікації атаки



A System Functional Concept Implements Control Loop Measurement
Системна функціональна концепція впровадження петлі управління вимірами

Рис. 2. Попередня обробка параметрів трафіка

Один з підходів виявлення вторгнень (атак) базується на тому, що всі вторгнення в загальному випадку характеризуються деякими шумами або певною індікацією [1]. В термінах мереж ці сигнали можна бачити у *TCP-RESET* пакетах, *ICMP* ехо-відподях або порту призначення недосяжних пакетів. Аналіз мережевого трафіка показує, що профілі таких сигналів, що належать намаганням вторгнень, чітко різняться порявняно з рутинними операціями або/та непереднаміреними помилками.

Моніторингом таких сигналів, що викликають підозру в розподілених мережах, вторгнення або намагання вторгнгень можна ефективно визначити. Визначення шляху атакувальника, котрий може бути використовуючим пародійовані (змінювані) адреси, може бути виконано завдяки новій техніці моніторингу патернів трафіка (ПТ).

Патерни трафіка можуть простежені через всю мережу. Для цієї мети запропонована система, що базується на *SNMP*-повідомленнях, котрі є «дружніми» для мережі, щоб співпрацювать на шляхах джерела вторгнення [2].

Акт *вторгнення* в загальному випадку включає:

1. Сканування «уразливого» потенційного приймача;

2. Експлуатація одного чи більшої кількості уразливих (потенційних приймачів), щоб отримати доступ до ресурсів системи приймача;

3. Внесення нелегальної активності на систему-приймач;

4. Руйнування/використання ресурсів приймача

Деякі з цих вторгнень є прихованими, але можуть бути зворотно трасованими. Інші використовують парадійовані джерела адрес, які можуть бути неприховано відвернутими сервісами атак.

Представляє інтерес витягнення характеристик, корельованих з характеристиками трафіка, що визначають трасу трафіка через мережу і тоді використовувати мережеву конфігураційну інформацію для визначення треків атакувальника. Доцільно зробити наголос на характеристиках, які не можуть бути зфальшованими або парадійованими.

Характеристики мережевих вторгнень. Загальний мережевий патерн користування вимагає, як правило, малої кількості сервісів – *HHTP*, *TELNET*, *FTP*, *SMTP*, *NFS*, *SNMP*, *NTP* та ін. Хоча *ICMP* хост/порт буквально власно присутні у мережі середнього розміру, профіль таких пакетів показує важливість патернів, коли очікується вторгнення. Розрахункова конфігурація мережа наведена на рис. 3.

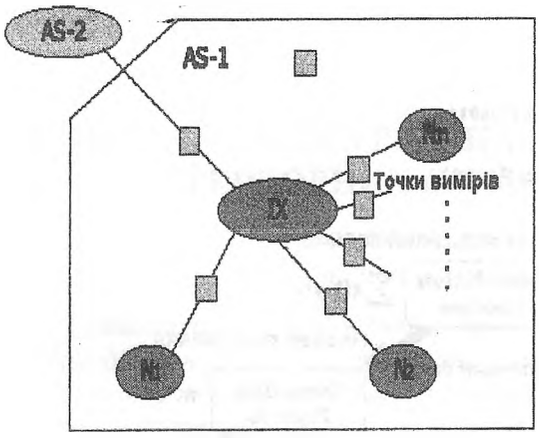


Рис. 3. Експериментальна мережа та точки вимірів

Рис. 4 показує кількість сервісів доступних клієнтам за період в один день. Ясно, що більшість клієнтів приєднують дуже малу кількість сервісів (< 3) – ліва частина графіка, права частина – клієнти, що мають необхідність доступу до великої кількості сервісів.

Зміст пакету може бути використаний для детального профілювання трафіка.

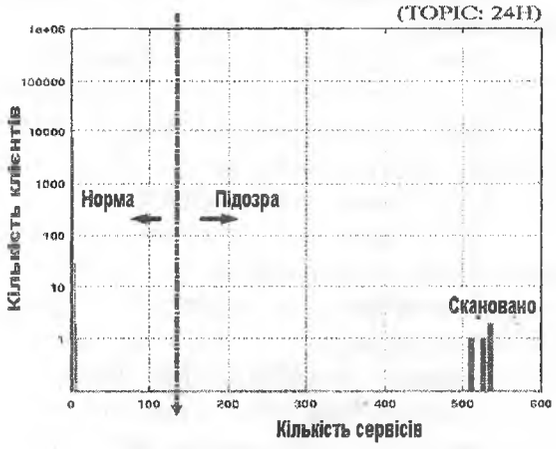


Рис. 4. Нормальні сигнали і такі, що викликають підозру

Заголовок протокола в кожному пакеті вміщує інформацію про джерело, приймач та протокол, що деталізує пакет.

Зразки (патерни) трафіка. Основи концепції сліду трафіка, базованого на патернах, наведена на рис. 5. Станція мережевого менеджменту (*NMS*) збирає відповідну інформацію через виміри на кожному з'єднанні.

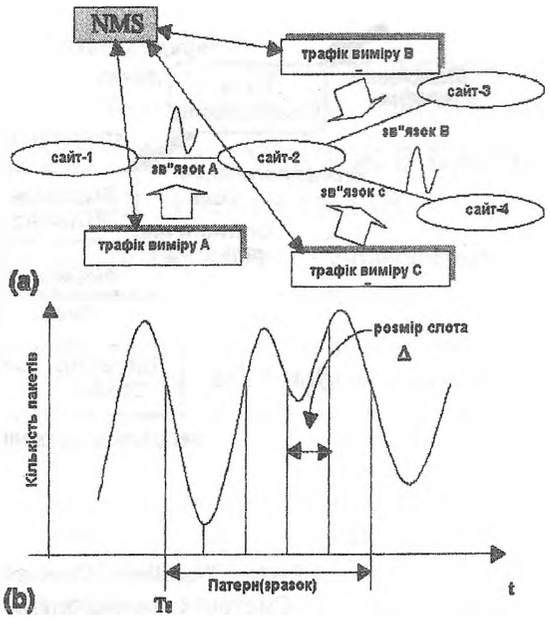


Рис. 5. Розподілене трасування патерно-базованого трафіка (*a*), патерн трафіка (*b*)

Відповідна інформація вміщує пакетні підрахунки для різних типів пакетів (*ICMP*, *TCP*, *UDP*) та підтипів (*ICMP*-ехо-вимога, *TCP-SYN*, *UDP*-ехо та ін.) *NMS* порівнює виміряні трафікові патерни та корелює їх. Ланцюг корельованих патернів вказує шлях (можливо парадійований) трафіка.

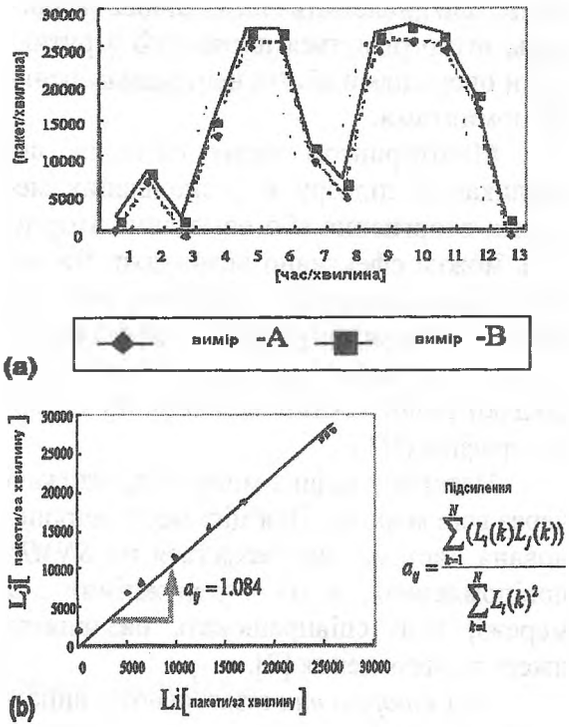


Рис. 6. Два пов'язаних патерни трафіка (*a*), кореляція двох патернів та підсилення

Визначення патерна трафіка. Моніторинг трафіка виконується у визначені часові інтервали (слоти часу). Патерн трафіка розглядається на певному розмаху часу, що вміщує інтегральну кількість часових слотів. Визначається ПТ такими параметрами: часовий розмір слота, розмір сегменту та виміряною величиною метрики сегменту (рис. 6). ПТ задається вектором

$$L = (\Delta, x_1, x_2, x_3, \ldots, x_N),$$

де x – виміряна величина метрики у слоті, сегмент вміщує N слотів розміром кожний D. Кожний ПТ має пов'язану часому мітку Ts, котра є стартовим часом сегменту. *Кореляція ПТ.* ПТ порівнюється, використовуючи класичну кореляційну техніку. Якщо два обо більше патернів з однією і тією ж часовою міткою подібні (рис.6-а), то це говорить про наявність одного й того ж трафіка.

Подібність ПТ визначається за формулою

$$r_{ij}(L_i, L_j) = \frac{1}{N\sigma_i\sigma_j} \times \sum_{k=1}^{N} (L_i(k) - \bar{L}_i)(L_i(k) - \bar{L}_i),$$

де Li, Lj – результати вимірів A та B відповідно, $\bar{L}$ – середнє, r – варіації відповідних статистик, $r_{ij} \rightarrow [0, 1]$.

Трасування шляхів. Кожний патерн визначається *NMS* з необхідністю показувати частки інформації разом з трасою шляху атаки. Трекова активність наведена на рис. 7.

Виділення патернових сегментів. Сегменти інтересів у патерні є частинами, що визначають певну аномальну активність. В загальному випадку зацікавленість патерна є кусками, отриманими з трафіка, шляхом використання порогів. На рис. 8 метрика спостереження, отримана порогом (*ThUp*), є симптомом події, що викликає зацікавленість. Метрика підрахунків записується в деталях. Коли метрика падає, використовується нижній поріг (*ThDn*).

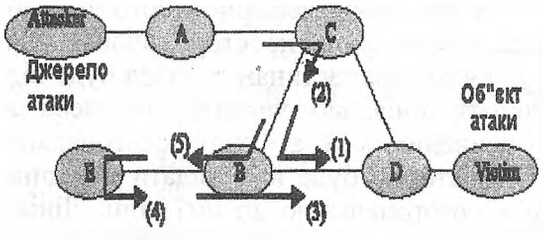


Рис. 7. Трасування шляху атаки

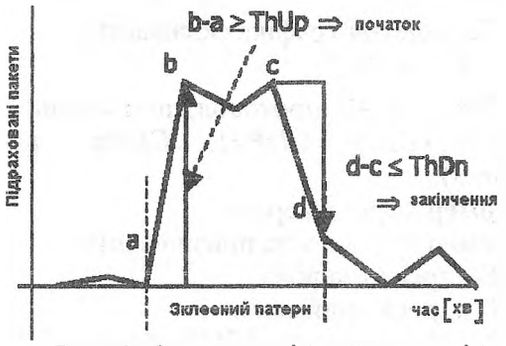


Рис. 8. Сегментаційний критерій

Послідовність підрахунків між *ThUp* та *ThDn* використовується мережевим менеджментом (*NMS*). На рис. 9 наведено загальну схему визначення вторгнень у КМ на підставі використання окремих сигнатур.

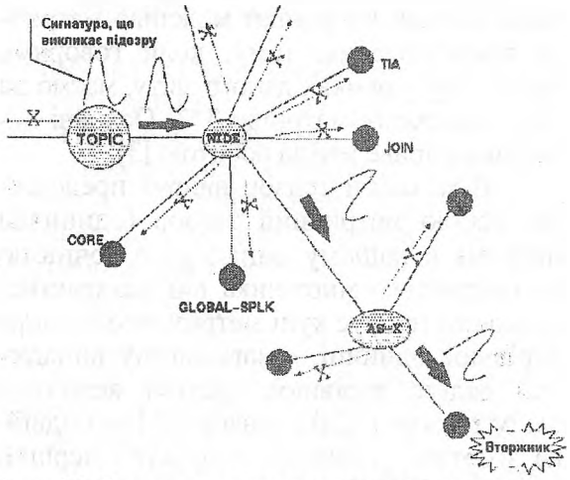


Рис. 10. Загальна схема визначення вторгнень у КМ на підставі використання окремих сигнатур

Тензорна методологія аналізу трафіка комп'ютерних мереж (кульова і девіаторна частина тензора й інваріанти)

Тензорний аналіз набув надзвичайно великого поширення у механіці, особливо у механіці деформованого середовища. Як показали досліди авторів, здобутки тензорного числення можуть бути ефективно використані для аналізу (ідентифікації) атак на комп'ютерні мережі на підставі аналізу трафіку. Як відомо, трафік описується системою 9 незалежних величин, що дає формальну можливість уявити його як двовалентний тензор (або тензор другого рангу. Як буде показано далі, парно рангові тензори можуть бути застосовані для аналізу захисту КМ в умовах невизначеності).

Параметри трафіка складають:
$\mathbf{x} = \{ x_i \}, i = 1,9$:
x_1 – ***Protocol ID*** протокол, пов'язаний з подією (*TCP*=0, *UDP*=1, *ICMP*=2, *unknown*=3);
x_2 – **номер** порта джерела;
x_3 – **номер** порта хоста призначення;
x_4 – ***IP*-адреса** джерела;
x_5 – ***IP*-адреса** приймача;
x_6 – ***ICMP Type*** **тип** *ICMP*-пакету (*Echo Request or Null*);
x_7 – ***ICMP Code*** кодове поле з *ICMP*-пакета (*None or Null*);
x_8 – ***Raw Data Length*** довжина даних в пакеті;
x_9 – ***Raw Data*** порція даних в пакеті

Для двовалентних тензорів у декартовій системі координат можливо матричне представлення, тому, коли говоримо про тензор у рамках даного звіту, маємо на увазі квадратну матрицю 3*3. Основні визначення приведені за роботою [3].

В механіці тензор напруг представляє собою метричний тензор (одинична матриця в нашому випадку), з точністю до скалярного множника він називається кульовим (радіус кулі метричного тензора дорівнює одиниці, в загальному випадку цей радіус дорівнює третині величини сліду тензора). Слід тензора T (чи подвійна згортка I лінійний інваріант I перший інваріант I *Sp* T) – для матричного представлення – сума діагональних елементів, у механіці має сенс потроєного гідростатичного тиску, в нашому випадку в залежності від обраної системи обліку параметрів трафіка він може мати інший зміст, якого можна надати у кожному випадку окремо. Зокрема, це може адресний простір трафіка.

Інваріанти тензора у ортогональному базисі. Як відомо з тензорного аналізу, завжди існує розкладання тензора на кульовий компонент і девіаторну компоненту. Семантично кульоіва частина тензора інтерпретується як його *незмінна* частина, а девіатор характеризує зміни. Це надзвичайно важливий елемент аналізу, бо при аналізі трафіка кульова частина може характеризувати нормальний стан, а девіатор присутність аномалій (зокрема, атаки).

Раніше зазначалось, що інваріанти тензора можуть бути представлені як певні функції від головних значень (власних векторів). Приведемо деякі співвідношення між різними системами інваріантів, ввівши таку систему позначень: *S1, S2, S3* – головні значення тензора *T*, *Dv* – девіатор тензора, $T = \{t_{ij}\}$, i, j=1,3. Тензор, записаний через інваріанти (головні значення), може бути записаний у вигляді $T = \begin{pmatrix} S1 & 0 & 0 \\ 0 & S2 & 0 \\ 0 & 0 & S3 \end{pmatrix}$. Якщо слід тензора (суму діагональних елементів тензора *T*) позначити через *Tr*[*T*], то девіатор цього тензора може бути відповідно визначений як:

$$Dv = T - \frac{Tr[T]}{3}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Доведено справедливість таких записів: для сліду квадрата девіаторної частини –

$$Tr[Dv * Dv] = ((S1 - S2)^2 + (S1 - S3)^2 + (S3 - S2)^2)/3;$$

для сліду квадрата девіаторної частини через тензор $-Tr[Dv*Dv=Tr[T*T]-Tr[T]^2/3$.

З тензорного аналізу відомо, що кульова та девіаторна частини тензора є ортогональними. У зв'язку з тим, що головні значення, або власні числа матриці є інваріантами, то будь-які комбінації з них також будуть інваріантами. Оберемо як два інваріанти такі вирази:

$$\sigma = \frac{1}{3}Tr[T] = \frac{1}{3}(S1 + S2 + S3);$$

$$\tau = \frac{1}{2}(Tr[T*T] - \frac{1}{3}Tr[T]^2)^{1/2} =$$

$$= (\frac{1}{2}Tr[Dv * Dv])^{1/2} =$$

$$= (\frac{1}{6}((S1 - S2)^2 + (S1 - S3)^2 + (S3 - S2)^2)^{1/2}.$$

В просторі (σ; τ) трафік, який уявляється тензором другого рангу, може бути уявлений як точка. Ця обставина відкриває додаткові можливості для аналізу аномальних станів.

З механіки деформованого середовища відомо, що у просторі головних напруг кульовим частинам тензора буде відповідати лінія, що однаково нахилена до координаних осей, а у силу ортогональності девіаторам буде відповідати площина, що є ортогональною до цієї лінії. Лінія і площина будуть проходити через початок координат і відповідно визначатися векто-

ром $\vec{n}=\{1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3}\}$. В нашому випадку простір головних напруг це простір трафіка. Маючи вектор $\vec{n}$, завжди можна визначити кут між цим вектором та тензором трафіка $(T*\vec{n})/(T*T)^{1/2} = Cos(\alpha)$.

Крім того, згідно до аналогій з механікою, можливо отримати ще декілька параметрів, які додатково можуть характеризувати трафік. Зокрема, проекція тензора трафіка на гідростатичну вісь (*pg*), довжина проекції на девіаторну площину (*pd*), довжина (*l*) вектора (трафіка):

$$pg = \left((1/\sqrt{3}\ 1/\sqrt{3}\ 1/\sqrt{3})*T*\begin{pmatrix}1/\sqrt{3}\\1/\sqrt{3}\\1/\sqrt{3}\end{pmatrix}\right)[1,1];$$

$$pd=\sqrt{\frac{1}{2}}\left(Tr[T*T]-\frac{1}{3}Tr[T]^2\right); l=(pg^2+pd^2)^{1/2}.$$

Таблиця 1. Базові вектори («атака присутня» – «атака відсутня»)

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	y
0	2314	80	1573638018	-1580478590	1	1	401	3758	0, 1
0	1611	6101	8801886082	-926176166	1	1	0	2633	1, 0

Результати комп'ютерного моделювання наведені нижче:
sigma1 = –5.2682e+008, tau1 = 2.4979e+018;
sigma2 =-3.08724511e+008, tau2 = 0. 85785e+018.

Навіть поверхневий аналіз результатів показує високу ефективність використання тензорної методології до визначення атаки – при наявності атаки маємо mod (sigma2) > mod (sigma1) та tau1 > tau2. Цей ефект доцільно застосовувати при нейромережевій класифікації станів КМ (трафіка) на предмет виявлення атаки у вигляді правила:
if mod (sigma2) > mod (sigma1) &
& tau1 > tau2 then y={1, 0}.

Алгоритм визначення аномальних станів КМ може мати такий вигляд:

1^0. На обраній сітці ((3×3) для стандартних умов або (3×3) × 3×3) (для умов невизначеності, коли окремі параметри трафіка уявляються як нечіткі змінні), *j, i* =1, 9), формується тензор (у матричному вигляді) трафіка.

2^0. Обчислюються стандартні та ортогональні інваріанти отриманого тензора.

$$I_A = \sum_{i=1}^{m} A_{ii} = trace\ (\underline{A}),$$

$$II_A = \sum_{i=1}^{m}\sum_{j=1}^{m} A_{ij}A_{ji} = trace\ (\underline{A}\cdot\underline{A}),$$

$$III_A = \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} A_{ij}A_{jk}A_{ki} = trace\ (\underline{A}\cdot\underline{A}\cdot\underline{A}).$$

3^0.Створюється база даних трафік-тензорів та множини їх інваріантів $\{I_1^p, I_2^p, I_3^p\}, (\sigma^p, \tau^p), p=1,P$, де P – кількість об'єктів (аномальних станів), які потрібно розпізнавати.

4^0. Визначаються можливі діапазони зміни інваріантів,

$$I_i^p \pm \Delta I_i^p, p=1,P; i=1,3.$$

5^0. Перевіряється належність отриманих інваріантів пред'явленого трафіка до БД інваріантів $(\forall p)\ \min\sum_i I_i^p - I_i^L)$, $p\in P$. Той трафік, на якому досягається мінімум інваріантів, приймається як саме той трафік, який потрібно розпізнати.

Класифікацію можна виконувати або за всіма інваріантами відразу, або послідовно, починаючи з першого і звертаючись до наступного в разі невиконання критерія.

Висновки

1. Для ідентифікації атак на КМ трафік доцільно уявляти як тензор другого рангу (в умовах невизначеності як тензор більш високих рангів).

2. Інваріанти тензора трафіка, зокрема,

3. Ортогональні інваріанти (σ^p, τ^p), $p=1,P$ первинного тензора, дозволяють практично однозначно ідентифікувати аномальний стан (зокрема, атаку) на КМ .

4. Застосування НМе для ідентифікації атак на КМ иоже бути зведено до задачі класифікації тензорів трафіка та інваріантів цього тензора.

Список літератури

1. *Myron L. Cramer, James Cannady, Jay Harrell.* Methods of Intrusion Detection using Control-Loop Measurement /Technology for Information Systems Security – May, 16, 1996.

2. *Glenn Mansreld, Kohei Ohta, Y. Takei, N. Kato, Y. Nemoto.* Computer Networks 34 (2000) 659±670 (Cyber Solutions Inc., 6-6-3, Minami Yoshinari, Aobaku, Sendai, Japan Graduate School of Information Sciences, Tohoku University, Sendai, Japan). www.elsevier.com/locate/comnet

3. *Ф. Пинежанинов.* Математика в механической прочности. http://pinega.da.ru/