УДК 004.056.53

Грищук Р.В., Мамарєв В.М., Корченко О.Г.

ВЕРИФІКАЦІЯ МЕТОДУ СКОРОЧЕННЯ РОЗМІРНОСТІ ПОТОКУ ВХІДНИХ ДАНИХ ДЛЯ МЕРЕЖНИХ СИСТЕМ ВИЯВЛЕННЯ АТАК

У статті подано результати верифікації методу скорочення розмірності потоку вхідних даних для мережних систем виявлення атак.

Ключові слова: система захисту інформації, система виявлення атак, потік вхідних даних, верифікація.

Постановка проблеми. Зі стрімкою інтеграцією інформаційно-телекомунікаційних систем (ІТС) в усі сфери життя сучасного суспільства все важче стає оцінити цінність інформації, що циркулює і зберігається в них, та збитки в результаті її втрат чи несанкціонованого доступу. Тому системи захисту інформації (СЗІ), такі як антивірусне програмне забезпечення, системи виявлення атак (СВА) тощо, стали невід'ємною частиною ІТС [1]. СЗІ є складними програмно-апаратними комплексами, що функціонують на основі спеціальних математичних методів і моделей. Вони є основним інструментом забезпечення безпеки інформації, тому помилки їх функціонування – недопустимі. Висока надійність згаданих систем забезпечується перевірками програмних кодів та верифікацією методів та моделей. Підтвердження відповідності (верифікація) здійснюється математичними методами згідно висунутих на етапі специфікації критеріїв ефективності функціонування СЗІ. Аналіз цих критеріїв за рядом метрик дозволяє оцінити якість реалізації окремих елементів та системи в цілому, що робить задачу верифікації новостворених методів та моделей актуальною.

Аналіз останніх досліджень і публікацій [2–4] та ін. показує значну увагу фахівців до питання верифікації методів та моделей. У літературі [2, 3] виділяють дві групи методів верифікації: з повним та неповним моделюванням. Вимогою для застосування методів першої групи є відомі специфікація і реалізація [2]. Це робить застосування методів з повним моделюванням практично неможливим для дослідження систем, які не існують або недоступні для проведення експериментів. Зазначеного недоліку позбавлені методи верифікації з неповним моделюванням, для застосування яких достатньою умовою є наявність специфікації [3]. Іншою перевагою використання методів другої групи є можливість переходу від аналізу абсолютних показників якості функціонування до аналізу ключових критеріїв, що зменшує розмірність задачі багатокритеріальної оптимізації з конфліктуючими критеріальми функціями.

Метою статті є верифікація методу скорочення розмірності потоку вхідних даних для мережних СВА.

Основні матеріали дослідження. Зважаючи на те, що метод скорочення розмірності потоку вхідних даних для мережних СВА [5] не є орієнтованим на конкретну мережну СВА, тобто не має конкретної реалізації, для задач його верифікації скористаємося методами з неповним моделюванням і перейдемо до аналізу частинних критеріїв. Згідно висунутих на етапі специфікації вимог [5], розроблений метод має сформувати вхідний потік даних скороченої розмірності $x_{S'}$, що забезпечить фіксоване значення ефективності функціонування мережних СВА E_S^{IDS}, і полягає в такому.

На першому кроці методу шляхом групування параметрів мережного з'єднання формується структурна схема ієрархії частинних критеріїв для потоку вхідних даних мережних СВА.

Другий крок полягає в оцінюванні значень частинних критеріїв та їх коефіцієнтів пріоритету, при цьому значення частинних критеріїв нижчого рівня визначаються приростом інформації відповідного параметру [6], а коефіцієнти пріоритету – кількістю включених в оцінку критерію вищого рівня частинних критеріїв нижчого. При цьому всі частинні критерії мають бути нормованими і рівновагомими для однієї групи.

На третьому кроці з використанням математичного апарату вкладених скалярних згорток здійснюється композиція груп частинних критеріїв нижчого рівня ієрархії.

Четвертий крок методу – композиція частинних критеріїв вищого рівня ієрархії. Якісні оцінки груп властивостей на кроках три та чотири отримують шляхом співставлення аналітичних оцінок з адаптованою до умов поставленої задачі інтервальною зворотною нормованою фундаментальною шкалою [7].

На п'ятому, заключному, кроці методу визначається скорочена розмірність потоку вхідних даних та здійснюється оцінка його ефективності. Для скорочення розмірності потоку вхідних даних застосовується метод [8], згідно якого здійснюється "відсіювання" неінформативних параметрів, приріст інформації яких нижче заданого порогового значення, й операції з кроків 1–4 повторюються для скорочених множин різної розмірності.

Оцінка ефективності полягає у порівнянні якісних оцінок композиції критеріїв першого рівня ієрархії для розрахованих значень повної x_s і скорочених x_{s_i} множин контрольованих параметрів. Операції з кроків 1–4 повторюються доти, доки якісна оцінка скороченої розмірності потоку вхідних даних належатиме до тієї ж самої категорії, що й повна множина.

З метою верифікації розробленого методу за модель потоку вхідних даних для мережних СВА обрано базу KDD99 [9].

На основі обраної моделі потоку вхідних даних *на першому кроці* сформуємо структурну схему ієрархії множини контрольованих параметрів (рис. 1).

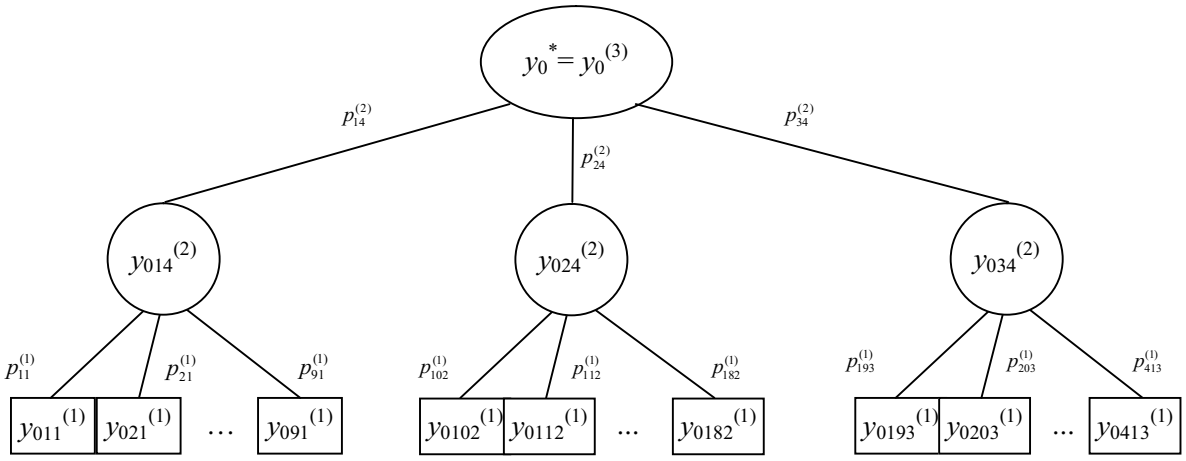

Рис. 1. Структурна схема ієрархії частинних критеріїв для множини контрольованих параметрів мережних СВА

Так, для вхідного потоку даних мережної СВА розрахунок кількісної $y_0^* = y_0^{(3)}$ і якісної оцінок здійснюється за сорок одним критерієм, що можуть бути об'єднані в три групи – група базових властивостей $y_{01}^{(2)}$, група властивостей контенту $y_{02}^{(2)}$ та група властивостей трафіку $y_{03}^{(2)}$. У свою чергу, групи базових властивостей та властивостей контенту оцінюються за дев'ятьма критеріями другого рівня, а група властивостей трафіку – двадцяти трьома. Сформовані групи не мають перехресних зв'язків. Вказані критерії нормовані і приведені до одного способу екстремізації – мінімізації. Критерії нижчого рівня залучаються до оцінювання критеріїв вищого з коефіцієнтами пріоритету $p_{ik}^{(j-1)}$, $j \in [2, m]$.

На другому кроці поставимо у відповідність критеріям нижчого рівня розраховані у [7] значення приросту інформації (табл. 1).

Коефіцієнти пріоритету $p_{ik}^{(j-1)}$ визначаються кількістю включених в оцінку критерію вищого рівня критеріїв нижчого, при цьому всі критерії нижчого рівня для однієї групи є нормованими, рівновагомими і набувають наступних значень:

$$p_{11}^{(1)} = p_{21}^{(1)} = p_{31}^{(1)} = p_{41}^{(1)} = p_{51}^{(1)} = p_{61}^{(1)} = p_{71}^{(1)} = p_{81}^{(1)} = p_{91}^{(1)} = \frac{1}{9} = 0{,}1111,$$

$$p_{102}^{(1)} = p_{112}^{(1)} = p_{122}^{(1)} = p_{132}^{(1)} = p_{142}^{(1)} = p_{152}^{(1)} = p_{162}^{(1)} = p_{172}^{(1)} = p_{182}^{(1)} = \frac{1}{9} = 0{,}1111,$$

$$p_{193}^{(1)} = p_{203}^{(1)} = p_{213}^{(1)} = p_{223}^{(1)} = p_{233}^{(1)} = p_{243}^{(1)} = p_{253}^{(1)} = p_{263}^{(1)} = p_{273}^{(1)} = p_{283}^{(1)} = p_{293}^{(1)} = p_{303}^{(1)} = \quad (1)$$

$$= p_{313}^{(1)} = p_{323}^{(1)} = p_{333}^{(1)} = p_{343}^{(1)} = p_{353}^{(1)} = p_{363}^{(1)} = p_{373}^{(1)} = p_{383}^{(1)} = p_{393}^{(1)} = p_{403}^{(1)} = p_{413}^{(1)} = \frac{1}{23} = 0{,}0434,$$

$$p_{13}^{(2)} = p_{23}^{(2)} = p_{33}^{(2)} = \frac{1}{3} = 0{,}3333.$$

Значення частинних критеріїв Таблиця 1

Група	№ параметру	Частинний критерій	Приріст інформації
Група базових властивостей	1	$y_{011}^{(1)}$	0,1257
	2	$y_{021}^{(1)}$	0,3877
	3	$y_{031}^{(1)}$	0,6732
	4	$y_{041}^{(1)}$	0,1340
	5	$y_{051}^{(1)}$	0,7241
	6	$y_{061}^{(1)}$	0,6127
	7	$y_{071}^{(1)}$	0,0702
	8	$y_{081}^{(1)}$	0,0702
	9	$y_{091}^{(1)}$	0,0702
Група властивостей контенту	10	$y_{0102}^{(1)}$	0,0735
	11	$y_{0112}^{(1)}$	0,0703
	12	$y_{0122}^{(1)}$	0,4858
	13	$y_{0132}^{(1)}$	0,0706
	14	$y_{0142}^{(1)}$	0,0705
	15	$y_{0152}^{(1)}$	0,0702
	16	$y_{0162}^{(1)}$	0,0730
	17	$y_{0172}^{(1)}$	0,0713
	18	$y_{0182}^{(1)}$	0,0705
Група властивостей трафіку	19	$y_{0193}^{(1)}$	0,0724
	20	$y_{0203}^{(1)}$	0,0702
	21	$y_{0213}^{(1)}$	0,0702
	22	$y_{0223}^{(1)}$	0,0730
	23	$y_{0233}^{(1)}$	0,7201
	24	$y_{0243}^{(1)}$	0,4328
	25	$y_{0253}^{(1)}$	0,1342
	26	$y_{0263}^{(1)}$	0,1375
	27	$y_{0273}^{(1)}$	0,0721
	28	$y_{0283}^{(1)}$	0,0736
	29	$y_{0293}^{(1)}$	0,1548
	30	$y_{0303}^{(1)}$	0,1551
	31	$y_{0313}^{(1)}$	0,2325
	32	$y_{0323}^{(1)}$	0,3848
	33	$y_{0333}^{(1)}$	0,2580
	34	$y_{0343}^{(1)}$	0,2356
	35	$y_{0353}^{(1)}$	0,2458
	36	$y_{0363}^{(1)}$	0,4733
	37	$y_{0373}^{(1)}$	0,3459
	38	$y_{0383}^{(1)}$	0,1506
	39	$y_{0393}^{(1)}$	0,1557
	40	$y_{0403}^{(1)}$	0,0871
	41	$y_{0413}^{(1)}$	0,1003

Третій крок. Розрахунок оцінки базових властивостей (другий рівень ієрархії). Адаптована рекурентна формула [7] має вигляд:

$$y_{01}^{(2)} = 1 - \frac{1}{\sum_{i=1}^{9} p_{i1}^{(1)} (1 - y_{0i1}^{1})^{-1}}. \quad (2)$$

Підставивши числові значення критеріїв та коефіцієнти їх ваги, отримаємо:

$$y_{01}^{(2)} = 1 - \left(\begin{array}{l} 0{,}1111\frac{1}{1-0{,}1257} + 0{,}1111\frac{1}{1-0{,}3877} + 0{,}1111\frac{1}{1-0{,}6732} + \\ +0{,}1111\frac{1}{1-0{,}134} + 0{,}1111\frac{1}{1-0{,}7241} + 0{,}1111\frac{1}{1-0{,}6127} + \\ +0{,}1111\frac{1}{1-0{,}0702} + 0{,}1111\frac{1}{1-0{,}0702} + 0{,}1111\frac{1}{1-0{,}0702} \end{array} \right)^{-1} = 0{,}452. \quad (3)$$

Співставивши отриману аналітичну оцінку з адаптованою інтервальною зворотною нормованою фундаментальною шкалою (далі – шкала), визначимо, що опис групою базових властивостей належить до категорії задовільної якості.

Вираз для кількісної оцінки властивостей контенту (друга група критеріїв) має вигляд:

$$y_{02}^{(2)} = 1 - \frac{1}{\sum_{i=10}^{18} p_{i2}^{(1)} (1 - y_{0i2}^{1})^{-1}}. \quad (4)$$

Підставивши числові значення, отримаємо:

$$y_{02}^{(2)} = 1 - \left(\begin{array}{l} 0{,}1111\frac{1}{1-0{,}0735} + 0{,}1111\frac{1}{1-0{,}0703} + 0{,}1111\frac{1}{1-0{,}4858} + \\ +0{,}1111\frac{1}{1-0{,}0706} + 0{,}1111\frac{1}{1-0{,}0705} + 0{,}1111\frac{1}{1-0{,}0702} + \\ +0{,}1111\frac{1}{1-0{,}073} + 0{,}1111\frac{1}{1-0{,}0713} + 0{,}1111\frac{1}{1-0{,}0705} \end{array} \right)^{-1} = 0{,}1475 \cdot \quad (5)$$

Співставлення аналітичної оцінки зі шкалою показує, що категорія якості опису стану ІТС групою властивостей контенту оцінюється як задовільна.

Вираз для розрахунку аналітичної оцінки властивості трафіку (другий рівень ієрархії), відповідно до сформованої структурної схеми ієрархії множини контрольованих параметрів, має вигляд:

$$y_{03}^{(2)} = 1 - \frac{1}{\sum_{i=19}^{41} p_{i3}^{(1)} (1 - y_{0i3}^{1})^{-1}}. \quad (6)$$

Підставивши числові значення, отримаємо:

$$y_{03}^{(2)} = 1 - \left(\begin{array}{l} 0{,}0434\frac{1}{1-0{,}0724} + 0{,}0434\frac{1}{1-0{,}0702} + 0{,}0434\frac{1}{1-0{,}0702} + 0{,}0434\frac{1}{1-0{,}073} + \\ +0{,}0434\frac{1}{1-0{,}7201} + 0{,}0434\frac{1}{1-0{,}4328} + 0{,}0434\frac{1}{1-0{,}1342} + 0{,}0434\frac{1}{1-0{,}1375} + \\ +0{,}0434\frac{1}{1-0{,}0721} + 0{,}0434\frac{1}{1-0{,}0736} + 0{,}0434\frac{1}{1-0{,}1548} + 0{,}0434\frac{1}{1-0{,}1551} + \\ +0{,}0434\frac{1}{1-0{,}2325} + 0{,}0434\frac{1}{1-0{,}3848} + 0{,}0434\frac{1}{1-0{,}258} + 0{,}0434\frac{1}{1-0{,}2356} + \\ +0{,}0434\frac{1}{1-0{,}2458} + 0{,}0434\frac{1}{1-0{,}4733} + 0{,}0434\frac{1}{1-0{,}3459} + 0{,}0434\frac{1}{1-0{,}1506} + \\ +0{,}0434\frac{1}{1-0{,}1557} + 0{,}0434\frac{1}{1-0{,}0871} + 0{,}0434\frac{1}{1-0{,}1003} \end{array} \right)^{-1} = 0{,}2915. \quad (7)$$

Згідно шкали розрахована якісна оцінка опису групою властивостей контенту – «задовільно».

На четвертому кроці здійснимо композицію частинних критеріїв першого рівня ієрархії. Вираз для розрахунку кількісної оцінки композиції критеріїв першого рівня ієрархії має вигляд:

$$y_0^* = y_0^{(3)} = 1 - \frac{1}{\sum_{i=1}^{3} p_{i4}^{(2)} (1 - y_{0i4}^{(2)})^{-1}}. \qquad (8)$$

Підставивши початкові та розраховані числові значення, отримаємо:

$$y_{04}^{(2)} = 1 - \frac{1}{0{,}3333\frac{1}{1-0{,}452} + 0{,}3333\frac{1}{1-0{,}1475} + 0{,}3333\frac{1}{1-0{,}2915}} = 0{,}3196. \qquad (9)$$

Якісна оцінка ефективності описання станів ІТС повною множиною параметрів оцінюється як «добре».

На п'ятому, заключному, кроці методу здійснимо оцінку його ефективності. Формування скорочених множин контрольованих параметрів виконано "відсіюванням" неінформативних параметрів, приріст інформації яких нижче заданого порогового значення, й повторено операції кроків 1–4. Результати розрахунків якісних оцінок композиції критеріїв першого рівня ієрархії для значень повної та скорочених множин контрольованих параметрів представлено в табл. 2.

Кількісні та якісні оцінки для повної та скорочених множин параметрів Таблиця 2

Розмірність множини параметрів	Оцінка групи базових властивостей		Оцінка групи властивостей контенту		Оцінка групи властивостей трафіку		Оцінка ефективності описання станів ІТС множиною параметрів	
	Кількісна	Якісна	Кількісна	Якісна	Кількісна	Якісна	Кількісна	Якісна
41	0,4520	Задовільна	0,1475	Висока	0,2915	Добра	0,3196	Добра
30	0,5453	Низька	0,2688	Добра	0,2817	Добра	0,3951	Добра
28	0,5453	Низька	0,2688	Добра	0,2984	Добра	0,3991	Добра
20	0,5852	Низька	0,4858	Задовільна	0,3396	Добра	0,4889	Задовільна
11	0,6330	Низька	0,4858	Задовільна	0,4887	Задовільна	0,5472	Низька

В результаті проведеного аналізу встановлено, що зменшення кількості контрольованих параметрів призводить до погіршення якості описання станів ІТС. Для зменшеної множини контрольованих параметрів, що належить до категорії якості не нижче «добре», оптимальна розмірність потоку вхідних даних мережних СВА – 28 контрольованих параметрів. Так, за умови застосування у СВА алгоритмів з обчислювальною складністю $O(n^2)$ (методи нейронних та імунних мереж, статистичного та кластерного аналізу), використання методу скорочення розмірності потоку вхідних даних для мережних систем виявлення атак дозволить скоротити кількість обчислень не менше ніж в

$$\frac{O(x_S^2)}{O(x_{S'}^2)} = \frac{O(41^2)}{O(28^2)} = 2{,}14 \text{ рази.}$$

Висновки. Верифікація методу скорочення розмірності потоку вхідних даних для мережних систем виявлення атак, на основі моделі KDD99, доводить його працездатність і

ефективність. Використання математичного апарату вкладених скалярних згорток професора Вороніна А. М. дозволяє скорочувати розмірність вхідних потоків мережних СВА навіть за умови недостатності експериментально-статистичних даних. Універсальність запропонованого методу пояснюється тим, що він не є орієнтованим на СВА будь-якого типу, а дозволяє сформувати скорочену множину інформативних параметрів, що відображає стани ІТС, без погіршення якості їх описання. За результатами аналізу встановлено, що застосування методу для алгоритмів СВА з обчислювальною складністю $O(n^2)$ дозволяє скоротити кількість обчислень не менш ніж в 2,14 рази, не погіршуючи якості описання станів ІТС.

ЛІТЕРАТУРА

1. Ленков С. В. Методы и средства защиты информации : монография [в 2-х т.] Т. 2. Информационная безопасность / С. В. Ленков, Д. А. Перегудов, В. А. Хорошко. – К. : Арий, 2008. – 344 с.
2. Кларк Э. М. Верификация моделей программ: Model Cheking / Э. М. Кларк, О. Грамберг, Д. Пелед [пер. с англ. / Под ред. Р. Смелянского]. – М. : МЦНМО, 2002. – 416 с.
3. Томашевський В. М. Моделювання систем / В. Томашевський. – К. : Видавнича група BHV, 2005. – 352 с.
4. Balci O. Verification, Validation and Accreditation of Simulation Models / O. Balci // Proceeding of the 29th conference on Winter simulation. – N.Y. :ACM Press, 1997. – P. 135-141;
5. Грищук Р. В. Постановка задачі розробки методики скорочення розмірності потоку вхідних даних для мережних систем виявлення атак / Р. В. Грищук, В. М. Мамарєв // Інформаційна безпека. – Луганськ : СНУ ім. В. Даля, 2011. – № 1 (5). – С. 74–78.
6. Грищук Р. В. Метод оцінювання інформативності параметрів потоку вхідних даних для мережних систем виявлення атак / Р. В. Грищук, В. М. Мамарєв // Системи обробки інформації. – Х. : ХУПС ім. І. Кожедуба, 2012. – №4 (102). – С. 103–108.
7. Воронин А. Н. Многокритериальные решения: модели и методы : монографія / А. Н. Воронин, Ю. К. Зиатдинов, М. В. Куклинский. – К. : НАУ, 2011. – 348 с.
8. Грищук Р. В. Диференціально-ігровий метод оцінювання ефективності систем захисту інформації / Р. В. Грищук // Сучасний захист інформації. – К. : ДУІКТ, 2012. – № 1 (10). – С. 40–44.
9. UCI Knowledge Discovery in Databases Archive : [Електронний ресурс]. – Режим доступу : http://kdd.ics.uci.edu.

Надійшла: 12.08.2012р.

Рецензент: д.т.н., проф. Скрипник Л.В.