«Системні технології» 2 (139) 2022 «System technologies»
DOI 10.34185/1562-9945-2-139-2022-02
УДК 004.942: 519.711.3

Д.С. Астахов, В.Б. Мазуренко, А.І. Федорович, Н.О. Лисенко

МЕТОДИ ВИЯВЛЕННЯ СТАТИСТИЧНИХ ЗМІН ХАРАКТЕРИСТИК МЕРЕЖЕВОГО ТРАФІКУ

Анотація. Досліджено інформативність критерію Шовене для виявлення викидів у вибірках випадкових величин із законом розподілу ймовірності, що відрізняється від нормального. Обчислювальний експеримент показав, що критерій Шовене має високу потужність для вибірок з симетричним, слабо асиметричним і сильно асиметричним законом розподілу ймовірності. Враховуючи цю властивість, даний критерій можна використовувати в системах виявлення вторгнень для підвищення рівня кібербезпеки інформаційних систем.

Ключові слова: інформаційна система, мережевий трафік, статистичні методи, кібербезпека.

Вступ. В інформаційних системах обмін даними між складовими об'єктами відбувається через мережеве з'єднання. Надійність мережевого з'єднання в багатьох випадках залежить від організації безпечного середовища передачі даних. Безпека даних, що передаються, має три складових – цілісність, автентичність та доступність. На практиці реалізація зазначених вимог здійснюється за допомогою програмно-апаратних комплексів та організаційних заходів.

Тобто, порушення безпеки інформаційної системи може бути викликано неправильною роботою обладнання або внаслідок втручання сторонньої людини. Перший тип причин долається за допомогою оновлення програмного забезпечення та/або за допомогою зміни налаштувань мережевого обладнання. Другий тип порушення безпеки більш складний. Для його виявлення потрібні нові більш досконалі інтелектуальні алгоритми. Вони розробляються на основі статистичних, нейронних, машинних та інших методів. В залежності від рівня абстракції, на якому вирішується задача виявлення втручання в роботу ІС, використовують відповідні методи для їх виявлення [2].

Мережевий трафік в ІС характеризується певними кількісними характеристиками. Ці характеристики мають певні статистичні властивості і закономірності. Тому, можна вважати, що статистичні характеристики мере-

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

жевого трафіку відображають робочий стан ІС. Це дає підставу для розробки нових і вдосконаленню існуючих методів виявлення змін статистичних характеристик мережевого трафіку. Якщо метрики мережевого трафіку роздивлятися як часовий ряд випадкових величин, то для їх дослідження застосовні будь-які методи математичної статистики [3].

Ціль дослідження. Перевірити інформативність критерію Шовене по виявленню викидів в вибірках випадкових величин з законом розподілу ймовірності, який відрізняється від нормального. Як звісно, на практиці ніколи не відомо, які статистичні закономірності мають величини, які спостерігаються. Для того, щоб застосувати той чи інший статистичний критерій для виявлення змін у випадковому процесі (часовому ряді випадкових величин) в багатьох випадках потрібне додаткове дослідження на належність до певного закону розподілу ймовірності. Для цього потрібен досить великий об'єм вибірки а значить і тривалий час для її отримання. Для вирішення оперативних задач це може бути неприємним. Тому бажано використовувати критерії виявлення статистичних відмінностей, які не чутливі до закону розподілу ймовірностей. Знайти такі критерії можна за допомогою обчислювального експерименту та математичного моделювання.

Постановка задачі. Розглянемо задачу виявлення викидів в коротких послідовностях ($n = 10...40$) випадкових величин методом Шовене [1]. Цей метод має достатню інформативність і простий в реалізації. Оскільки заявлено, що він добре спрацьовує для вибірок з нормальним законом розподілу ймовірності, треба дослідити його інформативність для вибірок з іншими видами розподілу. Для обчислювального експерименту були обрані три види законів – симетричний, слабо асиметричний і сильно асиметричний, - логістичний, Релея і експоненційний відповідно.

Викидами, як правило, стають мінімальні і максимальні значення вибірок *x(k)*. Для їх виявлення формують впорядковану вибірку $x^*(1) < x^*(2) < x^*(3) \ldots < x^*(i-1) < x^*(i) \ldots < x^*(n-1) < x^*(n)$, тоді $x^*(1) = x_{\min}$, $x^*(n) = x_{\max}$. Критерій Шовене вибірки *x(k),* перетвореної в упорядковану вибірку визначається за формулою:

$$y_1(j/max) = \frac{\left| x^*(n/j) - \bar{x}(j) \right|}{s(j)},$$

$$y_2(j/min) = \frac{\left| x^*(1/j) - \bar{x}(j) \right|}{s(j)},$$

де, $\bar{x}(j)=\frac{1}{n}\sum_{k=1}^{n}x_k$; $s^2(j)=\sigma^2=\frac{1}{n-1}\sum_{k=1}^{n}\left(x(k/j)-\bar{x}(j)\right)^2$; j – номер експерименту; k – номери вимірювання.

Рішення. Формуємо вибірки максимальних $y_1(j)$ та мінімальних $y_2(j)$ значень для експериментів j=1,2,...N; N=100000; довжини вибірок n=10, 25, 50. Емпіричні значення $y_1(j), y_2(j)$ порівняємо з критичними значеннями, які наведені в таблиці 1 [1].

$$y_1(j/n)\le y_0 \; ma \; y_2(j/n)\le y_0$$

Таблиця 1

Критичні значення (K*) критерія Шовене

n	4	5	6	10	15	25	50	100	300
K*	1,54	1,65	1,73	1,96	2,13	2,33	2,57	2,81	3,14

Математична модель прийняття рішення про стан вибірки $y(j/n)$ (без викидів)

$$y_1(j/n)\le K^* \; ma \; y_2(j/n)\le K^*$$

Запишеться у вигляді:

$$\operatorname{sgn}\left(y_1(j/n)-y_0(n)\right)=1,\ \operatorname{sgn}\left(y_2(j/n)-y_0(n)\right)=1,\ j=1,2,\ldots,N$$

Отож їх сума $\frac{K_1}{n}$ та $\frac{K_2}{n}$:

$$\frac{K_1}{n}=\frac{1}{n}\sum_{j=1}^{N}\operatorname{sgn}\left(y_1(j/n)-y_0(n)\right),$$

$$\frac{K_2}{n}=\frac{1}{n}\sum_{j=1}^{N}\operatorname{sgn}\left(y_2(j/n)-y_0(n)\right).$$

Це оцінки ймовірності прийняття рішень (по *x(max)* $\frac{K_1}{n}$ та по *x(min)* $\frac{K_2}{n}$), які за теоретичними дослідженнями вибірок випадкових величин за нормальним (логістичним) розподілом ймовірностей які повинні бути не більше ніж $\frac{1}{12n}$

$$P\left(K<K^*\right)<\frac{1}{12n}$$

Проведено обчислювальний експеримент за трьома видами законів розподілу: логістичний, експоненційний та Релея. Побудуємо гістограми

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

критерію та визначмо його основні статистичні характеристики. Параметри законів розподілу ймовірностей доберемо таким чином, щоб вони мали однакові значення математичних очікувань і, по можливості, однакові значення дисперсій. За основу візьмемо параметр $a = 5$ логістичного закону розподілу. Тоді параметр b для закону Релея дорівнює

$$b = a / \left(\sqrt{\pi / 2} \right) = 5 / 1.253 \approx 4 .$$

Відповідно параметр σ для логістичного закону дорівнює

$$\sigma = \sqrt{\frac{4-\pi}{2} b^2} = \sqrt{\frac{4-\pi}{2} 16} \approx 4 .$$

Для експоненційного закону розподілу параметр $\lambda = a = 5$.

За таких умов маємо наступні значення математичних очікувань і дисперсій для обраних законів розподілу ймовірностей.

Таблиця 2

	Логістичний	Релея	Експоненційний
$M[x]$	5	5	5
$D[x]$	15,5	15,5	25

В таблиці 3 наведено емпіричні ймовірності P^* прийняття рішення про відсутність викидів в вибірках випадкових величин.

Таблиця 3

n	Логістичний	Релея	Експоненційний
10	0,963	0,971	0,945
25	0,977	0,984	0,963
50	0,984	0,989	0,971

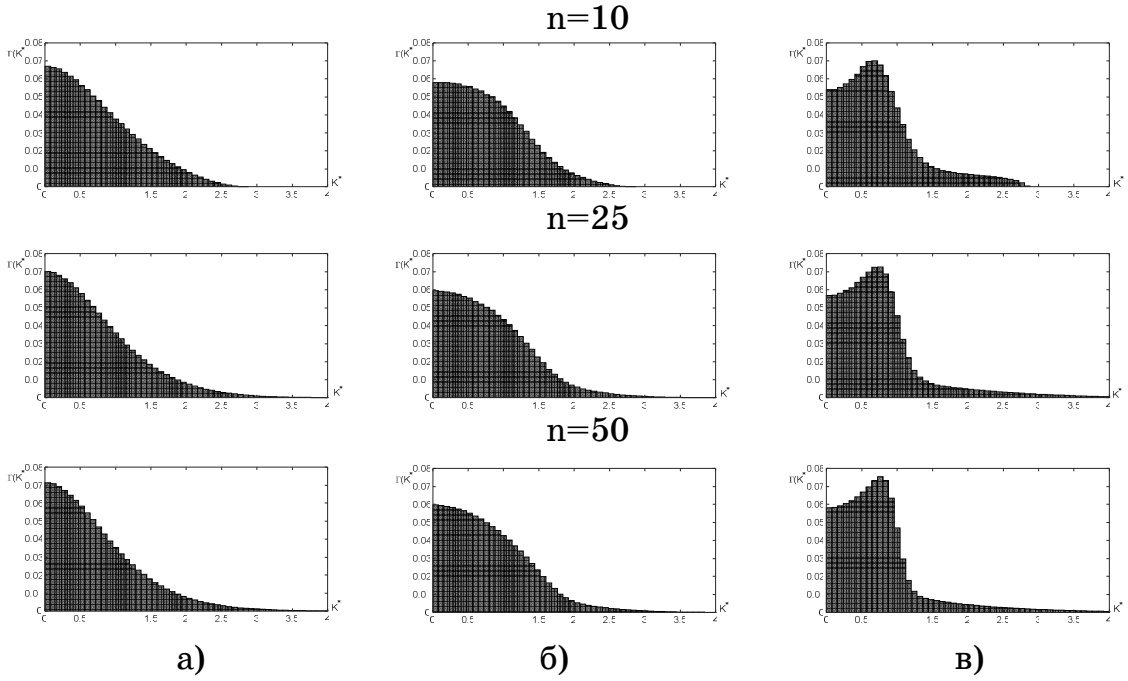

Рисунок 1 - Гістограми критерію Шовене для довжини вибірок $n = 10, 25, 50$ з логістичним а), Релеєвскім б), експоненційним в) законами розподілу ймовірностей

Висновки. Аналізуючи результат обчислювального експерименту, можна зробити висновок, що критерій Шовене, виявлення викидів в вибірках випадкових величин, має схожу статистику для законів розподілу ймовірностей від'ємних від нормального. Відмінності статистики спостерігаються для випадкових величин з експоненційним законом розподілу. В середині гістограми присутній «пагорб», але потім йде різкий спад. Ця особливість не впливає на результат прийняття правильного рішення. Тобто, цей критерій можна використовувати без додаткового дослідження тестової вибірки на «нормальність». До того ж він добре спрацьовує навіть на дуже коротких вибірках з $n = 10$.

Критерій Шовене можливо використовувати для виявлення викидів метрик мережевого трафіку в інформаційних системах. Він може входити до складу автоматизованих систем виявлення вторгнення (СВВ). Можливість роботи з вибірками малого розміру дозволяє скоротити час накопичування даних і тим самим зменшити час на виявлення змін стану системи. Що, в свою чергу, дозволить своєчасно приймати превентивні міри для запобігання зовнішньому втручанню в роботу ІС, тобто підвищується загальний рівень кібербезпеки системи.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

ЛІТЕРАТУРА

1. Кобзарь А. И. Прикладная математическая статистика/ А. И. Кобзарь. – М. : ФИЗМАТЛИТ, 2006. – 816 с.

2. Толюпа С.В., Штаненко С.С., Берестовенко Г.В. Класифікаційні ознаки систем виявлення атак та напрямки їх побудови. Збірник наукових праць ВІТІ № 3 – 2018.

3. Малайчук В.П., Клименко С.В., Астахов Д.С. Компьютерные информационные технологии обработки измерений в задачах наблюдения и контроля. «Системні технології» 4 (129) 2020.

REFERENCES

1. Kobzar A. I. Applied mathematical statistics/ A. And yu Kobzar - M. : FIZMATLIT, 2006. - 816 p.

2. Tolyupa S.V., Shtanenko S.S., Berestovenko G.V. Klasifikatsiyni signs of systems for detecting attacks that are direct xx pobudovi. Zbirnik naukovikh prats VITI No. 3 - 2018.

3. Malaychuk V.P., Klimenko S.V., Astakhov D.S. Computer information technologies of measurement processing in monitoring and control tasks. "System Technology" 4 (129) 2020.

Received 04.02.2022.
Accepted 07.02.2022.

Methods for detecting statistical changes in network traffic characteristics

Network flows of information systems (IS) are characterized by certain quantitative characteristics. They contain information about network load, the quality of communication between nodes, and many other service information. For example, an attacker can use this service information to prepare for a cyberattack. When an attack is already being carried out, network traffic is filled with additional atypical information. In other words, the values of its quantitative characteristics change. Therefore, quantitative indicators of these characteristics can indirectly monitor the atypical behavior of network nodes. For example, by the number of requests of the same type per unit of time. Such an event may occur, for example, when the network scanner is running or during a denial-of-service attack. To detect such events, special software packages are used, such as intrusion detection systems. These systems use a variety of algorithms in their work, which are based on statistical methods, neural networks, fuzzy logic Automata, and others. The type of mathematical processing depends on the complexity of the problem, the level of the protocol being observed, and the preferences of the detection system developer. Sometimes statistical methods for analyzing metrics can be simpler and faster to implement than others, because they do not contain a large number of mathematical operations. This can allow you to monitor the state of the IP in real time. And timely detection of changes in the state of the IP allows you to avoid malfunctions. One of the signs of a change in the system state is the release of values of quantitative indicators of network traffic parameters. Therefore, this task is urgent

and requires further development and improvement. In this paper, it is investigated that using the Chauvet criterion, it is possible to detect outliers with high probability in small time series with a probability distribution that differs from the normal one.

Астахов Дмитро Сергійович - старший викладач кафедри радіоелектронної автоматики, фізико-технічний факультет, Дніпровський національний університет імені Олеся Гончара.

Мазуренко Валерій Борисович - доцент кафедри радіоелектронної автоматики, фізико-технічний факультет, Дніпровський національний університет імені Олеся Гончара.

Федорович Ганна Ігорівна - доцент кафедри радіоелектронної автоматики, фізико-технічний факультет, Дніпровський національний університет імені Олеся Гончара.

Лисенко Наталія Олександрівна - доцент кафедри радіоелектронної автоматики, фізико-технічний факультет, Дніпровський національний університет імені Олеся Гончара.

Astakhov Dmitry Sergeevich - Oles Honchar Dnipro National University, Physical and Technical Faculty, Senior Lecturer of Chair of Radioelectronic Automation.

Mazurenko Valeriy Borisovich – Oles Honchar Dnipro National University, Physical and Technical Faculty, Doctor of Philosophy, docent of Chair of Radioelectronic Automation.

Fedorovych Anna Igorivna - Oles Honchar Dnipro National University, Physical and Technical Faculty, docent of Chair of Radioelectronic Automation.

Lysenko Nataliia Oleksandrivna - Oles Honchar Dnipro National University, Physical and Technical Faculty, docent of Chair of Radioelectronic Automation.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)