Scientific and Practical Cyber Security Journal (SPCSJ) 3(1): 58-65 ISSN 2587-4667 Scientific Cyber Security Association (SCSA)

ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ ВЫЯВЛЕНИЯ СЕТЕВЫХ АТАК ЗА СЧЕТ АЛГОРИТМОВ РАСПОЗНАВАНИЯ ВРЕДОНОСНОГО ТРАФИКА

Толюпа С.В. [1], Наконечный В.С [2]., Вялкова, В.И. [3], Войтенко И.Ю. [4]
1-4 Киевский национальный университет Тараса Шевченка, Киев, Украина

Вступление. Последние достижения информационно-коммуникационных и сетевых технологий, широкое применение Internet для обмена информации различной степени конфиденциальности между объектами критически важной инфраструктуры существенно повышают эффективность их функционирования. Однако, на ряду с этим наблюдается постоянный рост кибератак в виде вредоносного сетевого трафика, который злоумышленники могут использовать для компрометации информации в компьютерных сетях. Поэтому защита информационных ресурсов от кибератак является чрезвычайно важной и актуальной проблемой современности.

В настоящее время существуют различные инструменты и механизмы, направленные на обеспечение информационной безопасности. Известными решениями являются: межсетевые экраны, антивирусные программы, системы обнаружения вторжений и т. п. Но, к сожалению, эти решения не всегда являются достаточно эффективными. Поэтому была предложена и разработана родственная по свойствам к системе обнаружения вторжений (IDS) система предотвращения вторжений (IPS) - технология предупреждения сетевой безопасности, которая исследует потоки сетевого трафика для обнаружения и предотвращения кибератак [1].

Описание проблемы. Традиционные системы для обнаружения вредоносного трафика с целью предотвращения вторжений на информационную систему в основном используют "сигнатурный" метод, который требует определения уникального признака для каждого типа атаки. При этом каждая новая сигнатура сначала добавляется в банк эталонных признаков (базу данных) IPS, а затем, для дальнейшего распознавания и обнаружения сетевой атаки выполняется сравнение признаков входящего трафика с соответствующими эталонными признаками базы данных IPS.

Создание и обновление сигнатур обычно осуществляется с помощью анализа соответствующих экспертов, при этом существует несколько проблем, связанных с этим методом [2]:

- система должна быть сначала скомпрометирована для того, чтобы были известны основные признаки вредоносного трафика;
- для каждой новой кибератаки требуется определение новой сигнатуры.

Более того, в отдельных сценариях кибератак, IPS которая базируется на сигнатурном методе, не гарантирует достаточно быстрого обнаружения признаков вредоносного трафика, что связано с затраченным временем на распознавание одного признака и, как следствие, некоторые пакеты вредоносного трафика могут быть пропущены. Последнее обстоятельство может привести к скрытой компрометации информационной системы [3].

Как и каждый программный продукт, IPS требует значительных вычислительных ресурсов, а именно: больших объемов оперативной памяти и мощного процессора. На сегодня уже существует ряд алгоритмов обнаружения признаков вредоносного трафика (ОПВТ), по которым работают IPS. Однако, их применение существенно нагружает систему в целом, и гарантировать, что работа IPS будет максимально эффективной и наименее требовательной к вычислительным ресурсам, невозможно. Для этого алгоритмы

ОПВТ должны быть эффективными с точки зрения обеспечения вероятности правильного распознавания (ВПР) вредоносного трафика и работать в масштабе времени, приближенного к реальному.

Поэтому вопрос выбора эффективного с точки зрения быстродействия алгоритма ОПВТ является чрезвычайно актуальным.

Именно с этой целью в данной работе проведен сравнительный анализ возможных алгоритмов распознавания признаков вредоносного сетевого трафика, которые могут быть применены в перспективных системах IPS.

На (рис.1) представлена модель архитектуры работы IPS, в которой возможно применение алгоритмов обнаружения признаков вредоносного сетевого трафика.

Как известно [4], работа системы IPS зависит от эффективности методов распознавания, а именно от времени, которое тратится на эту процедуру. Поэтому важным вопросом является выбор наиболее эффективного алгоритма ОПВТ с точки зрения минимизации времени, которое затрачивается на процесс распознавания.

В настоящее время, по мнению авторов данной работы, существует несколько возможных методов ОПВТ, а именно [5]:

- алгоритм Кейпона;
- алгоритм расстояния Махаланóбиса;
- алгоритм Байесса;
- алгоритм теплового шума;
- корреляционный алгоритм.

Эффективность работы этих алгоритмов зависит от того, по какому закону распределено пространство векторов признаков. Именно поэтому анализ работы предложенных алгоритмов проведен исходя из того условия, что признаки обнаружения вредоносного сетевого трафика могут иметь различные законы распределения.

Рис. 1 – Архитектура работы IPS сигнатурного метода при использовании алгоритмов ОПВТ

Законом распределения вектора признаков распознавания является соотношение между возможными значениями случайных величин и соответствующими им вероятностями. В качестве наиболее вероятных законов распределения признаков вредоносного трафика могут выступать нормальный и закон распределения Лапласа. Учитывая возможность сложной текущей сетевой ситуации, законы распределения погрешностей измерения некоторых признаков можно считать равномерными. Поэтому, при дальнейшем анализе эффективности предлагаемых алгоритмов будем использовать указанные законы распределения вектора признаков: нормальный, равномерный и закон распределения Лапласа.

Нормальный закон распределения случайной величины широко применяется при решении практических задач [6]. Случайная величина x подлежит нормальному закону, если плотность вероятности этой случайной величины соответствует выражению (1).

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad (1)$$

где e = 2,71828 - основа натурального логарифма;

π = 3,14159;

m і σ - параметры распределения, определяемые по результатам испытаний.

График функции плотности вероятности $f(x)$ имеет вид (рис.2).

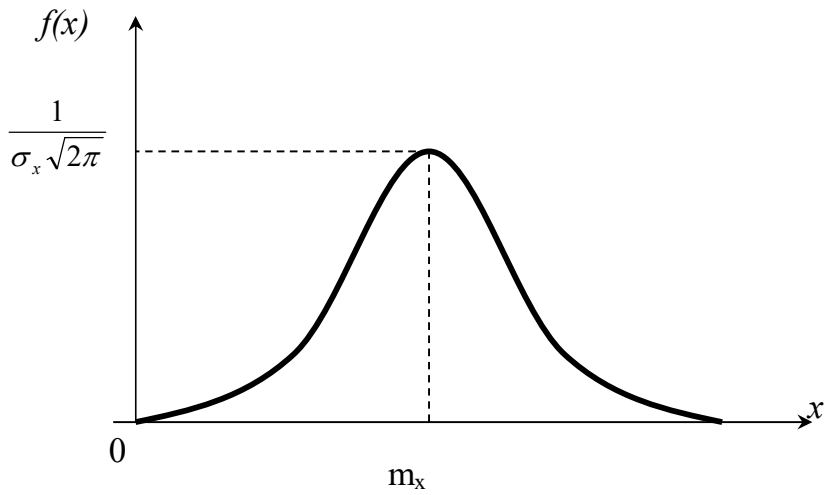

Рис.2 - График функции плотности вероятности при нормальном законе распределения

Закон распределения Лапласа (двойное экспоненциальное распределение). Случайная величина x имеет распределение Лапласа с параметрами (α, λ) и ($\lambda > 0$), если она имеет плотность распределения, соответствующую (2).

$$f(x) = \frac{l}{2} e^{-l|x-a|} \quad (2)$$

Плотность распределения случайной величины, которая распределена по закону распределения Лапласа, изображена на (рис. 3).

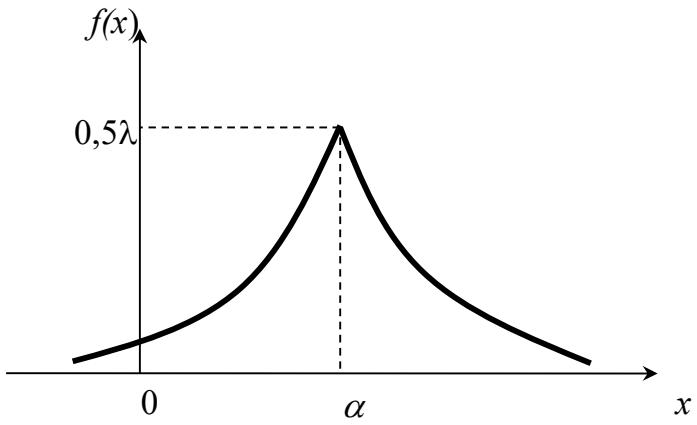

Рис. 3 - Плотность распределения случайной величины по закону распределения Лапласа

Если вероятность попадания случайной величины на интервал пропорциональна длине интервала и не зависит от расположения интервала на оси, то она имеет равномерный закон распределения. Плотность такого распределения представлена как (3):

$$f(x)=\begin{cases}0, x<a,\\ c, a\le x\le b,\\ 0, x>b\end{cases} \tag{3}$$

Непрерывная случайная величина *x* подлежит равномерному закону распределения на отрезке [*a, b*], если на этом отрезке плотность распределения случайной величины равна постоянной, а вне его равна нулю, таким образом функция имеет вид (4):

$$F(x)=\begin{cases}0, x<a,\\ \dfrac{x-a}{b-a}, 0<x<b,\\ 1, x\ge b.\end{cases} \tag{4}$$

Для случайной величины *X*, которая распределена по равномерному закону, дисперсия и математическое ожидание будут рассчитываться исходя из следующих выражений.

$$M[X]=\frac{b+a}{2} \qquad D[X]=\frac{(b-a)^2}{2}$$

График функции плотности вероятности распределения по равномерному закону [6] приведен на (рис. 4).

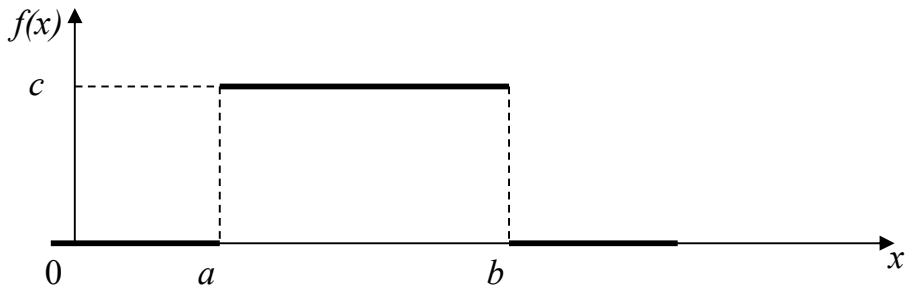

Рис. 4 - График функции плотности вероятности распределения по равномерному закону

Алгоритм Байеса в матричном виде [5, 7] представлен выражением (8)

$$K = \frac{e^{(S_{ei}^{*} P^{-1} S_{ei})}}{\sqrt{(2\pi)^{p} D}} \quad (5)$$

где P^{-1} - обратная корреляционная матрица;

$\pi = 3{,}14159$;

$\overline{S}_{ei}$ - усредненный вектор признаков i-го эталона;

$e = 2{,}71828$ - основание натурального логарифма.

Выражение для расчета алгоритма расстояния Махалано́биса будет иметь соответственно вид (6).

$$K = \arg\min_{i} (S - S_{ei})^{*} P_{i}^{-1} (S - S_{ei}) \quad (6)$$

где P^{-1} - обратная корреляционная матрица;

$\overline{S}_{ei}$ - усредненный вектор признаков i-го эталона.

В формулу для алгоритма "тепловой шум" входят те же переменные, что и для алгоритма Кейпона (7), однако значение оценочной корреляционной матрицы признаков берется в квадрате (8), для расчета которого необходимо больше машинного времени:

$$K = \arg\max_{i} \frac{1}{S^{*}_{ei} (P^{-1}) S_{ei}} \quad (7)$$

$$K = \arg\max_{i} \frac{1}{S^{*}_{ei} (P^{-1})^{2} S_{ei}} \quad (8)$$

где arg max - алгоритм максимума правдоподобия;

P^{-1} - обратная корреляционная матрица;

$\overline{S}_{ei}$ - усредненный вектор признаков i-го эталона.

Корреляционный алгоритм имеет следующее выражение (9).

$$K = \frac{1}{L} \frac{\sum_{L=1}^{L} (S_L - \overline{S}_L)(S_{ei} - \overline{S}_{ei})}{\sigma_s \sigma_{ei}} \quad (9)$$

где L – количество проведенных при распознавании испытаний;

$(S_L - \overline{S}_L)$ - L-й принимающий и усредненный по Р векторов признаков;

$\overline{S}_{ei}$ - усредненный вектор признаков i-го эталона;

σ_s та σ_{ei} - среднеквадратическое отклонение $(S_L - \overline{S}_L)$ и $(S_{ei} - \overline{S}_{ei})$.

Для сравнительной оценки быстродействия предлагаемых алгоритмов ОПВТ было проведено математическое моделирование для определения величины вероятности правильного распознавания (ВПР) от количества признаков распознавания (Р).

Значения Р не превышало 8. Они моделировались для каждого конкретного случая с помощью датчика случайных чисел с равномерным, нормальным и законом распределения Лапласа (рис. 5*а*, рис. 5*б*, рис. 5*в* соответственно).

На приведенных графиках использованы следующие обозначения для алгоритмов распознавания, подлежащих анализу: **МР** – расстояния Махаланóбиса; **Б** – Байеса; **К** – Корреляционный; **КП** – Кейпона; **ТШ** – "тепловой шум".

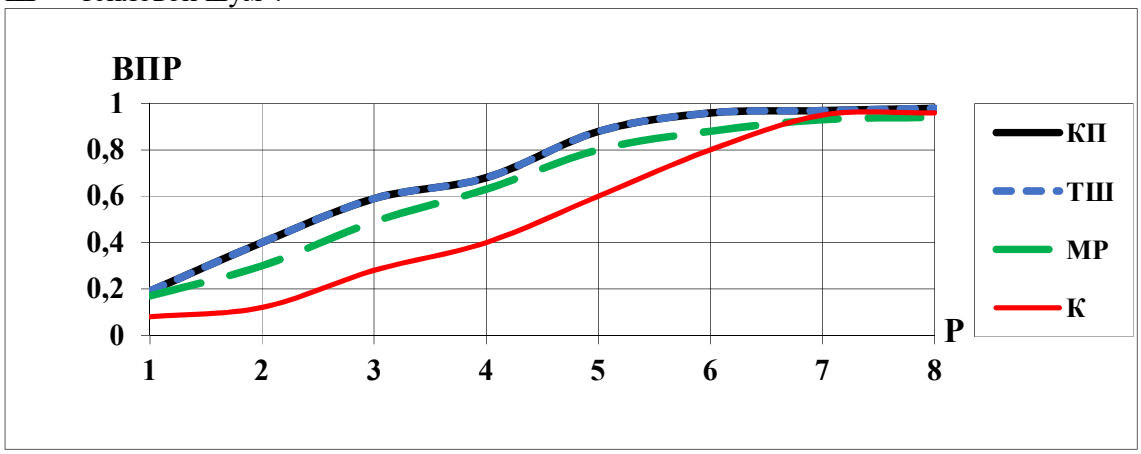

Рис.5*а* - Зависимость ВПР от количества признаков распознавания для равномерного закона распределения

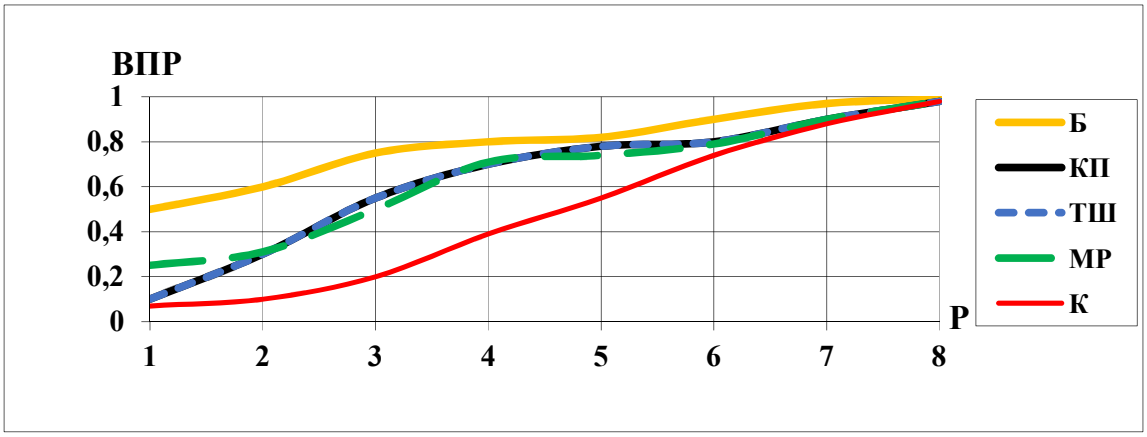

Рис.5*б* - Зависимость ВПР от количества признаков распознавания для нормального закона распределения

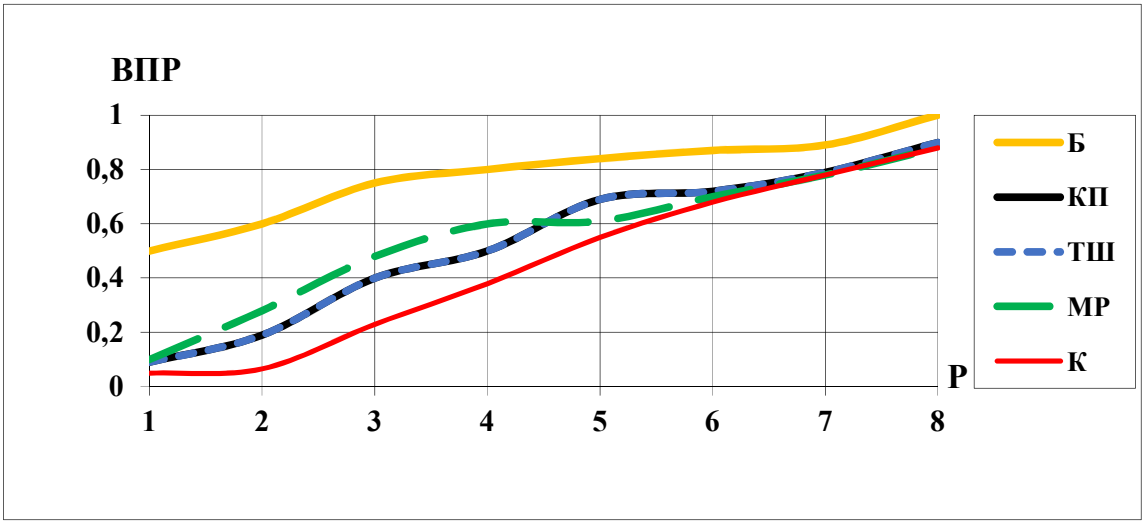

Рис 5*в* - Зависимость ВПР от количества признаков распознавания для закона распределения Лапласа

При равномерном законе распределения рис 5 *а* алгоритм Байеса не анализировался в связи с тем, что для его работы необходимо знание максимума значения закона распределения вектора признаков, который на всей своей протяженности имеет одинаковые значения, что можно наблюдать на (рис. 4).

Анализа полученных графиков (рис. 5*а*, рис. 5*б*, рис. 5*в*) показал, что алгоритм Байеса имеет достаточно высокие результаты уже при трех признаках распознавания, что наблюдается при нормальном и законе распределения Лапласа, однако применение алгоритма Байеса невозможно при равномерном законе распределения.

На графиках (рис. 5*а*, 5*б* и 5*в*) наблюдаются наилучшие показатели эффективности алгоритмов Кейпона и "тепловой шум". При этом анализ формул (7, 8) показал, что алгоритм Кейпона имеет значительное преимущество над алгоритмом "тепловой шум", так как при расчете первого проводится меньше математических вычислений, а именно возведение в квадрат корреляционной матрицы, что существенно уменьшает нагрузку на вычислительную систему [7].

Вывод

Таким образом, в данной работе, методами математического моделирования проведен анализ эффективности процесса распознавания признаков вредоносного трафика предложенными методами ОПВТ (5-9) с точки зрения их функционирования в условиях времени, приближенного к реальному.

Показано, что алгоритм Кейпона является наиболее эффективным по сравнению с другими методами и его применение в системе противодействия вторжений, позволит улучшить показатели эффективности IPS на величину до 5%.

Список использованной литературы

[1]. Lawson C., Hils A., Neiva C., "Defining Intrusion Detection and Prevention Systems" // Garthner report – 2016. – P. 5-17.

[2]. Viegas E., Santin A. O., Fran A. A., Jasinski R., Pedroni V. A., Oliveira, L. S.. "Towards an Energy-Efficient Anomaly-Based Intrusion Detection Engine for Embedded Systems" // IEEE Transactions on Computers – 2017. – P. 163–177.

[3]. Bezroukov, N., "Architectural Issues of Intrusion Detection Infrastructure in Large Enterprises" // Softpanorama Bulletin Vol. 17, No. 3 – 2010. – P. 3-17.

[4]. Eskin E, Lee W, Stolfo SJ. Modeling System Calls for Intrusion Detection with Dynamic Window Sizes. Proceedings of DISCEX II, 2001.

[5]. Марпл-младший С.Л. Цифровой спектральный анализ и его приложения // Москва: Мир, 1990 г. – 265 с.

[6]. Наконечний В.С. Аналіз ефективності та можливості застосування сучасних методів розпізнавання об’єктів радіолокаційного моніторингу. Науково-технічний журнал Зв'язок. - 2014. - №5. - С. 52-56.

[7]. В.С. Наконечний, С.В. Толюпа, І.Р. Пархомей, Н.В. Цьопа. Експериментальне дослідження надрозрізрювальних методів спектрального аналізу для задач пеленгації. Адаптивні системи автоматичного управління // Міжвідомчий науково-технічний збірник. — Київ: Національний технічний університет України “Київський політехнічний інститут”. – 2015. – Вип. 2(27). – с. 88-94.