УДК 681.324

АНАЛИТИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЯ ДЛЯ ИДЕНТИФИКАЦИИ СОСТОЯНИЯ ЛОКАЛЬНОЙ ВЫЧИСЛИТЕЛЬНОЙ СЕТИ

Зияутдинов В.С., Золотарева Т.А., Воронин И.В., Скуднев Д.М.
ФГБОУ ВО «Липецкий государственный педагогический университет имени П.П. Семенова-Тян-Шанского», Липецк, e-mail: scoudnev@lspu.lipetsk.ru

В данной статье рассматривается аналитическое представление статистического, сигнатурного и нейросетевого анализаторов, на основе которого, в комбинированном использовании, возможно реализовать идентификацию состояния локальной вычислительной сети. В аналитической модели сигнатурного анализатора сетевого трафика за основу взят анализ сигнатур – один из первых методов, который использовался для обнаружения проблем в работе компьютерных сетей, связанных с деятельностью злоумышленников. Аналитическая модель статистического анализатора сетевого трафика основана на том обстоятельстве, что в течение определенного интервала времени могут изменяться некоторые статистические характеристики потока пакетов. В этом случае методы обнаружения нарушений базируются на сравнении текущих характеристик потока пакетов с усредненными характеристиками за некоторый промежуток времени. В аналитической модели нейросетевого анализатора сетевого трафика используются искусственные нейронные сети Хемминга, предназначенные для распознавания класса принадлежности объекта.

Ключевые слова: анализатор трафика, сетевая ошибка, локальная вычислительная сеть

ANALYTICAL SUPPORT FOR THE INTELLIGENT SUPPORT OF DECISION-MAKING TO IDENTIFY THE STATE OF THE LOCAL COMPUTER NETWORK

Ziyautdinov V.S., Zolotareva T.A., Voronin I.V., Skudnev D.M.
Lipetsk State Pedagogical P. Semenov-Tyan-Shansky University, Lipetsk, e-mail: scoudnev@lspu.lipetsk.ru

This article discusses the analytical representation of the statistical, neural network and signature analyzers, based on which, in combined use, it is possible to realize the identification of the status of the local area network. In the analytical model, the signature analyzer of network traffic based on the analysis of signatures is one of the first methods that was used for the detection problem in computer networks related to the activities of criminals. Analytical model for statistical network traffic analyzer based on the fact that during a certain time interval may change some statistical characteristics of the packet flow. In this case, methods for the detection of violations based on the comparison of current characteristics of a stream of packets with mean characteristics for the certain period of time. In the analytical model, neural network analyzer network traffic using artificial neural networks Hamming is intended to recognize the class of belonging of the object.

Keywords: traffic analyzer, network error, local computer network

Несмотря на то, что существует множество приемов и инструментов обнаружения и устранения неполадок в локальной вычислительной сети (ЛВС), администратору необходимо осуществлять сбор данных о работе сети, проводить контроль, анализ и идентификацию всех основных сетевых процессов, т.е. выявлять сетевые проблемы. Такие нестандартные ситуации занимают много рабочего времени IT-специалистов. Для сокращения временных затрат и оптимизации работы администраторов необходимо применить современные технологии в комплексе с интеллектуальной системой поддержки принятия решений (СППР) для идентификации состояния ЛВС [1].

Аналитическая модель сигнатурного анализатора сетевого трафика

Анализ сигнатур – один из первых методов, который использовался для обнаружения проблем в работе компьютерных сетей (КС), связанных с деятельностью злоумышленников. Он базируется на простом понятии совпадения последовательности с эталонным образцом. Во входящем пакете просматривается байт за байтом и сравнивается с сигнатурой – характерной строкой программы, указывающей на наличие вредного трафика.

В качестве сетевой ошибки будем просматривать любую совокупность битовых критериев, которая не генерируется для решения какой-либо полезной задачи в ЛВС.

Структурная схема анализатора трафика представлена на рис. 1.

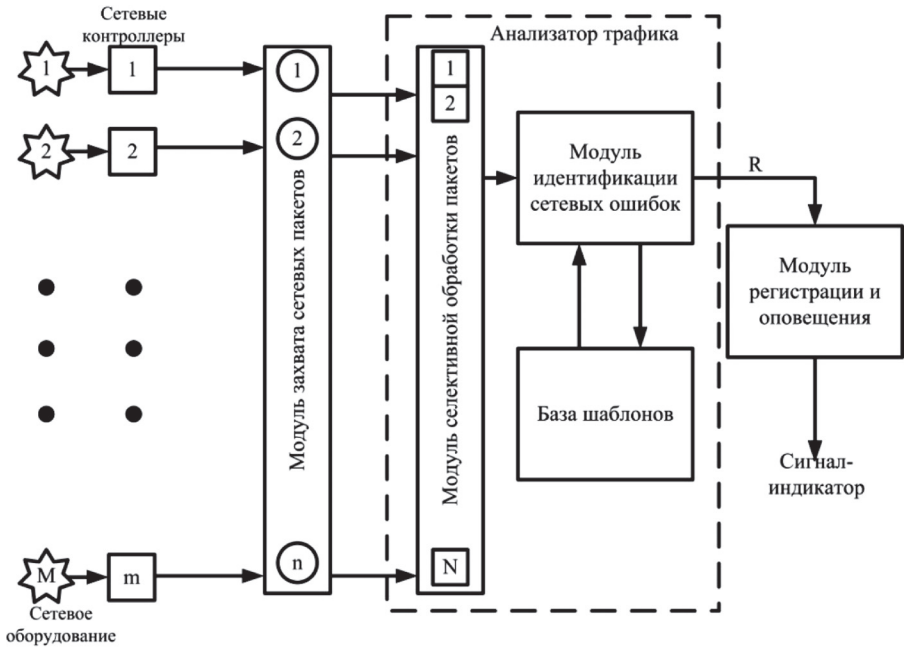

Рис. 1. Структурная схема сигнатурного анализатора сетевого трафика

Процедура анализа включает два этапа:

- селективный сбор фрагментов пакетов;
- распознавание ошибок по образцам.

Обозначим сетевой трафик как поток пакетов в виде множества

$$X = \{x_i\}_1^n,$$

где n – общее число пакетов. Основу образцов представим в виде множества A, объединяющего группы типов ошибок

$$A_j,\ j = \overline{1,g};$$

$$A = A_1 \cup A_2 \cup ... \cup A_\gamma = \bigcup_{j=1}^{\gamma} A_j,$$

где g – число групп; A_j – j-й кластер, являющийся множеством однотипных ошибок, $A_j = \{a_{jk}\}_1^K$; K – общее число ошибок в j-м кластере.

Ошибка считается найденной, если выполняется условие: $X \subseteq A$. На вход модуля регистрации подается сигнал R, который может принимать два значения («0» – есть совпадение, «1» – нет совпадения).

Аналитическая модель статистического анализатора сетевого трафика

Статистические методы идентификации сетевых проблем (СП) основаны на том обстоятельстве, что в течение определенного интервала времени (на протяжении суток, часов, минут) могут изменяться некоторые статистические характеристики потока пакетов. В этом случае методы обнаружения нарушений базируются на сравнении текущих характеристик потока пакетов с усредненными характеристиками за некоторый промежуток времени. Первые характеристики называются локальными, а вторые – глобальными.

Если локальные характеристики значительно отличаются от глобальных, то вполне возможна попытка злонамеренных действий, сбоев в работе аппаратуры, программного обеспечения по различным причинам.

Обобщенная схема статистического анализатора, реализующего данный метод идентификации, представлена на рис. 2.

Представим аналитическую модель статистического анализатора, основанного на выборе весовых функций для определения текущих статистических характеристик потока пакетов.

Предположим, что числовая величина A_i ($a_{\min} \le A_i \le a_{\max}$) представляет собой некоторое событие из потока событий, произошедшее в момент времени t_j, $j = \overline{1,n}$. Множество событий характеризуется средним значением $\overline{a}$ и дисперсией σ_a величины A.

В качестве статистической характеристики потока событий будем употреблять среднее арифметическое функции $f(X)$ от величины X

$$w(n) = \frac{1}{n}\sum_{i=1}^{n} f(A_i). \qquad (1)$$

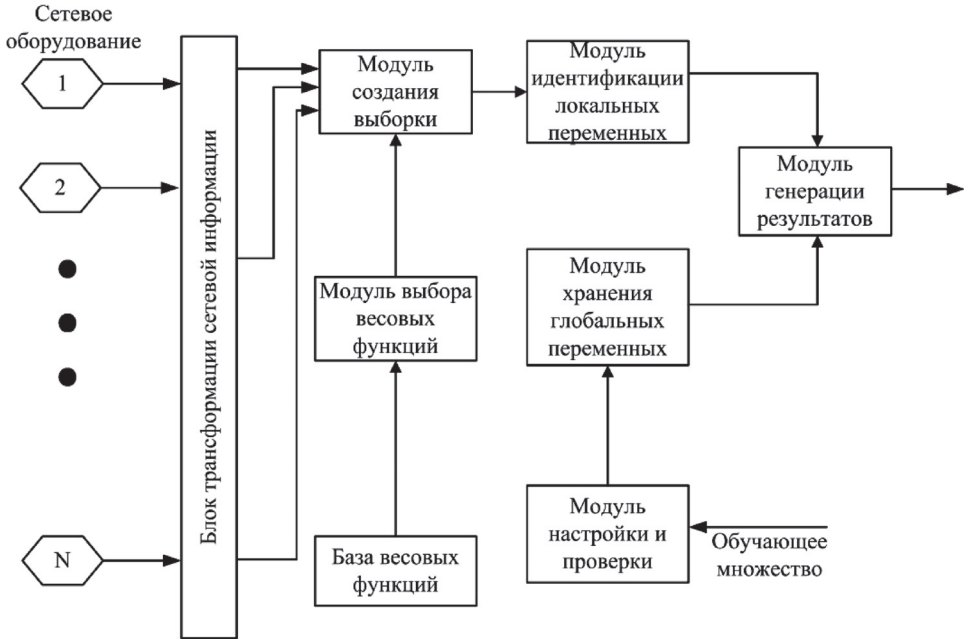

Рис. 2. Обобщенная схема статистического анализатора сетевого трафика

Для определения текущих характеристик будем определять не для всего потока N событий, а только для последних n событий. С этой целью введем понятие весовой функции $F(z)$ и значение текущих характеристик $W(N)$ определим как

$$W(N) = \sum_{j=1}^{N} \left(t_N - t_j\right) f\left(A_j\right). \qquad (2)$$

Значение довода характеристики $W(N)$ означает, что ее значение определяется вблизи N-го события потока, а размер выборки, для которой находится эта величина, определяется видом весовой функции $F(z)$.

Статистические характеристики потока пакетов ЛВС задаются видом функции $f(X)$ в выражении (2). Если функция имеет вид $f(X) = X$, то значение текущих характеристик $W_1(N) = \sum_{j=1}^{N} F\left(t_N - t_j\right) X_j$ соответствует среднему арифметическому величины X для последних пакетов сетевого трафика.

Если $F(X) = X^k$, то

$$W_k(N) = \sum_{j=1}^{N} F\left(t_N - t_j\right) X_j^k,$$

а это есть первый выборочный момент порядка k и т.д.

В случае применения критерия согласия λ^2 необходимо разбить величину на D интервалов

$$[x_{\min}, x_{\max}) = [x_0, x_1)\ [x_1, x_2) \ldots [x_{D-1}, x_D),$$

где $x_0 = x_{\min}$, $x_D = x_{\max}$, и подсчитать число попаданий величины X в тот или иной интервал [2].

Для учета числа событий, попадающих в интервал с номером D, определим функцию

$$\Phi_d(x) = \begin{cases} 1, & \text{если } X \in [x_{D-1}, x_D); \\ 0, & \text{в противном случае.} \end{cases}$$

Введем набор величин y_d $(1 \le d \le D)$,

где $$y_d(N) = \frac{1}{N} \Phi_d\left(X_i\right); \qquad (3)$$

$$\sum_{d=1}^{D} y_d(N) = 1. \qquad (4)$$

Общее число событий N определяется интервалом времени, в течение которого ведется наблюдение за трафиком. При увеличении числа событий N частоты $y_d(N)$ стремятся к P_d – возможностям попадания события в интервал с заданным номером и могут быть применены как глобальные (долговременные) характеристики потока пакетов сетевого трафика.

Для определения текущих (локальных) характеристик будем учитывать число попаданий событий в соответствующие

интервалы $[x_{min}, x_{max})$ не для целого потока пакетов, а только для n последних событий. Тогда значение локальных частот Y_d можно получить из выражения (2):

$$Y_d(N) = \sum_{i=1}^{N} F\left(t_N - t_i\right)\Phi_d\left(X_i\right); \qquad (5)$$

$$\sum_{d=1}^{D} Y_d(N) = 1. \qquad (6)$$

Аналитическая модель нейросетевого анализатора сетевого трафика

Искусственные нейронные сети (ИНС) позволяют более эффективно по сравнению с классическими подходами решать задачи в области обработки и распознавания различных образов [3]. Самая главная задача в применении ИНС для анализа сетевого трафика – это обучить ИНС правильно определять все проблемные события.

ИНС Хемминга предназначена для распознавания класса принадлежности объекта, заданного вектором X биполярных признаков (возможные значения признаков +1 и –1) размерности N. Предполагается, что имеются M классов, каждый из которых характеризуется своим эталонным представителем – объектом X_v, $v = 1, 2, ..., V$ [4, 5].

Эталонные образы и соответствующие векторы признаков хранятся в основе данных. Они отобраны экспертами для разных типов образов. На рис. 3 представлена схема обработки данных при применении нейросетевого классификатора Хемминга.

ИНС Хемминга принимает на N входов биполярные признаки объекта и после обработки данных активизирует один из K выходов, который указывает на класс принадлежности предъявленного на входе объекта.

Критерием отнесения объекта X к классу является квадрат расстояния между векторами X и X_q, $q = 1, 2, ..., Q$:

$$R(X, X_q) = \sum_{j=1}^{N}\left(x_j - x_{qj}\right)^2, \qquad (7)$$

где x_j и x_{qj} – j-й биполярный признак входного образа и q-го эталона соответственно, $j = \overline{1,N}$, $q = \overline{1,Q}$.

Простейшие преобразования $R(X, X_q)$ приводят к тому, что для нахождения k минимизация $R(X, X_q)$ по индексу k – номеру эталона – может быть заменена максимизацией скалярного произведения векторов X и X_q:

$$\max_q R\left(X, X_q^T\right) = X \cdot X_{q'}^T,$$

где индекс T означает транспонирование вектора.

Операцию определения скалярного произведения двух векторов реализует нейрон, потенциал которого определяется по формуле

$$p_k = \sum_{i=1}^{N} x_i x_{ki} = \sum_{i=1}^{N} x_i w_{ki},$$

где $w_{ki} = x_{ki}$ – синаптические коэффициенты k-го нейрона.

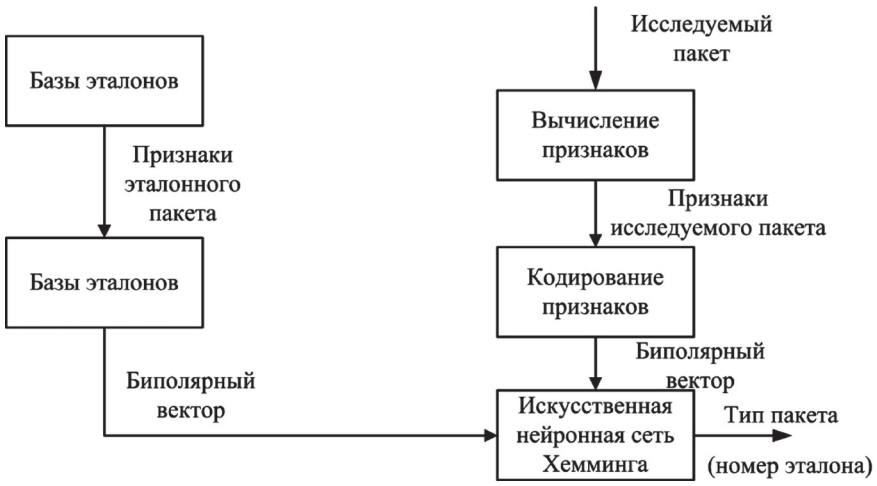

Рис. 3. Схема обработки данных в нейросетевом анализаторе Хемминга ИНС относит объект к классу q', если $\min_q R\left(X, X_q\right) = R\left(X, X_{q'}\right)$

Значит, характеристики эталонного объекта (образа) хранятся в ИНС Хемминга в форме синаптических коэффициентов нейрона. Для применения нейросетевого классификатора Хемминга необходимо все признаки образа, которые обычно представлены действительными или целыми числами, конвертировать в биполярный код. Тогда N будет означать общее число биполярных разрядов кода всех используемых для распознавания признаков [4].

ИНС Хемминга содержит столько нейронов, сколько эталонных образов хранится в базе данных. Пусть такое число будет обозначаться как M. Предполагается, что при каждой записи в базу эталонных образов (БЭО) нового типового образа выполняется детальное вычисление его вектора признаков. Значит, БЭО содержит не только образы, но и соответствующие векторы действительных чисел $X_q = \{x_{q1}, x_{q2}, ..., x_{qN}\}$, $q = \overline{1,M}$. Для успешного применения ИНС Хемминга все действительные признаки должны быть представлены своим двоичным кодом. Допустим, что длина кода каждого из признаков x_{qi}, $q = \overline{1,M}$, $i = \overline{1,N}$, равна J. Тогда общее число двоичных признаков, поступающих на вход ИНС Хемминга, можно обозначить как $N_0 = JN$. Обозначим двоичные признаки y_0, $i = \overline{1,N_0}$. Каждый из M нейронов ИНС Хемминга имеет N_0 синаптических коэффициентов, которые представляют код признаков соответствующего эталонного образа в базе данных. Заметим, что в основе эталонных образов хранятся не коды, а действительные числа x_{qi}, $q = \overline{1,M}$, $i = \overline{1,N}$. Коды выражаются в процедуре эксплуатации сети Хемминга. Такой подход дает возможность употреблять любую из имеющихся схем кодирования.

Заключение

Рассмотренные модели позволяют сделать вывод, что наилучшим подходом в создании современной системы идентификации ЛВС среднего предприятия является комбинированный подход. Он включает в себя хорошо зарекомендовавшие статистический метод, в дополнение к уже имеющимся сигнатурным системам, а в качестве самообучаемой модели применить нейросетевой анализатор сетевого трафика, основанный на оптимизированной нейросети Хемминга.

Список литературы

1. Епанешников А.М. Локальные вычислительные сети. – М.: Изд-во Диалог-МИФИ, 2005. – 224 с.

2. Золотарева Т.А. Алгоритмическая реализация интеллектуальной системы поддержки принятия решений для идентификации состояния локальной вычислительной сети // Вестник РГРТУ № 51. – Рязань, 2015.

3. Коваль С.А. Лингвистические проблемы компьютерной морфологии. – СПб.: Изд-во Санкт-Петербургского университета, 2005. – 152 с.

4. Круглов В.В. Искусственные нейронные сети. Теория и практика. Горячая Линия – М.: Телеком, 2009. – 382 с.

5. Рапопорт Г.Н. Биологический и искусственный разум. Часть 1. Сознание, мышление и эмоции. – М.: Либроком, 2011. – 184 с.

References

1. Epaneshnikov A.M. Lokalnye vychislitelnye seti. M.: Izd-vo Dialog-MIFI, 2005. 224 p.

2. Zolotareva T.A. Algoritmicheskaja realizacija intellektualnoj sistemy podderzhki prinjatija reshenij dlja identifikacii sostojanija lokalnoj vychislitelnoj seti Vestnik RGRTU no. 51, Rjazan 2015.

3. Koval S.A. Lingvisticheskie problemy kompjuternoj morfologii. Izdatelstvo Sankt-Peterburgskogo universiteta, 2005. 152 p.

4. Kruglov V.V. Iskusstvennye nejronnye seti. Teorija i praktika. Gorjachaja Linija Telekom, 2009. 382 p.

5. Rapoport G.N. Biologicheskij i iskusstvennyj razum. Chast 1. Soznanie, myshlenie i jemocii. Librokom, 2011. 184 p.