УДК 519.712.3

В.И. ГАСАНОВ[*]

ВЫЯВЛЕНИЕ АНОМАЛИЙ В СЕТЕВОМ ТРАФИКЕ НА ОСНОВЕ НЕЙРОСЕТЕВОГО МОДЕЛИРОВАНИЯ ДИНАМИКИ ИЗМЕНЕНИЯ ОБЪЁМОВ IP-ПАКЕТОВ

[*]Институт систем управления НАНА, г. Баку, Азербайджан

Анотація. *Розглядається нейромережевий метод прогнозування мережевого трафіка з метою виявлення і оцінки можливих аномалій в об'ємах IP-пакетів. У контексті системи управління мережевим трафіком пропонується нейромережева модель прямого розповсюдження сигналів, що дозволяє екстраполювати мережевий трафік і тим самим прогнозувати значення його об'ємів на короткостроковий період дії.*
Ключові слова: *інформаційна безпека, мережевий трафік, об'єм IP-пакета, аномалія, нейронна мережа.*

Аннотация. *Рассматривается нейросетевой метод прогнозирования сетевого трафика с целью обнаружения и оценки возможных аномалий в объёмах IP-пакетов. В контексте системы управления сетевым трафиком предлагается нейросетевая модель прямого распространения сигналов, позволяющая экстраполировать сетевой трафик и тем самым прогнозировать значения его объёмов на краткосрочный период действия.*
Ключевые слова: *информационная безопасность, сетевой трафик, объём IP-пакета, аномалия, нейронная сеть.*

Abstract. *It is considered a neural network method for predicting of network traffic in order to detect and evaluate possible anomalous into the volumes of IP packets. In the context of the network traffic management system, the feedforward neural network model is proposed, which allows to extrapolate network traffic and, thus, to predict the values of its volumes for a short-term period of validity.*
Keywords: *information security, network traffic, volume of IP packet, anomaly, neural network.*

1. Введение

Одной из главных причин, негативно влияющих на эффективность работы корпоративных сетей, являются аномалии в объёме сетевого трафика, которые могут быть вызваны случайными или преднамеренными действиями со стороны легитимных пользователей, неверной работой web-приложений, действиями злоумышленников и т.д. Поэтому при проектировании систем поддержки принятия решений в области управления сетевыми трафиками необходимо принимать меры по своевременному выявлению таких аномалий, поиску их источников и тем самым обеспечивать надёжное функционирование корпоративных сетей связи. Исходя из этой предпосылки, становится очевидны важность и актуальность разработки методов обнаружения аномальных пакетов и управления трафиком.

Одним из перспективных направлений в области выявления аномалий в сетевом трафике являются нейронечёткие методы моделирования и прогнозирования динамики изменения объёмов IP-пакетов. В частности, в работе [1] нами рассмотрен подход к моделированию сетевого трафика на основе нечёткого анализа изменения объёмов IP-пакетов, согласно которому в рамках системы управления сетевым трафиком были предложены несколько нечётких моделей, позволяющих экстраполировать сетевой трафик. Существуют и другие средства прогнозирования динамики изменения объёмов IP-пакетов в сетевом трафике, стандартные прогностические алгоритмы которых описаны в многочисленных публикациях, например, в [2–5]. Одним из таких способов является нейросетевое прогнозирование динамики сетевого трафика, который также способен обеспечить безопасность

и практическое выживание корпоративных сетей связи, а, значит, в целом обеспечить эффективность их функционирования как организационно-технических систем.

2. Постановка задачи

Методика выявления аномалий в сетевом трафике на начальном этапе подразумевает перехват входящего и исходящего трафиков. При этом в перехваченном трафике осуществляется поиск заголовков IP-пакетов, из которых извлекаются все необходимые атрибуты: объём IP-пакета, IP-адрес источника, IP-адрес назначения, дата получения IP-пакета, время получения IP-пакета. Полученная информация сохраняется в базе данных сетевой статистики.

Выбирая основным атрибутом объём IP-пакета, рассмотрим соответствующий динамический ряд реальных данных, поступающих с сетевого устройства, и осуществим его нейронное прогнозирование на основе циклического анализа. В случае значительного расхождения прогнозов объёмов IP-пакетов от соответствующих им реальных данных делается вывод об обнаружении аномалии и принимается решение о необходимости применения управляющих воздействий.

В качестве примера выберем динамический ряд изменения объёмов сетевого трафика с минимальным временным шагом между отсчётами данных в $\Delta t = 1$ мин., который был рассмотрен нами в [1] (рис. 1). Данный ряд отражает состав и количество данных, на основании которых будет производиться прогнозирование сетевого трафика.

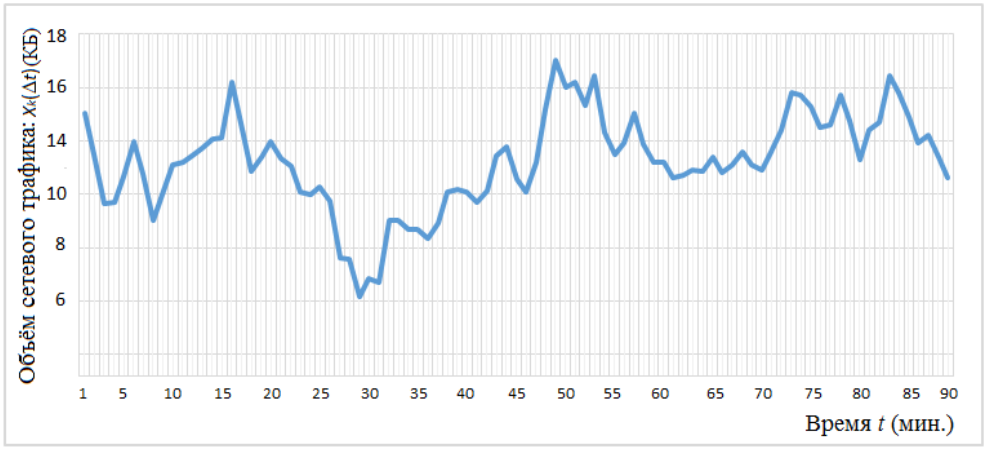

Рис. 1. Динамический ряд изменения объёма сетевого трафика

3. Нейросетевое моделирование сетевого трафика

При отсутствии адекватной модели сетевого трафика обработка динамического ряда

$$x = \{V(t)\}\,(t = t_1, t_2, \ldots, t_n), \tag{1}$$

в котором $V(t_k)$ $(k = 1 \div n)$ является значением объёма сетевого трафика на момент времени t_k, может служить эффективным способом анализа. Наличие динамического ряда (1) даёт возможность построить систему уравнений, воспроизводящую поведение сетевого трафика и тем самым предоставить прогнозы будущим значениям объёмов сетевого трафика: $\{V(t_{n+1}), V(t_{n+2}), \ldots\}$.

Существующие методы статистического анализа динамических рядов в основном опираются на исследования авторегрессионных моделей вида

$$V(t_k) = \{V(t_{k-1}), V(t_{k-2}), \ldots, V(t_{k-m})\},\ k \geq m+1. \tag{2}$$

В этом случае прогнозирование динамического ряда (1) сводится к типовой задаче нейросетевого моделирования, то есть к задаче нейронной аппроксимации непрерывной функции многих переменных по заданному набору обучающих образцов.

Если известна внешняя входная последовательность $\{V(t_k)\}$ и необходимо модифицировать её в иную наблюдаемую последовательность $\{Y(t_k)\}$, то, полагая саму систему нелинейной, можно представить причинно-следственную связь между входами и выходами в виде трёхслойной feedforward нейронной сети (рис. 2), которая, будучи нелинейной авторегрессионной моделью, на своем выходе индуцирует сигналы в виде

$$V(t_{k+1}) = \sum_{i=1}^{n} c_i \varphi[w_{ik} V(t_k) - \theta_i], k = \overline{1,m},\ k \geq m+1, \tag{3}$$

где n – число нелинейных нейронов в скрытом слое, w_{ik} и c_i – веса входных и выходных синоптических связей, соответственно, θ_i – пороговое значение (смещение) i-го нелинейного нейрона из скрытого слоя, $\varphi(\cdot)$ – нелинейная функция активации нейрона из скрытого слоя, например, сигмоидального вида: $\varphi(x) = 1/(1+e^{-x})$.

В общем виде задачу нейросетевого моделирования сетевого трафика как динамического ряда можно записать как

$$V(t_k) = F[V(t_{k-1}), V(t_{k-2}), ..., V(t_{k-m})], \tag{4}$$

где $F[\cdot]$ – нелинейная функция авторегрессии, которая реализуется посредством трёхслойной feedforward нейронной сети с топологической структурой, представленной на рис. 2.

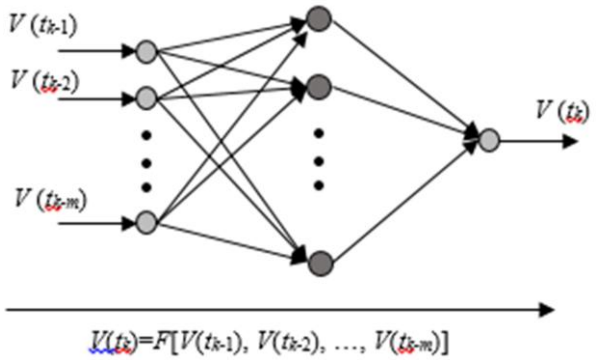

Рис. 2. Нейросетевая авторегрессионная модель прогнозирования сетевого трафика

Как и в традиционном случае, нейросетевое моделирование динамического ряда сетевого трафика на начальном этапе предусматривает формирование трёх совокупностей данных: обучающей, алидационной и тестовой. Обучающая совокупность используется для построения, собственно, нейронной сети – настройки её параметров, а именно, весов синоптических связей и порогов нейронов из скрытого слоя. Валидационная совокупность данных служит для выбора оптимальной топологической структуры сети, а тестовая – не используется при обучении и служит только для контроля достоверности прогнозов.

Для формирования обучающей выборки динамический ряд сетевого трафика, состоящий из n показателей объёмов IP-пакетов, разбивается на окна длиной d в следующем виде:

$$\begin{Bmatrix} V_1, V_2, ..., V_{d-1}, V_d \\ V_2, V_3, ..., V_d, V_{d+1} \\ ..... \\ V_k, V_{k+1}, ..., V_{d+k-1}, V_n \end{Bmatrix}.$$

В этом случае обучающая последовательность будет иметь вид

$$(V_j, V_{j+1}, ..., V_{d+j-1}) \to V_{d+j},\ j = 1, 2, \ ..., \ n-d$$

подразумевать, что при подаче на вход нейронной сети вектора $(V_j, V_{j+1}, ..., V_{d+j-1})$ сигнал V_{d+j} снимается с выхода.

Для оценки степени адекватности нейросетовой модели динамического ряда сетевого трафика можно применить сумму S погрешностей между фактическим объемом IP-пакета $V(t_i)$ и его прогнозируемым значением $F(t_i)$:

$$\xi = \sum_{i=0}^{S} e[V(t_i), F(t_i)], \tag{4}$$

где $e[\cdot,\cdot]$ – функция погрешности, которую, согласно [6, 7], можно представить в виде средней абсолютной ошибки, выраженной в процентах (MAPE – Mean Absolute Percentage Error):

$$MAPE = \frac{1}{S}\sum_{i=1}^{S}\frac{|V(t_i) - F(t_i)|}{V(t_i)} \times 100\% \tag{5}$$

или в виде среднеквадратичного отклонения (*MSE* – Mean Squared Error):

$$MSE = \frac{1}{S}\sum_{i=1}^{S}(V(t_i) - F(t_i))^2 . \tag{6}$$

4. Выявление аномалий в сетевом трафике

Как отмечалось выше, для определения объёма сетевого трафика в корпоративных сетях связи применяется информация, извлекаемая из состава IP-пакетов. В этом случае выявление возможных аномалий происходит путём сравнения текущего объёма IP-пакета с его прогнозным значением, то есть, иными словами, для текущего момента времени t сравниваются две величины: фактический объём IP-пакета $V(t)$ и его прогноз $F(t)$. Если разница между ними будет не меньше заранее установленного критического значения $\Delta V_{крит}$, то есть, если выполняется условие $\Delta V = |V(t) - F(t)| \geq \Delta V_{крит}$, то в этом случае, возможно, имеет место аномалия, вызванная случайными или преднамеренными действиями со стороны легитимных пользователей, неверной работой *web*-приложений, действиями злоумышленников и т.д.

В результате нейронного прогнозирования динамичского ряда сетевого трафика в нотации пакета MATLAB при длине окна $d = 3$ получены прогнозные значения объёмов IP-пакетов, которые сведены в табл. 1.

Таблица 1. Результаты нейронного прогнозирования динамического ряда сетевого трафика

Δt_k	Фактический объём IP-пакета	Нейро-прогноз	Δt_k	Фактический объём IP-пакета	Нейро-прогноз	Δt_k	Фактический объём IP-пакета	Нейро-прогноз
1	15,024		31	8,712	8,640	61	12,611	12,789
2	13,514		32	11,012	10,903	62	12,734	13,166
3	11,637		33	11,044	10,903	63	12,937	13,166
4	11,691	11,657	34	10,701	10,526	64	12,870	13,166
5	12,651	12,789	35	10,685	10,526	65	13,406	13,166
6	13,973	13,920	36	10,332	10,149	66	12,794	13,669
7	12,777	12,789	37	10,911	10,903	67	13,100	12,412
8	11,005	10,903	38	12,111	12,034	68	13,600	13,543
9	12,137	12,034	39	12,183	12,600	69	13,096	13,355
10	13,096	12,600	40	12,085	12,034	70	12,902	12,789

Продолж. табл. 1

11	13,183	14,109	41	11,684	12,034	71	13,606	13,166
12	13,441	12,789	42	12,158	12,034	72	14,401	13,669
13	13,748	13,543	43	13,455	13,543	73	15,803	15,806
14	14,091	13,637	44	13,787	13,920	74	15,704	15,806
15	14,123	14,297	45	12,570	13,637	75	15,297	15,429
16	16,186	14,486	46	12,096	12,034	76	14,497	14,674
17	14,633	14,674	47	13,186	13,166	77	14,598	13,732
18	12,848	13,732	48	15,211	14,109	78	15,701	15,806
19	13,379	13,543	49	17,030	16,937	79	14,773	14,674
20	13,987	13,669	50	16,012	16,183	80	13,313	13,166
21	13,336	13,637	51	16,202	16,183	81	14,403	14,297
22	13,071	13,166	52	15,320	15,429	82	14,708	14,674
23	12,113	12,789	53	16,450	16,560	83	16,432	16,560
24	11,988	12,034	54	14,298	14,297	84	15,825	15,806
25	12,284	12,034	55	13,495	13,543	85	14,911	15,051
26	11,761	11,657	56	13,920	13,355	86	13,951	13,920
27	9,620	9,772	57	15,045	13,637	87	14,197	13,732
28	9,595	9,772	58	13,862	13,920	88	13,421	14,486
29	8,169	8,263	59	13,188	13,732	89	12,619	13,355
30	8,837	9,018	60	13,183	13,166	90	11,736	12,412
MAPE					2,1630			
MSE					0,1977			

Графическая интерпретация нейронного прогнозирования динамичского ряда сетевого трафика в нотации пакета MATLAB при длине окна $d = 3$ представлен на рис. 3.

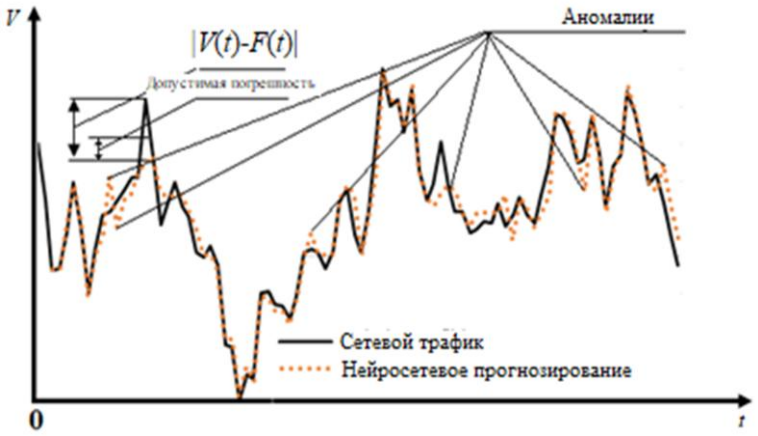

Рис. 3. Аномалии в сетевом трафике

В случае выявления аномалий в сетевом трафике результаты поиска переправляются в блок поиска источников аномалий. Установление источников аномалий происходит на основе имеющейся в наличии информации о IP-ракетах сетевого трафика, протекающего в корпоративной сети связи. После этого, собственно, оцениваются объёмы аномалий.

Помимо величины аномалии, при оценке применяется информация об источниках аномально высокого объема трафика, а также сведения, полученные от ответственного за принятие решений и экспертов. В результате вся приобретённая информация об аномалии обобщается и передается для дальнейшего рассмотрения со стороны ответственного за принятие решений. Оценка величины аномалии происходит на основе типичной для экспертных систем продукционной базы правил. Но это уже является предметом следующих исследований.

5. Заключение

Основанная на идее нейронного прогнозирования динамики изменения объёмов IP-пакетов и на наличии волатильности у сетевого трафика предложенная трёхслойная нейросетевая модель призвана решать задачу прогнозирования изменения объёма сетевого трафика. Это далеко не единственное средство в нейромоделировании динамических рядов. Существу-

ют и другие сети с отличными топологическими структурами, например, радиально-базисные функциональные нейронные сети, которые не менее эффективно решают класс подобных задач. Их можно строить для окон с различными длинами, то есть для описания внутренних причинно-следственных связей с более или менее высокими порядками. В итоге, среди них следует выбрать наиболее адекватную, чтобы полученные на её основе прогнозы могли бы сравниваться с реальными данными, поступающими с сетевого устройства. В случае их значительного расхождения делается вывод об обнаружении аномалии и принимается решение о необходимости применения управляющих воздействий.

СПИСОК ИСТОЧИКОВ

1. Прогнозирование сетевого трафика на основе нечёткого анализа изменения объёмов IP-пакетов // Известия Бакинского университета. – (Серия «Физико-математические науки»). – 2016. – № 4. – С. 123 – 132.
2. Ажмухамедов И.М. Повышение безопасности компьютерных систем и сетей на основе анализа сетевого трафика / И.М. Ажмухамедов, А.Н. Марьенков // Инфокоммуникационные технологии. – 2010. – Т. 8, № 3. – С. 106 – 108.
3. Ажмухамедов И.М. Обеспечение информационной безопасности компьютерных сетей на основе анализа сетевого трафика / И.М. Ажмухамедов, А.Н. Марьенков // Вестник АГТУ. – (Серия «Управление, вычислительная техника и информатика»). – 2011. – № 1. – С. 137 – 141.
4. Платов В.В. Исследование самоподобной структуры телетрафика беспроводной сети / В.В. Платов, В.В. Петров // Радиотехнические тетради. – 2004. – № 3. – С. 58 – 62.
5. Петров В.В. Структура телетрафика и алгоритм обеспечения качества обслуживания при влиянии эффекта самоподобия: дис. … канд. техн. наук: 05.12.13 / Петров В.В. – Москва, 2004. – 199 с.
6. Рзаев Р.Р. Интеллектуальный анализ данных в системах поддержки принятия решений / Рзаев Р.Р. – Verlag: LAP Lambert Academic Publishing GmbH & Co, 2013. – 130 с.
7. Моделирование временных рядов на основе нечёткого анализа данных / Р. Рзаев, Г. Шихалиева, М. Агамалыев [и др.] // Нечёткие системы и мягкие вычисления. – 2014. – Т. 9, № 1. – С. 39 – 86.

Стаття надійшла до редакції 04.04.2018