УДК 004.056, 519.812.3

А.В. Гирик, Г.П. Жигулин

ПРИНЦИПЫ ФОРМИРОВАНИЯ ПРОФИЛЯ НОРМАЛЬНОГО ФУНКЦИОНИРОВАНИЯ ОБЪЕКТОВ МОНИТОРИНГА В ЗАДАЧАХ ОБНАРУЖЕНИЯ СЕТЕВЫХ АНОМАЛИЙ

Гирик А.В., Жигулин Г.П. **Принципы формирования профиля нормального функционирования объектов мониторинга в задачах обнаружения сетевых аномалий.**
Аннотация. В статье предложен подход к формированию профилей нормального функционирования (ПНФ) объектов мониторинга, что является одним из этапов в задачах обнаружения сетевых аномалий. Показаны основные трудности, возникающие при формировании ПНФ по данным мониторинга, и способы их преодоления. В качестве математической основы для формирования ПНФ предлагается использовать итерационные методы, в частности, метод Шискина–Эйзенпресса. Представлены результаты экспериментальной проверки методов формирования ПНФ, сделаны выводы о возможности применения методов в задачах обнаружения сетевых аномалий.
Ключевые слова: сетевые аномалии, обнаружение вторжений, профиль нормального функционирования, информационная безопасность.

Guirik A.V., Zhigulin G.P. **Approach to creating normal functioning profile of monitored objects for network anomalies detection.**
Abstract. An approach to creating normal functioning profiles (NFP) of monitored objects is considered. NFP creation is one of the key steps in solving problems of network anomalies detection. Common issues of NFP creation and ways of overcoming these issues are considered. Iteration methods, Shiskin–Eisenpress method in particular, are proposed as a mathematical tool for NFP creation procedure. Described NFP creation method is verified on empirical network monitoring data and suggested suitable for network anomalies detection.
Keywords: network anomalies, intrusion detection, normal functioning profile, information security.

1. Введение. В настоящее время наблюдается активный рост количества информационных угроз и факторов, приводящих к нестабильному функционированию сетей передачи данных (СПД). Одним из компонетов обеспечения информационной защиты сетей стали программные комплексы, предназначенные для обнаружения вредоносной или подозрительной активности — системы обнаружения вторжений (СОВ). СОВ также решают задачи обнаружения аномалий в работе ключевых узлов СПД [1]. Аномалии представляют собой изменения в поведении показателей функционирования оборудования, которые либо свидетельствуют о том, что объект мониторинга вышел из режима стабильного функционирования, либо представляют собой существенные отклонения от априорно заданных закономерностей, либо являются исторически нехарактерными для рассматриваемых показателей.

Метод обнаружения аномалий в режиме реального времени заключается в следующем: выполняется мониторинг некоторого показателя безопасности, затем на основе накопленных данных строится прогноз, то есть рассчитывается, какие значения показатель примет в ближайшее время, после чего прогноз сравнивается с реальными значениями показателя и на основании определенных критериев (в первую очередь величины ошибки прогноза) принимается решение о наличии аномалии.

Простейшие методы анализа сетевой статистики (например, на предмет выхода значений показателей за границы заранее заданного диапазона), реализованные во многих популярных системах мониторинга, приходится признать несостоятельными применительно к задачам обнаружения аномалий в сетях с большим количеством узлов и высокой степенью агрегации потоков данных [2]. Потоки данных на уровне распределения и ядра сети обладают рядом свойств, в числе которых ярко выраженная периодичность и инерция тенденций. Следовательно, нужно выполнять анализ временных рядов показателей с учетом этих особенностей. В настоящей работе рассматривается метод обнаружения аномалий во временных рядах показателей функционирования объектов мониторинга, суть которого состоит в следующем:

1. Для выбранного показателя на основе его исторических значений или заранее известных закономерностей формируется «профиль нормального функционирования» — набор функций, описывающих поведение показателя в нормальных условиях (то есть при отсутствии аномалий). Профиль нормального функционирования (ПНФ) может задаваться в различных формах, в том числе в табличной и аналитической.
2. По мере поступления новых значений показателя формируется краткосрочный прогноз значений показателя на несколько шагов вперед. Полученный прогноз сравнивается с ПНФ на предмет отклонений. В случае, если прогноз по совокупности заданных критериев существенно отклоняется от ПНФ, формируется предупреждение о возможности преданомальной ситуации.
3. В дальнейшем фактические значения показателя сравниваются с ПНФ и прогнозом, вычисленным ранее, на предмет отклонений. В результате либо подтверждается наличие существенного отклонения фактических значений от ПНФ (обнаружение аномалии), либо выявляется ошибка прогнозиро-

вания, в результате чего происходит корректировка параметров метода построения краткосрочного прогноза.

4. В зависимости от того, были значения показателя классифицированы как аномальные или нет, и политики обновления ПНФ происходит корректировка ПНФ на основе этих значений.

2. Модель временных рядов показателей функционирования. Основной формой представления статистической информации, полученной в результате мониторинга показателя функционирования некоторого объекта, являются временные ряды (ВР) наблюдений, т.е. ряды динамики, у которых в качестве признака упорядочения берется время. Статистические методы исследования исходят из предположения о возможности представления уровней ряда в виде совместного появления нескольких компонент, отражающих закономерность и случайность развития процесса. В общем случае временной ряд показателя содержит четыре структурно образующих элемента: тренд, сезонную компоненту, циклическую компоненту и остаточную (случайную) компоненту [3].

Тренд представляет собой устойчивое изменение показателя в течение длительного времени. Он выражается аналитической функцией, которая используется для формирования прогнозных оценок. Сезонная компонента характеризует устойчивые колебания уровней. Она проявляется в некоторых показателях, которые представлены суточными или недельными данными. Циклическая компонента представляет колебания уровней с некоторым периодом. Тренд, сезонная и циклическая компоненты называются *регулярными*, или *систематическими* компонентами. Составная часть временного ряда, остающаяся после выделения из него регулярных компонент, называется *остаточной компонентой* ε_t. Она является обязательной частью любого временного ряда показателя, поскольку рассматриваемым процессам всегда сопутствуют небольшие изменения, вызванные слабым влиянием кратковременно действующих случайных факторов. Если систематические компоненты временного ряда выделены правильно, что собственно и необходимо при построении трендовой модели, то остаточная компонента будет характеризоваться случайностью изменения значений, соответствием нормальному закону распределения, равенством нулю математического ожидания и независимостью значений уровней друг от друга, т.е. отсутствием автокорреляции.

Проверка адекватности трендовых моделей основана на проверке наличия этих четырех свойств у остаточной последовательности. Если временной ряд представляется в виде суммы соответствующих компонент, то полученная модель носит название аддитивной и имеет вид:

$$X_t = U_t + V_t + C_t + \varepsilon_t . \quad (1)$$

Если компоненты умножаются, то получим мультипликативную модель:

$$X_t = U_t \cdot V_t \cdot C_t \cdot \varepsilon_t . \quad (2)$$

Основной целью статистического анализа временных рядов является изучение соотношения между закономерностью и случайностью в формировании значений уровней ряда, оценка количественной меры их влияния. Закономерности, объясняющие динамику показателя в прошлом, могут быть использованы для прогнозирования его значений в будущем, а учет случайности позволяет определить вероятность отклонения от закономерного развития и их возможную величину [4].

Формирование уровней ряда определяется закономерностями трех основных типов: инерцией тенденции, инерцией взаимосвязи между последовательными уровнями ряда и инерцией взаимосвязи между исследуемым показателем и показателями-факторами, оказывающими на него причинное воздействие.

В качестве типичной задачи, возникающей при анализе трафика, можно рассмотреть обнаружение аномалий в данных мониторинга загрузки некоторого канала связи. Для того, чтобы можно было применить описанный выше метод обнаружения аномалий, необходимо построить профиль нормального функционирования показателя (например, скорости передачи данных).

3. Формирование профиля нормального функционирования. Для некоторых классов задач ПНФ может задаваться как априорно известная зависимость, однако в большинстве случаев ПНФ нужно будет сформировать на основе исторических данных. ПНФ зависит от выбранной модели временного ряда. Далее будем считать, что модель ряда имеет вид (1) без сезонной компоненты $\hat{C}_t$. В этом случае в состав ПНФ войдут оценки компонентов ряда $\hat{U}_t$, и $\hat{V}_t$, наборы пороговых значений $\left\langle R_t^L, R_t^H \right\rangle$, исторические минимальные и максимальные значения $X_{\min}$ и $X_{\max}$, расчетные динамические характеристики.

ПНФ может быть получен в простейшем случае путем агрегирования данных (объединения *m* отсчетов в один) и вычисления среднего исторического значения показателя

$$\hat{X}_{(m,t,k)} = \frac{1}{N}\sum_{j=1}^{N}\left[\frac{1}{m}\sum_{i=1}^{m} X_{k+i}\right], \tag{3}$$

где N – количество рассматриваемых периодов продолжительностью T. Предполагается, что в каждом периоде одинаковое количество отсчетов и отсчеты получены через одинаковые интервалы времени τ (период опроса). Следует заметить, что в некоторых случаях при построении ПНФ требуется более сложный выбор периодов или участков периодов, например, если нужно исключить из рассмотрения известные аномальные участки. При построении ПНФ исходим из предположения, что в исторических данных ряда на обучающем наборе аномалий не содержится.

В настоящей работе решается задача построения более точного ПНФ, обеспечивающего меньшую относительную ошибку при сравнении с эмпирическими значениями ряда показателя. Предлагается использовать для построения ПНФ метод Шискина–Эйзенпресса [5], суть которого заключается в итерационной процедуре сглаживания ряда путем многократного применения скользящей средней вида

$$X_t = \frac{\sum_{i=-q}^{q} \alpha_i X_{t+i}}{W}. \tag{4}$$

В результате получаем предварительную оценку тренда $\hat{U}_t$. В методе Шискина–Эйзенпресса на втором и последующем этапах итерационной процедуры для выравнивания используются скользящие Спенсера.

Рассчитываются оценки тренда (5) и сезонной компоненты (6):

$$Y_j = \frac{\sum_{i=1}^{m}(X_{ij} - X'_{ij})}{m},\ \overline{Y} = \frac{\sum_{j=1}^{T} Y_j}{T},\ V_j = Y_j - \overline{Y},$$

$$\hat{U}_{ij}^{p} = S[p](X_{ij} - V_j^{p}), \tag{5}$$

$$\hat{V}_j^{p} = \frac{\sum_{i=1}^{m}(X_{ij} - \hat{U}_{ij}^{p})}{m}. \tag{6}$$

Полученные ПНФ можно рассматривать в качестве прогнозов значений показателя функционирования. При проверке на контроль-

ной выборке подсчитаем относительную ошибку прогноза для ПНФ1 и ПНФ2:

$$\varepsilon = \frac{|X - \hat{X}|}{X}. \quad (7)$$

ПНФ, построенный по методу Шискина–Эйзенпресса, точнее описывает поведение показателя функционирования, чем ПНФ, построенный путем простого вычисления средних исторических значений.

4. Заключение. Оперативное выявление сетевых аномалий является важной задачей, которая решается, как правило, специализированными системами обнаружения вторжений или системами управления сетью. Поскольку при защите сети нужно ориентироваться не только на известные в настоящее время угрозы, но и на возможные угрозы, а также учитывать особенности поведения показателей безопасности в конкретном сетевом окружении, решение указанной задачи представляет собой проблему как с точки зрения организации мониторинга информационных потоков в сети, так и с точки зрения анализа полученных данных.

В настоящей работе рассмотрен подход к повышению точности обнаружения сетевых аномалий методом сравнения оперативных прогнозов показателей функционирования с их ПНФ, для построения которого предлагается использовать адаптированный метод Шискина–Эйзенпресса.

Литература

1. *Лукацкий А.В.* Системы обнаружения атак // Сетевой. 2002. № 4. С. 45–48.
2. *Гирик А.В.* Применение методов многокритериального прогнозирования в сетевых системах обнаружения вторжений // Изв. высш. учебн. заведений. Приборостроение. 2009. № 05. С. 34–38.
3. *Гирик А.В.* Обнаружение информационных угроз безопасности передачи данных в телекоммуникационных сетях // Труды XV Всероссийской научно-методической конференции «Телематика», Тезисы докладов, 2008. С. 178–179.
4. *Жигулин Г.П., Серебров А.И., Яковлев А.Д.* Прогнозирование устойчивости и функционирования объектов с использованием теории игр и исследования операций. – СПб: СПбГУИТМО, 2004. 204 с.
5. *Федосеев В.В., Гармаш А.Н., Дайитбегов Д.М. и др.* Экономико-математические методы и прикладные модели: Учеб. пособие для вузов/ Под ред. В.В. Федосеева. – М.: ЮНИТИ, 1999. 391 с.

Гирик Алексей Валерьевич — аспирант кафедры МиПИУ НИУ ИТМО. Область научных интересов: системы обнаружения вторжений, сетевой мониторинг. Число научных

публикаций – 15. ИКВО, Кронверкский пр., 49., Санкт-Петербург, 197101, РФ; р.т.: +7(812)595-4132. Научный руководитель — Г.П. Жигулин.

Guirik Alexei Valerievich — post-graduate student of MaFIT Department, IFMO. Research interests: intrusion detection systems, network monitoring. Number of publications – 15. IKVO, Kronversky prospect, 49, Saint-Petersburg, 197101, Russian Federation; office phone +7(812)595-4132. Research supervisor — G.P. Zhigulin.

Жигулин Георгий Петрович — канд. тех. наук; декан ИКВО – зав. кафедрой МиПИУ НИУ ИТМО. Область научных интересов: интеллектуальные информационные системы. Число научных публикаций – более 70. ИКВО, Кронверкский пр., 49., Санкт-Петербург, 197101, РФ; р.т.: +7(812)595-4132.

Zhigulin Georgiy Petrovich — Candidate of Technical Sc., Dean, Head of Department of MaFIT. Research interests: intellectual information systems. Number of publications – over 70. IKVO, Kronversky prospect, 49, Saint-Petersburg, 197101, Russian Federation; office phone +7(812)595-4132.

Рекомендовано лабораторией автоматизации научных исследований СПИИРАН.
Статья поступила в редакцию 10.03.2013.

РЕФЕРАТ

Гирик А.В., Жигулин Г.П. **Принципы формирования профиля нормального функционирования объектов мониторинга в задачах обнаружения сетевых аномалий**.

В статье предложен подход к формированию профилей нормального функционирования (ПНФ) объектов мониторинга, что является одним из этапов в задачах обнаружения сетевых аномалий. ПНФ — набор функций, описывающих поведение показателя при отсутствии аномалий. Формирование ПНФ является одним из этапов метода обнаружения аномалий. В качестве типичной задачи, возникающей при анализе сетевого трафика, рассматривается обнаружение аномалий в данных мониторинга загрузки некоторого канала связи. Для построения ПНФ могут быть использованы различные подходы, например, агрегированные данных и вычисление среднего исторического значения показателя.

В настоящей работе в качестве математической основы для для формирования ПНФ предлагается использовать итерационные методы, в частности, метод Шискина–Эйзенпресса. Рассматривается временной ряд, содержащий тренд, периодическую и случайную компоненты. Получены оценки тренда и циклической компоненты.

При выполнении итеративной процедуры сглаживания применяются кривые Спенсера. Проверка на эмпирических данных мониторинга канала связи показывает, что ПНФ, построенный с использованием предложенного метода, дает меньшую относительную ошибку, чем ПНФ, построенный по средним историческим значениям показателя. Таким образом, применение описанного метода позволяет повысить вероятность обнаружения аномалий и снизить вероятность ложных срабатываний системы обнаружения вторжений, основанной на статистическом анализе данных сетевого мониторинга.

SUMMARY

Guirik A.V., Zhigulin G.P. **Approach to creating normal functioning profile of monitored objects for network anomalies detection**.

An approach to creating normal functioning profiles (NFP) of monitored objects is considered. NFP is a set of functions describing behavior of particular performance indicator in the absence of anomalous events. NFP creation is one of the key steps in solving problems of network anomalies detection. As a common problem of network traffic analysis anomalies detection in monitoring data of particular network link is considered. For NFP creation different methods can be used, for example, aggregation of data and mean historical values computation.

We propose using iteration methods, Shiskin–Eisenpress method in particular, as a mathematical tool for NFP creation procedure. Time series with trend and cyclic and random components is considered. We obtain evaluations for trend and cyclic component of time series.

Iteration smoothing procedure uses Spencer curves. Described NFP creation method is verified on empirical network monitoring data. It shows more accurate results than NFP creation method that uses data aggregation and mean historical values computation. Thus method proposed is suggested suitable for network anomalies detection since it can increase probability of anomaly detection and reduce the number of false alarms when used in intrusion detection system based on statistical analysis of network flows.