DOI 10.37882/2223–2966.2022.04.21

РАЗРАБОТКА МЕТОДА ПРОГНОЗИРОВАНИЯ И РАСПРЕДЕЛЕНИЯ ТРАФИКА СИСТЕМЫ СБОРА И АНАЛИЗА ИНФОРМАЦИИ

DEVELOPMENT OF A METHOD FOR FORECASTING AND TRAFFIC DISTRIBUTION OF AN INFORMATION COLLECTION AND ANALYSIS SYSTEM

K. Larionov

Summary. The article discusses forecasting methods and develops a new method of traffic forecasting and distribution for the load balancing subsystem in an existing information collection and analysis system. According to statistics, information security threats are most often directed at web applications, therefore, it is necessary to identify abnormal traffic flows as quickly as possible and redistribute them between existing servers in a timely manner. The method of traffic forecasting and distribution is based on the method of traffic control and distribution in a distributed information collection and analysis system.

Keywords: software, traffic, traffic distribution, distribution method, server, forecasting, traffic balancing.

Ларионов Константин Олегович
Аспирант, Оренбургский государственный университет, г. Оренбург
kostya12277@yandex.ru

Аннотация. В статье рассматриваются методы прогнозирования и разрабатывается новый метод прогнозирования и распределения трафика для подсистемы балансировки нагрузки в уже существующей системе сбора и анализа информации. По статистике чаще всего угрозы информационной безопасности направлены на веб-приложения, следовательно, необходимо как можно быстрее определять аномальные потоки трафика и своевременно перераспределять их между существующими серверами. Метод прогнозирования и распределения трафика основан на методе контроля и распределения трафика в распределенной системе сбора и анализа информации.

Ключевые слова: программное обеспечение, трафик, распределение трафика, метод распределения, сервер, прогнозирование, балансировка трафика.

По данным отчета компании Check Point Research каждая 61-я организация в мире становится жертвой киберпреступников, но, например, в России, в сравнении с 2020 годом, в 2021 году количество атак уменьшилось на 16% и всего 2,2% компаний пострадали от киберпреступлений. Среднее число кибератак в неделю на организацию в мире с января 2020 года по сентябрь 2021 года представлено на рисунке 1.

Для определения концепции исследования необходимо провести аналитический обзор современных публикаций на тему разработки метода прогнозирования и распределения трафика системы сбора и анализа информации.

Целью работы является эффективное распределение потока трафика на обрабатывающие сервера производства с использованием методов прогнозирования и статистического анализа данных.

Для достижения поставленной цели работы необходимо выполнить следующие задачи:

1. провести анализ литературы в направлениях прогнозирования и контроля сетевого трафика;
2. разработать математическую модель метода прогнозирования и распределения трафика на базе методов прогнозирования;

К публикациям, отражающим современный уровень исследуемой области относятся работы, описывающие методы равномерного распределения сетевой нагрузки. В частности, к данной теме можно отнести работы:

- Веретенникова П.В. [2];
- Кравченко С.М., Бойко Д.А. [4];
- Пальчевский Е.В., Халиков А.Р. [7–8];
- Мухизи С., Парамонов А.И. [5].

Аналогичными разрабатываемой контроля и распределения трафика системы сбора и анализа информации являются следующие научные разработки:

- научные работы [3,6];
- в диссертационных исследованиях [1,12];
- патентные разработки [9–11].

Аналитический обзор существующих научных достижений позволил определить особенности разрабатываемого метода, предполагаемую научную новизну, однако, в ряде задач необходимо полное исследование

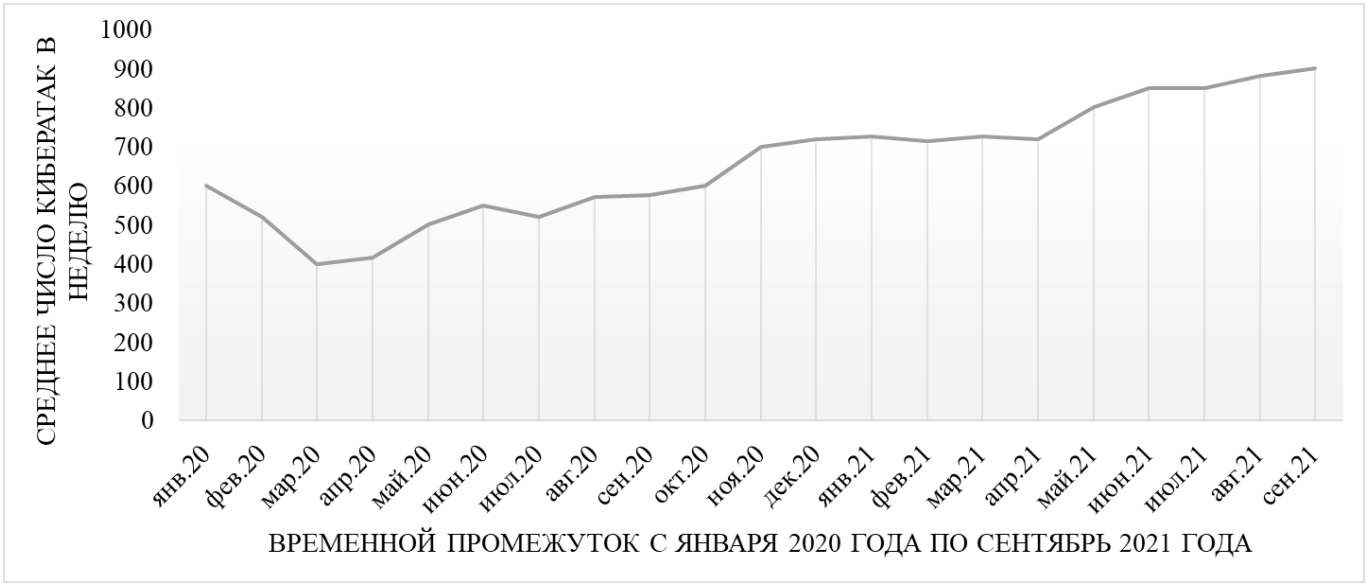

Рис. 1. Среднее число киберпреступных нападений в неделю на организацию в мире с января 2020 года по сентябрь 2021 года

предметной области и разработка метода прогнозирования распределения трафика.

Основываясь на методе контроля и распределения трафика системы сбора и анализа информации прогнозировать конкретный период хранения информации в системе контроля и распределения трафика, не затрагивая весь временной ряд, а используя только собранную информацию по конкретно одной базовой величине микропериода d. Таким образом, для реализации прогноза есть агрегированные данные в которых отражена история X по конкретному d со всех микропериодов t основного выбранного периода p.

Необходимо отметить что выбирать и разделять периоды необходимо только опираясь на проблемы и характеристики производства и его сезонности трафика.

Прогнозирование осуществляется 2 способами. Первый основной способ это с использованием формулы Фурье-анализа, где прогнозные величины повторяют начальное положение временного ряда, что подчеркивает период и его правильность.

Математическое описание модели полигармонического полинома представлено формулой 1.

$$X(t) = a_0 + \sum_{i=1}^{n}[a_i \cdot \cos(2 \cdot \pi \cdot K_i \cdot t / N) + b_i \cdot \sin(2 \cdot \pi \cdot K_i \cdot t / N)] + \varepsilon(t) + d_0 + d_1 \cdot t, \quad (1)$$

где:

N — число элементов исходного ряда;

n — число гармоник полигармонического полинома;

K_i — коэффициенты, определяющие номер гармонии;

$\varepsilon(t)$ — прогнозная оценка случайной компоненты;

d_0, d_1 — коэффициенты уравнения тренда;

t — порядковый номер элементов исходного ряда, t = 1,2, ….

Второй способ прогнозирования основан на определении типа временного ряда.

Определение типа временного ряда для прогнозирования выполняется на основе критерия аппроксимации R. Чем ближе к 1 тем точнее подобран график прогноза а значит и точнее подобран вид прогноза.

Получив линию прогноза агрегированного временного ряда с учетом всех периодов можно решить основную проблему принятия решения о распределении трафика с учетом нагрузки серверов обработки информации.

На рисунке 2 представлена структурная схема системы контроля и распределения трафика системы сбора и анализа информации.

Принятие решения о переадресации трафика на конкретный сервер исходит из прогнозирования временного ряда напрямую. Необходимо отметить, что принятие

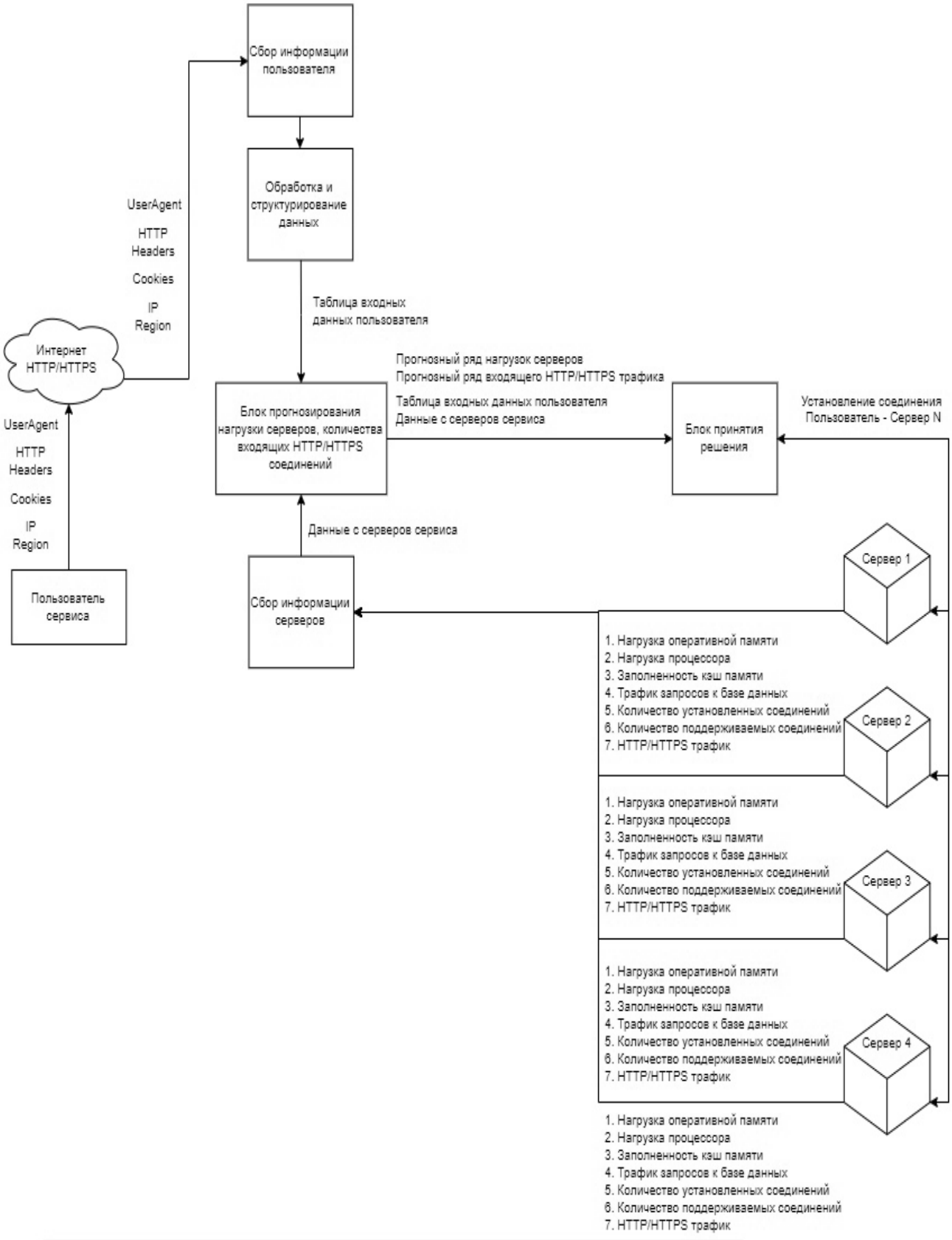

Рис. 2. Структурная схема распределенной информационной системы с использованием системы контроля и распределения трафика

решения и прогнозирование агрегированного ряда связанны конкретным периодом, а именно базовой величиной d микропериода t. Таким образом мы можем принимать решения только находясь в таком же по счету d что и выбранный ряд для прогнозирования.

Математическое описание процесса принятия решения о введении в эксплуатацию дополнительных серверов основывается на решении неравенства по формуле 2.

$$\gamma = a_i * \cos\left(2 * \pi * K_i * \frac{t}{N}\right) + \\ + b_i * \sin\left(2 * \pi * K_i * \frac{t}{N}\right), \\ z = \varepsilon(t) + d_0 + d_1 + t, \\ a_0 + \sum_{i=1}^{n} \gamma + z \geq \frac{C_{max}}{T_{max}} - \frac{C_i}{T_i} \quad (2)$$

где:

γ — гармоническая компонента из уравнения полигармонического полинома формула 1

z — трендовая компонента из уравнения полигармонического полинома формула 1

C_{max} — максимальная ширина пропускного окна сервера

C_i — ширина пропускного окна текущего, выбранного сервера для проверки

T_{max} — максимальная задержка при передачи пакета в секундах

T_i — задержка при передачи пакета в секундах выбранного текущего сервера для проверки

i — номер выбранного включенного сервера в момент времени

Критерий Дарбина — Уотсона (или DW-критерий) — статистический критерий, используемый для тестирования автокорреляции первого порядка элементов исследуемой последовательности. Наиболее часто применяется при анализе временных рядов и остатков регрессионных моделей.

Окончательно качество модели проверяется на величине статистики Дарбина-Уотсона d (Формула 3).

$$d = \frac{\sum_{t=2}^{N}(e_t - e_{t-1})^2}{\sum_{t=2}^{N} e_t^2}, \quad (3)$$

где:

N — число элементов исходного ряда;

e — значение автокорреляции полученное из разницы значений точек основного ряда от значений точек анализируемого исходного ряда;

t — порядковый номер элементов исходного ряда, t = 1,2, ….

Значение статистики Дарбина-Уотсона изменяется в диапазоне от 0 до 4. При этом dd = 2 указывает на отсутствие автокорреляции элементов временного ряда. Если dd меньше двух, то имеет место положительная автокорреляция, а больше двух — отрицательная.

В заключении необходимо отметить, что разработанный метод будет хорошо обрабатывать только те наборы временных рядов, в которых присутствует сезонность. Потому что в работе используется критерий остановки анализа прогнозного ряда, и метод Фурье как основной метод для прогнозирования базовых данных.

ЛИТЕРАТУРА

1. Бузинов, А.А., Модель и метод прогнозирования угроз информационной безопасности объектов на основе циклической динамики природной среды / Бузинов А.А.; Диссертация на соискание ученой степени кандидата технических наук, — Санкт-Петербург, — 2004;
2. Веретенников, П.В., Определение вида распределения нагрузки, изменяющейся по дням недели / Веретенников П.В.; — Санкт-Петербург: Издательство: Ижевский государственный технический университет имени М.Т. Калашникова — 2007, — С. 50–52;
3. Гадасин, Д.В., Применение модели бэкмена для распределения потоков в сетях с сегментной маршрутизацией / Гадасин Д.В., Пак Е.В.; — Москва: Издательство: Российское научно-техническое общество радиотехники, электроники и связи им. А.С. Попова — 2020, — С. 18–23;
4. Кравченко, С.М., Алгоритм оптимального распределения трафика / Кравченко С.М., Бойко Д.А.; — Красноярск: Издательство: Федеральное государственное бюджетное образовательное учреждение высшего образования «Сибирский Государственный университет науки и технологий имени академика М.Ф. Решетнева» (Красноярск) — 2020, — С. 408–409;
5. Мухизи, С., Метод классификации и приоритизации трафика в программно-конфигурируемых сетях / Мухизи С., Парамонов А.И.; — Санкт-Петербург: Издательство: Санкт-Петербургский Государственный Университет телекоммуникаций им. проф. М.А. Бонч-Бруевича, — 2019, — С. 64–70;
6. Немер, С., Процесс динамической маршрутизации разноприоритетного MPLS-трафика / Немер С.; — Москва: Издательство: ООО «Издательский дом медиа паблишер» — 2013, — С. 109–111;
7. Пальчевский, Е.В., Равномерное распределение сетевой нагрузки при атаке несанкционированным трафиком / Пальчевский Е.В., Халиков А.Р.; — Уфа: Издательство: Общество с ограниченной ответственностью Дендра (Уфа) — 2017, — С. 21–27;
8. Пальчевский, Е.В., Распределение сетевой нагрузки при DDOS-атаках /Пальчевский Е.В., Халиков А.Р.; — Уфа: Издательство: Общество с ограниченной ответственностью Дендра (Уфа) — 2017, — С. 34–38;

9. Патент 2018666578 Российская Федерация. «Моделирование функционирования распределенных систем обработки информации на базе трехуровневой клиент-серверной архитектуры» (МРСОИ) / Айеш А.Н., Михайлов В.К., Скоба А.Н.; заявитель и патентообладатель: федеральное государственное бюджетное образовательное учреждение высшего образования «Южно-Российский государственный политехнический университет (НПИ) имени М.И. Платова» заявл. 03.12.2018, опубл. 18.12.2018;
10. Патент 2019661408 Российская Федерация. Моделирование обслуживания трафика маршрутизаторами технологии интернета вещей (IOT) / Кутузов Д.В., Осовский А.В., Стукач О.В., Старов Д.В., Моторина Е.А.; заявитель и патентообладатель: Федеральное государственное бюджетное образовательное учреждение высшего образования «Астраханский государственный технический университет» заявл. 19.08.2019, опубл. 28.08.2019;
11. Патент 2703339 Российская Федерация. Способ моделирования процесса обоснования требований к системе мониторинга распределенных систем связи / Анисимов В.Г., Анисимов Е.Г., Гречишников Е.В., Кежаев В.А., Белов А.С., Сысуев С.Ю., Люборчук Ф.Н., Деров М.Н., Молоткова Б.Б., Сауренко Т.Н.; заявитель и патентообладатель: Федеральное государственное казенное военное образовательное учреждение высшего образования «Михайловская военная артиллерийская академия» Министерства Обороны Российской Федерации, заявл. 02.08.2018, опубл. 16.10.2019;
12. Пономарев, Д.Ю., Моделирование и оптимизация распределения трафика в телекоммуникационных сетях. Диссертация на соискание учёной степени доктора наук / Пономарев Д.Ю.; — Новосибирск: ФГБОУ ВО «Сибирский государственный университет науки и технологий им. акад. М.Ф. Решетнева», 2019.— 327 с.

Журнал «Современная наука: актуальные проблемы теории и практики»

г. Оренбург