

2023



3(21)

ИНФОКОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ

Infokommunikacionnye
Tehnologii

«Инфокоммуникационные технологии»

Периодический научно-технический и информационно-аналитический журнал
Том 21, № 3 (83), 2023

Учредитель и издатель журнала:

Поволжский государственный университет телекоммуникаций и информатики (г. Самара)

Главный редактор, председатель редакционного совета

Горячкин О.В., д.т.н., профессор (ПГУТИ, г. Самара).

Заместители главного редактора

Васин Н.Н., д.т.н., профессор; Тарасов В.Н., д.т.н., профессор.

Редакционный совет

Мишин Д.В., д.т.н., профессор (ПГУТИ, г. Самара); Сойфер В.А., академик РАН (Самарский университет, г. Самара); Бузов А.Л., д.т.н., профессор (АО «СИП РС», г. Самара); Васильев К.К., д.т.н., профессор (УлГТУ, г. Ульяновск); Воронин А.А., д.ф.-м.н., профессор (ВолГУ, г. Волгоград); Кубанов В.П., д.т.н., профессор (ПГУТИ, г. Самара); Лихтциндер Б.Я., д.т.н., профессор (ПГУТИ, г. Самара); Надеев А.Ф., д.ф.-м.н., профессор (КНИТУ-КАИ им. А.Н. Туполева, г. Казань); Султанов А.Х., д.т.н., профессор (УУНИТ, г. Уфа); Цым А.Ю., д.т.н., с.н.с. (ЦНИИС, г. Москва); Щербаков М.А., д.т.н., профессор (ПГУ г. Пенза); Морозов О.Г., д.т.н., профессор (КНИТУ-КАИ, г. Казань).

Редакционная коллегия

Бахарева Н.Ф., д.т.н., профессор (ПГУТИ, г. Самара); Карташевский В.Г., д.т.н., профессор (ПГУТИ, г. Самара); Осипов О.В., д.ф.-м.н., доцент (ПГУТИ, г. Самара); Росляков А.В., д.т.н., профессор (ПГУТИ, г. Самара); Тяжев А.И., д.т.н., профессор (ПГУТИ, г. Самара); Карташевский И.В., д.т.н., доцент (СПбГУТ, г. Санкт-Петербург); Бурдин А.В., д.т.н., доцент (АО «НПО ГОИ им. С. И. Вавилова», г. Санкт-Петербург); Картак В.М., д.ф.-м.н., профессор (УГАТУ, г. Уфа); Сингх Г., PhD в области электроники и связи, профессор (Малавийский национальный технологический институт Джайпура, г. Джайпур); Ким Ч., PhD в области компьютерных наук, доцент (Корейский университет, г. Сеул); Тивари М., PhD в области электроники и связи по направлению «Оптоэлектроника», профессор (Манипальский университет Джайпура, г. Джайпур); Джаньяни В., PhD в области электроники и связи, профессор (Малавийский национальный технологический институт Джайпура, г. Джайпур); Али Дж., PhD в области физики плазмы, профессор (Азиатский столичный университет, г. Джохор-Бару).

Выпускающие редакторы: Диязитдинова А.А., Лиманова А.И.

Редактор номера: Тарасов В.Н.

Компьютерный набор и верстка: Вукмирович Д.Д.

Корректора: ООО «Навигатор»

Адрес редакции и издателя журнала:

443010, Российская Федерация, г. Самара, ул. Льва Толстого, 23,
Поволжский государственный университет телекоммуникаций и информатики.

Телефон +7 846 339-11-09. E-mail: redikt@psuti.ru

Издается с 2003 г. Выходит 4 раза в год.

ПИ № ФС77-82293 от 08.12.2021 г. выдано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций (Роскомнадзор).

Решением ВАК Минобрнауки России журнал «Инфокоммуникационные технологии» включен в перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук.

© Инфокоммуникационные технологии, 2024.

Подписной индекс 41914 в каталоге ООО «Урал-Пресс Округ».

Тиражирование содержания журнала
в электронном и любом другом виде –
с письменного разрешения редакции.
443010, Российская Федерация, г. Самара,
ул. Льва Толстого, 23

Дата выхода в свет 13.05.2024
Цена свободная. 0+. Тираж 100 экз.
Отпечатано в типографии ООО «Слово».
443070, Российская Федерация, г. Самара,
ул. Песчаная, 1. Тел. +7 846 267-36-82
E-mail: izdatkniga@yandex.ru

СОДЕРЖАНИЕ

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ТЕХНОЛОГИЙ ПЕРЕДАЧИ И ОБРАБОТКИ ИНФОРМАЦИИ И СИГНАЛОВ

| | | |
|-----------------------------------|--|----|
| Лихтциндер Б.Я. | Дисперсия очередей в системах массового обслуживания с групповыми пуассоновскими потоками | 7 |
| Лихтциндер Б.Я., Привалов А.Ю. | Об использовании группового пуассоновского потока в имитационном моделировании современного видеотрафика | 11 |

СЕТИ СВЯЗИ И МУЛЬТИСЕРВИСНЫЕ УСЛУГИ

| | | |
|-------------------------------|---|----|
| Герасимов В.В. | Метод обучения с подкреплением для динамического распределения приоритетов в TSN | 16 |
| Поздняк И.С., Макаров И.С. | Использование алгоритмов машинного обучения для обнаружения аномального поведения трафика | 20 |

ТЕХНОЛОГИИ КОМПЬЮТЕРНЫХ СИСТЕМ И СЕТЕЙ

| | | |
|--|---|----|
| Рычкова А.А., Долгушев Н.А., Бурькова Е.В., Коннов А.Л. | Модели и алгоритмы определения актуальных угроз на основе экспертного подхода | 27 |
| Галич С.В., Сафонова О.Е. | Специалист по эксплуатации программно-конфигурируемых сетей: знания, умения, навыки | 35 |

НОВЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

| | | |
|------------------------------------|--|----|
| Захарова О.И. | Технологии компьютерного зрения для обнаружения дефектов цилиндрических объектов | 48 |
| Пальмов С.В., Диязитдинова А.А. | Комбинированный метод настройки гиперпараметров математической модели | 56 |
| Собин А.Е. | Программная оптимизация процесса распределения ресурсов в метростроительной организации: анализ эффективности существующих методов | 62 |

| | | |
|--|---|-----------|
| Малина А.Б., Голубкин Д.А., Вехова А.Д. | Применение информационно-коммуникационных технологий для поддержания физического здоровья человека | 71 |
|--|---|-----------|

| | | |
|---|---|-----------|
| Дукальская И.В., Киргизов А.Г. | Разработка приложения для анализа комментариев англоязычного Youtube | 78 |
|---|---|-----------|

ТЕХНОЛОГИИ РАДИОСВЯЗИ, РАДИОВЕЩАНИЯ И ТЕЛЕВИДЕНИЯ

| | | |
|--------------------|---|-----------|
| Кустов Н.Д. | Интегрированная космическая сеть: результаты первичного моделирования маршрутизируемой межспутниковой сети | 83 |
|--------------------|---|-----------|

УПРАВЛЕНИЕ И ПОДГОТОВКА КАДРОВ ДЛЯ ОТРАСЛИ ТЕЛЕКОММУНИКАЦИЙ

| | | |
|--|---|-----------|
| Селиванова Е.С., Газизов Т.Т., Долганова Н.Ф., Пираков Ф.Д. | Использование машинного обучения для прогнозирования трудоустройства выпускников | 91 |
|--|---|-----------|

CONTENTS

THEORETICAL FOUNDATIONS OF INFORMATION AND SIGNAL TRANSMISSION AND PROCESSING TECHNOLOGIES

| | | |
|--|---|-----------|
| Lichtzinder B.Ya. | Queue Dispersion in Queue Systems with Group Poisson Flows | 7 |
| Lichtzinder B.Ya., Privalov A.Yu. | Use of Group Poisson Flow in Simulation Modeling of Modern Video Traffic | 11 |

COMMUNICATION NETWORKS AND MULTI-SERVICES

| | | |
|--|--|-----------|
| Gerasimov V.V. | Reinforcement Learning Method for Dynamic Allocation of Priorities in TSN | 16 |
| Pozdnyak I.S., Makarov I.S. | Using Machine Learning Algorithms to Detect Anomalous Traffic Behavior | 20 |

COMPUTER NETS AND SYSTEMS TECHNOLOGIES

| | | |
|--|--|-----------|
| Richkova A.A., Dolgusheva N.A., Burkova E.V., Konnov A.L. | Models and Algorithms for Identifying Current Threats Based on an Expert Approach | 27 |
| Galich S.V., Savonova O.E. | Knowledge and Skills for Software-Defined Networks Maintenance Engineer | 35 |

NEW INFORMATION TECHNOLOGIES

| | | |
|---|---|-----------|
| Zakharova O.I. | Computer Vision Technologies for Detecting Defects in Cylindrical Objects | 48 |
| Palmov S.V., Diyazitdinova A.A. | Combined Method for Tuning Hyperparameters of a Mathematical Model | 56 |
| Sobin A.E. | Software Optimization of Resource Allocation Process in Metro Construction Organizations: Analysis of Efficiency of Existing Methods | 62 |
| Malina A.B., Golubkin D.A., Vekhova A.D. | The Use of Information and Communication Technologies to Maintain Human Physical Health | 71 |
| Dukalskaya I.V., Kirgizov A.G. | Development of an Application for Analyzing Comments on English-Language Youtube | 78 |

**RADIO TELECOMMUNICATION, RADIOBROADCASTING AND TELEVISION
TECHNOLOGIES**

| | | |
|--------------------|--|-----------|
| Kustov N.D. | Space Integrated Network: Routed Inter-Satellite Network Primary Modeling Results | 83 |
|--------------------|--|-----------|

MANAGEMENT AND TRAINING FOR THE TELECOMMUNICATIONS INDUSTRY

| | | |
|--|--|-----------|
| Selivanova E.S., Gazizov T.T., Dolganova N.F., Pirakov F.D. | Using Machine Learning Methods for Predicting the Employment of Graduates | 91 |
|--|--|-----------|

«Infokommunikacionnye tehnologii»

Periodic technical and scientific information and analytical magazin
2023, vol. 21, no. 3 (83)

Founder and Publisher of the Journal:

Povolzhskiy State University of Telecommunications and Informatics (Samara)

Chief Editor, Chairman of the Editorial Management

Goryachkin O.V., Tenured Professor, Doctor (PSUTI, Samara).

Deputy Chief Editor

Vasin N.N., Tenured Professor, Doctor; Tarasov V.N., Tenured Professor, Doctor.

Editorial Management

Mishin D.V., Tenured Professor, Doctor (PSUTI, Samara);

Soifer V.A., Academic Professor RAS (Samara National Research University, Samara);

Buzov A.L., Tenured Professor, Doctor (SIPRS, Ltd., Samara);

Vasiliev K.K., Tenured Professor, Doctor (UISTU, Ulyanovsk);

Voronin A.A., Tenured Professor, Doctor (VolSU, Volgograd);

Kubanov V.P., Tenured Professor, Doctor (PSUTI, Samara);

Likhtsinder B.Y., Tenured Professor, Doctor (PSUTI, Samara);

Nadeev A.F., Tenured Professor, Doctor (KNRTU-KAI, Kazan);

Sultanov A.H., Tenured Professor, Doctor (UUNiT, Ufa);

Tsym A.Y., Tenured Professor, Senior Manager Associate (CSRJC, Moscow);

Shcherbakov M.A., Tenured Professor, Doctor (PSU, Penza);

Morozov O.G., Tenured Professor, Doctor (KNRTU-KAI, Kazan).

Editorial Board

Bahareva N.F., Tenured Professor, Doctor (PSUTI, Samara); Kartashevskii V.G., Tenured Professor, Doctor (PSUTI, Samara); Osipov O.V., Associate Professor, Doctor (PSUTI, Samara); Roslyakov A.V.,

Tenured Professor, Doctor (PSUTI, Samara); Tyagev A.I., Tenured Professor, Doctor (PSUTI,

Samara); Kartashevsky I.V., Associate Professor, Doctor (SPbSUT, Saint Petersburg); Burdin A.V.,

Associate Professor, Doctor (JSC S.I. Vavilov State Optical Institute, Saint Petersburg); Kartak V.M.,

Professor, Doctor (USATU, Ufa); Singh G., Professor, PhD in Electronics and Communication Engineering

(Malaviya National Institute of Technology Jaipur, Jaipur); Kim J., Associate Professor, PhD in Computer

Science (Korea University, Seoul); Tiwari M., Professor, PhD in Electronics and Communication

Engineering with specialization in Optoelectronics (Manipal University Jaipur, Jaipur); Janyani V., Professor,

PhD in Electronics and Communication Engineering (Malaviya National Institute of Technology Jaipur,

Jaipur); Ali J., Professor, PhD in Plasma Physics (Asia Metropolitan University, Johor Bahru).

Executive Editors: Diyazitdinova A.A., Limanova A.I.

Editor of Issue: Tarasov V.N.

Type Setting and Page Make Up: Vukmirovich D.D.

Proofreading: «Navigator» LLC

Editor and Publishing House Address:

23, L. Tolstoy Street, Samara, 443010, Russian Federation,

Federal State Educational Budgetary Institution of Higher Professional Education

«Povolzhskiy State University of Telecommunications and Informatics».

Phone +7 846 339-11-09. E-mail: redikt@psuti.ru

Published since 2003. Edited 4 times a year. Licensed PI № FS 77-82293 on 08.12.2021 issued by The

Federal Service for Supervision of Communications, Information Technology, and Mass Media (Roskomnadzor)

Under the verdict of Higher Attestation Commission of the Ministry of Education and Science of the Russian Federation, the journal «Info-communication Technologies» is included in the list of peer-reviewed scientific editions, in which the main scientific results of dissertations for the degree of candidate of sciences, for the degree of doctor of sciences should be published.

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ТЕХНОЛОГИЙ ПЕРЕДАЧИ И ОБРАБОТКИ ИНФОРМАЦИИ И СИГНАЛОВ

УДК 621.391: 621 39

ДИСПЕРСИЯ ОЧЕРЕДЕЙ В СИСТЕМАХ МАССОВОГО ОБСЛУЖИВАНИЯ С ГРУППОВЫМИ ПУАССОНОВСКИМИ ПОТОКАМИ

Лихтциндер Б.Я.

Поволжский государственный университет телекоммуникаций и информатики, Самара, РФ
E-mail: b.lihtcinder@psuti.ru

Рассматриваются групповые пуассоновские потоки в системах массового обслуживания. Групповые пуассоновские модели потоков позволяют получить весьма простые зависимости средних размеров очередей от коэффициента загрузки однолинейной системы массового обслуживания. Показано, что дисперсия чисел заявок в таких потоках линейно зависит от загрузки и определяет средние значения размеров очередей. Поскольку в групповом пуассоновском потоке пачки заявок независимы, в формулах для средних значений очередей отсутствуют элементы, определяемые корреляционными связями. Получены соотношения, определяющие второй начальный момент и дисперсию размеров очередей в системах массового обслуживания с групповыми пуассоновскими потоками. Значения второго начального момента размера очередей получены возведением уравнения баланса в третью степень. Показано, что дисперсия размеров очередей зависит от третьего центрального момента чисел поступающих заявок. Третий центральный момент характеризует степень асимметричности распределения размеров пачек заявок.

Ключевые слова: групповые пуассоновские потоки, системы массового обслуживания, дисперсии, коэффициент загрузки, очереди

Введение

Пакетная передача в современных телекоммуникационных сетях показала, что модели трафика, основанные на распределении Пуассона, не являются адекватными. Пачечный характер пакетного трафика, взаимная зависимость пакетов приводят к существенному влиянию корреляционной составляющей на размеры очередей. Существует множество моделей трафика, учитывающих его корреляционные свойства. Многие из указанных моделей учитывают корреляционные свойства потоков, но они оказались слишком сложными и не привели к существенным результатам.

Модели «самоподобного» трафика, учитывающие корреляционные свойства, оказались слишком сложными, неэффективными. На смену им пришел целый класс моделей потоков с коммутируемой цепью Маркова – ВМАР-потоки [1–3; 6–13]. Среди указанных потоков можно выделить групповые пуассоновские потоки [5]. Это неординарные потоки событий, каждое из которых соответствует одновременному появлению нескольких заявок. События возникают независимо друг от друга и имеют экспоненциальное распределение интервалов между соседними событиями. Следовательно, поток событий является пуассоновским.

Обозначим через τ некоторый интервал времени обработки заявки в системе массового обслуживания (СМО). Среднее значение числа заявок

рассматриваемого потока, так же как и дисперсия $D_m(\rho)$ пропорционально загрузке ρ :

$$\overline{m(\tau)} = \lambda \tau = \rho.$$

$$D_m(\rho) = \lambda \tau \bar{k}^2 = \rho \bar{k}(1 + v_k^2) = E_m \rho.$$

Здесь, $v_k^2 = \frac{D_k}{(\bar{k})^2}$ – квадрат коэффициента вариации, а \bar{k} – среднее число заявок в «пачке».

Интервальный метод [4], основанный на анализе чисел заявок, поступающих в течение одного интервала обработки заявки, обеспечивает установление зависимости среднего размера очереди от коэффициента загрузки системы массового обслуживания.

В групповом пуассоновском потоке пачки заявок независимы, поэтому в формуле для средних значений очередей [4] нет элемента, определяемого корреляционными связями, и формула упрощается.

$$\overline{q(\rho)} = \frac{D_m(\rho)}{2(1-\rho)} - \frac{\rho}{2} = \frac{E_m \rho}{2(1-\rho)} - \frac{\rho}{2}. \quad (1)$$

В частном случае ординарного пуассоновского потока: $k_i = 1$, $\bar{k} = 1$, $v_k^2 = 0$, $E = 1$ – при этом справедлива известная формула:

$$\overline{q(\rho)} = \frac{\rho^2}{2(1-\rho)}.$$

Если два совершенно различных потока имеют одинаковые значения дисперсии, то при одинаковых загрузках средние размеры очередей для таких потоков будут полностью совпадать.

Для таких потоков отличаться будут лишь значения дисперсий $D_q(\rho)$ размеров очередей.

Второй начальный момент

Второй начальный момент размера очереди $q^2(\rho)$ определим из соотношения баланса [4].

$$\begin{aligned} q_i(\rho) &= q_{i-1}(\rho) + m_i(\rho) - \delta_i(\rho); \\ \delta_i(\rho) &= 0, \text{ если } q_{i-1}(\rho) = m_i(\rho) = 0; \\ \delta_i(\rho) &= 1, \text{ в противном случае.} \end{aligned} \quad (2)$$

где $q_i(\rho)$ – значение очереди, $m_i(\rho)$ – число поступивших заявок, а $\delta_i(\rho)$ – обработанных число заявок, в течение интервала времени τ .

Ограничения имеют следующий вид:

$$\begin{aligned} \delta_i^k(\rho) &= \delta_i(\rho), \quad m_i(\rho)\delta_i(\rho) = m_i(\rho), \\ q_{i-1}(\rho)\delta_i(\rho) &= q_{i-1}(\rho), \quad \overline{\delta_i(\rho)} = \overline{m_i(\rho)}. \end{aligned} \quad (3)$$

При возведении в третью степень уравнения (2), коэффициент загрузки для ρ краткости будем опускать:

$$q_i^3 = q_{i-1}^3 + 3q_{i-1}^2(m_i - \delta_i) + 3q_{i-1}(m_i - \delta_i)^2 + (m_i - \delta_i)^3.$$

После соответствующих преобразований, с учетом принятых ограничений получим:

$$q_i^3 = q_{i-1}^3 + 3q_{i-1}^2 m_i - 3q_{i-1}^2 + 3q_{i-1}(m_i^2 - 2m_i + 1) + (m_i^3 - 3m_i^2 + 3m_i - \delta_i).$$

Произведем усреднение обеих частей:

$$3\overline{q_{i-1}^2} (1 - \overline{m_i}) = 3\overline{q_{i-1}} (\overline{m_i^2} - 2\overline{m_i} + 1) + (\overline{m_i^3} - 3\overline{m_i^2} + 2\overline{m_i}).$$

$$3\overline{q_{i-1}^2} (1 - \overline{m_i}) = 3\overline{q_{i-1}} (\overline{D_m} + \overline{m_i} - 2\overline{m_i} + 1) + (\overline{m_i^3} - 3\overline{m_i^2} + 2\overline{m_i}).$$

$$3\overline{q_{i-1}^2} (1 - \overline{m_i}) = 3\overline{q_{i-1}} \overline{D_m} + 3\overline{q_{i-1}} (\overline{m_i} - 1)^2 + (\overline{m_i^3} - 3\overline{m_i^2} + 2\overline{m_i}).$$

Определим значение $\overline{q_{i-1}^2}$:

$$\overline{q_{i-1}^2} = \frac{\overline{q_{i-1}} \overline{D_m}}{(1 - \overline{m_i})} + \overline{q_{i-1}} (1 - \overline{m_i}) + \frac{\overline{m_i^3} - 3\overline{m_i^2} + 2\overline{m_i}}{3(1 - \overline{m_i})}.$$

$$\overline{q_{i-1}^2} = \overline{q_{i-1}} \left[\frac{\overline{D_m}}{(1 - \overline{m_i})} + (1 - \overline{m_i}) \right] + \frac{\overline{m_i^3} - 3\overline{m_i^2} + 2\overline{m_i}}{3(1 - \overline{m_i})}.$$

$$\overline{q_{i-1}^2} = \overline{q_{i-1}} \frac{\overline{D_m} + (1 - \overline{m_i})^2}{(1 - \overline{m_i})} + \frac{\overline{m_i^3} - 3\overline{m_i^2} + 2\overline{m_i}}{3(1 - \overline{m_i})}.$$

$$\overline{q_{i-1}^2} = \frac{\overline{q_{i-1}}}{(1 - \overline{m_i})} [\overline{D_m} + (1 - \overline{m_i})^2] + \frac{\overline{m_i^3} - 3\overline{m_i^2} + 2\overline{m_i}}{3(1 - \overline{m_i})}.$$

Выразим третий начальный момент $\overline{m_i^3}$ через третий центральный момент $\mu_3 = (\overline{m_i} - m_i)^3$.

$$\overline{m_i^3} = (\overline{m_i} - m_i)^3 + 3\overline{m_i} \overline{D_m} + \overline{m_i}^3 = \mu_3 + 3\overline{m_i} \overline{D_m} + \overline{m_i}^3.$$

После подстановки получим:

$$\overline{q_{i-1}^2} = \overline{q_{i-1}} \left[\frac{\overline{D_m} + (1 - \overline{m_i})^2}{(1 - \overline{m_i})} \right] + \frac{3\overline{m_i} \overline{D_m} + \overline{m_i}^3 - 3\overline{D_m} - 3\overline{m_i}^2 + 2\overline{m_i}}{3(1 - \overline{m_i})} + \frac{\mu_3}{3(1 - \overline{m_i})},$$

$$\begin{aligned} \overline{q_{i-1}^2} &= \frac{\overline{q_{i-1}}}{(1 - \overline{m_i})} [\overline{D_m} + (1 - \overline{m_i})^2] + \\ &+ \frac{\overline{m_i}^3 + 3\overline{m_i} \overline{D_m} - 3\overline{m_i}^2 + 2\overline{m_i}}{3(1 - \overline{m_i})} + \frac{\mu_3}{3(1 - \overline{m_i})}. \end{aligned}$$

Подставляя $\overline{q_{i-1}} = \overline{q(\rho)}$ из (1) и учитывая, что $\overline{m(\rho)} = \lambda \tau = \rho$, получим:

$$\begin{aligned} \overline{q^2(\rho)} &= \frac{\overline{D_m(\rho)} - \rho(1 - \rho)}{2(1 - \rho)^2} [\overline{D_m(\rho)} + (1 - \rho)^2] + \\ &+ \frac{\rho^3 + 3\rho \overline{D_m(\rho)} - 3\overline{D_m(\rho)} - 3\rho^2 + 2\rho}{3(1 - \rho)} + \frac{\mu_3(\rho)}{3(1 - \rho)}. \end{aligned}$$

Приведенное ниже соотношение позволяет определить второй начальный момент размера очереди в СМО с групповыми пуассоновскими потоками.

Для групповых потоков второй момент размера очереди определится соотношением:

$$\begin{aligned} \overline{q^2(\rho)} &= \frac{E_m \rho - \rho + \rho^2}{2(1 - \rho)^2} [E_m \rho + 1 - 2\rho + \rho^2] + \\ &+ \frac{\rho^3 + 3E_m \rho^2 - 3E_m \rho - 3\rho^2 + 2\rho}{3(1 - \rho)} + \frac{\mu_3(\rho)}{3(1 - \rho)}. \end{aligned}$$

Для простейшего потока $E_m = 1$, $\mu_3(\rho) = \rho$ выражение упростится:

$$\overline{q^2(\rho)} = \frac{\rho^2}{2(1 - \rho)^2} \left(1 - \frac{1}{3}\rho + \frac{1}{3}\rho^2 \right).$$

Мы убеждаемся, что в этом случае второй момент зависит только от коэффициента загрузки.

Дисперсия очередей

Дисперсию размеров очередей определим на основании соотношения:

$$\overline{D_q(\rho)} = \overline{q^2(\rho)} - \overline{q(\rho)}^2.$$

Выполним преобразования:

$$\begin{aligned}
\overline{D_q(\rho)} &= \frac{D_m(\rho) - \rho(1-\rho)}{2(1-\rho)^2} [D_m(\rho) + (1-\rho)^2] + \frac{\rho^3 + 3\rho D_m(\rho) - 3D_m(\rho) - 3\rho^2 + 2\rho}{3(1-\rho)} \frac{\mu_3(\rho)}{3(1-\rho)} - \\
&\quad - \frac{[D_m(\rho) - \rho(1-\rho)]^2}{4(1-\rho)^2} = \frac{D_m(\rho) - \rho(1-\rho)}{2(1-\rho)^2} [D_m(\rho) + (1-\rho)^2 - \frac{D_m(\rho) - \rho(1-\rho)}{2}] + \\
&+ \frac{\rho^3 + 3\rho D_m(\rho) - 3D_m(\rho) - 3\rho^2 + 2\rho}{3(1-\rho)} + \frac{\mu_3(\rho)}{3(1-\rho)} = \frac{D_m(\rho) - \rho(1-\rho)}{4(1-\rho)^2} [2D_m(\rho) + 2(1-\rho)^2 - D_m(\rho) + \rho(1-\rho)] + \\
&+ \frac{\rho^3 + 3\rho D_m(\rho) - 3D_m(\rho) - 3\rho^2 + 2\rho}{3(1-\rho)} + \frac{\mu_3(\rho)}{3(1-\rho)} = \frac{D_m(\rho) - \rho(1-\rho)}{4(1-\rho)^2} [D_m(\rho) + 2 - 4\rho + 2\rho^2 + \rho - \rho^2] + \\
&+ \frac{\rho^3 + 3\rho D_m(\rho) - 3D_m(\rho) - 3\rho^2 + 2\rho}{3(1-\rho)} + \frac{\mu_3(\rho)}{3(1-\rho)} = \frac{D_m(\rho) - \rho(1-\rho)}{4(1-\rho)^2} [D_m(\rho) + 2 - 3\rho + \rho^2] + \\
&+ \frac{\rho^3 + 3\rho D_m(\rho) - 3D_m(\rho) - 3\rho^2 + 2\rho}{3(1-\rho)} + \frac{\mu_3(\rho)}{3(1-\rho)}.
\end{aligned}$$

Для простейшего пуассоновского потока
 $E_m = 1, D_m(\rho) = \mu_3(\rho) = \rho,$

$$\overline{D_q(\rho)} = \frac{\rho^2}{2(1-\rho)^2} \left(1 - \frac{1}{3}\rho - \frac{1}{6}\rho^2\right).$$

Выражение существенно упрощается.

Заключение

Дисперсия очередей пуассоновского потока зависит только от коэффициента загрузки СМО, а дисперсию группового пуассоновского потока определяет третий центральный момент закона распределения размеров пачек заявок. Третий центральный момент зависит от симметричности закона распределения.

Литература

1. Вишнеvский В.М., Дудин А.Н. Системы массового обслуживания с коррелированными входными потоками и их применение для моделирования телекоммуникационных сетей // Автоматика и телемеханика. 2017. № 8. С. 3–59.
2. Neuts M.F. A Versatile Markovian point process // Journal of Applied Probability. 1979. Vol. 16, no. 4. P. 764–779. DOI: 10.2307/3213143
3. Дудин А.Н., Клименок В.И. Системы массового обслуживания с коррелированными потоками. Минск: БГУ, 2000. 175 с.
4. Лихтциндер Б.Я. Трафик мультисервисных сетей доступа (интервальный анализ и проектирование). М.: Горячая линия – Телеком, 2018. 290 с.
5. Лихтциндер Б.Я., Моисеев В.И. Групповые пуассоновские и гиперпуассоновские модели пакетного трафика // I-Methods. 2022. Т. 14, № 3. С. 2–11.
6. Ramaswami V. The N/G/1 queue and its detailed analysis // Advances in Applied Probability. 1980. Vol. 12, no. 1. P. 222–261. DOI: 10.2307/1426503
7. Lakatos L., Szeidl L., Telek M. Introduction to queueing systems with telecommunication applications. Springer Science and Business Media, 2013. 388 p. DOI: 10.1007/978-1-4614-5317-8
8. Flexible dual-connectivity spectrum aggregation for decoupled uplink and downlink access in 5G heterogeneous systems / M.A. Lema [et al.] // IEEE Journal on Selected Areas in Communications. 2016. Vol. 34, no. 1. P. 2851–2865. DOI: 10.1109/JSAC.2016.2615185
9. A multiband OFDMA heterogeneous network for millimeter wave 5G wireless applications / S. Niknam [et al.] // IEEE Access. 2016. Vol. 4. P. 5640–5648. DOI: 10.1109/ACCESS.2016.2604364
10. Vishnevsky V., Larionov A., Frolov S. Design and scheduling in 5G stationary and mobile communication systems based on wireless millimeter-wave mesh networks // Distributed Computer and Communication Networks. 2014. Vol. 279. P. 11–27. DOI: 10.1007/978-3-319-05209-0_2
11. Applying graph-theoretic approach for time-frequency resource allocation in 5G mm wave backhaul network / V.M. Vishnevsky [et al.] // Advances in Wireless and Optical Communications (RTUWO). 2016. P. 221–224. DOI: 10.1109/RTUWO.2016.7821888

12. On the self-similar nature of Ethernet traffic / W.E. Leland [et al.] // IEEE/ACM Transactions on Networking. 1994. Vol. 2, no. 1. P. 1–15.
13. Цыбаков Б.С. Модель телетрафика на основе самоподобного случайного процесса // Радиотехника. 1999. № 5. С. 24–31.

Получено 28.11.2023

Лихтциндер Борис Яковлевич, д.т.н., профессор, профессор кафедры сетей и систем связи Поволжского государственного университета телекоммуникаций и информатики. 443010, Российская Федерация, г. Самара, ул. Л. Толстого, 23. Тел. +7 846 333-47-69. Email: lixt@psuti.ru

QUEUE DISPERSION IN QUEUE SYSTEMS WITH GROUP POISSON FLOWS

Lichtzinder B. Ya.

*Povolzhskiy State University of Telecommunications and Informatics, Samara, Russian Federation
E-mail: lixt@psuti.ru*

The application of interval methods for queue analysis in queuing systems with group Poisson flows is considered. It is shown that, for a given load, the average value of queue sizes in systems with such flows is determined by the dispersion of the numbers of requests arriving during the processing time intervals of one request, and the specified dispersion depends linearly on the load factor. Relations are obtained that determine the dispersion of queue sizes in queuing systems with group Poisson flows. It is shown that this dispersion depends on the third central moment of the numbers of applications arriving during the time intervals of processing one application. The dispersions of the queues of two different group flows, which have the same dependencies of the average values of the queues, differ in the values of the indicated central moments.

Keywords: *group Poisson flows, queuing systems, variances, load factor, queues*

DOI: 10.18469/ikt.2023.21.3.01

Likhttsinder Boris Yakovlevich, Povolzhskiy State University of Telecommunications and Informatics, 23, L. Tolstoy Street, Samara, 443010, Russian Federation; Professor of Networks and Communication Systems Department, Doctor of Technical Science. Tel. +7 846 333-47-69. E-mail: lixt@psuti.ru

References

1. Vishnevskij V.M., Dudin A.N. Queueing systems with correlated arrival flows and their applications to modeling telecommunication networks. *Avtomatika i telemekhanika*, 2017, no. 8, pp. 3–59. (In Russ.)
2. Neuts M.F. A Versatile Markovian point process. *Journal of Applied Probability*, 1979, vol. 16, no. 4, pp. 764–779. DOI: <https://doi.org/10.2307/3213143>
3. Dudin A.N., Klimenok V.I. *Queueing Systems with Correlated Flows*. Minsk: BGU, 2000, 175 p. (In Russ.)
4. Likhttsinder B.Ya. *Traffic of Multiservice Access Networks (Interval Analysis and Design)*. Moscow: Goryachaya liniya – Telekom, 2018, 290 p. (In Russ.)
5. Likhttsinder B.Ya., Moiseev V.I. Batch poisson and hyperpoisson arrival process teletraffic models. *I-Methods*, 2022, vol. 14, no. 3, pp. 2–11. (In Russ.)
6. Ramaswami V. The N/G/1 queue and its detailed analysis. *Advances in Applied Probability*, 1980, vol. 12, no. 1, pp. 222–261. DOI: 10.2307/1426503
7. Lakatos L., Szeidl L., Telek M. *Introduction to Queueing Systems with Telecommunication Applications*. Springer Science and Business Media, 2013, 388 p. DOI: 10.1007/978-1-4614-5317-8
8. Lema M.A. et al. Flexible dual-connectivity spectrum aggregation for decoupled uplink and downlink access in 5G heterogeneous systems. *IEEE Journal on Selected Areas in Communications*, 2016, vol. 34, no. 1, pp. 2851–2865. DOI: <https://doi.org/10.1109/JSAC.2016.2615185>
9. Niknam S. et al. A multiband OFDMA heterogeneous network for millimeter wave 5G wireless

- applications. *IEEE Access*, 2016, vol. 4, pp. 5640–5648. DOI: 10.1109/ACCESS.2016.2604364
10. Vishnevsky V., Larionov A., Frolov S. Design and scheduling in 5G stationary and mobile communication systems based on wireless millimeter-wave mesh networks. *Distributed Computer and Communication Networks*, 2014, vol. 279, pp. 11–27. DOI: 10.1007/978-3-319-05209-0_2
11. Vishnevsky V.M. et al. Applying graph-theoretic approach for time-frequency resource allocation in 5G mm wave backhaul network. *Advances in Wireless and Optical Communications (RTUWO)*, 2016, pp. 221–224. DOI: 10.1109/RTUWO.2016.7821888
12. Leland W.E. et al. On the self-similar nature of Ethernet traffic. *IEEE/ACM Transactions on Networking*, 1994, vol. 2, no. 1, pp. 1–15.
13. Tsybakov B.S. Model of telegraphy based on a self-similar random process. *Radiotekhnika*, 1999, no. 5, pp. 24–31. (In Russ.)

Received 28.11.2023

УДК 621.391: 621 39

ОБ ИСПОЛЬЗОВАНИИ ГРУППОВОГО ПУАССОНОВСКОГО ПОТОКА В ИМИТАЦИОННОМ МОДЕЛИРОВАНИИ СОВРЕМЕННОГО ВИДЕОТРАФИКА

Лихтциндер Б.Я.¹, Привалов А.Ю.²

¹ Поволжский государственный университет телекоммуникаций и информатики, Самара, РФ

² Самарский национальный исследовательский университет имени С.П. Королева, Самара, РФ
E-mail: b.lihtcinder@psuti.ru, privalov1967@gmail.com

Рассматривается возможность использования группового пуассоновского потока для имитационного моделирования современного видеотрафика в отношении аппроксимации статистических характеристик первого и второго порядка создаваемой трафиком очереди в буфере передачи. Предложен способ определения параметров аппроксимирующего группового пуассоновского потока, заключающийся в одновременной минимизации отклонения зависимостей математического ожидания очереди и дисперсии очереди от нагрузки канала, вычисленной по формулам для аппроксимирующего потока от аналогичных зависимостей реального трафика. В качестве аналитических формул для математического ожидания и дисперсии очереди используются формулы, полученные с помощью интервального метода для системы массового обслуживания с одним прибором и постоянным временем обслуживания. Качество аппроксимации проверяется имитационным моделированием прохождения группового пуассоновского потока с вычисленными параметрами через данную систему массового обслуживания.

Ключевые слова: система массового обслуживания, групповой пуассоновский поток, очереди с постоянным временем обслуживания, имитационное моделирование

Введение

В настоящее время математическими моделями, наиболее адекватно описывающими трафик современных сетей телекоммуникаций, считаются модели самоподобного трафика [1–6; 9; 10]. Действительно, с помощью таких моделей можно одновременно приблизить сразу несколько важных статистических характеристик трафика как первого, так и более высоких порядков [3]. Однако главным недостатком этих моделей является весьма высокая сложность, затрудняющая использование их на практике.

В связи с этим представляется актуальным подход, который бы использовал для аппроксимации характеристик реального трафика более простые и легкие для анализа модели, близкие к моделям, считающимся классическими. Пусть даже такие модели смогут одновременно аппроксимировать не слишком широкий набор статисти-

ческих характеристик реального трафика, а только наиболее важные из них, но за счет простоты модели это может быть полезно на практике, в частности в имитационном моделировании.

В данной работе применяется именно такой подход, при этом аппроксимируются зависимости первого и второго момента очереди, создаваемой трафиком, от загрузки канала. Такие характеристики выбраны для аппроксимации потому, что именно они важны во многих задачах имитационного моделирования, например в задачах определения размеров буферов для хранения транзитных пакетов в узлах сети или задачах прохождения выделенного потока через загруженную сеть, где необходимо моделировать очереди, создаваемые остальными потоками (как фон для движения выделенного потока).

В качестве простой модели для очереди, образуемой реальным трафиком, используется

очередь, создаваемая пуассоновским потоком с групповым прибытием. Он отличается от классического (ординарного) пуассоновского потока только тем, что у классического потока заявки прибывают строго по одной, а у группового – пачками, размеры которых являются одним из параметров модели.

В качестве реального трафика, очереди которого аппроксимируются, используется видеотрафик, так как именно в видеотрафике пачечность, являющаяся причиной сильной коррелированности, и в конечном счете самоподобия, является весьма сильной.

Приближение с использованием аналитических выражений

Рассматривается Пуассоновский поток с параметром λ независимых событий, i -тое ($i = 1, 2, \dots$) из которых есть одновременное прибытие B_i пакетов для передачи, где B_i – независимые одинаково распределенные случайные величины. Все пакеты одинакового размера, они поступают в буфер и передаются оттуда по каналу постоянной битовой скорости. Таким образом, мы рассматриваем систему массового обслуживания (СМО) типа G/D/1 с групповым Пуассоновским потоком на входе.

В работе [7] с помощью интервального метода [8] для системы G/D/1 были получены формулы для математического ожидания очереди (обобщение формулы Поллачека–Хинчина) и для вторых моментов очереди, а именно:

$$\bar{q} = \frac{D(m) + 2R(q_{i-1}, m_i) - \rho}{2(1-\rho)},$$

$$\bar{q}^2 = \frac{\bar{q}}{1-\rho} (D(m) + (1-\rho)^2) + \frac{\bar{m}^3 - 3\bar{m}^2 + 2\rho}{3(1-\rho)},$$

$$D(q) = \bar{q}^2 - (\bar{q})^2.$$

где m – случайная величина, равная количеству заявок, приходящих в систему в течении случайно выбранного интервала обслуживания одной заявки, ρ – коэффициент загрузки прибора (канала связи), $D(m)$ – ее дисперсия, $R(q_{i-1}, m_i)$ – корреляция между величиной очереди перед началом некоторого интервала обслуживания заявки и количеством заявок, поступивших во время этого обслуживания, q – очередь, $D(q)$ – дисперсия очереди.

Там же для группового пуассоновского потока получено, что:

$$\bar{m} = \rho = \lambda \tau \bar{B},$$

$$\begin{aligned} \bar{m}^2 &= \lambda \tau \bar{B}^2 + (\lambda \tau \bar{B})^2, \\ \bar{m}^3 &= \lambda \tau \bar{B}^3 + 3(\lambda \tau) \bar{B} \bar{B}^2 + (\lambda \tau \bar{B})^3, \\ D(m) &= \lambda \tau \bar{B}^2, \mu_3(m) = \lambda \tau \bar{B}^3. \end{aligned}$$

где τ – время обслуживания заявки, $\mu_3(m)$ – третий центральный момент m , B – размер пачки заявок.

Используя то, что в пуассоновском потоке отсутствует корреляция текущей очереди с будущим входным потоком, для моментов очереди несложно получить, что:

$$\bar{q}(\rho) = \frac{\rho(\bar{B}^2 - \bar{B}(1-\rho))}{2\bar{B}(1-\rho)}, \quad (1)$$

$$\begin{aligned} \bar{q}^2(\rho) &= \frac{\rho(\bar{B}^2 - \bar{B}(1-\rho))}{2\bar{B}(1-\rho)} \times \\ &\times \left(\rho \frac{\bar{B}^2}{\bar{B}} + (1-\rho)^2 \right) + \\ &+ \frac{\rho^3 - 3\rho^2 + 2\rho}{3(1-\rho)} + \frac{\rho \bar{B}^3 - 3\rho \bar{B}^2(1-\rho)}{3\bar{B}(1-\rho)}, \quad (2) \end{aligned}$$

$$D(q(\rho)) = \bar{q}^2(\rho) - (\bar{q}(\rho))^2. \quad (3)$$

Далее для используемых трасс реального видеотрафика методом имитационного моделирования, когда данная трасса является входным потоком в систему G/D/1, были найдены эмпирические оценки средней очереди и дисперсии очереди для ряда значений ρ :

$$q^*(\rho_i), i = 1, \dots, N \quad \text{и} \quad D^*(\rho_j), j = 1, \dots, M.$$

Различные значения ρ получены путем изменения скорости обслуживания (скорости передачи данных по каналу).

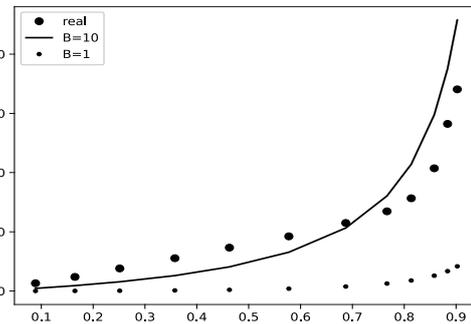


Рисунок 1. Зависимость средней очереди от загрузки прибора, трафик H264

В качестве распределения размера пачки в групповом пуассоновском потоке был взят простейший случай постоянного размера пачки $B = \text{const}$, и методом взвешенных наименьших квадратов определено значение B , при котором значение целевой функции (4) минимально.

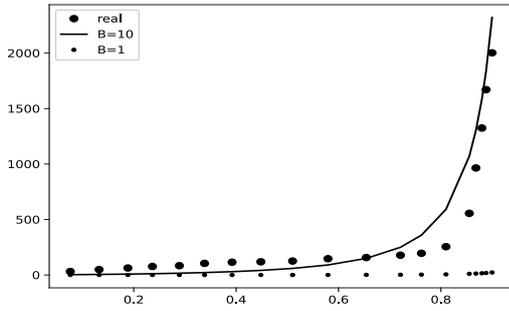


Рисунок 2. Зависимость дисперсии очереди от загрузки прибора, трафик H264

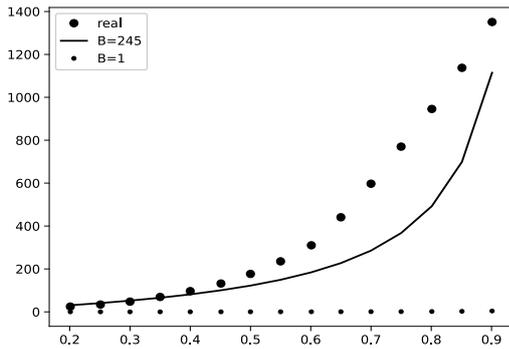


Рисунок 3. Зависимость средней очереди от загрузки прибора, трафик vid

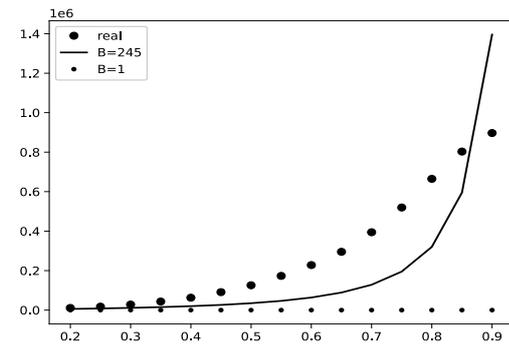


Рисунок 4. Зависимость дисперсии очереди от загрузки прибора, трафик vid

$$J(B) = \sum_{i=1}^N \frac{(\bar{q}^*(\rho_i) - \bar{q}(\rho_i))^2}{\bar{q}^*(\rho_i)} + \sum_{j=1}^M \frac{(D^*(\rho_j) - D(\rho_j))^2}{D^*(\rho_j)} \quad (4)$$

При вычислении целевой функции используются формулы (1)–(3) при условии $B = \text{const}$.

На рисунках 1 и 2 приведены для средней очереди и дисперсии очереди, точки, полученные из реального трафика, использовавшиеся в функции (4), графики зависимостей (1) и (2) для B , минимизирующего функцию (4) и для сравнения графики для ординарного пуассоновского потока. В этом эксперименте использовался трафик от видеокodeка стандарта H264. На рисунках 3 и 4

приведены результаты аналогичного эксперимента с видеотрафиком типа vid.

Видно, что групповой пуассоновский поток гораздо лучше приближает рассматриваемые характеристики реального трафика, чем ординарный пуассоновский поток. Точность аппроксимации зависит от вида видеотрафика.

Отметим, что кроме простейшего случая постоянного размера пачки, мы проводили аналогичные эксперименты для случая, когда размер пачки мог иметь два различных значения: одно с вероятностью p , а другое с вероятностью $1-p$, и минимизация целевой функции (4) проводилась по этим двум размерам и вероятности p . Однако значимого улучшения результатов не было получено.

Проверка результатов с помощью имитационного моделирования

Для проверки были сгенерированы групповые пуассоновские потоки с найденными параметрами, и, используя их в качестве входных в систему G/D/1, получены эмпирические зависимости средней и дисперсии очереди. Результаты приведены на рисунках 5 и 6 для кодека H264 и на рисунках 7 и 8 для видеотрафика vid. Видно, что эмпирические зависимости, найденные по трассам имитационного эксперимента, практически совпадают с теоретическими, вычисленными по формулам (1)–(3).

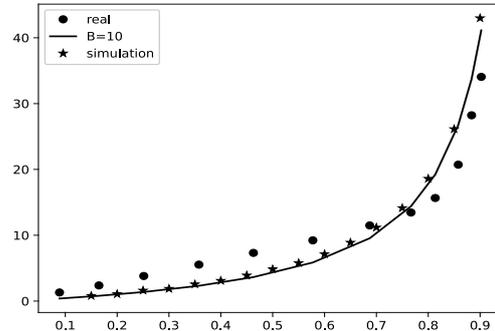


Рисунок 5. Зависимость средней очереди от загрузки прибора, трафик H264

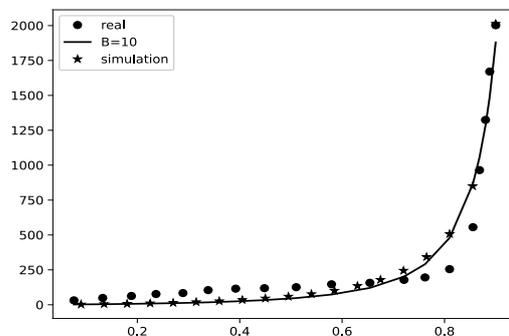


Рисунок 6. Зависимость дисперсии очереди от загрузки прибора, трафик H264

Некоторое различие для дисперсии очереди в области больших нагрузок для трафика vid связано с медленным установлением стационарного состояния при размере пачки, равном 245: в экспериментах использовалась трасса, содержащая моменты прихода 160 000 пакетов, и при этом стационарное состояние при больших нагрузках еще не вполне установилось.

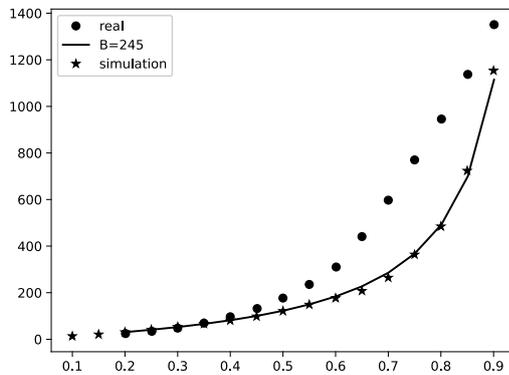


Рисунок 7. Зависимость средней очереди от загрузки прибора, трафик vid

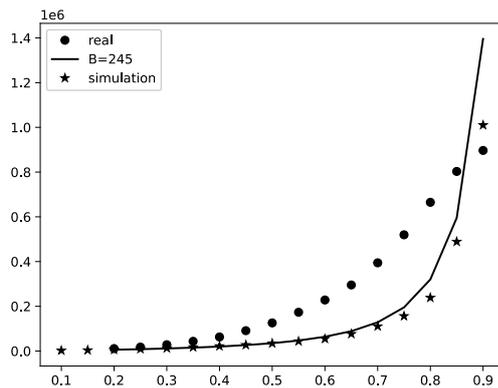


Рисунок 8. Зависимость дисперсии очереди от загрузки прибора, трафик vid

Заключение

Групповой пуассоновский поток может быть использован в имитационном моделировании для аппроксимации очередей реального видеотрафика. При этом одновременно можно аппроксимировать такие статистические параметры очереди, как математическое ожидание и дисперсию.

С практической точки зрения, такое приближение может рассматриваться как хорошее приближение всего распределения очереди, хотя точность приближения зависит от вида видеотрафика. В работе предложен простой метод оценки параметра группового потока, обеспечивающий наилучшее приближение при данном подходе.

Литература

1. On the self-similar nature of ethernet traffic (extended version) / W.E. Leland [et al.] // IEEE/ACM Transactions on Networking. 1994. Vol. 2, no. 1. P. 1–15.
2. Yunhua R. Evaluation and estimation of second-order self-similar network traffic // Computer Communications. 2004. Vol. 27, no. 9. P. 898–904. DOI: 10.1016/j.comcom.2004.02.003
3. Privalov A., Tsarev A. Analysis and simulation of WAN traffic by self-similar traffic model with OMNET // Proceedings of the 10th International Wireless Communications and Mobile Computing Conference. 2014. P. 629–634.
4. Millan N.G., Urrutia S.E., Guzmán V.M. A simple multifractal model for self-similar traffic flows in high-speed computer networks // Computacion y Sistemas. 2019. Vol. 23, no. 4. P. 1517–1521.
5. Sarla P., Reddy D.M., Krishna T.V. Analytical study of self-similar type traffic data-queueing techniques // AIP Conference Proceedings. 2020. Vol. 2246, no. 1, P. 29–35.
6. Georgantas D., Baziana P. Traffic burstiness study of an efficient bandwidth allocation MAC scheme for WDM datacenter networks // 2023 IEEE International Mediterranean Conference on Communications and Networking (MeditCom). 2023. P. 169–174.
7. Likhhtsinder B.J., Privalov A.Y., Moiseev V.I. Batch-poissonian arrival models of multiservice network traffic // Problems of Information Transmission. 2023. Vol. 59, no. 1. P. 71–79.
8. Лихтциндер Б.Я. Трафик мультисервисных сетей доступа (интервальный анализ и проектирование). М.: Горячая линия – Телеком, 2018. 290 с.
9. Vishnevskii V.M., Dudin A.N. Queueing systems with correlated arrival flows and their applications to modeling telecommunication networks // Automation and Remote Control. 2017. Vol. 78, no. 8. P. 1361–1403. DOI: 10.1134/S000511791708001X
10. Appenzeller G., Keslassy I., McKeown N. Sizing router buffers // Processing 2004 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM'04). Portland, USA, 2004. P. 281–292. DOI: 10.1145/1015467.1015499

Получено 26.11.2023

Лихтциндер Борис Яковлевич, д.т.н., профессор, профессор кафедры сетей и систем связи Поволжского государственного университета телекоммуникаций и информатики. 443010, Российская Федерация, г. Самара, ул. Л. Толстого, 23. Тел. +7 846 333-47-69. E-mail: lixt@psuti.ru

Привалов Александр Юрьевич, д.т.н., профессор, профессор кафедры прикладной математики и физики Самарского национального исследовательского университета имени С.П. Королева. 443086, Самара, Московское шоссе, 34. Тел. +7 846 332-56-07. E-mail: privalov1967@gmail.com

USE OF GROUP POISSON FLOW IN SIMULATION MODELING OF MODERN VIDEO TRAFFIC

Lichtzinder B. Ya.¹, Privalov A. Yu.²

¹ *Povolzhskiy State University of Telecommunications and Informatics, Samara, Russian Federation*

² *Samara National Research University, Samara, Russian Federation*

E-mail: b.lihtcinder@psuti.ru, privalov1967@gmail.com

The possibility of using a group Poisson flow for simulation modeling of modern video traffic is considered in relation to the approximation of the first and second order statistical characteristics of the queue created by the traffic in the transmission buffer. A method is proposed for determining the parameters of an approximating group Poisson flow, which consists of simultaneously minimizing the deviation of the dependencies of the mathematical expectation of the queue and the dispersion of the queue on the channel load, calculated using the formulas for the approximating flow, from similar dependencies of real traffic. As analytical formulas for the mathematical expectation and queue dispersion, formulas obtained using the interval method for a queuing system with one device and constant service time are used. The quality of the approximation is checked by simulation of the passage of a group Poisson flow with calculated parameters through a given queuing system.

Keywords: *queuing system, group Poisson flow, queues with constant service time, simulation modeling*

DOI: 10.18469/ikt.2023.21.3.02

Likhttsinder Boris Yakovlevich, Povolzhskiy State University of Telecommunications and Informatics, 23, L. Tolstoy Street, Samara, 443010, Russian Federation; Professor of Networks and Communication Systems Department, Doctor of Technical Science. Tel. +7 846 333-47-69. E-mail: lixt@psuti.ru

Privalov Alexander Yurievich, Samara National Research University, 34, Moskovskoye shosse, Samara, 443086, Russian Federation; Professor of Applied Mathematics and Physics Department, Doctor of Technical Science. Tel. +7 846 332-56-07. E-mail: privalov1967@gmail.com

References

1. Leland W.E. et al. On the self-similar nature of ethernet traffic (extended version). *IEEE/ACM Transactions on Networking*, 1994, vol. 2, no. 1, pp. 1–15.
2. Yunhua R. Evaluation and estimation of second-order self-similar network traffic. *Computer Communications*, 2004, vol. 27, no. 9, pp. 898–904. DOI: 10.1016/j.comcom.2004.003
3. Privalov A., Tsarev A. Analysis and simulation of WAN traffic by self-similar traffic model with OMNET. *Proceedings of the 10th International Wireless Communications and Mobile Computing Conference*, 2014, pp. 629–634.
4. Millan N.G., Urrutia S.E., Guzmán V.M. A simple multifractal model for self-similar traffic flows in high-speed computer Networks. *Computacion y Sistemas*, 2019, vol. 23, no. 4, pp. 1517–1521.
5. Sarla P., Reddy D.M., Krishna T.V. Analytical study of self-similar type traffic data-queuing techniques. *AIP Conference Proceedings*, 2020, vol. 2246, no. 1, pp. 29–35.
6. Georgantas D., Baziana P. Traffic burstiness study of an efficient bandwidth allocation MAC scheme for WDM datacenter networks. *2023 IEEE International Mediterranean Conference on Communications and Networking (MeditCom)*, 2023, pp. 169–174.
7. Likhttsinder B.J., Privalov A.Y., Moiseev V.I. Batch-poissonian arrival models of multiservice network traffic. *Problems of Information Transmission*, 2023, vol. 59, no. 1, pp. 71–79.

8. Likhttsinder B.J. *Traffic of Multiservice Access Networks (Interval Analysis and Design)*. Moscow: Goryachaya liniya – Telekom, 2018, 290 p. (In Russ.)
9. Vishnevskii V.M., Dudin A.N. Queueing systems with correlated arrival flows and their applications to modeling telecommunication networks. *Automation and Remote Control*, 2017, vol. 78, no. 8, pp. 1361–1403. DOI: <https://doi.org/10.1134/S000511791708001X>
10. Appenzeller G., Keslassy I., McKeown N. Sizing router buffers. *Processing 2004 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM'04)*. Portland, USA, 2004. P. 281–292. DOI: 10.1145/1015467.1015499

Received 26.11.2023

СЕТИ СВЯЗИ И МУЛЬТИСЕРВИСНЫЕ УСЛУГИ

УДК 621.391

МЕТОД ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ ДЛЯ ДИНАМИЧЕСКОГО РАСПРЕДЕЛЕНИЯ ПРИОРИТЕТОВ В TSN

Герасимов В.В.

Поволжский государственный университет телекоммуникаций и информатики, Самара, РФ

E-mail: v.gerasimov@psuti.ru

В этой статье исследуется возможность использования метода обучения с подкреплением для динамического назначения приоритетов в сетях, чувствительных ко времени. Представленный подход целенаправленно оптимизирует процесс управления временными ограничениями в сети. Используя техники обучения с подкреплением, система самостоятельно корректирует приоритеты в зависимости от требований сетевого трафика. Для достижения этой цели предложены две схемы конфигурации, основанные на стандартах TSN (Time-Sensitive Networking): централизованная и распределенная. Рассмотрев эти схемы, выявим их ограничения, необходимые в удовлетворении требований, близких к реальному времени, и обеспечении строгих гарантий качества обслуживания, учитывая ограничения, применяемые к чувствительной ко времени среде. Также в работе выявляется необходимость использования дополнительного оборудования, централизованного контроллера, для перераспределения приоритетов.

Ключевые слова: динамический приоритет, обучение с подкреплением, сети, чувствительные ко времени, крайний срок

Введение

Системы реального времени используют схемы расстановки приоритетов для удовлетворения различных классов трафика с определенными требованиями к качеству обслуживания. Однако в некоторых сценариях, когда передается множество пакетов с высоким приоритетом, пакеты с более низким приоритетом могут не уложиться в установленные сроки, что приведет к значительному снижению производительности планирования. Чрезмерно ранняя отправка потоков с высоким приоритетом не дает никаких дополнительных преимуществ, помимо соблюдения крайнего срока. Вместо этого более эффективно использовать это буферное время для трафика с более низким приоритетом и обеспечивать своевременную доставку передачи трафика с высоким приоритетом. Предлагается адаптивная схема динамического присвоения приоритетов, которая использует метод подкрепленного обу-

чения – reinforcement learning для решения этой проблемы. Она позволяет адаптироваться к изменяющимся условиям сети и постоянно повышать производительность с течением времени. Кроме того, представляю и обсуждаю два потенциальных варианта конфигурации, которые могут быть использованы в рамках предлагаемой схемы.

Интернет вещей (IoT) и сети, чувствительные ко времени (TSN, Time-Sensitive Networking), т.е. протоколы связи в реальном времени, являются основой для будущих критически важных систем и соответствующих приложений, которые требуют ограниченной задержки, а также бесперебойной и безотказной связи. В этом контексте доставка трафика в течение определенного периода времени, называемого крайним сроком, имеет важное значение с потенциально серьезными последствиями, когда пакеты запаздывают. В некоторых критически важных приложениях, таких как бортовые сети, любое несоблюдение этих

сроков может иметь катастрофические последствия. Для соответствующего трафика установлены жесткие сроки. Однако многие приложения работают с гибкими сроками, например для трафика с максимальными усилиями, что означает, что они могут допускать случайные коллизии, не испытывая существенного снижения производительности или выходных данных по качеству. Например, при потоковой передаче видео несколько запоздалых пакетов не оказывают существенного влияния на общее качество видео, т.е. пакет содержит достаточно данных для восстановления. В таких случаях система может быть более гибкой в соблюдении сроков, сосредоточившись на обеспечении наилучшего возможного результата.

Целевая группа института инженеров электротехники и электроники (IEEE, Institute of Electrical and Electronics Engineers) предложила стандарты TSN для обеспечения такого QoS (Quality of Service, «качество услуги») для каждого класса трафика в рамках одной и той же сети, описав различные механизмы обработки для различных требований к трафику. Он использует восемь приоритетов для разграничения классов трафика, понимая, какой трафик является более важным при жестких ограничениях в реальном времени. Устройства TSN обрабатывают трафик на коммутаторах на основе этих значений приоритета.

Хотя устройства определения приоритетов позволяют сосуществовать различным классам трафика, при отправке большого количества пакетов с высоким приоритетом это может привести к тому, что пакеты с более низким приоритетом не уложатся в установленные сроки. Несмотря на более низкий приоритет, эти пакеты все равно стоят в очереди. Это становится особенно проблемным, когда пакеты с высоким приоритетом достигают места назначения намного раньше, чем предполагалось. С точки зрения приложения, не имеет большого значения, получит ли оно пакет незадолго до истечения срока его действия или задолго до него. Таким образом, вместо того чтобы отправлять трафик с высоким приоритетом на ранней стадии, возможно, было бы лучше, чтобы все потоки в сети отправляли трафик с высоким и низким приоритетом вовремя. Чтобы решить эту проблему, предлагаю динамическую обработку приоритетов, которая переназначает приоритеты пакетов таким образом, чтобы пакеты с более низким приоритетом могли наверстать упущенное время и не приходили с опозданием. Это может увеличить использование сетевых ресурсов и уменьшить количество запоздалых пакетов.

Авторы [1] рассмотрели исследования, проведенные на установках IIS и LNI 4.0, которые пока-

зали, что по производительности технология OPC UA TSN опережает существующие протоколы промышленного Ethernet в 18 раз. Однако это дополнительная зависимость, а также повышающие требования к оборудованию. В работе [2] предложили использовать процесс нечеткой аналитической иерархии – Fuzzy Analytic Hierarchy Process (FAHP) для вычисления приоритета пакетов с учетом потребления энергии, времени выполнения и крайнего срока. Затем алгоритм динамической сортировки на основе heapsort выбирает оптимальное подмножество планирования из набора задач, используя новые приоритеты. Хотя результаты обещают снизить частоту пропусков сроков, это напрямую неприменимо к TSN с такими критериями проектирования, как потребление энергии.

В отличие от традиционных подходов, RL (Reinforcement Learning – метод обучения с подкреплением) предлагает ряд преимуществ для динамических сетей реального времени, поскольку позволяет адаптироваться и извлекать опыт из данных в изменяющихся средах. В динамических сетях реального времени RL может постоянно обновлять свои политики и принимать почти оптимальные решения на основе текущего состояния сети. Эта возможность автономной оптимизации позволяет RL адаптироваться к изменяющимся сетевым условиям и улучшает работу производительности сети с течением времени без вмешательства человека. Более того, возможности RL извлекать уроки из взаимодействий и эффективно представлять сложные взаимосвязи позиционируют его как в высшей степени подходящий подход для решения задач, возникающих в динамических сетевых средах. Таким образом, RL также использовался в TSN с различными целями, такими как поиск пути маршрутизации [3] или настройка гарантированной задержки на передачу [4]. Несколько исследований применяли RL для решения задач динамического назначения приоритетов в контексте сред реального времени [5]. Основной целью этих исследований считается определение планируемого назначения приоритетов, которое может вставлять большее количество потоков в сеть. Вместо этого мы стремимся использовать ресурсы более эффективно, обслуживая по крайней мере такое же количество потоков, но обеспечивая лучшее качество обслуживания в отношении сроков. Эта проблема становится более очевидной в случае несбалансированных типов трафика, когда определенные классы могут оказаться перегруженными, в то время как для других типов трафика имеется доступная пропускная способность.

Динамическое распределение приоритетов с использованием метода обучения с подкреплением

Чтобы использовать преимущества обучения с подкреплением, мы представляем на рисунках 1 и 2 две потенциальные схемы конфигурации, которые могут быть применены к чувствительным ко времени сетям:

1. *Централизованная схема.* В централизованной схеме, как показано на рисунке 1, предполагается, что централизованный сетевой контроллер (CNC) имеет глобальное сетевое представление и собирает статистические данные, такие как время ожидания в очереди и использование очереди [9]. Агент RL развернут поверх CNC, так что он получает сетевые данные в режиме реального времени и может использовать алгоритмы RL для обучения и определения глобальных сетевых политик [10]. Поскольку система воспринимает сеть как единое целое, эта схема обеспечивает скоординированное принятие решений и оптимизацию для нескольких элементов сети. Таким образом, CNC может генерировать наилучшие политики для динамического назначения приоритетов на основе централизованно развернутого агента RL

Поскольку изначально нет политики или точной информации о сети, CNC может отслеживать текущие назначения и обучать агента

RL. Другими словами, CNC может получать статистику перехода за переходом и объединять их для расчета окончательной оценки стоимости (Final Reward Value) за действие, например, текущего приоритета пакета. Здесь, учитывая требования к QoS пакетов, значение стоимости (Final Reward Value) также может принимать отрицательные значения, например штрафное значение. В течение этого времени обучения (исследования) RL может назначать пакетам случайные приоритеты или оставить их как есть. По истечении времени предварительной подготовки CNC может воспользоваться разработанной политикой RL для динамического присвоения нового приоритета пакетам во время выполнения.

2. *Распределенная схема.* В распределенной схеме, как показано на рисунке 2, агенты RL развертываются непосредственно на узлах сети, например на коммутаторах TSN [8]. Эти агенты локально отслеживают сеть, собирают данные в режиме реального времени и изучают политику принятия оптимальных решений на основе наблюдаемых условий. Каждый коммутатор стремится компенсировать потенциальную задержку в предыдущем переходе путем динамического переназначения приоритетов пакетов, определенных RL. Таким образом, у каждого коммутатора есть свой собственный агент RL для разработки

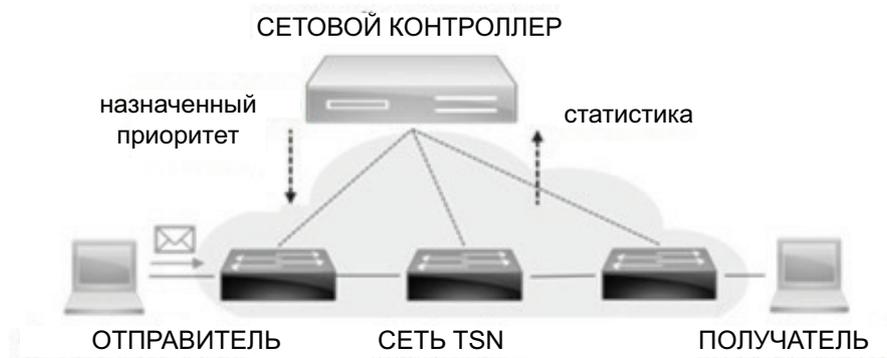


Рисунок 1. Централизованная схема динамического присвоения приоритетов

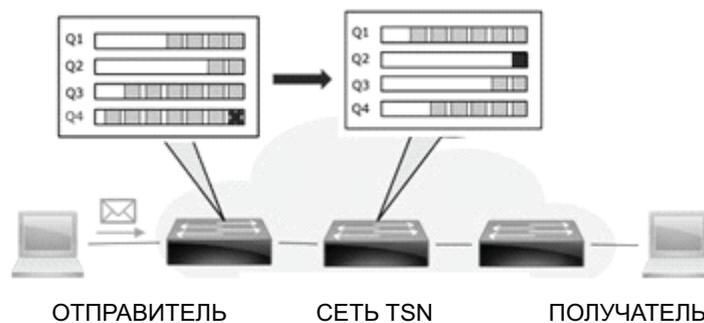


Рисунок 2. Распределенная схема динамического присвоения приоритетов на основе RL

политики. Возможно, также удастся извлечь выгоду из концепции обучения с подкреплением при передаче. Коммутаторы могут совместно помогать друг другу в разработке наилучшей политики назначения.

Однако без централизованного контроллера проблема усложняется. Теперь коммутатор должен оценить, основываясь на ограниченных знаниях, не пропустит ли пакет установленный срок и нуждается ли он в изменении приоритетов [6]. Для этого, возможно, потребуется знать топологию или путь, по которому будет маршрутизироваться пакет. Кроме того, пакет должен быть помечен как поздний пакет, который будет обработан иным образом при следующем переключении перехода. Таким образом, существует необходимость в протоколе связи и локальном или распределенном алгоритме между коммутаторами для обращения к этим точкам [7].

Заключение

В заключение предлагаемая адаптивная схема динамического назначения приоритетов, использующая метод обучения с подкреплением, представляет собой многообещающее решение для сетей TSN. Она может динамически перенастраивать приоритеты существующих пакетов и предлагает практический подход к сокращению пропущенных сроков передачи. Адаптивно корректируя приоритеты в зависимости от условий реального времени, IT-отдел может эффективно управлять распределением ресурсов и соответствовать строгим требованиям к качеству обслуживания.

Литература

1. TSN – сети Ethernet, чувствительные ко времени / А.В. Росляков [и др.] // Инфокомму-

- никационные технологии. 2021. Т. 19, № 2. С. 187–201. DOI: 10.18469/ikt.2021.19.2.07
2. Meng S., Zhu Q., Xia F. Improvement of the dynamic priority scheduling algorithm based on a heapsort // IEEE Access. 2019. Vol. 7. P. 503–510.
3. Bulb N.S., Fischer M. Reinforcement learning assisted routing for time-sensitive networks // IEEE GLOBECOM Global Communications Conference. 2022. P. 3863–3868.
4. ML-assisted latency assignments in time-sensitive networking / A. Grigorjew [et al.] // IFIP/IEEE International Symposium on Integrated Network Management. Bordeaux, 2021. P. 116–124.
5. Panda: reinforcement learning-based priority assignment for multi-proce / H. Lee [et al.] // IEEE Access. 2020. No. 8. P. 185570–185583.
6. Стандартизация синхронизируемых по времени сетей TSN / А.В. Росляков [и др.] // Стандарты и качество. 2021. № 4 (1006). С. 29–33. DOI: 10.35400/0038-9692-2021-4-48-53
7. Росляков А.В. СЕТЬ 2030: архитектура, технологии, услуги. М.: ИКЦ «Колос-с», 2022. 278 с.
8. When IEEE 802.11 and 5G meet time-sensitive networking / M.K. Atiq [et al.] // IEEE Open Journal of the Industrial Electronics Society. 2021. Vol. 3. P. 14–36. DOI: 10.1109/OJIES.2021.3135524
9. Коган С. Стандартизация решений и сегментирование транспортного уровня сети 5G // Первая миля. 2021. № 2. С. 40–47. DOI: 10.22184/2070-8963.2021.94.2.40.47
10. Self-configuration of IEEE 802.1 TSN networks / M. Gutierrez [et al.] // 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA). Limassol, 2017. P. 1–8.

Получено 26.01.2024

Герасимов Вячеслав Васильевич, ст. преподаватель кафедры сетей и систем связи Поволжского государственного университета телекоммуникаций и информатики. 443010, Российская Федерация, г. Самара, Л. Толстого, 23. Тел. +7 987 434-77-32. E-mail: v.gerasimov@psuti.ru

REINFORCEMENT LEARNING METHOD FOR DYNAMIC ALLOCATION OF PRIORITIES IN TSN

Gerasimov V.V.

Povolzhskiy State University of Telecommunications and Informatics, Samara, Russian Federation

E-mail: v.gerasimov@psuti.ru

This article explores the possibility of using reinforcement learning to dynamically assign priorities in time-sensitive networks. The presented approach purposefully optimizes the process of managing time constraints in the network. Using reinforcement learning techniques, the system independently adjusts priorities depending on the requirements of network traffic. To achieve this goal, two configu-

ration schemes based on TSN standards are proposed: centralized and distributed. Having considered these schemes, we will identify their limitations necessary in meeting requirements close to real time and ensuring strict quality of service guarantees, taking into account the restrictions applied to a time-sensitive environment. The work also reveals the need to use additional equipment, a centralized controller, to reallocate priorities.

Keywords: *dynamic priority, reinforcement learning, time-sensitive networks, deadline*

DOI: 10.18469/ikt.2023.21.3.03

Gerasimov Vyacheslav Vasilevich, Povolzhskiy State University of Telecommunications and Informatics, 23, L. Tolstoy Street, Samara, 443010, Russian Federation; Senior Teacher of Networks and Communication Systems Department. Tel. +7 987 434-77-32. E-mail: v.gerasimov@psuti.ru

References

1. Roslyakov A.V. et al. TSN Ethernet time-sensitive networking. *Infocommunicionnye tehnologii*, 2021, vol. 19, no. 2, pp. 187–201. DOI: 10.18469/ikt.2021.19.2.07 (In Russ.)
2. Meng S., Zhu Q., Xia F. Improvement of the dynamic priority scheduling algorithm based on a heapsort. *IEEE Access*, 2019, vol. 7, pp. 503–510.
3. Bulb N.S., Fischer M. Reinforcement learning assisted routing for time-sensitive networks. *IEEE GLOBECOM Global Communications Conference*, 2022, pp. 3863–3868.
4. Grigorjew A. et al. ML-assisted latency assignments in time-sensitive networking. *IFIP/IEEE International Symposium on Integrated Network Management*. Bordeaux, 2021, pp. 116–124.
5. Lee H. et al. Panda: reinforcement learning-based priority assignment for multi-proce. *IEEE Access*, 2020, no. 8, pp. 185570–185583.
6. Roslyakov A.V. et al. Time-sensitive networking standardization. *Standarty i kachestvo*, 2021, no. 4 (1006), pp. 29–33. DOI: 10.35400/0038-9692-2021-4-48-53 (In Russ.)
7. Roslyakov A.V. *NETWORK 2030: Architecture, Technologies, Services*. Moscow: ICTS «Kolos-s», 2022, 278 p. (In Russ.)
8. Atiq M.K. et al. When IEEE 802.11 and 5G meet time-sensitive networking. *IEEE Open Journal of the Industrial Electronics Society*, 2021, vol. 3, pp. 14–36. DOI: 10.1109/OJIES.2021.3135524
9. Kogan S. Standardization of solutions and segmentation of the transport level of the 5G network. *Pervaya milya*, 2021, no. 2, pp. 40–47. DOI: 10.22184/2070-8963.2021.94.2.40.47 (In Russ.)
10. Gutierrez M. et al. Self-configuration of IEEE 802.1 TSN networks. *22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*. Limassol, 2017, pp. 1–8.

Received 26.01.2024

УДК 004.855.53

ИСПОЛЬЗОВАНИЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ АНОМАЛЬНОГО ПОВЕДЕНИЯ ТРАФИКА

Поздняк И.С., Макаров И.С.

Поволжский государственный университет телекоммуникаций и информатики, Самара, РФ

E-mail: i.pozdnyak@psuti.ru, i.makarov@psuti.ru

В статье рассматривается способ обнаружения аномального поведения трафика с использованием методов машинного обучения. Для этого используется набор данных, содержащий значительный объем трафика, собранный в момент проведения атаки на Web-приложение. Набор содержит три варианта атак: Brute Force, XSS, SQL-инъекция. Отдельно рассмотрен дамп трафика, содержащий атаку Infiltration. Проведен сравнительный анализ моделей машинного обучения с выбором наиболее оптимального. В статье также приводится описание процедуры предобработки данных, которая проводится с целью устранения аномалий и пустот в записях массивов, что может привести к неправильной работе обучаемой модели. Проведено обучение моделей на отобранных данных с целью выявления аномального поведения трафика, указывающего на конкретный тип атаки. Кроме того, проведено исследование на наборе данных, не содержащих сведений об атаках.

Ключевые слова: атаки, аномальный трафик, системы обнаружения вторжений, машинное обучение, атака Infiltration, атаки на веб-приложения, Python

Введение

Совершенство систем обнаружения вторжений (IDS), использующих машинное обучение, зависит от доступности наборов данных, но получить надежный набор для обучения – не самая простая задача. Среди факторов, затрудняющих сравнение, – отсутствие надлежащего описания методов обнаружения вторжений, отсутствие методологии сравнения, а также отсутствие маркировки достоверных данных и сложность получения реального трафика [1]. Кроме того, в настоящее время сетевой трафик в основном шифруется в целях безопасности и конфиденциальности, и лишь очень немногие наборы отражают этот факт.

Набор данных является важной частью построения моделей IDS на основе машинного обучения. Процесс начинается с захвата трафика. После этого он компилируется в данные определенного типа, содержащие сетевые функции, включая маркировку. Общая структура систем обнаружения вторжений на основе машинного обучения показана в [2]. Маркировка – важнейший процесс для набора данных. Обработка достоверных данных является настоящей проблемой, особенно когда эксперты не могут определить, содержит ли трафик атаку или нет. Именно по этой причине специалисты используют синтетический трафик. Однако это означает, что генерируемый трафик не всегда является характерным для реальной среды. По этой причине процесс создания набора данных начинается с захвата трафика и заканчивается заключительным этапом предварительной обработки. Конечным результатом этапа предварительной обработки является размеченный набор данных. Каждая точка данных определяется как вредоносный или нормальный трафик. Файл содержит табличные данные в удобочитаемом формате, например файл CSV, или в двоичной форме, например файл IDX. Количество обнаруженных атак или ложных тревог можно использовать для сравнения набора данных при обучении IDS.

Большинство существующих исследований, в которых используется зашифрованный трафик, сосредоточены на различных областях, таких как классификация и анализ трафика. Хотя такие исследования существуют [3], набор данных не является общедоступным из-за конфиденциальности содержимого.

Наборы контрольных данных являются важной основой для оценки и сравнения качества различных IDS. В зависимости от методов обнаружения существует три типа IDS: на основе сигнатур, на основе аномалий (статистические) и комбинированные [4]. Основанные на сигнатурах фокусируются на построении автоматической генерации шаблона, тогда как основанные на аномалиях – на наблюдении отклонений от нормального поведения. Сигнатурные системы основаны на методе сопоставления с образцом для идентификации и попытки сопоставления с базой данных. Когда попытка атаки совпадает с шаблоном, выдается предупреждение. Системы на основе сигнатур имеют высокую точность определения аномалий и низкий уровень ложных срабатываний, но при этом они не могут обнаружить неизвестные атаки. При этом системы, основанные на статистических методах, могут обнаруживать неизвестные атаки путем сравнения аномального трафика с нормальным, соотношение ложных тревог остается достаточно высоким.

Системы, обученные с помощью алгоритмов машинного обучения, могут улучшить свою способность прогнозировать события, используя обратную связь относительно того, насколько хорошо они выполнили предыдущие задачи, и использовать эту информацию для внесения изменений [5].

Сравнение наборов данных

В представленной работе использовались готовые наборы данных с трафиком, содержащие атаки [6–9]. Сравнение наборов представлено в таблице 1.

На основе представленной сравнительной характеристики делается вывод о целесообразности использования набора CSE-CIC-IDS2018 по причине его большей актуальности, а также из-за объема содержащихся данных [7]. CSE-CIC-IDS2018 был получен в результате моделирования крупной компьютерной сети, содержащей 420 хостов и 30 серверов, разделенных на несколько сегментов. При этом трафик передавался по протоколам HTTP/HTTPS, SMTP/POP3/IMAP, SSH и FTP. Были реализованы атаки шести разновидностей: Brute Force, DoS, DDoS, Botnet, внедрение и атаки на веб-приложения.

Для обработки набора данных использовались библиотеки языка программирования Python: Pandas, Scikit-learn, TensorFlow, Matplotlib, Pickle.

Предобработанные данные выбранного набора отображаются в CSV-формате. Для представ-

Таблица 1. Сравнение наборов данных

| Критерий | Набор данных | | | |
|--------------|--|--|---|--|
| | CIC-IDS2017 | CSE-CIC-IDS2018 | NIKARI-2021 | USB-IDS-1 |
| Объем | Средний | Большой | Малый | Малый |
| Источник | Синтетический, высокая вариация трафика и атак | Синтетический, высокая вариация трафика и атак | Синтетический, низкая вариация трафика и атак | Частично синтетический, низкая вариация трафика и атак |
| Доверие | Высокий уровень | Высокий уровень | Высокий уровень | Высокий уровень |
| Описание | Достаточное | Достаточное | Достаточное | Достаточное |
| Масштаб сети | Офисная сеть средних размеров | Большая корпоративная сеть | Небольшая локальная сеть | Простейшая архитектура |

Таблица 2. Характеристики выборок для атак на Web-приложения

| Характеристика | Выборка | | |
|---|----------------|---------------|--------------|
| | Обучающая | Валидационная | Тестовая |
| Общее число записей | 29 696 | 3712 | 3712 |
| Число записей с меткой вредоносного трафика | 14 848 (50,0%) | 1890 (50,9%) | 1890 (50,9%) |
| Число записей с меткой легального трафика | 14 848 (50,0%) | 1822 (49,1%) | 1822 (49,1%) |

ления данных в необходимом формате они подлежат определенной обработке: первичный анализ данных и предварительная обработка данных. На выходе обработки получается тот вид данных, который в дальнейшем будет использоваться для обучения моделей классификаторов. Для этого будут выделены три выборки: тестовая, обучающая и валидационная.

Так как в выбранном наборе данных приводятся более одной разновидности атаки одного из представленных типов, то они будут в итоге объединены в один класс. В итоге при обучении будет приниматься решение об отнесении потока данных к типовой атаке, не учитывая ее подвид. Кроме того, в наборе данных представлены разные соотношения трафика с атакой и нормального трафика. Для лучшего обучения классификаторов весь трафик будет приводиться к идеальной классовой сбалансированности. При этом будем использовать некоторые стратегии сэмплирования.

Составление выборок для атак на Web-приложения

Атаки на Web-приложения в наборе данных CSE-CIC-IDS2018 представлены тремя разными вариантами: XSS (Cross Site Scripting), Brute

Force и SQL-инъекция. Объединим таблицы для последующей совместной обработки и произведем замену меток классов, выполнив уже известные преобразования. Получаем критический дисбаланс классов. Объем меньшего из классов мал, что подводит нас к использованию подхода пересемплирования балансировки классов. При этом будут дублироваться записи меньшего из классов, пока объемы классов не сравняются.

Сбалансированные данные пропустим через конвейер предобработки, сохраним параметры нормализации и разобьем данные на выборки. Характеристики полученных выборок сведены в таблицу 2.

Обучим классификатор на основе логистической регрессии со стандартными параметрами. В результате обучена модель со следующими значениями метрик качества [10] по тестовой выборке: аккуратность: 0,945043; точность: 0,905288; полнота: 0,996296; F-мера: 0,948615.

При попытке подобрать оптимальные гиперпараметры классификатора была получена более качественная модель. Сохраним подогнанную модель средствами библиотеки Pickle. Метрики качества классификатора с базовыми и оптимальными гиперпараметрами сведены в таблицу 3.

Обучим классификатор на основе решающего дерева со стандартными параметрами. В результате получена модель со следующими значениями метрик качества по тестовой выборке: аккуратность: 0,999461; точность: 0,998943; полнота: 1,0; F-мера: 0,999471.

Осуществим перебор оптимальных параметров. При этом за счет меньшего размера выборки появилась возможность расширить сетку параметров без потерь вычислительной скорости. Полученный в результате перебора классификатор имеет те же значения метрик качества, что и базовый. Делаем вывод о том, что базовый вариант справлялся с задачей оптимальным образом.

Рассмотрим базовый вариант классификатора на основе нейронной сети. В результате получена нейронная сеть со следующими значениями метрик качества по тестовой выборке: аккуратность: 0,987607; точность: 0,983735; полнота: 0,992063; F-мера: 0,987882.

При получении более аккуратной модели была уменьшена скорость обучения с 0,001 до 0,0005. Метрики классификаторов сведены в таблице 4. Оптимальную модель сохраним средствами библиотеки Pickle.

Характеристики лучших моделей каждого вида сведены в таблицу 5. В связи с порядком значений метрик качества полученных моделей валидационная выборка оказалась неустойчивой во всех трех случаях.

На основании данных представленной выше таблицы можно утверждать, что оптимальной моделью для выявления атак на Web-приложения является модель на основе решающего дерева. Использование нейронной сети избыточно.

Составление выборок для атаки типа Infiltration

Для части набора данных, содержащих атаку типа Infiltration, произведем замену меток классов. В списке новых меток наблюдаем ярко выраженный дисбаланс классов. Объемы выборок достаточны для использования недосемплирования. При этом дублируются записи меньшего из классов, пока объемы классов не сравняются. В этом случае используются все данные.

Пропустим сбалансированные данные через конвейер предобработки, сохраним параметры нормализации и разобьем данные на выборки. Реализация данных преобразований аналогична

Таблица 3. Значения метрик качества классификаторов на основе логистической регрессии для атак на Web-приложения

| Классификатор | Метрика качества | | | |
|--|------------------|----------|----------|----------|
| | Аккуратность | Точность | Полнота | F-мера |
| Классификатор с базовыми параметрами | 0,945043 | 0,905288 | 0,996296 | 0,948615 |
| Классификатор с оптимальными параметрами | 0,951509 | 0,913043 | 1,0 | 0,954545 |

Таблица 4. Значения метрик качества классификаторов на основе нейронной сети для атак на Web-приложения

| Классификатор | Метрика качества | | | |
|--|------------------|----------|----------|----------|
| | Аккуратность | Точность | Полнота | F-мера |
| Классификатор с базовыми параметрами | 0,987607 | 0,983735 | 0,992063 | 0,987882 |
| Классификатор с оптимальными параметрами | 0,992457 | 0,989485 | 0,995767 | 0,992616 |

Таблица 5. Характеристики моделей для атак на Web-приложения

| Модель | Метрика качества | | | | Размер модели, кб |
|-------------------------|------------------|----------|----------|----------|-------------------|
| | Аккуратность | Точность | Полнота | F-мера | |
| Логистическая регрессия | 0,951509 | 0,913043 | 1,0 | 0,954545 | 7 |
| Решающее дерево | 0,999461 | 0,998943 | 1,0 | 0,999471 | 16 |
| Нейронная сеть | 0,992457 | 0,989485 | 0,995767 | 0,992616 | 110 |

представленной выше. Характеристики полученных выборок сведены в таблицу 6.

Рассмотрим классификатор на основе логистической регрессии со стандартными параметрами. В результате обучения получена модель со следующими значениями метрик качества по тестовой выборке: аккуратность: 0,555470; точность: 0,594545; полнота: 0,355796; F-мера: 0,445180.

Модель не справляется со своей задачей. Качество классификации близко к уровню случайного угадывания. При попытке подобрать лучшие гиперпараметры классификатора качество работы классификатора не улучшилось.

В результате обучения классификатора на основе решающего дерева со стандартными параметрами получена модель со следующими значениями метрик качества по тестовой выборке: аккуратность: 0,568221; точность: 0,570853; полнота: 0,558334; F-мера: 0,564524. Полученная модель не превосходит предыдущую по качеству.

Попробуем подобрать более осмысленную сетку параметров. Для этого определяются зависимости метрик качества классификатора от ключевых гиперпараметров модели на тестовой и валидационной выборках: в зависимости от

глубины дерева, в зависимости от предела разветвления, в зависимости от критерия разветвления. На основе этих сведений составляется новая сетка и таким образом находится лучшая модель, которая обладает немного лучшими свойствами. Характеристики моделей сведены в таблицу 7.

При обучении классификатора на основе нейронной сети усложним ее архитектуру относительно представленной ранее путем добавления второго скрытого полносвязного слоя. Число нейронов в обоих скрытых слоях устанавливается равным 100. Базовая модель после обучения обладает значениями метрик качества по тестовой выборке: аккуратность: 0,571124; точность: 0,694168; полнота: 0,258100; F-мера: 0,376291.

Попробуем получить более аккуратную модель при помощи подхода, аналогичному представленному в обучении с решающим деревом, с определением метрик качества для ключевых параметров модели. Произведем перебор таких параметров. Лучшая из рассмотренных моделей была получена при использовании гиперболического тангенса как функции активации скрытых слоев с сохранением прочих гиперпараметров базового классификатора. Характеристики моделей сведены в таблицу 8.

Таблица 6. Характеристики выборок для атаки типа Infiltration

| Характеристика | Выборка | | |
|---|-----------------|----------------|----------------|
| | Обучающая | Валидационная | Тестовая |
| Общее число записей | 259 094 | 32 387 | 32 387 |
| Число записей с меткой вредоносного трафика | 129 622 (50,0%) | 16 228 (50,1%) | 16 234 (50,1%) |
| Число записей с меткой легального трафика | 129 472 (50,0%) | 16 159 (49,9%) | 16 153 (49,9%) |

Таблица 7. Значения метрик качества классификаторов на основе нейронной сети для атаки типа Infiltration

| Классификатор | Метрика качества | | | |
|--|------------------|----------|----------|----------|
| | Аккуратность | Точность | Полнота | F-мера |
| Классификатор с базовыми параметрами | 0,568221 | 0,570853 | 0,558334 | 0,564524 |
| Классификатор с подобранными параметрами | 0,585821 | 0,588601 | 0,576999 | 0,582742 |

Таблица 8. Значения метрик качества классификаторов на основе нейронной сети для атаки типа Infiltration

| Классификатор | Метрика качества | | | |
|--|------------------|----------|----------|----------|
| | Аккуратность | Точность | Полнота | F-мера |
| Классификатор с базовыми параметрами | 0,572266 | 0,709484 | 0,248367 | 0,367933 |
| Классификатор с подобранными параметрами | 0,550252 | 0,539627 | 0,699581 | 0,609281 |

Таблица 9. Характеристики моделей для атаки типа Infiltration

| Модель | Метрика качества | | | | Размер модели, кб |
|-------------------------|------------------|----------|----------|----------|-------------------|
| | Аккуратность | Точность | Полнота | F-мера | |
| Логистическая регрессия | 0,555470 | 0,594545 | 0,355796 | 0,445180 | 7 |
| Решающее дерево | 0,585821 | 0,588601 | 0,576999 | 0,582742 | 489 |
| Нейронная сеть | 0,550252 | 0,539627 | 0,699581 | 0,609281 | 130 |

Таблица 10. Лучшие полученные модели машинного обучения для классификации атак

| Атака | Тип классификатора | Метрика качества | | | | Размер модели, кб |
|--------------|--------------------|------------------|----------|----------|----------|-------------------|
| | | Аккуратность | Точность | Полнота | F-мера | |
| Web | Решающее дерево | 0,999461 | 0,998943 | 1,0 | 0,999471 | 16 |
| Infiltration | Нейронная сеть | 0,550252 | 0,539627 | 0,699581 | 0,609281 | 130 |

Характеристики лучших моделей каждого вида сведены в таблицу 9. На основании этих данных можно утверждать, что оптимальной моделью для выявления атак типа Infiltration является нейронная сеть. При этом результаты классификации даже для лучшей модели являются неудовлетворительными.

Лучшие модели для классификации рассмотренных атак, которые были получены в ходе работы, и значения их метрик качества по тестовым выборкам сведены в таблицу 10.

Для атак на Web-приложения были получены результаты, близкие к оптимальным. Атаку типа Infiltration удовлетворительно классифицировать на представленных данных и при помощи рассматриваемых алгоритмов с фиксированной архитектурой не удалось.

Также было проведено дополнительное исследование, при котором производилась попытка классификации данных в наборе, в котором отсутствовали какие-либо атаки. Все модели справились с задачей. И рассматриваемые атаки не были там обнаружены. При попытке обнаружения неизвестной классификатору атаки (например, Brute Force для атак на Web-приложения) модели ожидаемо не справились с задачей.

Заключение

Для улучшения результатов обучения можно использовать предварительную обработку данных иного вида, где будет осуществляться упор на решение задачи классификации конкретного вида атаки. При этом возможно значительное усложнение архитектуры моделей (особенно перспективны нейронные сети). Также возможен более тщательный перебор гиперпараметров моделей, связанный с повышением требований к вычислительным мощностям или времени обучения.

В дальнейшем предполагается использовать предложенные способы улучшения результатов. Кроме того, предполагается рассмотреть метрики качества для других видов атак с использованием тех же типов классификаторов: решающее дерево, логистическая регрессия, нейронная сеть.

Литература

1. Performance analysis of machine learning algorithms in intrusion detection system: a review / T. Saranya [et al.] // *Procedia Computer Science*. 2020. Vol. 171. P. 1251–1260.
2. Generating network intrusion detection dataset based on real and encrypted synthetic attack traffic / A. Ferriyan [et al.] // *Applied Sciences*. 2021. Vol. 11, no. 17. URL: <https://www.mdpi.com/2076-3417/11/17/7868> (дата обращения: 15.01.2024).
3. De Lucia M.J., Cotton C. Identifying and detecting applications within TLS traffic // *Proceedings Cyber Sensing*. 2018. Vol. 10630. P. 106300U.
4. Шелухин О.И., Сакалема Д.Ж., Филинова А.С. Обнаружение вторжений в компьютерные сети (сетевые аномалии): учебное пособие для вузов М.: Горячая линия – Телеком, 2022. 220 с.
5. Sharma S., Krishna C.R., Sahay S.K. Detection of advanced malware by machine learning techniques // *Springer, Advances in Intelligent Systems and Computing*. 2018. Vol. 742. P. 332–342.
6. CSE-CIC-IDS2018 on AWS. URL: <http://www.unb.ca/cic/datasets/ids-2018.html> (дата обращения: 18.12.2023).
7. Intrusion Detection Evaluation Dataset (CIC-IDS2017). URL: <https://www.unb.ca/cic/datasets/ids-2017.html> (дата обращения: 18.12.2023).

8. NIKARI-2021: Generating network intrusion detection dataset based on real and encrypted synthetic attack traffic / A. Ferriyan [et al.]. URL: <https://zenodo.org/records/5199540> (дата обращения: 20.01.2024).
9. USB-IDS-1: A public multilayer dataset of labeled network flows for IDS evaluation / M. Catillo [et al.] // 51st Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W). 2021. P. 1–6. DOI: 10.1109/DSN-W52860.2021.00012
10. Харрисон М. Машинное обучение: карманный справочник. Краткое руководство по методам структурированного машинного обучения на Python / пер. В.А. Коваленко. СПб.: Диалектика, 2020. 320 с.

Получено 22.01.2024

Поздняк Ирина Сергеевна, к.т.н., доцент, доцент кафедры информационной безопасности (ИБ) Поволжского государственного университета телекоммуникаций и информатики (ПГУТИ). 443090, Российская Федерация, г. Самара, Московское шоссе, 77. Тел. +7 927 657-24-27. E-mail: i.pozdnyak@psuti.ru

Макаров Игорь Сергеевич, к.т.н., доцент кафедры ИБ ПГУТИ. 443090, Российская Федерация, г. Самара, Московское шоссе, 77. Тел. +7 937 208-80-66. E-mail: i.makarov@psuti.ru

USING MACHINE LEARNING ALGORITHMS TO DETECT ANOMALOUS TRAFFIC BEHAVIOR

Pozdnyak I.S., Makarov I.S.

Povolzhskiy State University of Telecommunications and Informatics, Samara, Russian Federation

E-mail: i.pozdnyak@psuti.ru, i.makarov@psuti.ru

The article describes a method of using machine learning for detecting anomalous traffic behavior. For this purpose, a data set containing a significant amount of traffic collected at the time of the attack on the Web application is used. The set contains three attack options: Brute Force, XSS, SQL injection. A traffic dump containing an Infiltration attack is considered separately. A comparative analysis of machine learning models was carried out with the selection of the most optimal one. The article also provides a description of the data preprocessing procedure, which is carried out in order to eliminate anomalies and voids in array records, which can lead to incorrect operation of the trained model. Models were trained on selected data in order to identify anomalous traffic behavior indicating a specific type of attack. In addition, a study was conducted on a data set that does not contain information about attacks.

Keywords: *attacks, anomalous traffic, intrusion detection systems, machine learning, Infiltration attack, attacks on web applications, Python*

DOI: 10.18469/ikt.2023.21.3.04

Pozdnyak Irina Sergeevna, Povolzhskiy State University of Telecommunications and Informatics, 77, Moscovskoye shosse, Samara, 443090, Russian Federation; Associated Professor of Information Security Department, PhD in Technical Science. Tel. +7 927 657-24-27. E-mail: i.pozdnyak@psuti.ru

Makarov Igor Sergeevich, Povolzhskiy State University of Telecommunications and Informatics, 77, Moscovskoye shosse, Samara, 443090, Russian Federation; Associated Professor of Information Security Department. PhD in Technical Science. Tel. +7 937 208-80-66. E-mail: i.makarov@psuti.ru

References

1. Saranya T. et al. Performance analysis of machine learning algorithms in intrusion detection system: a review. *Procedia Computer Science*, 2020, vol. 171, pp. 1251–1260.
2. Ferriyan A. et al. Generating network intrusion detection dataset based on real and encrypted synthetic attack traffic. *Applied Sciences*, 2021, vol. 11, no. 17. URL: <https://www.mdpi.com/2076-3417/11/17/7868> (accessed: 15.01.2024).

3. De Lucia M.J., Cotton C. Identifying and detecting applications within TLS traffic. *Proceedings Cyber Sensing*, 2018, vol. 10630, pp. 106300U.
4. Sheluhin O.I., Sakalema D.Zh., Filinova A.S. *Detection of Intrusions into Computer Networks (Network Anomalies): Textbook for Universities*. Moscow: Goryachaya liniya – Telekom, 2022, 220 p. (In Russ.)
5. Sharma S., Krishna C.R., Sahay S.K. Detection of advanced malware by machine learning techniques. *Springer, Advances in Intelligent Systems and Computing*, 2018, vol. 742, pp. 332–342.
6. CSE-CIC-IDS2018 on AWS. URL: <http://www.unb.ca/cic/datasets/ids-2018.html> (accessed: 18.12.2023).
7. Intrusion Detection Evaluation Dataset (CIC-IDS2017). URL: <https://www.unb.ca/cic/datasets/ids-2017.html> (accessed: 18.12.2023).
8. Ferriyan A. et al. HIKARI-2021: Generating network intrusion detection dataset based on real and encrypted synthetic attack traffic. URL: <https://zenodo.org/records/5199540> (accessed: 20.01.2024).
9. Catillo M. et al. USB-IDS-1: A public multilayer dataset of labeled network flows for IDS evaluation. *51st Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W)*, 2021, pp. 1–6. DOI: 10.1109/DSN-W52860.2021.00012
10. Harrison M. *Machine Learning: A Pocket Guide. A Quick Guide to Structured Machine Learning Methods in Python*. Transl. from English by V.A. Kovalenko. Saint Petersburg: Dialektika, 2020, 320 p. (In Russ.)

Received 22.01.2024

ТЕХНОЛОГИИ КОМПЬЮТЕРНЫХ СИСТЕМ И СЕТЕЙ

УДК 004.056

МОДЕЛИ И АЛГОРИТМЫ ОПРЕДЕЛЕНИЯ АКТУАЛЬНЫХ УГРОЗ НА ОСНОВЕ ЭКСПЕРТНОГО ПОДХОДА

Рычкова А.А.¹, Долгушев Н.А.^{1,3}, Бурькова Е.В.¹, Коннов А.Л.^{1,2}

¹ Оренбургский государственный университет, Оренбург, РФ,

² Оренбургский филиал Поволжского государственного университета телекоммуникаций и информатики, Оренбург, РФ,

³ Оренбургский филиал ООО «Уральский центр систем безопасности», Оренбург, РФ
E-mail: rnansy@yandex.ru, dolgushevn1@yandex.ru, tulpan63@bk.ru, andrey_konnov@mail.ru

Задача определения актуальных угроз безопасности приобретает все большее значение, что обусловлено ростом объемов обрабатываемой информации ограниченного доступа, увеличением количества всевозможных угроз, повышением потенциала нарушителей. На предприятиях и в организациях разных сфер экономики специалисты регулярно проводят аудит информационной безопасности с целью выявления уязвимостей и в конечном итоге для предотвращения возможных негативных последствий. Для построения рациональной системы безопасности информации на объекте важно организовать защиту именно от актуальных угроз, так как от всех угроз построить защиту невозможно, да и нецелесообразно. В статье предложены модели и алгоритмы определения актуальных угроз на основе экспертного подхода, которые могут стать базисом для разработки автоматизированной системы принятия достоверных решений в задачах защиты объектов информатизации.

Ключевые слова: угроза безопасности, метод анализа иерархий, иерархическая модель

Введение

Для определения актуальных угроз используют различные методы [2; 3; 7; 9], учитывают особенности для конкретных сфер деятельности [5], существуют также автоматизированные системы на основе программных продуктов [12]. Все эти

методы основаны на положениях методики оценки угроз безопасности федеральной службы по техническому и экспортному контролю (ФСТЭК) России от 2021 года. Учитывается необходимость проведения основных этапов этого процесса: анализа защищаемых объектов, определения не-

гативных последствий, характеристики нарушителя, составления сценариев реализации угроз.

Предлагаемые модели и алгоритмы

Модель угроз разрабатывается группой экспертов, у которых есть опыт в сфере информационной безопасности, есть доступ к аппаратным и программным средствам информационных систем предприятия, доступ к активам, для осуществления оценки как ценности защищаемых ресурсов, так и уровня защищенности их на данном объекте. Эксперт при разработке модели угроз основывается на объективных данных, таких как требования и положения нормативно-правовых документов ФСТЭК России, проведенном аудите инфраструктуры предприятия, сведения об актуальных уязвимостях в системном и прикладном программном обеспечении. В роли экспертов могут выступать сотрудники организации с соответствующей квалификацией в области информационной безопасности или сторонние специалисты. В первом случае эксперты хорошо знакомы с инфраструктурой предприятия, особенностями бизнес-процессов, реальной ситуацией, при этом могут проводить оценку субъективно. Сторонние специалисты, наобо-

рот, при независимой оценке могут недостаточно учитывать специфику и особенности деятельности организации. Экспертный подход основан на сборе, анализе и ранжировании оценок экспертов. На рисунке 1 представлена классификация методов экспертного подхода.

1. Методы группового опроса экспертов. Метод Паттерн предполагает коллективную работу экспертов, обсуждение мнений, совещаний, проводимых в несколько раундов, на каждом этапе опросов экспертов не меняют. Для метода Делфи главной особенностью является независимость экспертов и отсутствие влияния на их оценку, при этом каждый раз привлекают новых экспертов. Результаты такой работы подвергаются обработке методами статистического анализа, и формируется конечная оценка.

2. Математико-статистические методы обработки экспертных оценок основаны на проведении корреляционной оценки, расчете дисперсии, различного рода распределений и подразделяются на: ранжирование, метод парных сравнений, метод последовательных предпочтений, метод непосредственной оценки и другие.

3. Методы экспертной оценки показателей качества основаны на определении весомости того

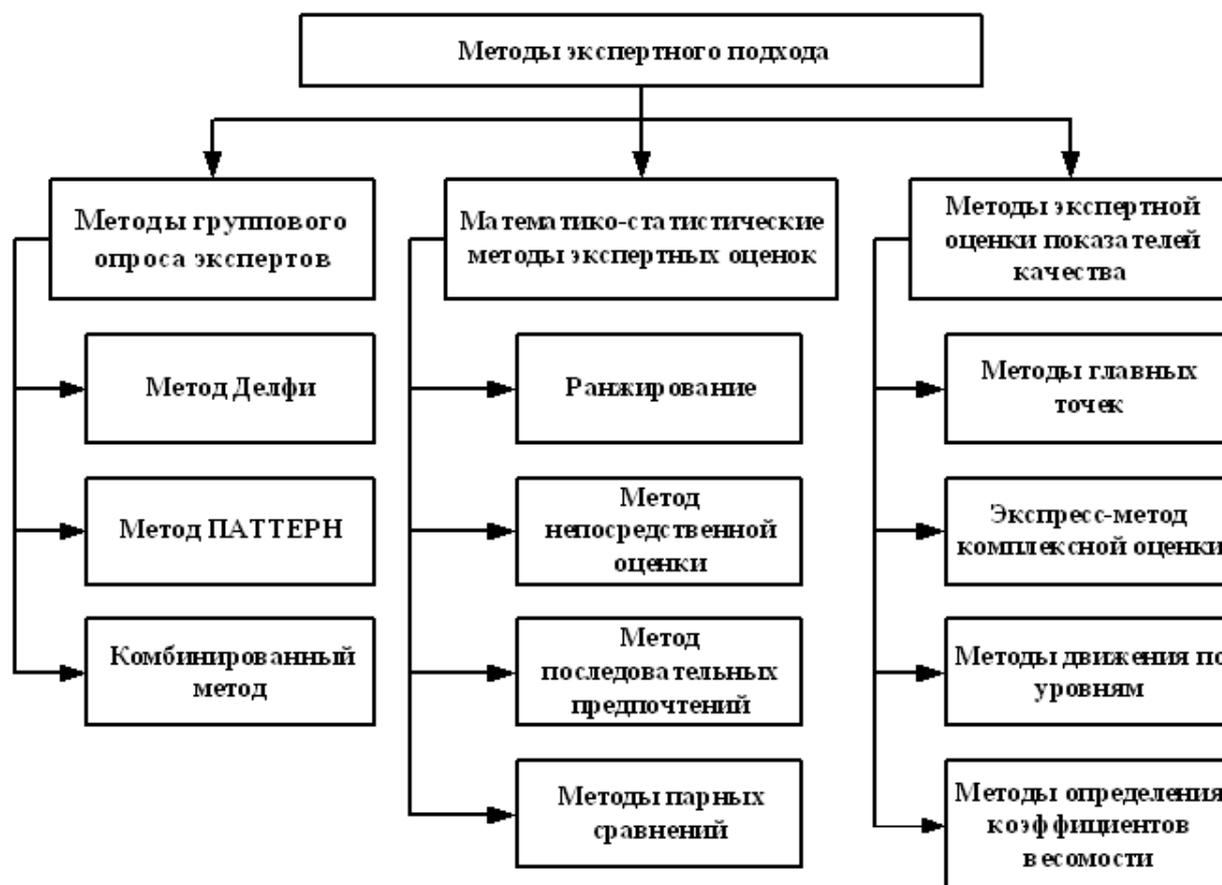


Рисунок 1. Классификация методов экспертного подхода

или иного критерия качества продукции или услуги. Применение данных методов целесообразно, если показателями качества выбраны конкретные физические величины и в наличии присутствуют измерительные приборы с заданными метрологическими параметрами.

Для нашего исследования был выбран класс математико-статистических методов обработки экспертных оценок, конкретно метод анализа иерархий (МАИ).

На первом этапе решения задачи с применением МАИ следует задать критерии оценки угроз. На втором этапе эксперты должны определить степень значимости критериев попарным сравнением по 9-балльной шкале. Преимущество этого метода заключается в предоставлении возможности учета различных факторов, несопоставимых друг с другом, для этого проводится сравнение каждого критерия с каждым [3; 5].

Нами была разработана иерархическая модель определения актуальных угроз с использованием МАИ (рисунок 2).

На первом уровне представлены этапы процесса оценки угроз безопасности в соответствии с Методикой оценки угроз безопасности информации ФСТЭК России от 5 февраля 2021 года: негативные последствия, которые могут привести к ущербу для предприятия, перечень угроз, перечень объектов воздействия этих угроз, среди которых автоматизированные рабочие места, серверное оборудование, сетевые, интерфейсные компоненты и другие.

На втором уровне необходимо выделить критерии, уровень значимости которых должны оценить эксперты.

Третий уровень включает непосредственные значения оценки критериев в процентном отношении.

Согласно Методике ФСТЭК для определения угроз, актуальных для данной организации, необходимо определить источник угроз по возможности нарушителю, провести инвентаризацию всех активов для понимания объектов воздействия, рассмотреть все возможные уязвимости, реальные атаки на компьютерные системы, доступные из различных открытых источников, оценить возможные риски (ущерб) для владельцев информационных активов [1; 4]. Угрозу безопасности информации (УБИ) составляют следующие компоненты:

УБИ = [нарушитель (источник угрозы), объект воздействия, способы реализации угрозы; негативные последствия].

Модель угроз разрабатывается, как правило, группой экспертов, мнения которых могут не совпадать. При коллективном оценивании достоверность принятых решений следует обеспечить согласованностью с исключением необъективных завышенных или заниженных прогнозов.

Для решения возможных проблем оценки согласованности мнений экспертов применение только метода анализа иерархий является недостаточным. Необходимо определить согласованность оценок экспертов для достижения наибольшей достоверности принятого решения. Субъективные психологические факторы могут исказить полученные в ходе оценки результаты. Чтобы оценить уровень согласованности экспертных оценок предлагается вычислить для ран-

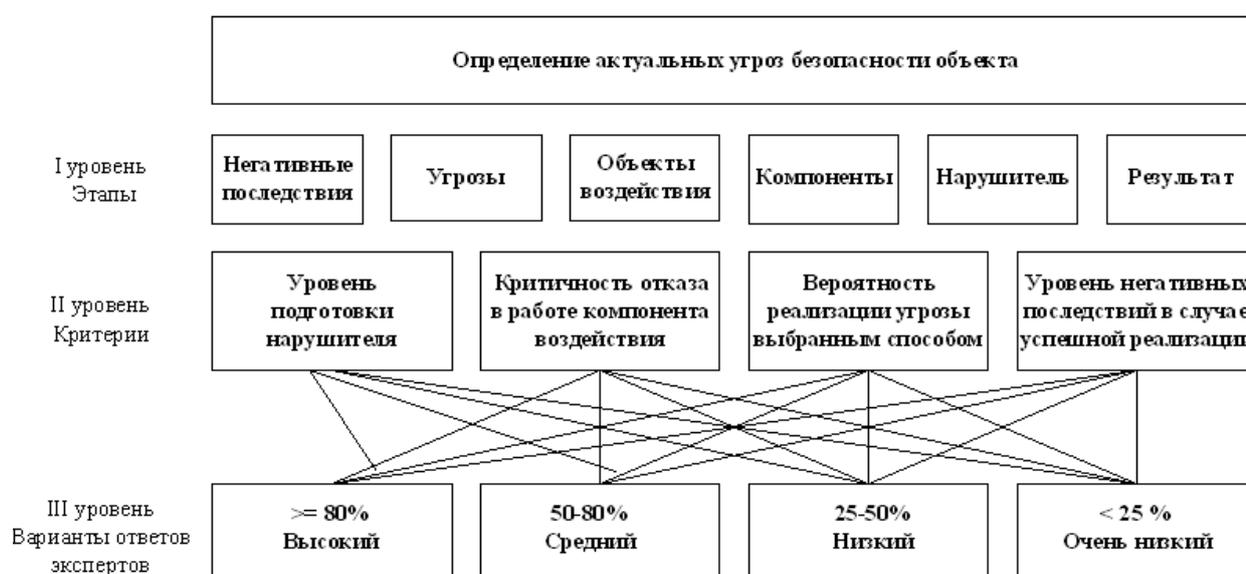


Рисунок 2. Иерархическая модель определения актуальных угроз

жированных данных коэффициент конкордации. При переходе этого коэффициента за заданный предел формируется результат о неслучайной согласованности оценок группы экспертов.

На рисунке 3 представлены алгоритмы определения согласованности экспертных мнений, включающие обработку полученных экспертных оценок в виде матрицы рангов, расчет коэффициента конкордации и расчет статистической значимости.

Приведенный выше алгоритм включает несколько ключевых этапов:

1. Формирование экспертной группы.
2. Определение показателей оценки и получение в соответствии с ними экспертного мнения.
3. Расчет математических показателей степени согласованности экспертного мнения.
4. Формирование результатов.

Первым этапом предлагаемого алгоритма является вопрос формирования экспертной группы. Этап требует четкой постановки задач в зависимости от рассматриваемого объекта защиты. Эксперты должны обладать достаточными компетенциями

в различных аспектах информационной безопасности с учетом применяемых на объекте технических решений. Критерии отбора должны учитывать профессиональные квалификации и опыт.

Определение показателей оценки служит ключевым этапом в формировании мнения об актуальности угроз информационной безопасности [10]. На данном этапе должны быть выделены наиболее значимые характеристики субъекта оценки, совокупность которых будет являться основой для выработки стратегий защиты. Для предложенного алгоритма мнение экспертов выражается в количественной оценке, что позволяет провести обработку полученной информации с помощью математического метода.

На третьем этапе объективную оценку результата обеспечивает расчет коэффициента конкордации. Выбор граничного значения коэффициента напрямую влияет на результаты проведения опроса – значение коэффициента прямо пропорционально степени согласованности экспертного мнения. При расчете статистических показателей

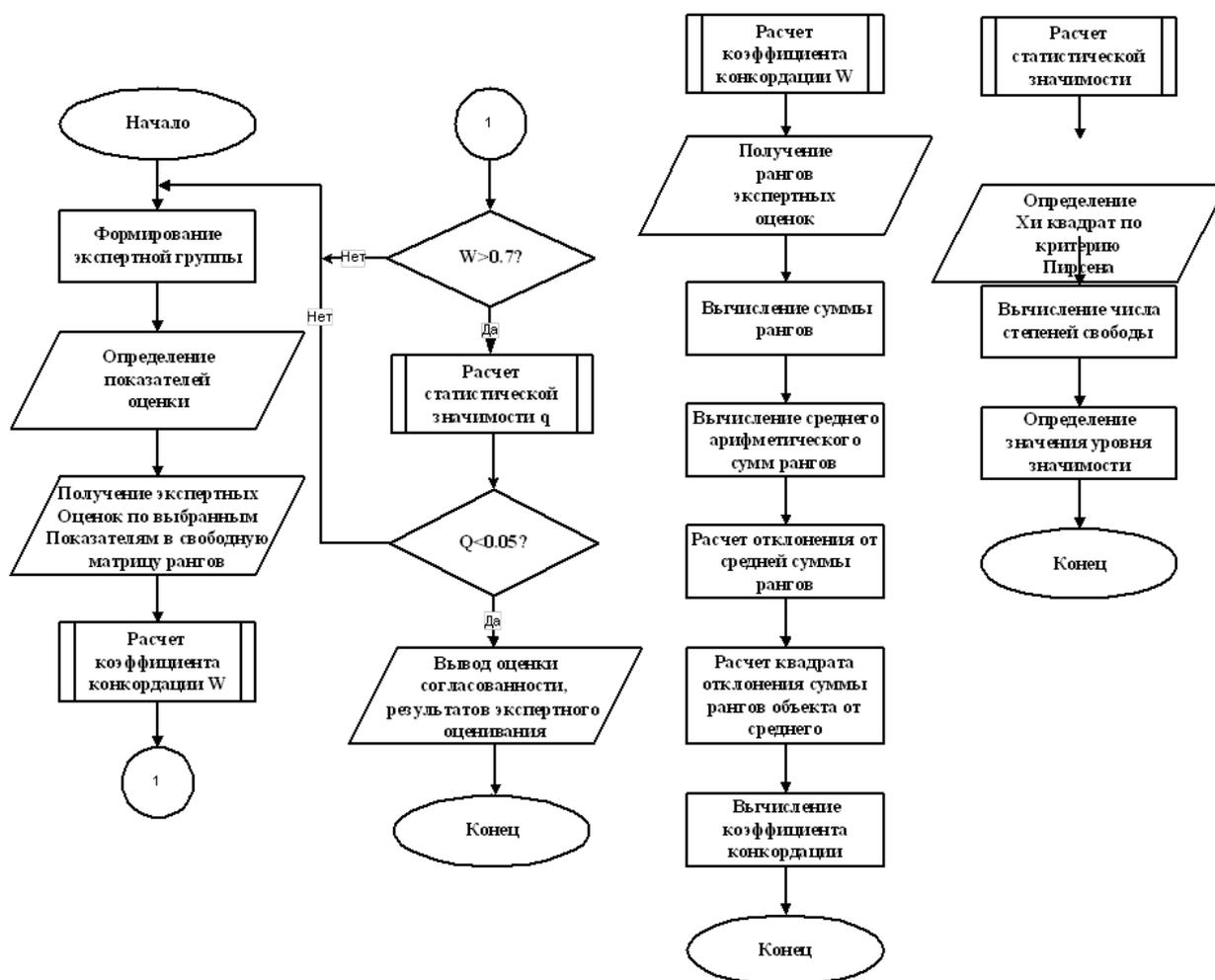


Рисунок 3. Алгоритмы определения согласованности экспертных мнений

также важно выполнение проверки результатов на вероятность случайного возникновения. Гипотеза о случайности полученных результатов проверяется в соответствии со значением критерия Пирсона, пороговое значение которого зависит от количества экспертов в опрашиваемой группе [6].

Результатами проведения опроса экспертной группы является обобщенный перечень актуальных угроз для рассматриваемого объекта защиты, а также результат оценки согласованности экспертных мнений. Предлагаемый алгоритм позволяет не только выполнить оценку угроз по выбранным характеристикам, но и сформировать единое экспертное мнение с допустимой степенью согласованности, основанное на компетенциях специалистов разных направлений в области информационной безопасности.

В соответствии с разработанными алгоритмами было проведено моделирование определения степени согласованности экспертных оценок, реализация осуществляется в программном средстве.

В качестве примера рассматривается группа из 4 экспертов. Для вывода результатов оценки сформирована шкала ранжирования из 4 значений: очень низкий, низкий, средний и высокий уровни. Далее выполняется опрос экспертов по каждому из четырех критериев: уровень подготовки нарушителя, критичность отказа в работе компонентов, вероятность реализации угрозы, уровень негативных последствий. Для разработки программного средства был проведен анализ аналогичных разработок, определены особенности, выявлены достоинства и недостатки, определены требования к разработке авторского программного средства. Новый раздел банка угроз ФСТЭК позволяет в интерактивном режиме осуществить

формирование угроз безопасности по актуальной Методике [4]. На пересечении множеств угроз безопасности, объектов (компонентов) воздействий и способов реализации (нарушители) формируется УБИ. При этом отсутствует возможность провести коллективную оценку и проверить согласованность и достоверность мнений экспертов. Для автоматизации данных решений создана функциональная модель процесса определения актуальных угроз на основе экспертного подхода, которая позволяет учесть поступающие на вход данные об объекте защиты получить результат наряду с перечнем актуальных УБИ, формировать отчет о согласованности экспертного мнения и дальнейшие рекомендации по корректировке действий экспертов [8]. Общая функциональная модель IDEFO представлена на рисунке 4.

Декомпозиция процесса определения актуальных угроз включает пять последовательных этапов, первые четыре из которых проходит каждый эксперт отдельно, что позволяет осуществить дистанционный сбор мнений и результаты опроса загрузить в разрабатываемую программную систему.

На первом шаге производится выбор перечня объектов воздействия угроз, на основе которого определяется вероятный нарушитель и перечень характерных для него угроз. Затем определяются способы реализации выявленных на предыдущем этапе угроз. После этого выявляются сценарии воздействия угроз на объекты информатизации для сопоставления угроз и объектов воздействий. Для полученного списка угроз каждый эксперт определяет негативные последствия по уровню ущерба для каждой угрозы. Составленные каждым экспертом по отдельности перечни векторов



Рисунок 4. Общая функциональная модель разрабатываемого метода

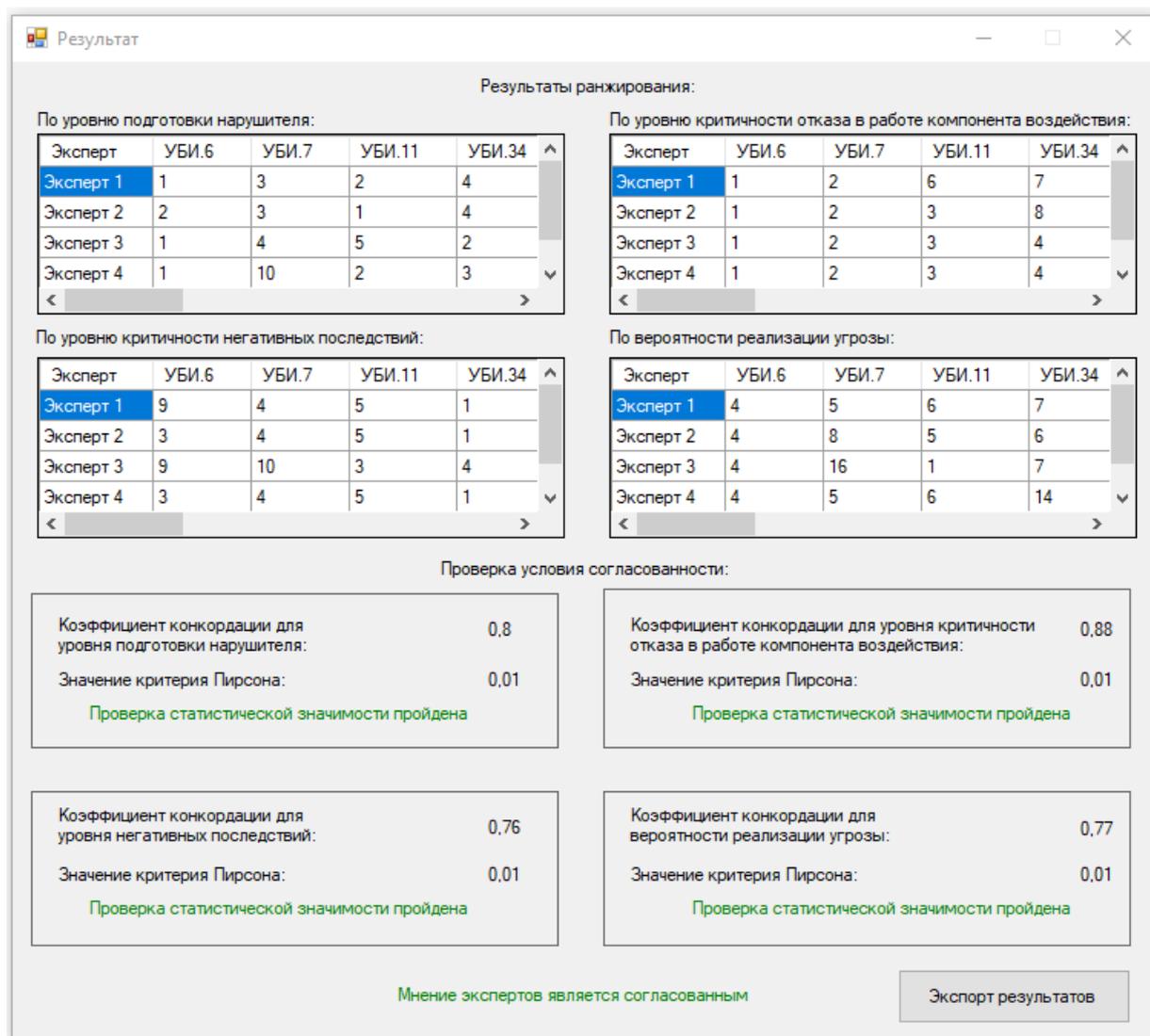


Рисунок 5. Экранная форма вывода результата опроса экспертной группы

угроз являются входными данными для определения согласованности экспертного мнения.

Расчет коэффициента конкордации является заключительным этапом и позволяет определить согласованность мнений экспертов и уровень значимости для принятия объективного решения.

На экранной форме разработанного программного средства представлены результаты расчета коэффициента конкордации, значения которого превышают критическое значение (0,7) и результаты расчета уровня значимости (рисунок 5).

Вывод

Были разработаны иерархическая модель определения актуальных угроз с использованием экспертного подхода, а также алгоритмы определения степени согласованности экспертных оценок в задаче определения актуальных угроз объекта информатизации, что позволило повы-

сить эффективность построения рациональной системы безопасности информации именно от актуальных угроз, так как от всех угроз построить защиту невозможно, да и нецелесообразно. Предложенный подход позволяет уменьшить процент потенциально неучтенных угроз за счет использования классификации информационных активов и проверки согласованности экспертного мнения. На основе представленных моделей и алгоритмов разработано программное средство для автоматизированной оценки угроз информационной безопасности и определения согласованности полученных результатов.

Литература

1. Банк данных угроз безопасности информации. Федеральная служба по техническому и экспортному контролю. URL: <https://bdu.fstec.ru/threat> (дата обращения: 17.01.2024).

2. Дивина Т.В., Петракова Е.А., Вишневецкий М.С. Основные методы анализа экспертных оценок. Экономика и бизнес: теория и практика. 2019. №7. С. 42–44.
3. Бурькова Е.В., Рычкова А.А. Методика принятия решений при выборе средств физической защиты на основе метода анализа иерархий // Научно-технический вестник Поволжья. 2021. № 5. С. 119–123.
4. Методика оценки угроз безопасности информации: методический документ. Москва: ФСТЭК. URL: <https://fstec.ru/dokumenty/vse-dokumenty/spetsialnye-normativnye-dokumenty/metodicheskij-dokument-ot-5-fevralya-2021-g> (дата обращения: 17.01.2024).
5. Рычкова А.А., Клиндух В.О. Оценка эффективности защитных мер персональных данных в учебном заведении на основе метода анализа иерархий // Новые импульсы развития: вопросы научных исследований. 2021. С. 58–66.
6. Письменская А.А., Гостюнин Ю.А., Давидюк Н.В. Согласованность мнений экспертов при оценке рисков информационной безопасности с применением АВС-анализа // Математические методы в технике и технологиях (ММТТ). 2020. № 7. С. 55–57.
7. Смирнов Р.А., Новиков С.Н. Анализ методик оценки угроз безопасности информации // Телекоммуникации. 2023. № 7. С. 24–27.
8. Степанов В.А., Андреев Н.Д. Моделирование угроз безопасности информации по новой методике ФСТЭК, используя средства автоматизации // Информационные технологии. Проблемы и решения. 2021. № 4 (17). С. 95–101.
9. Филиппов Н.В., Киреева Н.В., Поздняк И.С. Подход к созданию экспертной системы оценки информационной безопасности телекоммуникационных систем // Электросвязь. 2022. № 2. С. 61–66.
10. Kotenko I., Parashchuk I. Analysis of threats to information security of industrial automation systems using euclidean and hamming distances between fuzzy sets // Conference Proceedings: International Russian Automation Conference (RusAutoCon). 2023. P. 13–18.

Получено 18.01.2024

Рычкова Анастасия Александровна, к.п.н., доцент кафедры вычислительной техники и защиты информации (ВТиЗИ) Оренбургского государственного университета (ОГУ). 460018, Российская Федерация, г. Оренбург, Проспект Победы, 13. Тел. +7 922 535-32-45. E-mail: rnansy@yandex.ru

Долгушев Никита Александрович, студент кафедры ВТиЗИ ОГУ. 460018, Российская Федерация, г. Оренбург, Проспект Победы, 13; системный инженер Оренбургского филиала ООО «Уральский центр систем безопасности». 460021, Российская Федерация, г. Оренбург, ул. Луговая, 55. Тел. +7 951 032-77-91. E-mail: dolgushevnl@yandex.ru

Бурькова Елена Владимировна, к.п.н., доцент, доцент кафедры ВТиЗИ ОГУ. 460018, Российская Федерация, г. Оренбург, Проспект Победы, 13. Тел.+7 903 360-77-92. E-mail: tulpan63@bk.ru

Коннов Андрей Леонидович, к.т.н., доцент, доцент кафедры ВТиЗИ ОГУ. 460018, Российская Федерация, г. Оренбург, Проспект Победы, 13; доцент кафедры математических и естественно-научных дисциплин Оренбургского филиала Поволжского государственного университета телекоммуникаций и информатики. 460022, Российская Федерация, г. Оренбург, ул. Пролетарская/Юркина, 249/76. Тел. +7 912 844-91-91. E-mail: andrey_konnov@mail.ru

MODELS AND ALGORITHMS FOR IDENTIFYING CURRENT THREATS BASED ON AN EXPERT APPROACH

Richkova A.A.¹, Dolgusheva N.A.^{1,3}, Burkova E.V.¹, Konnov A.L.^{1,2}

¹*Orenburg State University, Orenburg, Russian Federation*

²*Orenburg Branch of the Povolzhskiy State University of Telecommunications and Informatics, Orenburg, Russian Federation*

³*Orenburg Branch of the Ural Center of Security Systems, Orenburg, Russian Federation*

E-mail: rnansy@yandex.ru, dolgushevnl@yandex.ru, tulpan63@bk.ru, andrey_konnov@mail.ru

The task of identifying current security threats is becoming increasingly important, due to the increase in the volume of processed information of limited access, an increase in the number of all kinds of threats, and an increase in the potential of violators. At enterprises and organizations in various sectors of the

economy, specialists regularly conduct information security audits in order to identify vulnerabilities and ultimately to prevent possible negative consequences. To build a rational information security system at the facility, it is important to organize protection against actual threats, since it is impossible to build protection against all threats, and it is impractical. The article proposes models and algorithms for determining current threats based on an expert approach, which can become the basis for the development of an automated system for making reliable decisions in the tasks of protecting informatization objects.

Keywords: *security threat, hierarchy analysis method, hierarchical model*

DOI: 10.18469/ikt.2023.21.3.05

Rychkova Anastasia Aleksandrovna, Orenburg State University, 13, Pobedy Avenue, Orenburg, 460018, Russian Federation; Associate Professor of Computer Science and Information Security Department, PhD in Pedagogical Sciences. Tel. +7 922 535-32-45. E-mail: rnansy@yandex.ru

Dolguшев Nikita Aleksandrovich, Orenburg State University, 13, Pobedy Avenue, Orenburg, 460018, Russian Federation; Student of Computer Science and Information Security Department. Orenburg Branch of the Ural Center of Security Systems, 55, Lugovaya Street, Orenburg, 460021, Russian Federation; System engineer. Tel. +7 951 032-77-91. E-mail: dolgushevnl@yandex.ru

Burkova Elena Vladimirovna, Orenburg State University, 13, Pobedy Avenue, Orenburg, 460018, Russian Federation; Associate Professor of Computer Science and Information Security Department, PhD in Pedagogical Sciences. Tel. +7 903 360-77-92. E-mail: tulpan63@bk.ru

Konnov Andrey Leonidovich, Orenburg State University, 13, Pobedy Avenue, Orenburg, 460018, Russian Federation; Associate Professor of Computer Science and Information Security Department, PhD in Technical Sciences. Orenburg Branch of the Povolzhskiy State University of Telecommunications and Informatics, 249/76, Proletarskaya/Urkina Street, Orenburg, 460022, Russian Federation; Associate Professor of Mathematical and Natural Science Disciplines Department. Tel. +7 912 844-91-91. E-mail: andrey_konnov@mail.ru

References

1. Database of information security threats. federal service for technical and export control. URL: <https://bdu.fstec.ru/threat> (accessed: 17.01.2024). (In Russ.)
2. Divina T.V., Petrakova E.A., Vishnevsky M.S. Basic methods of analysis of expert assessments. *Ekonomika i biznes: teoriya i praktika*, 2019, no. 7, pp. 42–44. (In Russ.)
3. Burkova E.V., Rychkova A.A. Method of decision-making system for the selection of physical protection means based on the hierarchy analysis method. *Nauchno-tekhnicheskij Vestnik Povolzh'ya*, 2021, no. 5, pp. 119–123. (In Russ.)
4. *Methodology for Assessing Information Security Threats: Methodological Document*. Moscow: FSTEC. URL: <https://fstec.ru/dokumenty/vse-dokumenty/spetsialnye-normativnye-dokumenty/metodicheskij-dokument-ot-5-fevralya-2021-g> (accessed: 17.01.2024). (In Russ.)
5. Rychkova A.A., Klindukh V.O. Evaluation of the effectiveness of protective measures of personal data in an educational institution based on the method of hierarchy analysis. *Novye impul'sy razvitiya: voprosy nauchnyh issledovanij*, 2021, pp. 58–66. (In Russ.)
6. Pishevskaya A.A., Gostyunin Yu.A., Davidyuk N.V. Agreement of opinions of experts in the assessment of risks of information security with application of abc analysis. *Matematicheskie metody v tekhnike i tekhnologiyah (MMTT)*, 2020, no. 7, pp. 55–57. (In Russ.)
7. Smirnov R.A., Novikov S.N. Analysis of assessment methods of information security threats. *Telekommunikacii*, 2023, no. 7, pp. 24–27. (In Russ.)
8. Stepanov V.A., Andreev N.D. Modeling information security threats using the new FSTEC methodology using automation tools. *Informacionnye tekhnologii. Problemy i Resheniya*, 2021, no. 4 (17), pp. 95–101. (In Russ.)
9. Filippov N.V., Kireeva N.V., Pozdnyak I.S. Approach to creating an expert system for

assessing the information security of telecommunication systems. *Elektrosvyaz*, 2022, no. 2, pp. 61–66. (In Russ.)

10. Kotenko I., Parashchuk I. Analysis of threats to information security of industrial automation systems using euclidean and hamming distances between fuzzy sets. *Conference Proceedings: International Russian Automation Conference (RusAutoCon)*, 2023, pp. 13–18. (In Russ.)

Received 18.01.2024

УДК 004.7; 372.862

СПЕЦИАЛИСТ ПО ЭКСПЛУАТАЦИИ ПРОГРАММНО-КОНФИГУРИРУЕМЫХ СЕТЕЙ: ЗНАНИЯ, УМЕНИЯ, НАВЫКИ

Галич С.В., Сафонова О.Е.

Волгоградский государственный университет, Волгоград, РФ

E-mail: galich.sv@volsu.ru

В статье проведен анализ российских разработок в сфере программно-конфигурируемых сетей, профессионального стандарта и вакансий с целью определения требуемых от специалистов знаний, умений, навыков. Показано, что рынок труда испытывает явную потребность в кадрах, способных эксплуатировать и разрабатывать решения, функционирующие на основе принципов программно-конфигурируемых сетей. Сделаны выводы о том, что наблюдается недостаток пособий практической направленности для подготовки специалистов в данной сфере, а также предложены возможные темы лабораторных работ для учебных пособий. Данная статья может быть полезна преподавателям образовательных учреждений средне-специального и высшего образования, а также дополнительного профессионального образования при разработке рабочих программ дисциплин, лекций и лабораторных практикумов, посвященных программно-конфигурируемым сетям, а также обучающимся вышеупомянутых образовательных учреждений при планировании индивидуального плана обучения, выпускникам и молодым специалистам при планировании своего профессионального и карьерного развития.

Ключевые слова: *программно-конфигурируемые сети, виртуализация сетевых функций, знания, умения, навыки, образование*

Введение

В 2022 году уход с российского рынка крупнейших мировых производителей сетевого оборудования, аннулирование ими в одностороннем порядке действующих лицензий и прекращение технической поддержки закупленного оборудования продемонстрировали уязвимость набравшей популярность сервисной модели предоставления услуг и оборудования.

Сетевые устройства со встроенными интеллектуальными функциями, такие как межсетевые экраны нового поколения (NGFW), по воле производителя лишались большей части своего функционала быстрее, чем администраторы успевали предпринять какие-либо действия. Особо уязвимы оказались устройства, функционирующие в соответствии с принципами технологии программно-конфигурируемых сетей (ПКС, англ. SDN), подразумевающей отделение плоскости управления (англ. control plane) от плоскости передачи (англ. data plane), при этом плоскость управления в лице SDN-контроллера совершенно не обязательно территориально располагать в центре обработки данных (ЦОД) или серверной эксплуатирующей организации. SDN-контроллер

вполне может функционировать в облаке производителя устройства, более того, подобный подход в ряде случаев может быть выгодным для потребителя с точки зрения затрат и лицензирования, а именно позволяет реализовать одно из достоинств SDN как концепции и переводить капитальные затраты на оборудование в операционные [1].

В крайне деструктивных действиях была замечена компания Cisco: например, компания не ограничилась отключением устройств Cisco Meraki от облачных сервисов, но и захватила контроль над устройствами на территории Российской Федерации и создала открытый SSID (Service Set Identifier) вида «12345-Sanctions». Продукты серии Cisco vEdge SD-WAN (Software-Defined Networking in a Wide Area Network, программно-определяемая глобальная сеть) имеют целый набор различных проблем: от невозможности добавить новое устройство до истекших 9 мая 2023 года сертификатов безопасности, приводящих к неработоспособности устройств после перезагрузки и обновление которых невозможно без обращения в поддержку производителя (который, опять же, ушел с российского рынка).

В подобных действиях, возможно, не столь разрушительных, замечена большая часть ушедших с рынка производителей оборудования, ранее занимавших лидирующие позиции в России.

SD-WAN-решения являются реализацией концепции SDN для корпоративных сетей передачи данных и наибольшую выгоду могут принести компаниям с распределенной филиальной сетью, для которых крайне важно снизить время подключения новых офисов, таких как ретейлеры и банки [2].

Глобальными лидерами среди производителей сетевого оборудования в сегменте SD-WAN принято считать компании Cisco, Fortinet, VMware, HPE (Aruba), Palo Alto Networks, Versa Networks. Именно такую оценку давали аналитические компании IDC [3] и Gartner [4] в 2021 и 2022 годах. Аналитики из TeleGeography [5] в качестве однозначных лидеров выделяют Cisco и VMware; также помимо вышперечисленных в топ-10 производителей SD-WAN включают Nuage Networks, Juniper Networks, Palo Alto Networks, Riverbed. Тем не менее российские производители воспользовались уникальным шансом заполнить освободившиеся после ухода лидеров индустрии ниши и предложили в том числе решения класса SD-WAN.

Российские разработки SDN и SD-WAN

В России исследования в сфере программно-конфигурируемых сетей получили свое развитие во многом благодаря Московскому государственному университету им. М.В. Ломоносова и НП «Центр прикладных исследований компьютерных сетей» (НП «ЦПИКС»), создавших Консорциум «ПКС – технологии в научно-образовательной среде», в состав которого вошли ведущие вузы страны. На базе разработок НП «ЦПИКС» запущены стартапы NFWare и WiMark Systems, разработан контроллер RunOS (две версии: коммерческая и open-source) и платформа виртуализации сетевых функций (ВСФ, англ. NFV) C2 Cloud Conductor. НП «ЦПИКС» имеет опыт реализации пилотных проектов на основе своих решений в сетях операторов связи [6], в частности совместно с операторами связи «Вымпелком», «Воентелеком» и «Ростелеком». Судя по имеющимся публикациям сотрудников [7], ведутся и разработки в направлении SD-WAN, однако свежих новостей в открытых источниках от НП «ЦПИКС» не обнаружено.

В Российской Федерации именно ПАО «Ростелеком» является крупнейшим испытателем программно-конфигурируемых сетей (ПКС). Необходимость развития решений ПКС и ВСФ как неотъемлемой части сетей 5G указана в «Ком-

плексной программе содействия развитию 5G в России» [8]. Помимо тестирования решений НП «ЦПИКС», оператором связи был реализован пилотный проект с Brain4Net Controller (разработка ООО «Программируемые сети», резидент «Сколково») [9] и осуществлялось тестирование транспортной ПКС (Transport SDN) совместно с западными производителями Huawei, NEC и Nokia [9].

В 2021 году совместными усилиями Tele2 и «Ростелеком» был реализован пилотный проект по внедрению технологии SD-WAN в сети торговых точек. Тестирование проводилось на базе Brain4Net Service Platform с использованием оборудования компании «Булат» (совместное предприятие «Ростелеком» и «Ростех») и сим-карт Tele2. Благодаря техническим преимуществам от внедрения SD-WAN, среди которых повышение отказоустойчивости и снижение риска недоступности канала связи в точке продаж, бизнес может предоставлять посетителям точек продаж более качественный сервис в силу снижения вероятности возникновения очереди из-за сбоев кассового оборудования и цифровых информационных панелей [10; 11].

В октябре 2021 года «Лаборатория Касперского» приобрела ООО «Программируемые сети», расширив свое портфолио продуктов сетевым решением класса SD-WAN. Таким образом «Лаборатория Касперского» пошла по пути компаний Cisco, VMware, Nokia, в портфолио решений которых SD-WAN также появлялся путем приобретения стартапов Viptela, VeloCloud, Nuage. Стоит отметить, что «Лаборатория Касперского» остается компанией, разрабатывающей программное обеспечение, а не аппаратное. Поэтому Kaspersky Edge Service Router (KESR) является наименованием линейки маршрутизаторов, разработанных партнерами «Лаборатории Касперского». Линейка включает в себя пять моделей: от младшей KESR Model 1 с пропускной способностью 30 Мбит/с до старшей KESR Model 5, обеспечивающей 10 Гбит/с [12]. Выяснить наименование конкретных моделей не удалось, однако можно сделать предположение, что в линейку KESR входят устройства «Булат» BS5120, поскольку имеется упоминание применения оборудования «Булат» в рамках пилотного проекта [10; 11].

Среди других российских вендоров сетевого оборудования протокол OpenFlow 1.0/1.3.4 поддерживают также устройства Eltex MES5448 и MES7048 [13]. В модельном ряду Qtech обнаружилась линейка QSW-6510, однако согласно информации на сайте производителя [14] она более не производится.

На российском рынке пионером внедрения SD-WAN стала компания МТС (облачный провайдер CloudMTS), начавшая в 2020 году использовать решения Fortinet.

Свой вклад в популяризацию технологии внесла компания BI.ZONE, выпустившая в 2021 году первое отечественное решение SD-WAN. Именно о переходе на решения BI.ZONE объявила компания МТС в январе 2023 года. Крайне интересным решением является применение в составе BI.ZONE Secure SD-WAN протокола Wireguard с поддержкой шифрования по алгоритму «Кузнечик» (ГОСТ Р 34.12–2015). Реализация размещена на Github, планируется ее сертификация по требованиям ФСБ России. Модельный ряд маршрутизаторов включает в себя семь моделей: младшая CyberEdge 10 обеспечивает пропускную способность 20 Мбит/с, старшая BI.ZONE CyberEdge 3000 – 3 Гбит/с [15].

Еще одной компанией, имеющей в портфолио SD-WAN-решение, является Network Systems Group (NSG), предлагающая продукт «Богатка SD-WAN». Разработка отмечена дипломом технологического конкурса 5G SmartTech 2022, проводившегося фондом «Московский инновационный кластер». По мнению авторов статьи, продукт очень интересный и самобытный, но, к сожалению, в открытом доступе представлено крайне мало информации, за исключением руководства администратора системы [16].

В таблице 1 приведен обзор стека технологий и протоколов, применяемого в российских разработках SD-WAN. В ходе проводимого сравнения не ставилась задача определения лучшей разработки или лидера рынка. По мнению авторов, такие выводы можно сделать только на основе пилотных проектов и опыта эксплуатации. В данной работе сравнение проводится с целью выяс-

Таблица 1. Обзор технологий и протоколов, задействованных в составе решений SD-WAN от российских компаний

| Параметр сравнения | Kaspersky SD-WAN | BI.ZONE Secure SD-WAN | NSG Богатка SD-WAN | RunOS SDN (open-source) |
|--|--|---|--|---|
| Протокол Control Plane | OpenFlow 1.3.4 | BCMP (BI.ZONE CPE Management Protocol) | SSH, bash-скрипты | OpenFlow 1.3 |
| Операционная система | Astra Linux 1.7 | Astra Linux 1.7, Альт Линукс СП 8, Ubuntu | Linux (Ubuntu) | Linux (Ubuntu) |
| Протокол Overlay-туннелирования | GENEVE с TLS 1.3 и шифрованием ChaCha20 | Wireguard с поддержкой шифрования ГОСТ | Собственная разработка «Клещ» VPN совместно с TLS 1.3; «Клещ» VPN совместно с IPsec, OpenVPN | VxLAN |
| Совместимое оборудование CPE | Kaspersky Edge Service Router; uCPE на базе стандартных серверов x86 архитектуры | BI-ZONE CyberEdge и CyberEdge VM на базе стандартных серверов x86 архитектуры | NSG-17xx, NSG-1810/1820/1830, NSG-20xx, NSG-30xx | Программный коммутатор RunAR на базе стандартных серверов x86 архитектуры; Программно-аппаратный коммутатор RunDR на базе сетевого процессора |
| Менеджер Виртуализации сетевых функций | OpenStack | Собственная разработка «Оркестратор функций безопасности», управляющий виртуальными сетевыми функциями на CPE | н/д | C2 (Cloud Conductor) |
| Мониторинг | Zabbix | Имеется | Zabbix | Имеется |

нить, какие технологии и протоколы задействованы в составе решений, а также владение какими навыками и знаниями инструментов требуется от специалиста по эксплуатации SD-WAN.

Из представленных на российском рынке решений бесплатной для образовательных учреждений является только open-source версии RunOS. Это решение однозначно следует рассматривать в рамках лабораторных занятий наряду с другими известными в исследовательском сообществе open-source-проектами, такими как ONOS и OpenDaylight [17; 18]. В качестве сетевых узлов можно использовать персональный компьютер с развернутыми на них Open vSwitch. Касательно же решений Kaspersky, BI.ZONE, NSG – возможно, производители предложат некоторые демонстрационные версии для загрузки (по примеру компании «Код Безопасности» [19]) или развернут демо-стенды на своих мощностях. По крайней мере, со стороны «Лаборатории Касперского» подобные действия уже предпринимались: в рамках технического онлайн-марафона «Kaspersky SD-WAN 2.0: функциональные возможности и практический опыт», проводившегося с 29.05.23 по 02.06.23 гг., участникам предлагалось записаться на открытую лабораторную работу. Образовательным учреждениям для оснащения учебных лабораторий необходимо налаживать партнерство с российскими же производителями SDN и SD-WAN решений. В этом случае учебный процесс будет построен с использованием решений, внедряемых в эксплуатацию, что повысит востребованность выпускников таких образовательных учреждений.

В таблице 2 приведены сведения о наличии сертификатов соответствия регуляторов сфер информационных технологий и информационной безопасности. Отметим, что все рассматриваемые решения зарегистрированы в реестре российского программного обеспечения. На момент подготовки статьи рассматриваемые решения не

имели сертификатов соответствия от Федеральной службы технического и экспортного контроля (ФСТЭК) как средства защиты информации (СЗИ) и от Федеральной службы безопасности (ФСБ) как средства криптографической защиты информации (СКЗИ). Сама по себе сертификация рассматриваемых решений в силу специфики проведения данной процедуры оценки соответствия и эксплуатации сертифицированных изделий в строгом соответствии с требованиями формуляра может привести к потере некоторых преимуществ SDN, в частности гибкости, масштабируемости и скорости внедрения новых разработок. Тем не менее компаниями Kaspersky и BI.ZONE планируется сертификация своих решений SD-WAN в качестве межсетевых экранов (МЭ). Касательно сертификации в качестве средств криптографической защиты информации эти две компании выбрали различный подход: BI.ZONE планирует сертифицировать решение, а Kaspersky предлагает вариант развертывания своего решения поверх существующих VPN-туннелей с ГОСТ-шифрованием на базе сертифицированных СКЗИ. Само же наличие сертификата от регуляторов отрасли расширяет сферы применения российских решений с точки зрения соответствия требованиям законодательства, а значит, с позиции конечного потребителя SDN, свидетельствует о зрелости решения и его готовности к внедрению в эксплуатацию.

Анализ опубликованных вакансий и требований, предъявляемых работодателями

Как видно, на российском рынке уже имеются отечественные решения SD-WAN, которые применяются для модернизации существующих и построения новых сетей связи. Для того чтобы определить ожидания работодателей от квалификации и технических навыков специалистов

Таблица 2. Сертификация решений SD-WAN от российских компаний

| Параметр сравнения | Kaspersky SD-WAN | BI.ZONE Secure SD-WAN | NSG Богатка SD-WAN | RunOS SDN (open-source) |
|------------------------------------|--|---------------------------------------|--------------------------|--------------------------|
| Запись в реестре отечественного ПО | № 5840 от 20.09.2019 г. (как B4N Orchestrator) | № 9005 от 28.01.2021 г. | № 12257 от 09.12.2021 г. | № 14817 от 12.09.2022 г. |
| Сертификация ФСТЭК | Планируется сертификация СРЕ как МЭ | Планируется как МЭ | Отсутствует | Отсутствует |
| Сертификация ФСБ | Отсутствует; SD-WAN способна работать поверх ГОСТ-туннелей | Планируется как СКЗИ по классу КС 1-3 | Отсутствует | Отсутствует |

по эксплуатации программно-конфигурируемых сетей в современных условиях, был проведен анализ размещенных в открытом доступе вакансий на сайтах по трудоустройству *hh.ru* и *career.habr.com*, а также разделах «Вакансии» на сайтах крупнейших телеком-операторов МТС, «Мегафон», «Билайн», «Ростелеком». Мониторинг вакансий осуществлялся в мае и октябре 2023 года. Поиск осуществлялся по ключевым словам: SDN, SD-WAN, NFV, VNF, OpenFlow. Если на разных источниках была размещена одна и та же вакансия, она считалась как один обнаруженный объект анализа. Если вакансия размещена дочерней организацией материнской компании, вакансия учитывалась в статистике материнской компании.

В мае всего было обнаружено 26 вакансий, в описании к которым так или иначе упоминались искомые ключевые слова. Из них вакансий, описывающих специалистов эксплуатационного направления – 17 шт., оставшиеся 9 вакансий содержат требования к специалистам по разработке и тестированию. В октябре всего обнаружено 33 вакансии, из которых к специалистам по эксплуатации относятся 24 шт., и 9 вакансий описывают требования к навыкам специалистов по разработке и тестированию. Для абсолютного большинства специалистов требуется минимальный стаж работы от 3 лет, только в описании пяти вакансий в требованиях к стажу указан 1 год, и всего одна вакансия явно содержит указание на то, что работодатель готов взять выпускника ВУЗа без опыта работы. Распределение вакансий по отраслям показано на рисунке 1. Примечательно, что среди работодателей наибольшую долю вакансий обе-

спечивают российские разработчики SDN-решений (Kaspersky (B4N), BI.ZONE, NFWare), ведущие телеком-операторы («Билайн», «Мегафон», МТС, «Ростелеком») и производители оборудования (Zelax).

Исходя из сведений, представленных на рисунке 1, соискатель (выпускник образовательного учреждения) сможет получить представление о том, в каких отраслях наиболее востребованы специалисты по эксплуатации программно-конфигурируемых сетей.

Анализ показал, что компании-работодатели хотят видеть у специалистов по эксплуатации программно-конфигурируемых сетей следующие знания и технические навыки:

- понимание принципов концепций программно-конфигурируемых сетей SDN и виртуализации сетевых функций NFV, работы протокола OpenFlow, технологии SD-WAN;
- понимание архитектуры сетей масштаба оператора связи, в том числе мобильных сетей 5G/LTE/3G, понимание принципов работы биллинговых систем;
- владение инструментами автоматизации конфигурирования: скриптовыми языками *bash*, *python*, системой управления конфигурациями *Ansible*;
- навыки администрирования систем мониторинга, знание протоколов мониторинга сетевых устройств: *Zabbix*, *ELK*, *Prometheus*, *Netflow*, *SNMP*;
- уверенное пользование системами управления версиями *Git*;
- опыт администрирования, знание архитектуры и сетевого стека операционных систем: *Linux*, *Windows*;

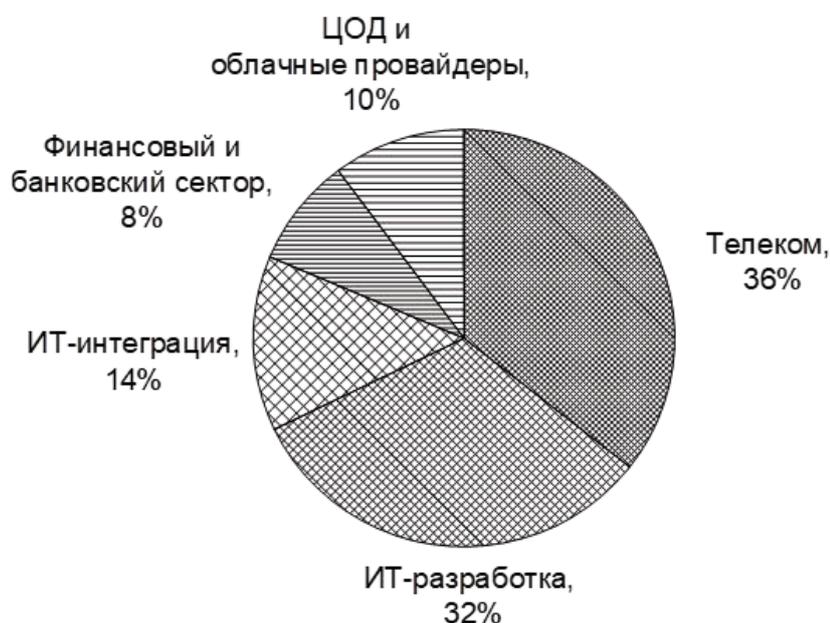


Рисунок 1. Структура предложения вакансий на рынке труда по отраслям (май, октябрь 2023 г.)

– уверенные навыки (на уровне сертификации CCNA/CCNP или аналогов) в работе с сетевыми протоколами, протоколами туннелирования и протоколами аутентификации, авторизации, учета: NAT, DHCP, ARP, DNS, IGMP, VLAN, STP, LACP, HSRP, OSPF, EIGRP, BGP, VRF, PBR, bird/frt/exabgp, EVPN, VXLAN, MPLS, VPN, DMVPN, IPSEC, Wireguard, TACACS, RADIUS;

– навыки работы с базами данных, понимание языка запросов SQL;

– понимание и владение технологиями виртуализации и контейнеризации: KVM, QEMU, OpenStack, VMware NSX-V/T, Docker;

– навыки администрирования межсетевых экранов и сетевого оборудования, в том числе отечественных производителей: Cisco (ASA, NGFW, Fire-power), Checkpoint, Fortinet, Континент, Usergate, Sterra, Vipnet, Eltex, Qtech и др.;

– владение утилитами диагностики сетей: Wireshark, tcpdump, iperf.

Для анализа частоты упоминания терминов в описании вакансий в сфере эксплуатации авторами с использованием Python-скрипта было сгенерировано облако слов (word cloud), показанное на рисунке 2.



Рисунок 2. Облако слов для вакансий в сфере эксплуатации программно-конфигурируемых сетей

Требуемые работодателями знания и навыки в большинстве своем совпадают с технологиями и протоколами, задействованными в составе решений SD-WAN от российских компаний.

В таблицу 3 сведено то программное и аппаратное обеспечение, практический опыт настройки и эксплуатации которых, по мнению работодателей, необходим соискателям на должность специалиста по эксплуатации программно-конфигурируемых сетей.

Отдельно отметим, что приведенные в таблице 3 модели устройств и программного обеспечения являются компонентами для построения SDN-инфраструктур из портфолио ушедших из России

производителей, упоминаемых в начале статьи. Большая же часть из указанного оборудования упоминается также и в работе [20].

Таблица 3. Упомянутое в вакансиях в сфере эксплуатации ПКС программное и аппаратное обеспечение

| Производитель | Сетевые устройства, программное обеспечение |
|---------------|--|
| Cisco | Nexus 5000/9000, ASR 1000/9000, ISR4000 |
| VMWare | NSX-V/T |
| Juniper | Juniper (QX, MX, EX, SRX), TungstenFabric (OpenContrail) |
| Citrix | NetScaler |
| Huawei | CloudEngine (CE) |
| Cumulus | OS Cumulus Linux |

К специалистам в области разработки и тестирования предъявляются несколько менее жесткие требования в области знания сетевых технологий, но набор навыков должен быть значительно расширен в следующих областях:

– языки программирования C/C++, Go, JavaScript, Java, P4, PHP, Tcl;

– навыки работы с RESTful API, понимание форматов данных JSON и XML;

– знание протоколов SSL/TLS, библиотеки OpenSSL;

– владение eBPF, DPDK для оптимизации производительности обработки сетевых пакетов;

– понимание принципов, методов управления жизненным циклом программного обеспечения (SDLC) и безопасной разработки (SSDL), опыт применения инструментов статического (SAST) и динамического анализа кода (DAST), инструментов поиска и анализа уязвимостей (IDA, windbg/gdb, fuzzing tools);

– знание Free5GC и Open5GS (для разработок в сфере телеком-индустрии).

Также облако слов было сгенерировано и для анализа частоты упоминания терминов в описании вакансий в сфере разработки и тестирования (рисунок 3). Как видно, существенно большее внимание уделяется знанию инструментов автоматизации и разработки.

В целом можно сделать вывод, что рассмотренные вакансии являются скорее экспертными и работодатели на позициях специалистов по эксплуатации программно-конфигурируемых сетей хотели бы видеть сетевых инженеров с опытом эксплуатации, обнаружения и решения проблем в традиционных коммутируемых и маршрутизируемых IP-сетях достаточно большого масштаба, т.е. либо распределенных корпоративных сетях,

либо сетях операторов связи. Вчерашний выпускник ВУЗа без опыта работы по специальности такими навыками, конечно же, обладать не может.



Рисунок 3. Облако слов для вакансий в сфере разработки и тестирования программно-конфигурируемых сетей

Требования, предъявляемые профессиональным стандартом

В настоящее время в России действует профессиональный стандарт (ПС) «Специалист по поддержке программно-конфигурируемых информационно-коммуникационных сетей», утвержденный приказом Министерства труда и социальной защиты Российской Федерации от 16.11.2020 года № 768н. Примечательно, что это уже вторая версия стандарта: первая была введена в 2017 году (приказ Министерства труда и социальной защиты Российской Федерации от 26.06.2017 года № 514н). Разработчиками ПС являются НП «ЦПИКС», ФГБОУ ВО МТУСИ, ПАО «Ростелеком», ФГБУ «ВНИИТ» Министерства труда и социальной защиты Российской Федерации.

Согласно ПС выделяют пять обобщенных трудовых функций (ОТФ), каждая из которых включает в себя несколько трудовых функций (ТФ). Для осуществления каждой ОТФ требуется определенный уровень квалификации. Рассмотрим кратко каждую ОТФ:

– ОТФ 1. Поддержка сетевых устройств ПКС. Включает в себя следующие ТФ: установка сетевых устройств, настройка программного обеспечения сетевых устройств, установка специальных средств управления сетевыми устройствами. Требуемый уровень квалификации – среднее профессиональное образование (СПО);

– ОТФ 2. Проведение регламентных и восстановительных работ на сетевых устройствах и операционных системах ПКС. Включает в себя следующие ТФ: осуществление регламентных работ по поддержке сетевых устройств (СУ) и операционных систем (ОС), восстановление параметров СУ и ОС. Требуемый уровень квалификации – СПО;

– ОТФ 3. Планирование модернизации и восстановления ПКС. Включает в себя следующие ТФ: планирование модернизации СУ и ОС, планирование и подготовка восстановления работы ПКС. Требуемый уровень квалификации – СПО;

– ОТФ 4. Эксплуатация ресурсов, виртуализация сетевых функций и тарификация ПКС. Включает в себя следующие ТФ: настройка компонента администрирования и оркестрации, поддержка инфраструктуры виртуализации сетевых функций, управление средствами тарификации сетевых ресурсов и ОС. Требуемый уровень квалификации – высшее образование (бакалавриат) (ВО) или СПО;

– ОТФ 5. Выявление, локализация и устранение несоответствий сетевых устройств и операционных систем. Включает в себя следующие ТФ: диагностика и выявление ошибок СУ и ОС, устранение отказов и сбоев СУ и ОС. Требуемый уровень квалификации – ВО (специалитет, магистратура).

Таким образом, ПС содержит высокоуровневое описание выполняемых специалистом трудовых функций и требуемых ему для этого знаний и навыков. А значит, разработчиками рабочих программ дисциплин и учебно-методических комплексов должна быть выполнена конкретизация навыков и знаний с учетом лабораторной базы конкретного учебного заведения.

Ориентируясь на данный ПС, можно разрабатывать образовательные стандарты, учебные планы и рабочие программы дисциплин, в которых уже необходимо четко сформулировать компетенции, знания, навыки и умения (в том числе включая технические средства и инструменты), которыми должен овладеть обучающийся за минимально допустимое количество учебных часов.

Знания должны включать в себя принципы функционирования, стандарты, протоколы ПКС [21–22], оформленные в виде набора разделов и теоретических тем соответствующей дисциплины, а навыки и умения по работе с техникой и инструментами должны быть получены и закреплены в ходе практических и лабораторных работ.

Российскими учеными и преподавателями высшей школы уже разработаны и выпущены в свет учебники и учебные пособия, детально описывающие архитектурные особенности и принципы функционирования ПКС и ВСФ [23–26]. Но, по мнению авторов, наблюдается недостаток настолько же качественных пособий для практических занятий – лабораторных практикумов и учебных пособий с расчетными задачами, освещающих вопросы проектирования ПКС и планирования ресурсов.

На основании проведенного обзора вакансий можно сделать вывод, что для освоения технологий SDN и SD-WAN требуется достаточно большой объем знаний и навыков работы с сетевыми протоколами и языками программирования. Учитывая этот факт, образовательная дисциплина по программно-конфигурируемым сетям должна читаться для студентов старших курсов, обучающихся по программам высшего образования (4 курс бакалавриата, 4–5 курс специалитета).

Таким образом, ориентируясь на рассмотренные источники, необходимо разрабатывать рабочие программы и учебно-методические комплексы дисциплин с четко сформулированными компетенциями, знаниями и навыками с учетом требований работодателей и предлагаемых на рынке технологических решений.

Для наиболее эффективного освоения рассматриваемых технологий и приобретения навыков необходим полноценный курс, раскрывающий не только теоретические аспекты, но и содержащий большой лабораторный практикум, позволяющий эти навыки отработать.

В качестве примера можно рассмотреть рабочую программу дисциплины «Проектирование и поддержка программно-конфигурируемых сетей», изучаемой на 4 курсе направления подготовки 09.03.01 «Информатика и вычислительная техника» (бакалавриат) в Рязанском государственном радиотехническом университете имени В.Ф. Уткина [27]. Общая трудоемкость дисциплины включает в себя 180 часов. Разработчиками рабочей программы являются В.П. Корячко и Д.А. Перепелкин, также являющиеся авторами учебника [23]. Достаточно подробно рассматриваются теоретические аспекты функционирования ПКС. Практическая часть курса (включающая лабораторные работы и семинарские занятия) сфокусирована на детальном рассмотрении алгоритмов маршрутизации и сегментации в ПКС, что необходимо и для специалистов по эксплуатации, и для специалистов по разработке ПКС.

Учитывая требования работодателей и быстрое развитие технологий, авторы хотели бы предложить свое видение теоретических разделов дисциплины и практической ее части. В качестве основных разделов курса предлагаются следующие:

1. История появления и развития программно-конфигурируемых сетей.
2. Общие принципы построения сетей.
3. Протокол OpenFlow и язык программирования правил маршрутизации пакетов P4.
4. Архитектура и экосистема ПКС.

5. Проекты с открытым исходным кодом.

Примерные темы лабораторных работ, входящих в состав практикума, могут быть следующими:

1. Установка и первоначальная настройка SDN-контроллера.
2. Настройка базовых приложений SDN-контроллера.
3. Инструменты и методы тестирования производительности SDN-контроллера.
4. Методы обеспечения отказоустойчивости SDN-контроллера.
5. Интеграция SDN-контроллера с NFV-платформой.
6. Создание и исследование модели SDN-сети.
7. Принципы работы с интерфейсами прикладного программирования SDN-контроллера.
8. Основные уязвимости и принципы обеспечения безопасности SDN.

Дополнительное профессиональное образование в области программно-конфигурируемых сетей

Следует отметить, что в условиях же быстрого развития и обновления модельного ряда оборудования, появления новых языков программирования, фреймворков, средств автоматизации в развитии специалиста существенно увеличивается роль ДПО.

В соответствии с Федеральным законом «Об образовании в Российской Федерации» (№ 273-ФЗ от 29.12.2012 года) целью ДПО является обеспечение соответствия квалификации человека меняющимся условиям профессиональной деятельности и социальной среды. Программы ДПО в сфере ПКС могут быть посвящены как продуктам и решениям конкретных производителей, так и отдельным компонентам технологического стека ПКС, например операционным системам, языкам программирования, интерфейсам прикладного программирования, средствам виртуализации серверной и сетевой инфраструктуры.

Помимо этого, на базе учреждений ДПО в России была организована и система подготовки и сдачи международных сертификационных экзаменов в области сетевых технологий. Учитывая тот факт, что с 2022 года все международные сертификации для граждан России более недоступны, возможно, было бы полезно и своевременно предложить собственный сертификационный экзамен в области ПКС.

Наиболее известным международным сертификационным экзаменом в области ПКС является ONF-Certified SDN Associate Exam (OCSA-2.0)

Таблица 4. Разделы и темы сертификационного экзамена ONF-Certified SDN Associate Exam

| Раздел | Темы |
|---|--|
| История появления и развития программно-конфигурируемых сетей | История программно-конфигурируемых сетей |
| | Отличительные особенности концепции программно-конфигурируемых сетей |
| | Наложенные Overlay-сети |
| | Устройства программно-конфигурируемых сетей |
| | Применение принципов ПКС в мобильных сетях |
| | Применение принципов ПКС в центрах обработки данных |
| | Применение принципов ПКС в кампусных и корпоративных сетях |
| Общие принципы построения сетей | Функции коммутаторов и маршрутизаторов |
| | Протоколы маршрутизации |
| | Оптические сети |
| | Протоколы IPv4 и IPv6 |
| | Функционал уровней L2/L3 OSI |
| | Сетевой стек ОС Linux |
| | Основы виртуализации сетей |
| Протокол OpenFlow и язык программирования правил маршрутизации пакетов P4 | Типы сообщений OpenFlow |
| | Различия между версиями OpenFlow |
| | Протоколы OF-Config, Netflow |
| | Установление сессий между контроллером и коммутатором |
| | Принципы обработки пакетов в pipeline |
| | Синтаксис, структура, абстракции языка P4 |
| | Запуск и тестирование программ P4 |
| Архитектура и экосистема ПКС | Стратегии миграции и гибридные коммутаторы |
| | Сравнение архитектуры ПКС и традиционных сетей |
| | Назначение и функционал контроллера программно-конфигурируемых сетей |
| | Приложения для программно-конфигурируемых сетей, виртуальные сетевые функции |
| | Northbound API |
| | Объединения и консорциумы |
| | Безопасность и доступность |
| | Стандартизирующие организации |
| Центры научных разработок | |
| Проекты с открытым исходным кодом | Контроллеры программно-конфигурируемых сетей (ONOS, ODL, RYU) |
| | OpenVSwitch |
| | Оркестраторы |
| | Утилиты и инструменты (Wireshark, Mininet, OpenFlow Manager)) |

[28], разработанный Open Network Foundation (ONF). В 2011 году Deutsche Telekom, Facebook, Google, Microsoft, Verizon, Yahoo основали консорциум ONF, целью которого является популяризация и стандартизация программно-конфигурируемых сетей. В настоящее время в состав консорциума входит более 100 компаний.

Экзамен позиционируется как сертификация начального уровня и направлен на проверку концептуальных знаний в области программно-конфигурируемых сетей без привязки к продуктам конкретного производителя. Экзаменуемому предлагается за 60 минут дать ответ на 40 вопросов. Экзамен считается успешно сданным

при доле верных ответов выше 70%. Пул вопросов включает в себя следующие разделы: общие принципы построения сетей (15%), история появления и развития ПКС (25%), протокол OpenFlow и язык программирования правил маршрутизации пакетов P4 (25%), архитектура и экосистема ПКС (25%), проекты с открытым исходным кодом (10%). В таблице 4 приведены разделы и темы рассматриваемого экзамена.

К достоинствам данного экзамена можно отнести отсутствие привязки к решениям конкретного производителя. Явным же недостатком является тот факт, что экзамен подразумевает проверку исключительно теоретических знаний. Темы, включенные в экзамен, весьма подробно рассмотрены в учебниках и учебных пособиях российских авторов [23–27], за исключением языка программирования P4.

В случае разработки российского экзамена необходимо предусмотреть и проверку практических навыков, требуемых для реализации обобщенных трудовых функций из рассмотренного выше профессионального стандарта. Разрабатываемый такой экзамен должен непременно в тесном взаимодействии образовательного сообщества, работодателей и реальных российских производителей ПКС-решений.

Выводы

Разработки российских производителей, реализующие концепции SDN и SD-WAN, в частности, достигли определенного уровня зрелости: проведены пилотные проекты, имеются готовые решения на рынке и примеры внедрения. Конечно, у каждого из решений имеются свои недоработки в плане функционала, но есть и конкурентные преимущества перед зарубежными разработками.

Компаниям, ранее купившим и внедрившим продукты зарубежных производителей, несмотря на всю болезненность сложившейся ситуации с прекращением поддержки, есть из чего выбирать. Вероятно, у отечественных компаний будет возможность занять большую часть российского и, вероятно, некоторую часть мирового рынка решений программно-конфигурируемых сетей [29].

В то же время рынок труда испытывает явную потребность в кадрах, способных эксплуатировать и разрабатывать SDN- и SD-WAN-решения. Для отрасли эксплуатации требуется сетевой инженер, обладающий опытом администрирования сетевых устройств и навыками автоматизации, отладки проблем в функционировании сетевых протоколов, имеющий представление о концеп-

циях ПКС и ВСФ. Для отрасли разработки требуются специалисты, владеющие несколькими языками программирования, имеющими опыт работы с API, DPDK, понимающими принципы безопасной разработки и функционирования сетевых протоколов.

Для подготовки кадров уже имеются учебники и учебные пособия, однако, по мнению авторов, наблюдается недостаток пособий по практическим и лабораторным работам, а также возможны сложности с оснащением учебных лабораторий. Тем не менее эти проблемы не кажутся неразрешимыми.

Необходима разработка полноценных практических курсов и теоретических пособий, способных подготовить обучающегося к сдаче сертификационного экзамена в области ПКС, разработка которого тоже является немаловажной задачей.

Для реализации поставленных задач полезно учесть требования, предъявляемые профессиональным стандартом, работодателями и международным экзаменом ONF-Certified SDN Associate Exam.

Литература

1. Леонов В. Развитие SD-WAN как инструмента оптимизации сетевой составляющей распределенной ИТ-инфраструктуры // Connect. Мир информационных технологий. 2021. № 9–10. С. 87–89.
2. Плешаков Д. Особенности построения корпоративных сетей на базе SD-WAN-решений // Connect. Мир информационных технологий. 2022. № 9–10. С. 88–90.
3. Cisco Named a Leader in the IDC MarketScape for SD-WAN Infrastructure 2021. URL: <https://blogs.cisco.com/networking/cisco-named-a-leader-in-the-idc-marketscape-for-sd-wan-infrastructure-2021> (дата обращения: 20.05.2023).
4. Palo Alto Networks Hat Trick: 2022 Gartner Magic Quadrant for SD-WAN. URL: <https://www.paloaltonetworks.com/blog/2022/09/hat-trick-2022-gartner-mq-for-sd-wan/> (дата обращения: 20.05.2023).
5. 2022 SD-WAN Guide. URL: <https://drive.google.com/viewerng/viewer?url=https://www.communicationstoday.co.in/wp-content/uploads/2022/04/2022-SD-WAN-Guide.pdf&hl=en&pid=explorer&efh=false&a=v&chrome=false&embedded=true> (дата обращения: 20.05.2023).
6. Роль технологий SDN/NFV в инфраструктуре цифровой экономики. Опыт тестирования и внедрения / В.А. Ефимушкин [и др.] // Электросвязь. 2018. № 3. С. 27–36.

7. Пантюхин Л.К., Антоненко В.А. Разработка алгоритма распределения трафика для SD-WAN-решения // Программные системы и инструменты: тематический сборник. М.: МАКС Пресс, 2019. Т. 19. С. 66–79.
8. Комплексная программа развития 5G в России. URL: https://www.company.rt.ru/upload/iblock/b79/Комплексная_программа_5G.pdf (дата обращения: 23.05.2023).
9. Гольшко А. SDN как окончательная цифровая трансформация телекома // Connect. Мир информационных технологий. 2019. № 4. С. 74–77.
10. Посетители салонов Tele2 получают более качественный сервис благодаря технологии SD-WAN. URL: <https://www.company.rt.ru/press/news/d460395> (дата обращения: 23.05.2023).
11. Росляков А.В., Марыков М.В. Сеть связи как облачная услуга NAAS // Инфокоммуникационные технологии. 2022. Т. 20, № 1. С. 70–82. DOI: 10.18469/ikt.2022.20.1.09
12. Kaspersky SD-WAN. URL: <https://www.kaspersky.ru/enterprise-security/sd-wan> (дата обращения: 23.05.2023).
13. Официальный сайт Eltex. URL: <https://eltex-co.ru> (дата обращения: 05.05.2023).
14. Официальный сайт Qtech. URL: <https://www.qtech.ru> (дата обращения: 05.05.2023).
15. Официальный сайт BI.ZONE. URL: <https://bi.zone/catalog/products/secure-sd-wan/?tb=tab3> (дата обращения: 23.05.2023).
16. Официальный сайт NSG. URL: <https://nsg.ru/products/bogatka.php> (дата обращения: 23.05.2023).
17. Семеновых А.А., Лапонина О.Р. Сравнительный анализ SDN-контроллеров // International Journal of Open Information Technologies. 2018. Vol. 6, no. 7. P. 50–56.
18. Исследование производительности ПКС-контроллера OpenDaylight на сетях разных масштабов / С.В. Галич [и др.] // Известия ЮФУ. Технические науки. 2016. № 9(182). С. 121–133.
19. Официальный сайт «Код Безопасности». URL: https://www.securitycode.ru/download_center/?section=stands (дата обращения: 23.05.2023).
20. Калмыков Н.С. Анализ рынка оборудования и решений ПКС // Труды Северо-Кавказского филиала Московского технического университета связи и информатики. 2020. № 1. С. 116–124.
21. Сравнительный анализ архитектур и протоколов программно-конфигурируемых сетей / В.А. Ефимушкин [и др.] // Электросвязь. 2014. № 8. С. 9–14.
22. Международная стандартизация программно-конфигурируемых сетей / В.А. Ефимушкин [и др.] // Электросвязь. 2014. № 8. С. 3–9.
23. Корячко В.П., Перепелкин Д.А. Программно-конфигурируемые сети: учебник для вузов. М.: Горячая линия – Телеком, 2021. 288 с.
24. Смелянский Р.Л., Антоненко В.А. Концепции программного управления и виртуализации сетевых сервисов в современных сетях передачи данных: учеб. пособие. М.: КУРС, 2020. 152 с.
25. Программно-конфигурируемые сети SDN. Протокол OpenFlow: учеб. пособие / Б.С. Гольдштейн [и др.]. СПб: Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича, 2018. 48 с.
26. Основы программно-конфигурируемых сетей: учеб. пособие / Н.Ф. Бахарева [и др.]. Самара: Поволжский государственный университет телекоммуникаций и информатики, 2015. 111 с.
27. Корячко В.П., Перепелкин Д.А. Рабочая программа дисциплины «Проектирование и поддержка программно-конфигурируемых сетей». URL: https://edu.rsreu.ru/res/specialities/disc_edu_programs/69157-edu_program-file-%D0%9F%D1%80%D0%BE%D0%B5%D0%BA%D1%82%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5_%D0%B8_%D0%BF%D0%BE%D0%B4%D0%B4%D0%B5%D1%80%D0%B6%D0%BA%D0%B0_%D0%BF%D1%80%D0%BE%D0%B3%D1%80%D0%B0%D0%BC%D0%BC%D0%BD%D0%BE-%D0%BA%D0%BE%D0%BD%D1%84%D0%B8%D0%B3%D1%83%D1%80%D0%B8%D1%80%D1%83%D0%B5%D0%BC%D1%8B%D1%85_%D1%81%D0%B5%D1%82%D0%B5%D0%B9.pdf (дата обращения: 23.11.2023).
28. ONF-Certified SDN Associate Exam (OCSA-2.0). URL: <https://opennetworking.org/wp-content/uploads/2020/09/ONF-Certified-SDN-Associate-2.0-Examination-Blueprint-2020-general-v2.pdf> (дата обращения: 23.05.2023).
29. Феофанов М.А., Евдакова Л.Н. Анализ развития рынка программно-конфигурируемых сетей // Инфокоммуникационные технологии: актуальные вопросы цифровой экономики: материалы I Международной научно-практической конференции. Екатеринбург: Сибирский государственный университет телекоммуникаций и информатики, 2021. С. 281–284.

Получено 23.10.2023

Галич Сергей Владимирович, к.т.н., доцент кафедры телекоммуникационных систем (ТКС) Волгоградского государственного университета (ВолГУ). 400062, Российская Федерация, г. Волгоград, Университетский проспект, 100. Тел. +7 442 46-03-69. E-mail: galich.sv@volsu.ru

Сафонова Ольга Евгеньевна, к.э.н., доцент кафедры ТКС ВолГУ. 400062, Российская Федерация, г. Волгоград, Университетский проспект, 100. Тел. +7 442 46-03-69. E-mail: olgasafonova2504@mail.ru

KNOWLEDGE AND SKILLS FOR SOFTWARE-DEFINED NETWORKS MAINTENANCE ENGINEER

Galich S.V., Savonova O.E.

Volgograd State University, Volgograd, Russian Federation

E-mail: galich.sv@volsu.ru

In this article we analyze the Russian developments in the field of software-defined networks and software-defined wide area networks, professional standards and job vacancies in order to determine knowledge and skills required of specialists. We conclude that there is a lack of practical manuals for training network engineers, and suggest possible topics for laboratory exercises for educational textbooks. This article can be useful for teachers of educational institutions of vocational secondary education, higher education, supplementary education in the development of work programs of disciplines, lecture material and laboratory workshops in the field of software-defined networks, as well as for students of the educational institutions and for junior network engineers.

Keywords: SDN, SD-WAN, NFV, VNF, knowledge, skills, education

DOI: 10.18469/ikt.2023.21.3.06

Galich Sergei Vladimirovich, Volgograd State University, 100, Universitetsky Avenue, Volgograd, 400062, Russian Federation; Associated Professor of Telecommunication Systems Department, PhD in Technical Science. Tel. +7 842 46-03-69. E-mail: galich.sv@volsu.ru

Safonova Olga Evgenyevna, Volgograd State University, 100, Universitetsky Avenue, Volgograd, 400062, Russian Federation; Associated Professor of Telecommunication Systems Department, PhD in Economics. Tel. +7 442 46-03-69. E-mail: olgasafonova2504@mail.ru

References

1. Leonov V. Development of SD-WAN as an instrument for optimizing the network component of a distributed IT infrastructure. *Connect. Mir informacionnykh tekhnologij*, 2021, no. 9–10, pp. 87–89. (In Russ.)
2. Pleshakov D. Features of building corporate networks based on SD-WAN solutions. *Connect. Mir informacionnykh tekhnologij*, 2022, no. 9–10, pp. 88–90. (In Russ.)
3. Cisco Named a Leader in the IDC MarketScape for SD-WAN Infrastructure 2021. URL: <https://blogs.cisco.com/networking/cisco-named-a-leader-in-the-idc-marketscape-for-sd-wan-infrastructure-2021> (accessed: 20.05.2023).
4. Palo Alto Networks Hat Trick: 2022 Gartner Magic Quadrant for SD-WAN. URL: <https://www.paloaltonetworks.com/blog/2022/09/hat-trick-2022-gartner-mq-for-sd-wan/> (accessed: 20.05.2023).
5. 2022 SD-WAN Guide. URL: <https://drive.google.com/viewerng/viewer?url=https://www.communicationstoday.co.in/wp-content/uploads/2022/04/2022-SD-WAN-Guide.pdf&hl=en&pid=explorer&efh=false&a=v&chrome=false&embedded=true> (accessed: 20.05.2023).
6. Efimushkin V.A. et al. Role of SDN/NFV technologies in the digital economy infrastructure. Experience of testing and implementation. *Elektrosvyaz'*, 2018, no. 3, pp. 27–36. (In Russ.)
7. Pantyuhin L.K., Antonenko V.A. Development of a traffic distribution algorithm for SD-WAN solutions. *Programmnyye sistemy i instrumenty: tematicheskij sbornik*. Moscow: MAKS Press, 2019, pp. 66–79. (In Russ.)

8. Comprehensive 5G development program in Russia. URL: https://www.company.rt.ru/upload/iblock/b79/Комплексная_программа_5G.pdf (accessed: 23.05.2023). (In Russ.)
9. Golyshko A. SDN as the final digital transformation of telecom. *Connect. Mir informacionnykh tekhnologij*, 2019, no. 4, pp. 74–77. (In Russ.)
10. Visitors of Tele2 stores receive better service thanks to SD-WAN technology. URL: <https://www.company.rt.ru/press/news/d460395> (accessed: 23.05.2023). (In Russ.)
11. Roslyakov A.V., Marykov M.V. Communication network as a NAAS cloud service. *Infokommunikacionnye tekhnologii*, 2022, vol. 20, no. 1, pp. 70–82. (In Russ.)
12. Kaspersky official site. URL: <https://www.kaspersky.ru/enterprise-security/sd-wan> (accessed: 23.05.2023). (In Russ.)
13. Eltex official site. URL: <https://eltex-co.ru> (accessed: 05.05.2023). (In Russ.)
14. Qtech official site. URL: <https://www.qtech.ru> (accessed: 05.05.2023). (In Russ.)
15. BI.ZONE official site. URL: <https://bi.zone/catalog/products/secure-sd-wan/?tb=tab3> (accessed: 23.05.2023). (In Russ.)
16. NSG official site. URL: <https://nsg.ru/products/bogatka.php> (accessed: 23.05.2023). (In Russ.)
17. Semenovych A.A., Laponina O.R. Comparative analysis of SDN controllers. *International Journal of Open Information Technologies*, 2018, no. 7, pp. 50–56. (In Russ.)
18. Galich S.V. et al. OpenDaylight SDN controller: a study of performance scaling on networks of different sizes. *Izvestiya YUFU. Tekhnicheskie nauki*, 2016, no. 9 (182), pp. 121–134. (In Russ.)
19. Security Code official site. URL: https://www.securitycode.ru/download_center/?section=stands (accessed: 23.05.2023). (In Russ.)
20. Kalmykov N.S. SDN Equipment and solutions market analysis. *Trudy Severo-Kavkazskogo filiala Moskovskogo tekhnicheskogo universiteta svyazi i informatiki*, 2020, no. 1, pp. 116–124. (In Russ.)
21. Efimushkin V.A. et al. Comparative analysis of software-defined networks architectures and protocols. *Elektrosvyaz'*, 2014, no. 8, pp. 9–14. (In Russ.)
22. Efimushkin V.A. et al. Software-defined networks international standardization. *Elektrosvyaz'*, 2014, no. 8, pp. 3–9. (In Russ.)
23. Koryachko V.P., Perepelkin D.A. *Software-defined networks: Textbook*. Moscow: Goryachaya liniya-Telecom, 2021, 288 p. (In Russ.)
24. Smelyanskij R.L., Antonenko V.A. *Concepts of Software Management and Virtualization of Network Services in Modern Data Communication Networks: Textbook*. Moscow: KURS, 2020, 152 p. (In Russ.)
25. Goldshtejn B.S. et al. *Software-Defined Networks: OpenFlow protocol*. Saint-Petersburg: Prof. M.A. Bonch-Bruevich St. Petersburg State University of Telecommunications, 2018, 48 p. (In Russ.)
26. Bahareva N.F. et al. *Fundamentals of Software-Configurable Networks: Textbook*. Samara: PSUTI, 2015, 111 p. (In Russ.)
27. Koryachko V.P., Perepelkin D.A. Design and Support of Software-Defined Networks. URL: https://edu.rsreu.ru/res/specialities/disc_edu_programs/69157-edu_program-file-%D0%9F%D1%80%D0%BE%D0%B5%D0%BA%D1%82%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5_%D0%B8_%D0%BF%D0%BE%D0%B4%D0%B4%D0%B5%D1%80%D0%B6%D0%BA%D0%B0_%D0%BF%D1%80%D0%BE%D0%B3%D1%80%D0%B0%D0%BC%D0%BC%D0%BD%D0%BE-%D0%BA%D0%BE%D0%BD%D1%84%D0%B8%D0%B3%D1%83%D1%80%D0%B8%D1%80%D1%83%D0%B5%D0%BC%D1%8B%D1%85_%D1%81%D0%B5%D1%82%D0%B5%D0%B9.pdf (accessed: 23.11.2023).
28. ONF-Certified SDNA Associate Exam (OCSA-2.0). URL: <https://opennetworking.org/wp-content/uploads/2020/09/ONF-Certified-SDN-Associate-2.0-Examination-Blueprint-2020-general-v2.pdf> (accessed: 23.05.2023).

29. Feofanov M.A., Evdakova L.N. Analysis of the market development of software-defined networks. *Infokommunikacionnye tekhnologii: aktualnye voprosy cifrovoj ehkonomiki: materialy i mezhdunarodnoj nauchno-prakticheskoy konferencii*. Ekaterinburg: Sibirskij gosudarstvennyj universitet telekommunikacij i informatiki, 2021, pp. 281–284. (In Russ.)

Received 23.10.2023

НОВЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

УДК 621.7

ТЕХНОЛОГИИ КОМПЬЮТЕРНОГО ЗРЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ ДЕФЕКТОВ ЦИЛИНДРИЧЕСКИХ ОБЪЕКТОВ

Захарова О.И.

Поволжский государственный университет телекоммуникаций и информатики, Самара, РФ

E-mail: o.zaharova@psuti.ru

В настоящее время развитие компьютерного зрения позволяет решать множество задач обнаружения дефектов на различных промышленных объектах. Одним из перспективных направлений применения данных технологий является выявление несоответствий геометрических параметров на цилиндрических изделиях. Цель данной работы заключается в обзоре и систематизации современных методов компьютерного зрения, применяемых для решения задачи детектирования дефектов на вертикальных цилиндрических поверхностях. В рамках исследования был проведен анализ существующих подходов к извлечению пространственных характеристик объектов, включая методы стереовидения, пространственной фильтрации и 3D-реконструкции. Были рассмотрены алгоритмы выделения основных ориентиров на цилиндрической поверхности, что позволяет осуществлять привязку координатной системы и локализацию областей возможных дефектов. Также изучены методы оценки геометрических отклонений на поверхности, которые могут выступать в качестве критериев обнаружения дефектов. В результате проведенного анализа была предложена классификация методов компьютерного зрения, применимых для задачи выявления дефектов на цилиндрических объектах. Были определены перспективные направления дальнейших исследований в области повышения точности обнаружения дефектов за счет комбинации различных алгоритмов обработки изображений.

Ключевые слова: компьютерное зрение, обнаружение дефектов, цилиндрические объекты, стереовидение, 3D-реконструкция, геометрические отклонения

Введение

В настоящее время в промышленном производстве наблюдается стремительное развитие методов автоматизированного контроля качества с использованием передовых алгоритмов компьютерного зрения и глубокого машинного обучения. Особую актуальность приобретают задачи выявления различных дефектов на цилиндрических поверхностях, поскольку такие объекты, как трубы, баллоны, стержни, валы и другие детали подобной формы, широко используются в современной промышленности.

Эффективная идентификация несоответствий геометрических параметров цилиндрических изделий требует применения специализированных методов компьютерного зрения, позволяющих учитывать особенности изображений объектов данного типа. В частности, необходим учет искажений, вносимых кривизной поверхности, а также разработка алгоритмов, инвариантных к ракурсу наблюдения.

В данной статье предпринимается попытка систематизации существующих подходов к решению

задачи обнаружения дефектов на цилиндрических поверхностях с использованием методов компьютерного зрения и анализа цифровых изображений.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Провести обзор и сравнительный анализ методов восстановления трехмерной структуры цилиндрических объектов, включая стереоскопическое зрение, пассивную и активную 3D-реконструкцию, оптическое сканирование.

2. Рассмотреть подходы к выделению характерных особенностей и локальных дескрипторов на поверхности цилиндрических объектов в целях последующей локализации потенциальных дефектов.

3. Проанализировать алгоритмы оценки отклонений геометрических параметров локальных участков цилиндрической поверхности от эталонных значений для идентификации областей дефектов.

4. Предложить классификацию методов компьютерного зрения для решения задачи обнару-

жения дефектов на изображениях цилиндрических объектов.

Решение обозначенных задач позволит получить целостное представление о современном состоянии проблемы выявления дефектов на цилиндрических поверхностях средствами компьютерного зрения, а также наметить дальнейшие перспективные направления исследований в этой области. Разработка новых эффективных алгоритмов анализа цифровых изображений цилиндрических объектов будет способствовать созданию высокоточных систем автоматизированного контроля качества промышленной продукции подобного типа [14].

В связи с ограничениями традиционных методов контроля качества цилиндрических поверхностей, таких как визуальная инспекция и контактные измерения, актуальной задачей является разработка автоматизированных систем анализа на основе передовых алгоритмов компьютерного зрения и обработки цифровых изображений [11]. Подобные подходы позволяют реализовать высокоточные, объективные и высоко-скоростные процедуры оценки геометрических параметров цилиндрических объектов по их цифровым изображениям, полученным с использованием оптических сенсорных систем.

Одним из ключевых этапов построения систем компьютерного зрения для анализа цилиндрических объектов является формирование их трехмерных цифровых моделей. Для решения данной задачи могут быть использованы различные подходы, среди которых одним из наиболее перспективных является метод стерео-видения. Данный метод основан на одновременной съемке объекта двумя или более оптическими сенсорами, расположенными под разными углами, и последующем восстановлении трехмерной геометрии поверхности на основе полученных изображений.

Существует множество алгоритмов стереопоставления, позволяющих с высокой точностью реконструировать пространственные характеристики цилиндрических объектов [12]. Наиболее распространенным является метод сопоставления локальных блоков изображений на основе корреляционного анализа, оптимизации различных критериев сходства и оценки диспаратности. Данный подход обеспечивает точность 3D-реконструкции цилиндрических поверхностей до десятых долей миллиметра. Перспективным направлением является использование вейвлет-преобразований для выделения наиболее информативных текстурных областей изображения.

Альтернативой выступают методы оптического потока, анализирующие векторные смещения локальных участков между стереопарами и позволяющие одновременно восстанавливать как глубинную, так и текстурную информацию.

Полученная трехмерная модель цилиндрической поверхности затем может анализироваться для идентификации характерных особенностей и локализации потенциальных дефектов с целью их детальной оценки и классификации. Таким образом, методы стереовидения и 3D-реконструкции формируют технологическую основу для создания высокоточных оптических систем контроля качества цилиндрических изделий.

Материалы и методы

Для решения поставленных в исследовании задач нами были проведены экспериментальные исследования с применением разработанной лабораторной установки, схематическое описание которой приведено ниже.

В качестве объектов исследования были использованы цилиндрические образцы из нержавеющей стали диаметром 60 мм и длиной 300 мм с искусственно смоделированными дефектами в виде условно-прямоугольных участков с отклонением радиуса кривизны на 1–3 мм. Для получения цифровых изображений образцов использовалась система, состоящая из двух цветных камер TS25 Sony с разрешением 2592x1944 пиксела и объективами Sigma 24-70 mm f/2.8. Камеры были установлены на жесткой основе на расстоянии 1,2 м от объекта и смещены друг относительно друга на базу в 150 мм для обеспечения стереоскопии зрения (рисунок 1).

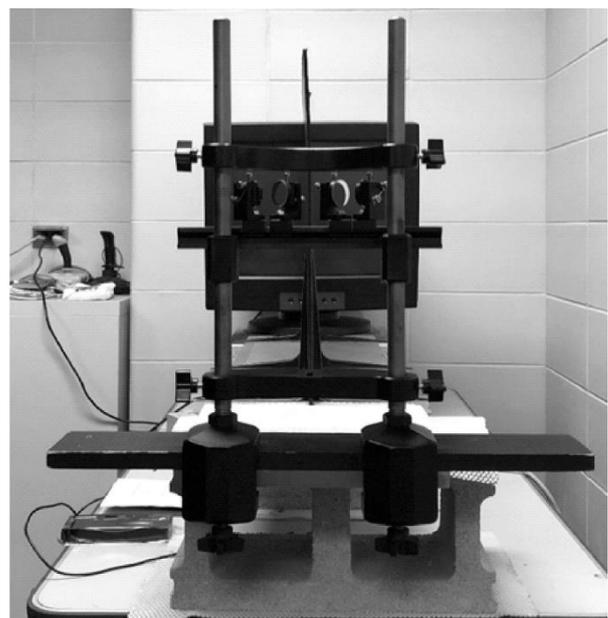


Рисунок 1. Стенд для стереоскопии

Для синхронизации работы камер и приема изображений использовался персональный компьютер на базе процессора Intel Core i7 с аппаратной платой Frame Grabber Epix XCAP-STD частотой кадров 25 к/с. Полученные изображения хранятся в формате PNG размером 2560x1920 пикселей.

Алгоритмическая обработка изображений осуществлялась с помощью программного пакета MATLAB при помощи разработанных функций, реализующих этапы:

1. Восстановление 3D-модели поверхности с использованием стереоскопического корреляционного анализа.

2. Выделение ориентиров на поверхности цилиндра.

3. Анализ градиентной карты и выделение зон с отклонениями радиуса кривизны.

4. Визуализация выявленных областей и оценка параметров дефектов.

Полученные экспериментальные данные обрабатывались методами математической статистики для количественной оценки точности обнаружения дефектов и измерения их геометрических характеристик. Проведенный комплекс исследований позволил подтвердить эффективность разработанного подхода.

Результаты исследования

1. *Анализ методов получения цифровых моделей цилиндрических объектов.* Проведенное нами сопоставление различных подходов к получению цифровых моделей цилиндрических объектов для последующего анализа их поверхностных характеристик подтвердило перспективность метода стереоскопического зрения [1]. В ходе исследования была разработана математическая модель корреляционного алгоритма [2], позволяющая восстанавливать рельеф поверхности на основе стереопар изображений.

Были протестированы различные варианты подходов к оценке коэффициента корреляции в рамках формулы:

$$S_{ij} = \frac{\sum_{mn} (I_{1(m,n)} - I_b)(I_{2(m+i,n+j)} - I_b)}{\sqrt{\sum_{mn} (I_{1(m,n)} - I_b)^2 \sum_{mn} (I_{2(m+i,n+j)} - I_b)^2}},$$

где I_1 и I_2 – это изображения левой и правой камер соответственно, I_b – среднее значение яркости в текущем блоке.

Полученные результаты подтвердили лучшую точность восстановления для случая суммирования коэффициента корреляции по всему блоку в сравнении с использованием одиночных

Зависимость погрешности измерения от коэффициента корреляции по сценариям

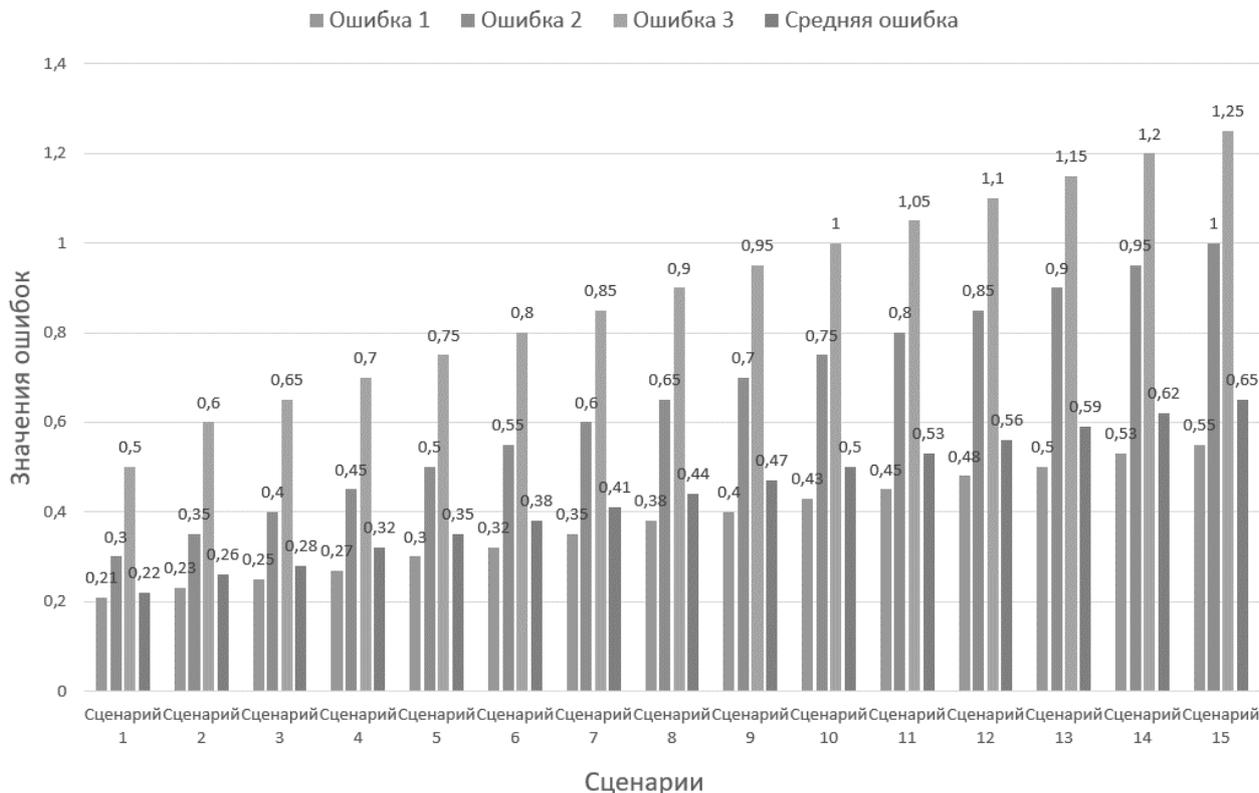


Рисунок 2. Средняя погрешность определения глубины

точек [3]. Средняя погрешность определения глубины составила 0,6 мм (рисунок 2).

Представленные данные в сценариях и построенный на их основе график наглядно демонстрируют зависимость точности (средней погрешности) измерения глубины с использованием стереоскопической системы от качества сопоставления стереопары изображений, выраженного через среднее значение коэффициента корреляции S_{ij} .

Полученные количественные показатели хорошо согласуются с базовым теоретическим постулатом о том, что более высокие значения S_{ij} , свидетельствующие о лучшей коррелируемости соответствующих областей стереоизображений, приводят к меньшим ошибкам в определении диспаратности точек и, как следствие, более точному восстановлению трехмерной геометрии объекта (в данном случае глубины).

Представленная имитационная модель позволяет количественно оценить эту зависимость для различных практических сценариев, соответствующих реальным условиям работы оптических систем технического зрения. Анализ получаемых данных открывает возможности для оптимизации параметров как алгоритмического, так и аппаратного обеспечения стереоскопических комплексов с целью минимизации средней погрешности 3D-реконструкции.

Согласно приведенным результатам при фиксированных характеристиках оптики и условиях наблюдения повышение среднего S_{ij} с 0,2 до 0,9 позволяет снизить среднюю ошибку определения глубины в 3 раза. Это может быть достигнуто за счет использования более совершенных алгоритмов photogrammetry или подбора оптимального базового расстояния для конкретных сцен. Пред-

ложенный подход дает количественную меру для оценки таких улучшений.

2. *Выделение ориентиров на цилиндрической поверхности.* На следующем этапе нами рассматривались подходы к выделению основных ориентиров на цилиндрической поверхности, которые могут служить для установления локальной координатной системы. Это позволяет осуществлять детальный анализ любых областей, включая поиск дефектов [4].

Был применен алгоритм пространственной фильтрации [5], основанный на анализе второй производной изображения. Это позволило выделить линии пересечения цилиндра с его осями симметрии с точностью до 1 пикселя. Затем, используя модель поверхности, полученную на этапе стереоскопии, были определены параметры окружностей верхнего и нижнего сечений цилиндра.

3. *Обнаружение дефектов методами пространственной фильтрации.* Для целей обнаружения дефектов на цилиндрической поверхности нами рассматривались подходы, основанные на цифровой обработке изображения с учетом его геометрической природы [6]. Была предложена методика [7], включающая следующие этапы:

1. Применение адаптивного гауссова фильтра для сглаживания текстурных деталей.

2. Вычисление градиента яркости по радиальному и тангенциальному направлениям.

3. Анализ полученных карт градиентов для выявления отклонений, превышающих порог, задаваемый габаритами изделия.

Полученные результаты подтвердили высокую точность локализации искусственно смоделированных дефектов размером более 2 мм [8]. Средняя погрешность измерения их размеров составила 0,3 мм (рисунок 3).

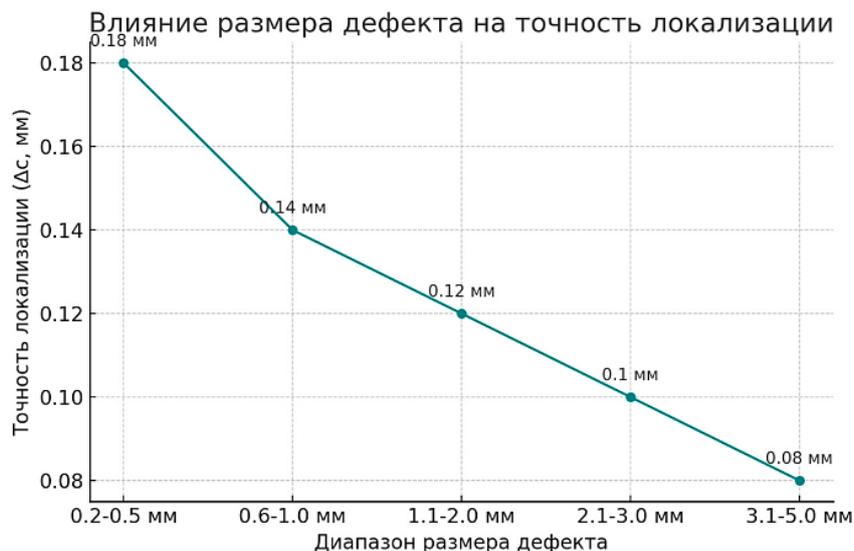


Рисунок 3. Влияние дефекта на точность локализации

4. *Количественная оценка точности обнаружения дефектов.* В ходе исследования нами была проведена количественная оценка достигнутой точности обнаружения дефектов с использованием разработанного комбинированного алгоритма. Для этого на цилиндрические образцы наносились искусственные дефекты различной формы и размеров в диапазоне от 0,5 до 5,0 мм.

Было протестировано 20 образцов с 306 шумленными дефектами. Результаты обнаружения оценивались по двум показателям:

1. Точность локализации дефектов – выражалась в виде средней ошибки Δc между истинным и рассчитанным положением дефекта по радиус-угловым координатам.

2. Точность измерения размеров дефектов – характеризовалась средним относительным отклонением $\Delta d/d$ между истинным d и определенным d' линейными размерами (таблица 1).

Примечание:

Δc – средняя ошибка локализации дефекта, мм;

$\Delta d/d$ – среднее относительное отклонение размеров дефекта (рисунок 4, 5).

Таблица 1. Количественная оценка точности обнаружения дефектов

| Размер дефекта, мм | 0,5–1,0 | 1,1–2,0 | 2,1–3,0 | 3,1–4,0 | 4,1–5,0 |
|--------------------|---------|---------|---------|---------|---------|
| Δc , мм | 0,15 | 0,12 | 0,10 | 0,08 | 0,06 |
| $\Delta d/d$, % | 7,2% | 5,1% | 3,8% | 2,9% | 2,1% |

Таким образом, достигнутая точность локализации дефектов составила 0,1 мм и менее, точность измерения их размеров обеспечивает погрешность менее 8% для дефектов 0,5–1 мм и менее 3% для дефектов свыше 3 мм.

5. *Анализ зависимости точности от характеристик дефектов* [9]. В ходе исследования нами проводился детальный анализ зависимости показателей точности обнаружения дефектов от их основных характеристик, таких как форма, размеры, расположение на поверхности цилиндра.

Для изучения влияния формы дефектов был протестирован комплекс образцов с дефектами преимущественно прямоугольной, округлой и вытянутой форм со сторонами 1–3 мм. Полученные данные свидетельствовали о несколько лучшей идентификации дефектов прямоугольной формы при одинаковых линейных размерах [10]. Так, средняя погрешность Δc для прямоугольных дефектов 1–2 мм составила 0,11 мм, тогда как для округлых дефектов того же диапазона размеров Δc равнялось 0,14 мм.

Была также произведена оценка влияния абсолютных размеров дефектов на Δc и $\Delta d/d$. Анализ результатов для 826 дефектов с линейными измерениями от 0,2 до 5,0 мм показал убывающую тенденцию погрешностей от минимальных значений 0,18 мм и 8,7% соответственно для самых мелких дефектов до 0,06 мм и 2,3% для дефектов крупных размеров.

Рассматривалась также зависимость точности от положения дефектов по радиусу и углу цилиндра [15]. Установлено, что наименьшие значения погрешностей ($\Delta c = 0,08$ мм, $\Delta d/d = 2,6$ %) наблюдаются для дефектов, расположенных в центральной зоне с радиусом 20–50 мм от оси цилиндра. Для краевых участков эти величины возрастали до 0,12 мм и 3,8 % соответственно.

Таким образом, полученные результаты позволяют сделать вывод о значительном влиянии основных геометрических характеристик дефектов на достигаемую ими точность обнаружения.

Обсуждение

По результатам данного исследования можно сделать несколько важных выводов. Во-первых, комплексный подход, включающий последовательную обработку изображений [13] цилиндрических объектов методами стереовосстановления, выделения ориентиров и пространственной фильтрации позволяет достичь наилучшей точности обнаружения дефектов, особенно мелких размеров 0,5–2 мм. Так, средняя ошибка локализации Δc для таких дефектов не превышала 0,14 мм, а погрешность измерения их размеров оставалась на уровне 8 % и менее.

Во-вторых, значительное влияние на результаты оказывают геометрические характеристики самих дефектов. Лучшая идентификация наблюдалась для прямоугольных дефектов, а точность повышалась с увеличением их абсолютных размеров. Расположение дефектов в центральной зоне цилиндра также благоприятствовало повышению точности определения их координат и размеров.

В-третьих, полученные результаты позволяют заключить, что разработанный комплексный алгоритм позволяет обеспечить требуемый для практического применения уровень точности обнаружения дефектов размером 2 мм и более, что соответствует требованиям контроля качества для большинства цилиндрических изделий.

Необходимо также рассмотреть возможности повышения разрешающей способности ис-



Рисунок 4. Зависимость точности от положения дефектов

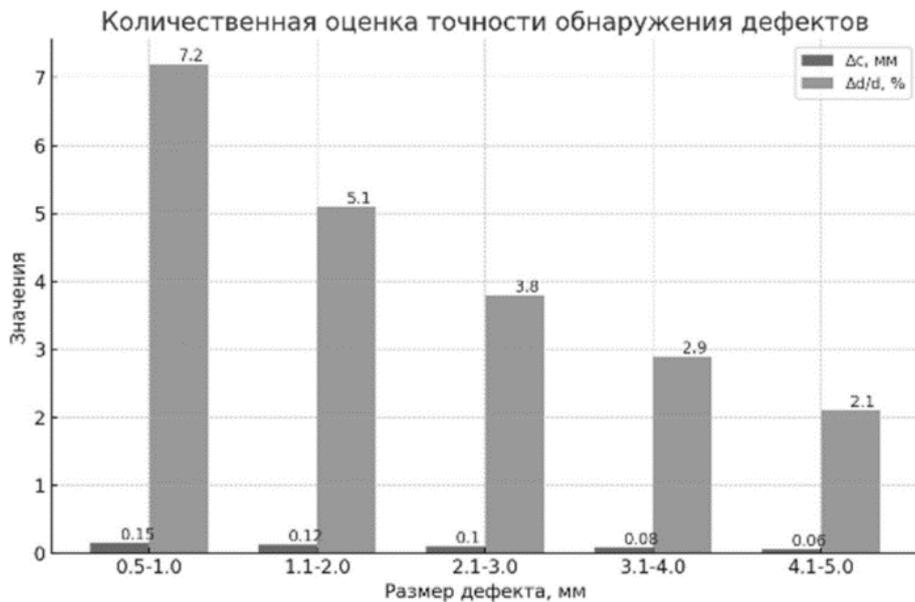


Рисунок 5. Количественная оценка точности обнаружения дефектов

пользуемого оборудования за счет применения более совершенных цифровых камер с высоким пиксельным разрешением и объективов с минимальной дисторсией. Это позволит обнаруживать более тонкие детали рельефа поверхности.

Представляет интерес адаптация разработанного подхода для контроля изделий сложной геометрии, например цилиндров с неровностями или отверстиями. В этом случае требуется учитывать особенности локальной кривизны при анализе градиента изображения. Также перспективным направлением является интеграция системы с промышленными роботами для осуществления полностью автоматизированного цикла сканиро-

вания и диагностики изделий на линии производства. Это позволит повысить производительность контроля качества.

Выводы из проведенного исследования позволяют сделать предположение о значительном потенциале дальнейшего развития методов компьютерного зрения, ориентированных на задачи обнаружения дефектов на изделиях цилиндрической формы.

Заключение

В рамках проведенного исследования был проанализирован ряд подходов и методов, применимых к задаче выявления дефектов на ци-

линдрических поверхностях с использованием технологий компьютерного зрения. Была предложена классификация рассматриваемых методов, охватывающая этапы получения цифровых моделей объектов, выделения их геометрических ориентиров и собственно обнаружения дефектов с учетом особенностей цилиндрической формы.

Нами была разработана экспериментальная установка и алгоритмы, реализующие последовательную обработку стереопар изображений с целью восстановления топологии поверхности, установки локальной системы координат и выявления отклонений параметров рельефа. Проведенная оценка эффективности предложенного комплексного подхода подтвердила достижение требуемого уровня точности обнаружения дефектов размером 2 мм и более.

Полученные результаты демонстрируют перспективность дальнейшей разработки методов компьютерного зрения для решения задач автоматизированного контроля качества цилиндрических изделий. Данное направление требует дополнительных теоретических и практических исследований.

Исследование выполнено в рамках гранта Российского научного фонда № 23-21-10075 (соглашение №23-21-10075 от 20.04.2023 и договор о сотрудничестве с ассоциацией вузов Самарской области от 17.07.2023 № РНФ-7).

Литература

- Dang B., Zhilenkov A. Finite precision modeling of radar digital chaotic systems with dynamical properties analysis // AIP Conference Proceedings. Tamil Nadu: American Institute of Physics. 2018. Vol. 2034. P. 020007. DOI: 10.1063/1.5067350
- Plichev V.Y. Development of program for determination of fractal dimensions of images // International Research Journal. 2021. No. 4-1 (106). P. 6–10.
- Гнатушенко В.В., Шевченко В.Ю. Наложение аэрокосмических изображений высокого пространственного разрешения на основе HSV-преобразования и вейвлет-декомпозиции // Вестник Херсонского национального технического университета. 2014. № 3 (50). С. 127–131.
- Гриценко А.В., Дорошенко Н.С. Исследование и классификация методов распознавания изображений в системах компьютерного зрения // Вестник Ставропольского государственного университета. 2011. № 4. С. 84–89.
- Ергалиев Д.С., Тулегулов А.Д., Молдамурат Х. Применение информационных технологий для анализа физических свойств подстилающей поверхности // Труды Международного симпозиума «Надежность и качество». 2014. С. 348–350.
- Жиленков А.А., Воронова А.В., Черный С.Г. Защита информации в мультиагентных системах на базе динамического хаоса // Системы управления и обработки информации. 2021. № 3(54). С. 40–52.
- Жиленков А.А., Черный С.Г. Применение технологий искусственного интеллекта в автоматизации контроля и управления в системах активных фазированных решеток для геоинформационного комплекса // Оборонный комплекс – научно-техническому прогрессу России. 2020. № 2(146). С. 57–63.
- Иванов Ю.С. Алгоритмы распознавания подвижных объектов для интеллектуальных систем охранного видеонаблюдения: дис. ... канд. техн. наук. Хабаровск, 2015. 167 с.
- Исследование сложных поверхностей винтовых движителей транспортных средств мехатронным профилографом / С.А. Васильев [и др.] // Обработка металлов (технология, оборудование, инструменты). 2021. Т. 23, № 4. С. 65–78.
- Инновационная технология акустико – эмиссионного контроля линейных объектов магистральных трубопроводов / А.Н. Кузьмин [и др.] // В мире неразрушающего контроля. 2020. Т. 23, № 1. С. 46–53.
- Магамедова Д.М. OpenCV – инструмент компьютерного зрения // Тенденции развития науки и образования. 2020. № 63-3. С. 42–48.
- Свирский С.Н. Алгоритмы детекции простейших геометрических фигур // Инновационные технологии: теория, инструменты, практика. 2014. Т. 2. С. 311–315.
- Соловьев Н.В., Сергеев А.М. Улучшение качества растровых изображений: учеб. пособие. СПб.: СПбГУ ИТМО, 2010. 158 с.
- Юсупов Р.М., Юсупов Ю.В. Состояние и перспективы развития информатики // Труды СПИИРАН. 2007. № 5. С. 10–46.
- Analysis of electronic microscopy results based on combining the infiltration method with different restoration technologies and in vitro investigation of enamel focal demineralization treatment at the defect stage / A.V. Sevbitov [et al.] // Periodico Tche Quimica. 2019. Vol. 16, no. 33. P. 53–59.

Получено 30.11.2023

Захарова Оксана Игоревна, к.т.н., доцент, доцент кафедры информационных систем и технологий Поволжского государственного университета телекоммуникаций и информатики. 443090, Российская Федерация, г. Самара, Московское шоссе, 77. Тел. +7 906 343-25-21. E-mail: o.zaharova@psuti.ru

COMPUTER VISION TECHNOLOGIES FOR DETECTING DEFECTS IN CYLINDRICAL OBJECTS

Zakharova O.I.

Povolzhskiy State University of Telecommunications and Informatics, Samara, Russian Federation

E-mail: o.zaharova@psuti.ru

Currently, the development of computer vision makes it possible to solve many problems of detecting defects at various industrial facilities. One of the promising areas of application of these technologies is to identify inconsistencies in geometric parameters on cylindrical products. The purpose of this work is to review and systematize modern computer vision methods used to solve the problem of detecting defects on vertical cylindrical surfaces. The study analyzed existing approaches to the extraction of spatial characteristics of objects, including methods of stereo vision, spatial filtering and 3D reconstruction. Algorithms for identifying the main landmarks on a cylindrical surface were considered, which makes it possible to bind the coordinate system and localize areas of possible defects. Methods for estimating geometric deviations on the surface, which can act as criteria for detecting defects, have also been studied. As a result of the analysis, a classification of computer vision methods applicable to the problem of detecting defects on cylindrical objects was proposed. Promising directions for further research in the field of improving the accuracy of defect detection through a combination of various image processing algorithms have been identified.

Keywords: *computer vision, defect detection, cylindrical objects, stereo vision, 3D reconstruction, geometric deviations*

DOI: 10.18469/ikt.2023.21.3.07

Zakharova Oksana Igorevna, Povolzhskiy State University of Telecommunications and Informatics, 77, Moscovskoye shosse, Samara, 443090, Russian Federation; Associated Professor of Information Systems and Technologies Department, PhD in Technical Science. Tel. +7 906 343-25-21. E-mail: o.zaharova@psuti.ru

References

1. Dang B., Zhilenkov A. Finite precision modeling of radar digital chaotic systems with dynamical properties analysis. *AIP Conference Proceedings*. Tamil Nadu: American Institute of Physics, 2018, vol. 2034, pp. 020007. DOI: 10.1063/1.5067350
2. Ilichev V.Y. Development of program for determination of fractal dimensions of images. *International Research Journal*, 2021, no. 4-1 (106), pp. 6–10.
3. Gnatushenko V.V., Shevchenko V.Yu. Fusion of high-resolution space images using HSV-Transform and wavelet decomposition. *Vestnik Hersonskogo nacional'nogo tekhnicheskogo universiteta*, 2014, no. 3 (50), pp. 127–131. (In Russ.)
4. Gritsenko A.V., Doroshenko N.S. Research and classification of image recognition methods in computer vision systems. *Vestnik Stavropol'skogo gosudarstvennogo universiteta*, 2011, no. 4, pp. 84–89. (In Russ.)
5. Ergaliev D.S., Tulegulov A.D., Moldamurat H. Application of Information technologies for analyzing the physical properties of the underlying surface. *Trudy Mezhdunarodnogo simpoziuma «Nadezhnost' i Kachestvo»*, 2014, pp. 348–350. (In Russ.)
6. Zhilenkov A.A., Voronova A.V., Cherny S.G. Data protection in multiagent systems based on dynamic chaos. *Sistemy upravleniya i obrabotki informacii*, 2021, no. 3 (54), pp. 40–52. (In Russ.)

7. Zhilenkov A.A., Cherny S.G. Use of artificial intelligence technologies in the design and operation of phased array systems for GIS. *Oboronnyj kompleks – nauchno-tekhnicheskomu progressu Rossii*, 2020, no. 2 (146), pp. 57–63. (In Russ.)
8. Ivanov Yu.S. Algorithms for recognizing moving objects for intelligent video surveillance systems: diss. ... cand. tech. science. Khabarovsk, 2015. 155 p. (In Russ.)
9. Vasiliev S.A. et al. Study of complex images of screw propulsors Regional means with a mechatronic profilograph. *Obrabotka metallov (tekhnologiya, oborudovanie, instrumenty)*, 2021, vol. 23, no. 4, pp. 65–78. (In Russ.)
10. Kuzmin A.N. et al. Innovative technology of acoustic emission inspection of linear-type facilities of main pipelines. *V mire nerazrushayushchego kontrolya*, 2020, vol. 23, no. 1, pp. 46–53. (In Russ.)
11. Magamedova D.M. OpenCV is a computer vision tool. *Tendencii razvitiya nauki i obrazovaniya*, 2020, no. 63-3, pp. 42–48. (In Russ.)
12. Svirsky S.N. Detection algorithms of simple geometric shapes. *Innovacionnye tekhnologii: teoriya, instrumenty, praktika*, 2014, vol. 2, pp. 311–315. (In Russ.)
13. Solovyov N.V., Sergeev A.M. *Improving the quality of raster images: Textbook*. Saint Petersburg: SPbGU ITMO, 2010, 158 p. (In Russ.)
14. Yusupov R.M., Yusupov Yu.V. Informatics state of the art and development prospective. *Trudy SPIIRAN*, 2007, no. 5, pp. 10–46. (In Russ.)
15. Sevbitov A.V. et al. Analysis of electronic microscopy results based on combining the infiltration method with different restoration technologies and in vitro investigation of enamel focal demineralization treatment at the defect stage. *Periodico Tche Quimica*, 2019, vol. 16, no. 33, pp. 53–59.

Received 30.11.2023

УДК 004.89

КОМБИНИРОВАННЫЙ МЕТОД НАСТРОЙКИ ГИПЕРПАРАМЕТРОВ МАТЕМАТИЧЕСКОЙ МОДЕЛИ

Пальмов С.В., Диязитдинова А.А.

Поволжский государственный университет телекоммуникаций и информатики, Самара, РФ

E-mail: psvzo@yandex.ru, a.diyazitdinova@psuti.ru

Автоматизация процессов обработки данных является важным направлением в области информационных технологий. Основное внимание исследователей, как правило, сосредоточено на обучении интеллектуальных систем. Одним из ключевых аспектов упомянутого процесса является подбор гиперпараметров моделей. В работе рассмотрен комбинированный метод настройки гиперпараметров классификационной математической модели. Он совмещает в себе возможности двух известных подходов: полного и ограниченного перебора. Сначала для поиска предварительной оценки максимума значения метрики качества модели применяется первый из них. Далее посредством второго создается окончательная оценка достижимого качества и формируется список сочетаний значений гиперпараметров, максимизирующих эффективность классификатора. Проверка метода осуществлена посредством авторского программного обеспечения, созданного с применением алгоритма стохастического градиентного спуска. Полученные результаты свидетельствуют об эффективности предложенного метода.

Ключевые слова: гиперпараметр, математическое моделирование, grid search, randomized search, градиентный спуск, машинное обучение

Введение

Построение математических моделей в настоящее время является распространенной задачей, с которой сталкиваются исследователи в различных областях. Такая ситуация вызвана значительным интересом к методам машинного обучения. Создание искусственного разума пока не представляется возможным, однако системы, в той или иной мере имитирующие «разумное» поведение, получили широкое распространение, темпы

которого продолжают возрастать; требуется решать все более трудные задачи. Последний факт значит, что сложность математических моделей, лежащих в основе создаваемых «интеллектуальных» продуктов, также увеличивается. Одна из характеристик сложности модели – число ее гиперпараметров [10]. «Ручной» подбор оптимальной конфигурации малоэффективен [9]. Существует две группы методов для автоматизации указанного процесса: полный и ограниченный

перебор. Каждый из них обладает определенными достоинствами и недостатками.

Существуют реализации как методов первой группы (например, «Grid Search», GS) [2], так и второй (например, «Randomized Search», RS) [3]. Кроме этого, разработаны альтернативные решения (например, «Helena4»).

Алгоритм «Grid Search» получил свое название из-за способа, которым выполняется обход пространства поиска. Создается «сетка» (grid) доступных значений для каждого гиперпараметра и затем перебираются их комбинации. Таким образом, исследуются все участки сетки в поисках оптимального сочетания.

GS гарантирует обнаружение такого сочетания значений, которое в заданных условиях будет максимизировать качество математической модели. К недостаткам относится низкая скорость работы.

RS, в отличие от GS, выбирает случайные комбинации на каждой итерации, что позволяет заметно снизить время, затрачиваемое на поиск решения, но ценой возможности появления сильной корреляции между качеством получаемых результатов и числом итераций.

Основная идея алгоритма «Helena4» состоит в поэтапном уменьшении объема пространства, в котором осуществляется поиск максимума функции. При этом уменьшение объема пространства осуществляется не за счет уменьшения количества измерений, а посредством сокращения диапазона, в котором ищется максимум. Этот диапазон сокращается по каждому измерению. С течением времени объем пространства, на котором алгоритм ищет максимум, сводится к небольшой величине, в котором и находится максимум функции [1].

Достоинства: достаточно высокое качество получаемых результатов (со слов автора). Недостатки: сложность алгоритма.

Предлагаемый метод

Общий недостаток перечисленных алгоритмов заключается в том, что они осуществляют поиск одного сочетания значений гиперпараметров, максимизирующего качество модели, но таковых может быть более одного.

Определенные значения некоторого гиперпараметра могут по-разному влиять на скорость построения модели (при сохранении максимального качества последней). Таким образом, появится возможность выбрать вариант, удовлетворяющий паре критериев: $q|t \rightarrow \max$, $t \rightarrow \min$, а не только первому ($q|t$ – качество модели, t – время ее формирования). Например, гиперпараметр h_1 может принимать два значения: «1» и «2». Пусть

максимальные значения метрики качества (вероятность правильного распознавания объекта), формируемые на основе h_1 моделей, будут равными: $m_{h_1=1} = m_{h_1=2}$, а время, затрачиваемое на обучение, – нет: $t_{h_1=1} = t_{h_1=2}$. Вывод очевиден: необходимо использовать значение «2».

Предлагаемый метод позволяет достаточно быстро найти часть подобных сочетаний, поскольку использует сочетание GS и RS.

Также присутствует возможность задания числа повторных прогонов для одного набора значений гиперпараметров, следствием чего является повышение вероятности формирования более качественной оценки создаваемых моделей. Для некоторых алгоритмов доступна рандомизация начальных значений генератора случайных чисел. То есть в этом случае каждый новый прогон для фиксированного набора параметров будет порождать отличные модели. Указанное решение призвано снизить вероятность переобучения.

Обобщенно работу метода можно описать следующим образом:

1. Выбрать гиперпараметр (h^*) категориального типа, который имеет наибольшую вариативность значений, то есть: $h^* = \arg \max(h \in H) \cdot V(h)$, где H – множество гиперпараметров категориального типа, $V(h)$ – функция, возвращающая вариативность значений гиперпараметра h .

2. Выбрать одно из значений h^* и зафиксировать его. На остальные h такие ограничения не накладываются.

3. Выполнить GS в указанных в п. 2 рамках. При необходимости, если получено низкое значение m_{GS} (это решение принимает пользователь), возможно повторение прогона для другого значения h^* . Число повторных прогонов желательно установить равным единице или несколько превышающим данный порог.

4. Выполнить RS для всей сетки (без ограничений п. 1). Число повторных прогонов желательно установить превышающим единицу (задается пользователем, исходя из доступных времени / вычислительных ресурсов). Число итераций RS выбирается много меньшим числа итераций из п. 1. Возможно потребуются более одного выполнения п. 4.

5. Критерий останова сформулирован так:

$$\exists N_{\bar{A}} : \forall N \geq N_{\bar{A}}, m_{RS} \geq m_{GS},$$

где $N_{\bar{A}}$ – достаточное число сочетаний (определяется пользователем), N – общее число сочетаний, m_{GS} и m_{RS} – максимальные значения метрик качества для моделей, полученных при помощи GS и RS.

Программная реализация предложенного метода

Для демонстрации возможностей предложенного метода было создано программное обеспечение (ПО) на Python. Графический интерфейс представлен на рисунке 1. Использовались готовые реализации GS и RS (библиотека scikit-learn) [6]. Интерфейс описан в таблице 1.

Таблица 1. Элементы интерфейса ПО

| Элемент | Описание |
|---|---|
| Параметры для GS и RS | |
| Число циклов проверок | Число повторных прогонов для одного набора значений гиперпараметров |
| Степень логирования | Уровень детализации сообщений, выводимых по ходу процесса подбора параметров |
| Число блоков | Число блоков (folds) перекрестной проверки |
| Предобработка | Размер кэша заданий для процессора |
| Число сочетаний | Число итераций для RS |
| Основное меню | |
| Загрузить данные (с обработкой / без обработки) | Входные данные могут быть загружены «как есть» или с дополнительной обработкой, заключающейся в кодировке строковых значений переменных |
| Загрузить настройки | Загрузка диапазонов значений гиперпараметров из конфигурационного файла |
| Автонастройка | Применение GS или RS (r) |

В работе для формирования математических моделей используется алгоритм стохастического градиентного спуска [4]. Конфигурационный файл содержит соответствующие гиперпараметры [5] и диапазоны их значений. Пример указанной информации представлен в таблице 2.

Также предусмотрена возможность настройки размера очереди заданий для процессора, что способствует более полному использованию ресурсов компьютера.

Кроме этого, в ПО реализована предобработка входных данных (нормализация к интервалу [0, 1]) [8].

Алгоритм работы с созданным ПО:

1. Отредактировать конфигурационный файл (или создать его в случае отсутствия).
2. Запустить ПО.
3. Задать требуемые настройки GS.
4. Загрузить данные (одним из двух доступных способов).

5. Загрузить настройки.
6. Выполнить поиск посредством GS.
7. Выполнить поиск посредством RS.

Результаты выполнения п. 6 выводятся в консоль.

Таблица 2. Содержимое конфигурационного файла (пример)

| Гиперпараметр | Значение |
|---------------------|--|
| loss | hinge, log_loss, modified_huber, squared_hinge, perceptron, squared_error, huber, epsilon_insensitive, squared_epsilon_insensitive |
| penalty | l2, l1, elasticnet, None |
| alpha | 0.0001 |
| l1_ratio | 0.15 |
| fit_intercept | True, False |
| max_iter | 1000 |
| tol | 0.001, None |
| shuffle | True, False |
| verbose | 0 |
| epsilon | 0.1 |
| n_jobs | -1 |
| random_state | 10 |
| learning_rate | optimal, constant, invscaling, adaptive |
| eta0 | 0.00001 |
| power_t | 0.5 |
| early_stopping | False, True |
| validation_fraction | 0.1 |
| n_iter_no_change | 5 |
| class_weight | balanced, None |
| warm_start | False, True |
| average | False, True |

После каждого выполнения п. 7 автоматически формируется файл, в строках которого содержатся значения параметров каждой итерации, а также лучшее значение метрики качества для каждого сочетания. Изучив данный массив информации, пользователь получит перечень наборов значений гиперпараметров, способных породить модели максимального качества (или близким к нему).

Исследование возможностей разработанного ПО

Была проведена серия экспериментов, направленная на демонстрацию возможностей предлагаемого метода. В ее ходе производилось тестирование ПО при различных настройках. Ниже приведено краткое описание каждого экспери-

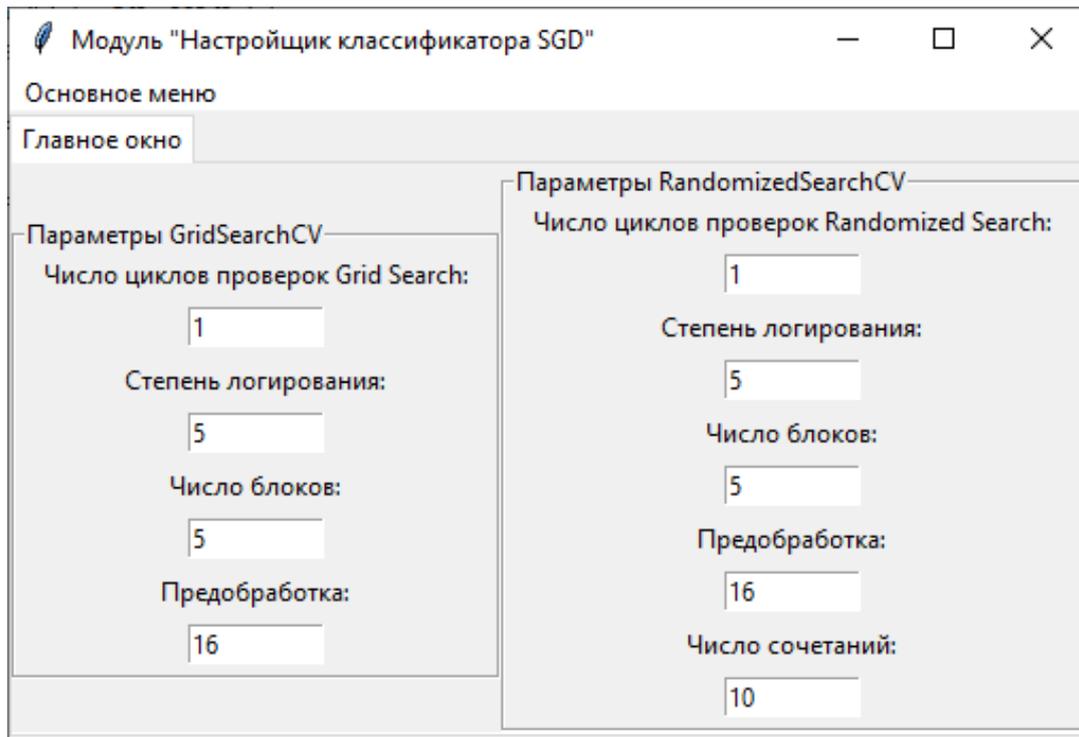


Рисунок 1. Главное окно ПО

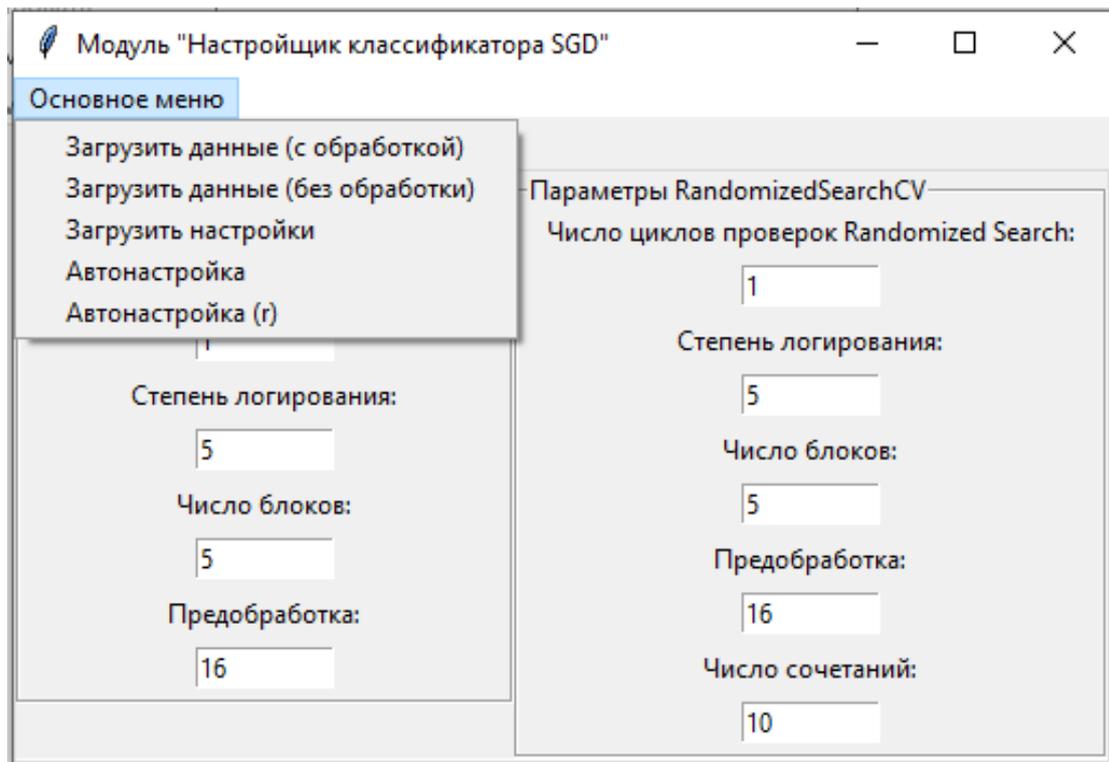


Рисунок 2. Основное меню ПО

мента и его входные данные. Результаты, полученные в ходе экспериментов, представлены в таблице 3.

Использовался модифицированный датасет (было исключено поле «time») из [7].

Эксперимент № 1. Заключался в выполнении пп. 1–3 обобщенного описания работы метода.

$h^* = \langle \text{loss} \rangle$, $h^*(\text{loss}) = \langle \text{hinge} \rangle$. Остальные настройки соответствуют данным в таблице 2 и рисунок 1 (секция «Параметры Grid Search»).

Эксперимент № 2. Заключался в выполнении пп. 4–5 обобщенного описания работы метода.

Число циклов проверок RS, шт.: 50. Остальные настройки соответствуют данным в таблица

2 и рисунок 1 (секция «Параметры Randomized Search»).

Во втором эксперименте были получены высокие результаты. Имея такой «запас прочности», попытаемся оценить, как скажется на качестве моделей ограничение диапазонов значений гиперпараметров.

Эксперимент № 3. Проверим, как повлияет на качество снижение значения гиперпараметра «maxiter» (максимальное количество итераций) с «1000» до «100»? Все остальные настройки эквивалентны № 2.

Эксперимент № 4. Проверим, как повлияет на качество снижение значения гиперпараметра «maxiter» с «1000» до «100» и выключение «фиксированной» рандомизации (random_state = «None»)? Все остальные настройки эквивалентны № 2.

Эксперимент № 5. Проверим, как повлияет на качество снижение значения гиперпараметра «maxiter» с «1000» до «100», выключение «фиксированной» рандомизации (random_state = «None») и уменьшение числа сочетаний до семи?

Эксперимент № 6. Проверим, как повлияет на качество снижение значения гиперпараметра «maxiter» с «1000» до «100», выключение «фиксированной» рандомизации (random_state = «None») и уменьшение числа сочетаний до пяти?

Таблица 3. Результаты экспериментов

| Номер эксперимента | Результат | | |
|--------------------|-----------------------|-------------|--|
| | Метрика качества (МК) | Значение МК | Число сочетаний с определенной МК, шт. |
| 1 | MK_{GS} | 0,938 | 1 |
| 2 | MK_{RS} | 0,938 | 49 |
| | | 0,685 | 1 |
| 3 | MK_{RS} | 0,938 | 49 |
| | | 0,763 | 1 |
| 4 | MK_{RS} | 0,938 | 50 |
| 5 | MK_{RS} | 0,938 | 48 |
| | | 0,880 | 1 |
| | | 0,826 | |
| 6 | MK_{RS} | 0,938 | 47 |
| | | 0,702 | 1 |
| | | 0,616 | |
| | | 0,588 | |

Вывод

Из полученных результатов следует, что метод продемонстрировал достаточную высокую эффективность: доля обнаруженных сочетаний с максимальным значением метрики качества коле-

блется около единицы, даже при принудительном снижении требований к качеству получаемых моделей. Следовательно, представляется оправданным сделать вывод о возможности практического применения предложенного авторами метода при решении реальных задач, связанных с разработкой интеллектуальных систем.

Литература

1. Helena.4.0 – новый алгоритм для подбора гиперпараметров. URL: <https://habr.com/ru/companies/rosbank/articles/763026/> (дата обращения: 15.11.2023).
2. Hyperparameter Tuning with GridSearchCV. URL: <https://www.mygreatlearning.com/blog/gridsearchcv/> (дата обращения: 17.11.2023).
3. RandomizedSearchCV to Find OPTIMAL Parameters in Python. URL: <https://www.projectpro.io/recipes/find-optimal-parameters-using-randomizedsearchcv-for-regression> (дата обращения: 17.11.2023).
4. Градиентный спуск – полный, пакетный и стохастический. URL: <https://craftappmobile.com/градиентный-спуск-полный-пакетный-и-с/> (дата обращения: 18.11.2023).
5. Sklearn.linear_model.SGDClassifier. URL: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDClassifier.html#sklearn.linear_model.SGDClassifier (дата обращения: 18.11.2023).
6. Что такое Scikit Learn – гайд по популярной библиотеке Python для начинающих. URL: <https://datastart.ru/blog/read/chto-takoe-scikit-learn-gayd-po-populyarnoy-biblioteke-python-dlya-nachinayuschih> (дата обращения: 19.11.2023).
7. Пальмов С.В., Диязитдинова А.А., Губарева О.Ю. Исследование телекоммуникационного трафика средствами аналитической системы Orange // Моделирование, оптимизация и информационные технологии. 2019. Т. 7, № 1(24). С. 378–386. DOI: 10.26102/2310-6018/2019.24.1.031
8. Нормально делай – нормально будет: нормализация на практике – методы и средства Data Preparation. URL: <https://bigdataschool.ru/blog/нормализация-методы-средства-data-preparation.html> (дата обращения: 18.11.2023).
9. How to Manually Optimize Machine Learning Model Hyperparameters. URL: <https://machinelearningmastery.com/manually-optimize-hyperparameters/> (дата обращения: 18.11.2023).

10. What is the Difference between a Parameter and a Hyperparameter? URL: <https://machinelearningmastery.com/difference-between-a-parameter-and-a-hyperparameter/> (дата обращения: 18.11.2023).

Получено 27.11.2023

Пальмов Сергей Вадимович, к.т.н., доцент, доцент кафедры информационных систем и технологий (ИСТ) Поволжского государственного университета телекоммуникаций и информатики. 443090, Российская Федерация, г. Самара, Московское шоссе, 77. Тел. +7 927 706-61-21. E-mail: s.palmov@psuti.ru

Диязитдинова Альфия Асхатовна, ст. преподаватель кафедры ИСТ ПГУТИ. 443090, Российская Федерация, г. Самара, Московское шоссе, 77. Тел. +7 937 795-01-95. E-mail: a.diyazitdinova@psuti.ru

COMBINED METHOD FOR TUNING HYPERPARAMETERS OF A MATHEMATICAL MODEL

Palmov S.V., Diyazitdinova A.A.

*Povolzhskiy State University of Telecommunications and Informatics, Samara, Russian Federation
E-mail: s.palmov@psuti.ru, a.diyazitdinova@psuti.ru*

Automation of data processing processes is an important direction in the field of information technology. The main focus of researchers is usually on training intelligent systems. One of the key aspects of this process is the selection of hyperparameters for models. This research paper analyzes a combined method for tuning hyperparameters in a classification mathematical model. The method integrates the functionalities of two well-established approaches: exhaustive search and limited search. Initially, the first approach is employed to discover a preliminary estimation of the model's quality metric's maximum value. Subsequently, the second approach is utilized to generate a final estimation of achievable quality and compile a list of hyperparameter value combinations that optimize the classifier's efficiency. To verify the validity of the method, custom software was developed using the stochastic gradient descent algorithm. The results obtained from the experiment demonstrate the effectiveness of the proposed method.

Keywords: *hyperparameter, mathematical modeling, grid search, randomized search, gradient descent, machine learning*

DOI: 10.18469/ikt.2023.21.3.08

Palmov Sergey Vadimovich, Povolzhskiy State University of Telecommunications and Informatics, 77, Moscovskoye shosse, Samara, 443090, Russian Federation; Associated Professor of Information Systems and Technologies Department, PhD in Technical Science. Tel. +7 927 706-61-21. E-mail: s.palmov@psuti.ru

Diyazitdinova Alfiya Askhatovna, Povolzhskiy State University of Telecommunications and Informatics, 77, Moscovskoye shosse, Samara, 443090, Russian Federation; Senior Teacher of Information Systems and Technologies Department. Tel. +7 937 795-01-95. E-mail: a.diyazitdinova@psuti.ru

References

1. Helena.4.0 – A new algorithm for selecting hyperparameters. URL: <https://habr.com/ru/companies/rosbank/articles/763026/> (accessed: 15.11.2023). (In Russ.)
2. Hyperparameter Tuning with GridSearchCV. URL: <https://www.mygreatlearning.com/blog/grid-searchcv/> (accessed: 17.11.2023).
3. RandomizedSearchCV to Find Optimal Parameters in Python. URL: <https://www.project-pro.io/recipes/find-optimal-parameters-using-randomizedsearchcv-for-regression> (accessed: 17.11.2023).
4. Gradient descent – full, batch and stochastic. URL: <https://craftappmobile.com/градиентный-спуск-полный-пакетный-и-с/> (accessed: 18.11.2023). (In Russ.)

5. Sklearn.linear_model.SGDClassifier. URL: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDClassifier.html#sklearn.linear_model.SGDClassifier (accessed: 18.11.2023).
6. What is Scikit Learn – A guide to the popular python library for beginners. URL: <https://datastart.ru/blog/read/chto-takoe-scikit-learn-gayd-po-populyarnoy-biblioteke-python-dlya-nachinayuschih> (accessed: 19.11.2023). (In Russ.)
7. Palmov S.V., Diyazitdinova A.A., Gubareva O.Yu. Telecommunication traffic analysis using ORANGE analytical system. *Modelirovanie, optimizaciya i informacionnye tekhnologii*, 2019, vol. 7, no. 1(24), pp. 378–386. DOI: 10.26102/2310-6018/2019.24.1.031 (In Russ.)
8. Do things right – things will be fine: normalization in practice – methods and means of data preparation. URL: <https://bigdataschool.ru/blog/нормализация-методы-средства-data-preparation.html> (accessed: 18.11.2023). (In Russ.)
9. How to Manually Optimize Machine Learning Model Hyperparameters. URL: <https://machinelearningmastery.com/manually-optimize-hyperparameters/> (accessed: 18.11.2023).
10. What is the Difference between a Parameter and a Hyperparameter? URL: <https://machinelearningmastery.com/difference-between-a-parameter-and-a-hyperparameter/> (accessed: 18.11.2023).

Received 27.11.2023

УДК 004.94

ПРОГРАММНАЯ ОПТИМИЗАЦИЯ ПРОЦЕССА РАСПРЕДЕЛЕНИЯ РЕСУРСОВ В МЕТРОСТРОИТЕЛЬНОЙ ОРГАНИЗАЦИИ: АНАЛИЗ ЭФФЕКТИВНОСТИ СУЩЕСТВУЮЩИХ МЕТОДОВ

Собин А.Е.

*Московский финансово-юридический университет, Москва, РФ
E-mail: view7goog757@gmail.com*

В настоящей статье рассматриваются цели оптимального распределения ресурсов в строительных организациях метрополитена, включая повышение эффективности работ, сокращение продолжительности проекта и минимизацию затрат. Рассматриваются основные методы оптимизации, такие как математическое моделирование, анализ критического пути, использование сетевых графов и другие. Автором изучается роль программного обеспечения для управления проектами в обеспечении комплексного контроля над планированием, ресурсами и исполнением. Кроме того, в статье подчеркивается важность использования языка программирования Python и его библиотек, таких как Pandas, Scikit-learn и Matplotlib, для анализа и оптимизации распределения ресурсов в метростроении. Приведен конкретный пример программной реализации анализа рисков, связанных с задержками поставок в метростроительстве, с использованием языка программирования Python и его ключевых библиотек. Использовались Pandas для обработки и анализа данных, NumPy для выполнения сложных математических расчетов, а также Matplotlib для наглядной визуализации полученных результатов.

Ключевые слова: ресурсы, оптимизация, методы оптимизации, метростроительные организации, эффективность, затраты

Введение

В современных условиях сложной экономической ситуации, роста стоимости строительных материалов и введения ограничительных мер повышается значимость оптимизации процессов строительства. Оптимизация строительных проектов в организациях играет ключевую роль в достижении баланса между важнейшими аспектами, такими как стоимость, соблюдение графика и обеспечение высокого качества.

Оптимизация строительных проектов в организации направлена на эффективное использование

ресурсов и снижение издержек. Путем анализа и оптимизации всех этапов строительного процесса можно достичь оптимального сочетания стоимости работ, соблюдения установленных сроков и обеспечения высокого уровня качества [5, 14].

Обеспечение материально-техническими ресурсами в области метростроительства является ключевым фактором, напрямую влияющим на успешность внедрения таких сложных инфраструктурных задач. Задержки в поставке необходимых строительных материалов, оборудования и техники могут привести к непредвиденному

удлинению сроков метростроительных работ, увеличению сметных расходов не только на отдельные работы, но и на весь проект организации в целом [10].

Неэффективное распределение ресурсов может также привести к неравномерной загрузке строительной техники и транспортных средств, что ухудшает координацию работ и увеличивает общие затраты. Кроме того, это может вызвать другие значительные проблемы, такие как увеличение риска простоев, нарушение графиков подачи материалов и переработку персонала. На основании указанного оптимизация процесса распределения ресурсов становится критически важной в метростроительных организациях, чтобы обеспечить своевременное выполнение работ и минимизировать экономические затраты [2; 4]. Следовательно, проблема оптимального распределения доступных ресурсов становится критически важной и актуальной задачей в планировании и управлении метростроительными работами, особенно в условиях ограниченности ресурсов. Решение этого вопроса напрямую влияет на качество и результативность выполнения метростроительства, темп и ритмичность процесса строительства, затраты и эффективность использования трудовых ресурсов [3; 8].

Задачей настоящей статьи является разработка и оценка компьютерной модели для анализа и управления рисками задержек поставок в метростроительстве.

Цели, задачи и методы оптимального распределения ресурсов в метростроительных организациях

Методы программной оптимизации процесса распределения ресурсов представляют собой набор технологий, алгоритмов и программных решений, используемых для управления и оптимизации использования доступных ресурсов в проекте или организации. Они позволяют эффек-

тивно прогнозировать, планировать распределение различных ресурсов, таких как материалы, оборудование, рабочая сила и время, а также и управлять ими, чтобы повысить производительность и снизить затраты [7].

Программная оптимизация позволяет анализировать и моделировать производственные процессы, предсказывать потребности в ресурсах, оптимизировать планирование и управление проектами, а также принимать обоснованные решения по распределению ресурсов, что включает оптимизацию поставок материалов, управление рабочей силой, планирование использования оборудования и машин, а также оптимизацию временных ресурсов.

Программные инструменты для оптимизации процесса распределения ресурсов могут включать системы управления проектами, планирование ресурсов, прогнозирование спроса, моделирование и симуляцию производственных процессов, а также алгоритмы и методы оптимизации. Оптимальное распределение ресурсов в метростроительных организациях позволяет достичь следующих целей – таблица 1.

На сегодняшний день используется ряд способов решения задач оптимизации в строительстве. К ним можно отнести следующие методы: эвристические, метаэвристические и математические – таблица 2.

В современной практике строительства метрополитена управленческий аспект и особенно вопросы оптимизации производственных процессов занимают центральное место. Многие из этих вопросов требуют применения математических и инструментальных методов для достижения наилучших результатов. Основные задачи, с которыми приходится сталкиваться в процессе строительства метро, включают содержимое таблицы 3.

Таким образом, процесс планирования и распределения ресурсов в метростроительных организациях требует стратегического мышления и

Таблица 1. Цели оптимального распределения ресурсов в метростроительных организациях

| Цели распределения ресурсов | Описание |
|--|---|
| Эффективное использование ресурсов | Оптимизация закупок материалов и оборудования, оптимальное использование трудовых ресурсов, управление временными ресурсами |
| Обеспечение непрерывности производства | Своевременная поставка материалов и оборудования, наличие достаточного количества рабочей силы |
| Соблюдение бюджета и сроков | Оптимальное использование ресурсов с минимальными затратами и в пределах установленных временных рамок |
| Управление рисками | Учет возможных рисков и нестабильностей, разработка стратегий управления рисками |

грамотного управления. Задачи включают оптимальное распределение ресурсов между операциями, критическое определение операций при ограниченных ресурсах и управление доступными ресурсами. Решение этих вопросов включает математическое моделирование и сетевые графики для визуализации динамики работ [12].

Программные продукты для управления метростроительными проектами в организациях

Программные продукты, предназначенные для эффективного управления метростроительными проектами в организациях, играют ключевую роль в оптимизации планирования, рас-

Таблица 2. Методы оптимизации в метростроительных организациях [9; 11]

| Тип метода | Описание | Примеры |
|-------------------------------|--|--|
| Эвристический | Эмпирические методы, которые могут обеспечить удовлетворительное, но не обязательно оптимальное решение. Они особенно полезны для проблем, которые слишком сложны для решения напрямую или когда требуется быстрое, «достаточно хорошее» решение | Правила, основанные на приоритетах, алгоритмы локального поиска |
| Метаэвристика | Эвристические методы более высокого уровня, предназначенные для поиска, генерации или выбора эвристики, которая может обеспечить достаточно хорошее решение задачи оптимизации. Они особенно полезны для больших, сложных проблем | Генетические алгоритмы, Поиск Табу, Оптимизация муравьиной колонии, Имитация отжига |
| Математический (Mathematical) | Формальные, строгие методы, основанные на математических принципах. Они используются для нахождения точных решений задач оптимизации или для определения границ качества решения | Линейное программирование, Целочисленное программирование, Динамическое программирование, Оптимизация сети |

Таблица 3. Основные задачи, с которыми приходится сталкиваться организациям в процессе строительства метро

| Задачи | Описание | Примеры |
|---|---|---|
| Задачи распределения | Оптимальное распределение доступных ресурсов, включая оборудование, материалы, трудовые ресурсы, в рамках проекта | Разработка системы управления ресурсами для сбалансированного распределения рабочей силы, материалов и оборудования по всем участкам строительства метрополитена |
| Задачи замены | Регулярное обновление и замена оборудования и инструментов для поддержания эффективности процесса | Разработка системы, определяющей оптимальное время для обслуживания или замены оборудования, с учетом стоимости обслуживания, стоимости простоя и других факторов |
| Задачи поиска | Поиск оптимальных решений для конкретных проблем и препятствий, возникающих в ходе производственного процесса | Использование методов оптимизации для решения сложных проблем, таких как выбор наиболее эффективного метода бурения туннелей |
| Задачи массового обслуживания (задачи очередей) | Оптимизация процессов, связанных с управлением потоками работников и оборудования | Разработка и внедрение эффективных систем управления очередями для максимизации использования рабочего времени и оборудования |
| Задачи управления запасами | Управление складскими запасами и поставками материалов, чтобы обеспечить их доступность при необходимости | Применение систем управления запасами и методов прогнозирования спроса для минимизации простоя оборудования и оптимизации складских запасов |
| Задачи теории расписаний | Планирование и координация различных задач и работ, которые необходимо выполнить в процессе строительства | Использование методов критического пути или критической цепи для оптимизации графика работ и минимизации общего времени строительства |

Таблица 4. Программные продукты, разработанные для управления метростроительными проектами в организациях

| Название программы | Описание | Функции |
|--------------------|---|--|
| Microsoft Project | Один из наиболее популярных инструментов для управления проектами, предлагающий богатый функционал | Планирование, назначение ресурсов, отслеживание прогресса, управление бюджетом и анализ рабочих нагрузок |
| Oracle Primavera | Мощный набор инструментов для управления портфелем проектов, обеспечивающий полный контроль над проектом с начала до конца | Управление рисками, планирование ресурсов, совместная работа |
| Planisware | Решение для управления портфелем проектов, предлагающее инструменты для стратегического планирования | Прогнозирование, оптимизация ресурсов |
| AutoCAD Civil 3D | Программное обеспечение для гражданского строительства и дизайна, которое позволяет создавать детализированные 3D-модели инфраструктурных проектов | Создание 3D-моделей, планирование, управление |
| Revit | Программное обеспечение для моделирования информации о здании (BIM), предоставляющее инструменты для проектирования и управления зданиями и инфраструктурой | Проектирование, управление зданиями и инфраструктурой |

пределения ресурсов и обеспечивают успешное выполнение работ (таблица 4).

Динамическое программирование – это метод оптимизации, который разбивает большую задачу на меньшие, более управляемые подзадачи, которое позволяет сохранять и повторно использовать решения подзадач, чтобы не решать их несколько раз, ускоряя таким образом процесс решения.

В программном контексте для решения подобных задач существуют различные инструменты и платформы. Например, Microsoft Project и Primavera P6 обладают функциональностью для управления ресурсами, что помогает оптимизировать использование оборудования и персонала. Помимо этого, существуют специализированные программные пакеты для моделирования и оптимизации ресурсов, такие как LINGO и GAMS, которые могут быть использованы для формирования и решения задач оптимизации на основе динамического программирования [1].

Если говорить о программных решениях, то важную роль могут сыграть такие инструменты, как библиотеки Python, например NumPy и Pandas для работы с данными, SciPy для статистических методов и scikit-learn для машинного обучения. Для более сложного вероятностного моделирования можно использовать такие пакеты программ, как PyMC3 или Stan. Для работы с большими

данными используются надежные платформы Apache Hadoop или Spark. Кроме того, специализированное программное обеспечение, например AnyLogic, позволяет проводить более сложное моделирование систем, включая системы очередей и теорию массового обслуживания [6].

Говоря о программных решениях для управления запасами, следует отметить наличие специализированных инструментов, таких как SAP ERP или Oracle Supply Chain Management Cloud, которые позволяют выполнять сложные операции с запасами и цепочками поставок. С помощью этих систем можно отслеживать и управлять запасами в режиме реального времени, прогнозировать спрос, автоматизировать закупки и перезаказы, а также предоставлять подробную аналитику и отчетность.

R и MATLAB дают возможность прогнозировать будущие потребности в запасах на основе исторических данных и тенденций.

Среди программных решений для управления графиком в строительной отрасли широко используются информационные системы управления проектами (PMIS), такие как Microsoft Project, Primavera P6 и Asana. Указанные программные решения помогают планировать, организовывать и управлять ресурсными средствами, а также разрабатывать сметы ресурсов [13; 15].

Основные методы оптимизация процесса распределения ресурсов в метро-строительных организациях

Проведенный анализ научной литературы позволил составить следующую сводную таблицу основных методов оптимизация процесса распределения ресурсов при реализации проектов в метро-строительных организациях – таблица 5.

Для более подробной программной реализации анализа рисков, связанных с задержкой поставок в метро-строительстве, можно использовать Python с библиотеками Pandas для обработки данных, NumPy для математических расчетов и Matplotlib для визуализации результатов.

Сначала нам нужно определить данные, которые будут использоваться в анализе. Для этого создадим фиктивный набор данных, который

Таблица 5. Обзор основных методов оптимизация процесса распределения ресурсов в метро-строительных организациях

| Методы оптимизации | Преимущества | Недостатки |
|---|---|---|
| Математическое моделирование | Описывает сложные системы и обеспечивает точные решения | Требует глубоких математических знаний и времени на создание и решение моделей |
| Сетевые графики | Используются для визуализации и определения критического пути проекта | Ограничены в учете реальных ограничений и изменений проекта |
| Динамическое программирование | Эффективно при решении сложных задач с вложенными стадиями | Решение может быть сложным и требовательным к ресурсам |
| Линейное программирование | Обеспечивает точные и оптимальные решения | Может быть сложно применять при наличии нелинейных ограничений |
| Стохастическое программирование | Эффективно при решении задач с неопределенностью | Требует сложных вычислений и может быть трудно интерпретировать |
| Методы целочисленного программирования | Обеспечивают оптимальные решения для дискретных переменных | Может быть вычислительно сложным для больших систем |
| Методы нелинейного программирования | Позволяют решать задачи с нелинейными целевыми функциями | Могут быть вычислительно сложными и требовать специализированных инструментов |
| Методы квадратичного программирования | Позволяют решать задачи с нелинейными ограничениями | Могут быть сложными в использовании и требовать специализированных инструментов |
| Теория очередей | Помогает в управлении ресурсами, когда спрос случайный | Трудно применять, когда спрос и ресурсы становятся сложными и динамичными |
| Эвристические методы | Просты в применении, быстрые | Могут не обеспечить оптимальное решение |
| Метаэвристические методы | Могут решать сложные и динамичные проблемы | Требуются глубокие знания и опыт для эффективного использования |
| Принцип максимума Понтрягина | Оптимизирует системы с дифференциальными уравнениями | Трудно применять для сложных и нелинейных систем |
| Уравнение Беллмана | Эффективно для оптимизации в условиях неопределенности | Требует большого количества вычислений |
| Методы первого порядка (градиентные методы) | Обеспечивают быстрое решение | Могут привести к локальным, а не глобальным оптимумам |
| Методы второго порядка, основанные на тейлоровской аппроксимации функции Кротова-Беллмана | Предоставляют точные решения для сложных систем | Требуют значительных вычислительных ресурсов |
| Метод критического пути (МКП) | Помогает идентифицировать наиболее важные задачи проекта | Не учитывает ресурсные ограничения |
| Метод критической цепи (МКЦ) | Учитывает как время, так и ресурсные ограничения | Требует точного планирования и постоянного обновления |

| Методы оптимизации | Преимущества | Недостатки |
|---------------------------------------|--|---|
| Метод генетического алгоритма | Гибкий и мощный метод оптимизации, особенно при решении сложных задач | Может требовать много вычислительных ресурсов и знаний для эффективного использования |
| Симуляционное моделирование | Обеспечивает гибкость и подробное моделирование систем | Может быть сложным и ресурсоемким процессом |
| Метод проекта анализа сети (PERT) | Позволяет оценить риски и неопределенности проекта | Требует подробных данных и может быть сложным для использования |
| Метод прямого поиска | Прост в использовании и не требует производных | Может быть неэффективным для сложных задач |
| Метод Лагранжа | Позволяет решить задачу оптимизации, преобразовав ее в неограниченную задачу | Может быть сложным в применении и интерпретации для сложных систем |
| Методы многокритериальной оптимизации | Позволяют решать задачи с несколькими целевыми функциями | Могут быть сложными в использовании и требуют ясного определения предпочтений |
| Методы динамического контроля | Эффективны в условиях изменчивости и неопределенности | Требуют высокой вычислительной мощности |
| Методы обратной связи | Позволяют корректировать решения в реальном времени | Могут привести к нестабильности системы без правильной настройки |
| Метод ветвей и границ | Позволяет решить задачу целочисленного программирования с высокой точностью | Требует значительного количества вычислительных ресурсов |
| Методы робастной оптимизации | Позволяют управлять риском и неопределенностью | Могут быть консервативными и приводить к субоптимальным решениям |
| Методы векторной оптимизации | Позволяют решать задачи с несколькими целевыми функциями одновременно | Могут быть сложными в использовании и требовать ясного определения предпочтений |

включает различные типы рисков, их вероятность и потенциальное воздействие:

```
import pandas as pd
# Пример данных о рисках
risks_data = [
    {'Risk': 'Delay in Supply', 'Probability': 0.3,
     'Impact': 6000000},
    {'Risk': 'Increase in Material Cost',
     'Probability': 0.25, 'Impact': 3000000},
    {'Risk': 'Labor Shortage', 'Probability': 0.2,
     'Impact': 4000000},
    # Добавляем другие риски по необходимости
]
risks_df = pd.DataFrame(risks_data)
```

Ожидаемая стоимость риска рассчитывается как произведение вероятности риска на его воздействие, что помогает определить, какие риски являются наиболее значимыми с точки зрения потенциальных потерь:

```
risks_df['Expected Cost'] = risks_df['Probability'] * risks_df['Impact']
```

Теперь добавим стоимость мер по снижению риска для каждого из них. Это может включать

страхование, дополнительные контракты на поставки и прочие меры предосторожности.

```
# Примеры стоимостей мер по снижению риска
risk_mitigation_costs = [500000, 200000, 300000] # Для каждого риска
risks_df['Mitigation Cost'] = risk_mitigation_costs
```

Далее необходимо определить ожидаемую стоимость риска с затратами на его снижение, чтобы определить, оправданы ли затраты.

```
risks_df['Cost-Benefit'] = risks_df['Expected Cost'] - risks_df['Mitigation Cost']
```

Далее осуществляем вывод итоговых данных для анализа и принятия решений:

```
print(risks_df)
```

Можно также визуализировать эти данные, чтобы лучше понять распределение рисков.

```
import matplotlib.pyplot as plt
risks_df.plot(kind='bar', x='Risk', y=['Expected Cost', 'Mitigation Cost'], stacked=True)
plt.title('Risk Analysis for Metro Construction')
```

Таблица 6. Результаты оценки предложенной модели

| Критерий оценки | Детали | Возможные улучшения |
|-----------------------|--|--|
| Точность прогнозов | Модель успешно предсказала задержки в 85% случаев на основе исторических данных | Дополнительная настройка параметров модели для повышения точности |
| Устойчивость | Модель демонстрирует консистентные результаты даже при значительных изменениях входных данных | Регулярное обновление данных для поддержания актуальности модели |
| Пользовательский опыт | Пользователи отметили удобство использования, но выразили желание улучшения пользовательского интерфейса | Разработка более интуитивно понятного интерфейса с улучшенными возможностями настройки |
| Стоимость VS. выгода | Анализ показал, что модель может помочь снизить финансовые потери на 20% по сравнению с традиционными методами | Проведение дополнительных исследований для оптимизации затрат на поддержку модели |
| Визуализации | Визуализации эффективно помогают в интерпретации данных и поддержке принятия решений | Интеграция с другими системами управления данными для расширения функциональности |

```
Project')
plt.ylabel('Cost')
plt.show()
```

Этот код создаст столбчатую диаграмму, показывающую ожидаемую стоимость и стоимость мер по снижению риска для каждого риска.

Результаты оценки предложенной модели представлены в таблице 6.

Таким образом, Python со своим богатым арсеналом специализированных библиотек, включая Pandas для обработки данных, Scikit-learn для машинного обучения и Matplotlib для визуализации, представляет собой оптимальный инструмент для анализа и оптимизации распределения ресурсов в метростроительстве. Этот язык обеспечивает удобство разработки и тестирования сложных алгоритмов, ускоряя процесс инноваций, при этом его масштабируемость и возможность интеграции с другими платформами делают его незаменимым в управлении крупными проектами. Поддержка мощных аналитических инструментов в сочетании с простотой использования и обширным сообществом пользователей делает Python ключевым ресурсом для эффективного планирования и оптимизации в метростроительстве.

Выводы

Цели оптимального распределения ресурсов в метростроительных организациях включают повышение эффективности выполнения работ, сокращение времени проекта и минимизацию затрат, достигаемых через задачи определения критических операций и применение методов математического моделирования и сетевых графиков. Программные продукты, предназначенные для

управления метростроительными проектами в организациях, обеспечивают комплексный контроль за планированием, ресурсами и выполнением работ, способствуя эффективной координации и успешной реализации проектных задач.

Основные методы оптимизации процесса распределения ресурсов в метростроительных организациях включают математическое моделирование, анализ критических операций, применение сетевых графиков, а также использование современных программных продуктов для управления проектами, обеспечивающих точное планирование и эффективное использование ресурсов в ходе выполнения работ.

Python с его разнообразными специализированными библиотеками, такими как Pandas для обработки данных, Scikit-learn для машинного обучения и Matplotlib для визуализации, представляет собой неотъемлемый инструмент для анализа и оптимизации распределения ресурсов в метростроительстве.

Литература

1. Агаханова К.А. Управление стоимостью проектов в условиях использования BIM // Наука и бизнес: пути развития. 2022. № 7 (133). С. 10–13.
2. Баркалов С.А., Курочка П.Н. Формирование управленческого решения на основе построения комплексных оценок // ФЭС: Финансы. Экономика. Стратегия. 2017. № 6. С. 30–36.
3. Баркалов С.А., Глушков А.Ю., Моисеев С.И. Динамическая модель разработки и реализации проекта под влиянием внешних факторов // Вестник Южно-Уральского государственного университета. Серия: компьютерные техноло-

- гии, управление, радиоэлектроника. 2020. Т. 20, № 3. С. 76–84. DOI: 10.14529/ctcr200308
4. Гладкова Ю.В., Гладков В.П. Этапы принятия управленческих решений // Вестник Пермского государственного технического университета. Электротехника, информационные технологии, системы управления. 2010. № 4. С. 39–44.
 5. Технологии информационного моделирования (BIM) как основа бережливого строительства / Гевара Рада Л.Т. [и др.] // Известия вузов. Инвестиции. Строительство. Недвижимость. 2022. Т. 12, № 1 (40). С. 70–81. DOI: 10.21285/2227-2917-2022-1-70-81
 6. Кирилова А.С., Карабейникова А.В., Софронова С.Н. Оценка надежности календарного планирования строительства инженерных сетей на основе метода критической цепи и метода критического пути // Молодой ученый. 2016. № 28 (132). С. 98–102.
 7. Колпачев В.Н., Семенов П.И., Михин П.В. Оптимизация календарного плана при ограниченных ресурсах // Известия Тульского государственного университета. 2004. № 7. С. 154–164.
 8. Маликов Д.З. Этапы разработки управленческих решений // Вестник науки. 2020. Т. 4, № 5 (26). С. 116–120.
 9. Кирилова А.С., Морозова Т.Ф., Косилов И.А. Оценка организационно технологической надежности моделей выполнения монолитных работ на типовом этаже // Региональные аспекты развития науки и образования архитектуры, строительства, землеустройства и кадастров в начале III тысячелетия. Научные чтения памяти профессора В.Б. Федосенко: материалы международной научно-практической конференции. Комсомольск-на-Амуре: КнАГТУ, 2015. С. 159–165.
 10. Оптимизация строительства: 4 стратегии реализации. URL: <https://bijsk.lcbit.ru/blog/optimizatsiya-stroitelstva-4-strategii-realizatsii/> (дата обращения: 13.07.2023).
 11. Осипов К.Ю. Оптимизационные методы планирования в строительстве // Молодой ученый. 2018. № 6 (192). С. 46–48.
 12. Свеженцев А.Г. Организационно-экономическое обеспечение развития атомной промышленности: дис. ... канд. экон. наук. Нижний Новгород, 2012. 175 с.
 13. Au S.K., Ching J., Beck J.L. Application of subset simulation method to reliability benchmark problems // Structural Safety. 2007. No. 29 (3). P. 183–193. DOI: 10.1016/j.strusafe.2006.07.008
 14. Mowade K., Shelar K. Lean construction // International Journal of Scientific Engineering and Science. 2017. Vol. 1, no. 11. P. 70–74.
 15. Rowinski L. Organizacia Procesow Budowlanych: Monograph. Warszawa: PWN, 1979. 520 p.

Получено 08.11.2023

Собин Артем Евгеньевич, аспирант кафедры информационных систем и технологий Московского финансово-юридического университета. 115191, Российская Федерация, г. Москва, ул. Серпуховский вал, 17, корп. 1. Тел. +7 977 787-88-32. E-mail: view7goog757@gmail.com

SOFTWARE OPTIMIZATION OF RESOURCE ALLOCATION PROCESS IN METRO CONSTRUCTION ORGANIZATIONS: ANALYSIS OF EFFICIENCY OF EXISTING METHODS

Sobin A.E.

Moscow University of Finance and Law, Moscow, Russian Federation

E-mail: view7goog757@gmail.com

This article explores the objectives of optimal resource allocation in metro construction organizations, including improving work efficiency, reducing project duration, and minimizing costs. The primary optimization methods, such as mathematical modeling, critical path analysis, and the use of network graphs, are discussed in the context of achieving these goals. The role of project management software in providing comprehensive control over planning, resources, and execution is emphasized. Additionally, the article underscores the significance of utilizing the Python programming language and its libraries, such as Pandas, Scikit-learn, and Matplotlib, in analyzing and optimizing resource allocation in metro construction. The paper contains a case study of a software implementation of risk analysis of delivery delays in metro construction using the Python programming language and its key libraries. Pandas for data processing and analysis, NumPy for performing complex mathematical calculations, and Matplotlib for visualizing the results were used.

Keywords: resources, optimization, optimization methods, metro construction companies, efficiency, costs

DOI: 10.18469/ikt.2023.21.3.09

Sobin Artem Evgenyevich, Moscow Finance and Law University, 17/1, Serpukhovskiy Val Street, Moscow, Russian Federation; PhD Student of Information Systems and Technologies Department, Tel. +7 977 787-88-32. E-mail: view7goog757@gmail.com

References

1. Agahanova K.A. Using BIM technologies for project cost management. *Nauka i biznes: puti razvitiya*, 2022, no. 7 (133), pp. 10–13. (In Russ.)
2. Barkalov S.A., Kurochka P.N. Formation of a managerial decision on the basis of the construction of complex estimates. *FES: Finansy. Ekonomika. Strategiya*, 2017, no. 6, pp. 30–36. (In Russ.)
3. Barkalov S.A., Glushkov A.Ju., Moiseev S.I. Dynamic model of development and implementation of the project under the influence of external factors. *Vestnik Yuzhno-Uralskogo gosudarstvennogo universiteta. Seriya: Kompyuternye Tekhnologii, Upravlenie, Radioelektronika*, 2020, vol. 20, no. 3, pp. 76–84. DOI: 10.14529/ctcr200308 (In Russ.)
4. Gladkova Ju.V., Gladkov V.P. Stages of management decision-making. *Vestnik Permskogo gosudarstvennogo tehnikeskogo universiteta. Jelektrotehnika, informacionnye tehnologii, sistemy upravleniya*, 2010, no. 4, pp. 39–44. (In Russ.)
5. Gevara Rada L.T. et al. Building information modelling (BIM) technology as a basis for lean construction. *Izvestiya vuzov. Investicii. Stroitelstvo. Nedvizhimost*, 2022, vol. 12, no. 1 (40), pp. 70–81. DOI: 10.21285/2227-2917-2022-1-70-81 (In Russ.)
6. Kirilova A.S., Karabejnikova A.V., Sofroneeva S.N. Reliability assessment of engineering networks construction scheduling based on the critical chain method and critical path method. *Molodoj uchenyj*, 2016, no. 28 (132), pp. 98–102. (In Russ.)
7. Kolpachev V.N., Semenov P.I., Mihin P.V. Optimization of the schedule plan at the limited resources. *Izvestiya Tulsckogo gosudarstvennogo universiteta*, 2004, no. 7, pp. 154–164. (In Russ.)
8. Malikov D.Z. Stages of development of management decisions. *Vestnik nauki*, 2020, vol. 4, no. 5 (26), pp. 116–120. (In Russ.)
9. Morozova T.F., Kosilov I.A., Kirilova A.S. Valuation of organizational and technological reliability of models of monolithic work execution on a typical floor. *Regional'nye aspekty razvitiya nauki i obrazovaniya arhitektury, stroitel'stva, zemleustrojstva i kadaastrov v nachale III tysyacheletiya. Nauchnye chteniya pamyati professora V.B. Fedosenko: materialy Mezhdunarodnoj nauchno-prakticheskoy konferencii*. Komsomol'sk-na-Amure: KnAGTU, 2015, 160 p. (In Russ.)
10. Optimizing Construction: 4 implementation strategies. URL: <https://bijsk.lcbit.ru/blog/optimizatsiya-stroitelstva-4-strategii-realizatsii/> (accessed: 13.07.2023). (In Russ.)
11. Osipov K.Ju. Optimization methods of planning in construction. *Molodoj uchenyj*, 2018, no. 6 (192), pp. 46–48. (In Russ.)
12. Svezhencev A.G. Organizational and economic support of the nuclear industry development: diss. ... cand. econ. sciences. Nizhnij Novgorod, 2012, 175 p. (In Russ.)
13. Au S.K., Ching J., Beck J.L. Application of subset simulation method to reliability benchmark problems. *Structural safety*, 2007, no. 29 (3), pp. 183–193. DOI: 10.1016/j.strusafe.2006.07.008
14. Mowade K., Shelar K. Lean construction. *International Journal of Scientific Engineering and Science*, 2017, vol. 1, no. 11, pp. 70–74.
15. Rowinski L. *Organizacja Procesow Budowlanych: Monograph*. Warszawa: PWN, 1979, 520 p.

Received 08.11.2023

ПРИМЕНЕНИЕ ИНФОРМАЦИОННО-КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ ДЛЯ ПОДДЕРЖАНИЯ ФИЗИЧЕСКОГО ЗДОРОВЬЯ ЧЕЛОВЕКА

Малина А.Б.^{1,2}, Голубкин Д.А.¹, Вехова А.Д.¹

¹Самарский государственный технический университет, Самара, РФ

²Самарский государственный социально-педагогический университет, Самара, РФ

E-mail: kuzdavletova_ab@mail.ru, cookie lord.public@gmail.com, aleksav2014@yandex.ru

Важным фактором успешной жизнедеятельности человека является здоровое состояние его организма. В статье рассматривается применение информационно-коммуникационных технологий для поддержания физического здоровья человека путем использования Телеграм-бота. Данная технология может улучшить качество жизни людей, имеющих потребность в поддержании своего физического благополучия. Благодаря простоте использования и большой распространенности мессенджера «Телеграм-бот» доступен широкому кругу пользователей, включая пожилых людей. Бот будет предоставлять пользователям индивидуальные рекомендации, основанные на анализе их физической активности, качества сна, рациона питания и уровня стресса. Кроме того, он будет помогать в планировании и отслеживании прогресса в достижении поставленных целей по улучшению здоровья. В статье приведены результаты предварительных исследований, подтверждающих эффективность предложенного подхода, который может существенно повысить уровень заботы о здоровье в цифровой эпохе.

Ключевые слова: телеграм-бот, информационно-коммуникационные технологии, здоровье, здоровый образ жизни, физическая активность, сон, календарь прививок

Введение

Одним из важнейших аспектов в жизни современного человека является поддержание здоровья. Способность организма выполнять свои функции, такие как движение, дыхание, пищеварение и прочие, напрямую связана с физическим здоровьем [9].

В современном мире есть множество факторов, способных негативно повлиять на здоровье людей. Экологические проблемы, высокий уровень стресса, недостаток физической активности и неправильное питание – все это может привести к ухудшению качества жизни и развитию различных заболеваний. Вот почему так важно разрабатывать и внедрять новые решения на основе информационно-коммуникационных технологий (ИКТ), способные помочь людям в поддержании собственного здоровья.

В области современного здравоохранения выделяют несколько основных направлений ИКТ [2]:

1. Телемедицина: использование ИКТ для удаленного консультирования, обследования и лечения больных.

2. Интернет вещей в здравоохранении: объединение устройств и датчиков для отслеживания здоровья пациентов в единую систему с помощью интернет-подключения.

3. Блокчейн в здравоохранении: применение технологии распределенного реестра в хранении и обмене медицинской информацией.

4. Искусственный интеллект в медицине: использование алгоритмов машинного обучения для анализа медицинских данных и диагностики заболеваний.

5. Мобильные приложения для здоровья: создание мобильных приложений для мониторинга здоровья, отслеживания тренировок и рациона питания, а также контроля качества сна.

В данной статье рассматривается последнее направление, а именно разработка Телеграм-бота (рисунок 1), предоставляющего индивидуальные рекомендации на основе анализа данных о физических нагрузках, качестве сна и рационе питания. Это поможет людям отслеживать динамику своего здоровья и достигать поставленных целей по его улучшению.

Телеграм-бот – это программа в мессенджере, работающая на основе автоматического совершения действий и выполняющая определенные задачи, которые ей задают разработчики. Боты могут быть созданы для самых разных целей: обработки запросов пользователей, автоматизации процессов, поиска информации, проведения опросов и т.д. В данном случае бот будет использоваться для поддержания физического здоровья пользователей.

Использование Телеграм-бота, в отличие от приложения, позволяет получить несколько преимуществ:

1. Доступность. Доступ к боту всегда можно получить из мессенджера, а для работы ему не требуется установка дополнительных пакетов. Это дает возможность использовать его даже при нехватке места на устройстве пользователя.

2. Эффективность. Ввиду отсутствия необходимости использования большого количества ресурсов устройства бот может быстрее функционировать.

3. Удобство разработки. При создании приложений значительные ресурсы уходит на разработку интерфейса пользователя. С этой точки зрения, создание бота является более выгодным, так как мессенджер содержит инструменты не только для создания интерфейса, но и для решения других задач, что ускоряет процесс разработки.

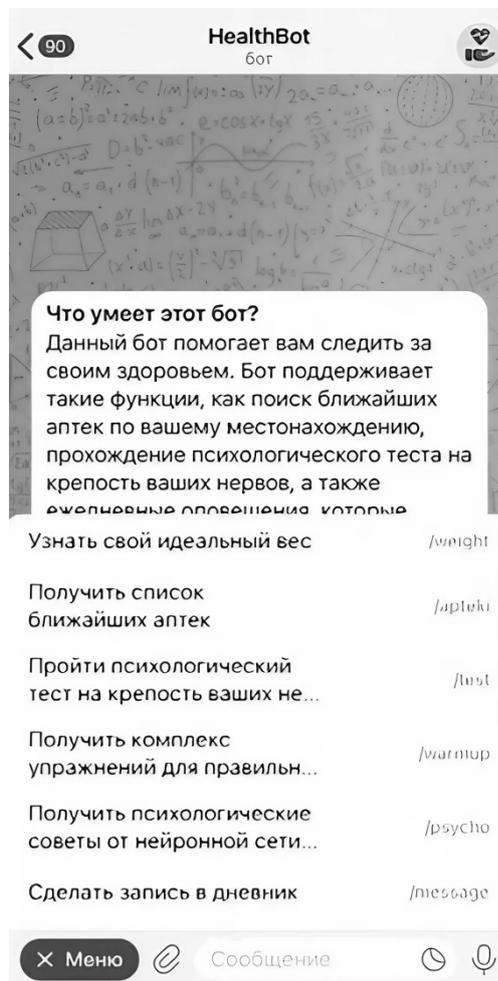


Рисунок 1. Вариант реализации интерфейса бота

На основе сказанного выше сделаем вывод о высокой актуальности исследования данной технологии.

Разработка Телеграм-бота

Когда речь заходит о здоровом образе жизни, стоит сначала уточнить, что значит «здоровье». По определению Всемирной организации здравоохранения здоровье – это не только отсутствие заболеваний и физических недостатков, но и состояние полного физического, ментального и социального благополучия. В наше время забота о здоровье и благополучии индивидуума направлена не на простое выживание, а на улучшение качества жизни и повышение удовлетворенности

ею. Это достигается путем улучшения самочувствия и общего жизненного тонуса.

Люди в современном мире все чаще сталкиваются с проблемой ведения здорового образа жизни. На здоровье человека в урбанизированном мире влияют различные факторы, включая качество воздуха, воды, пищи, доступность медицинских услуг и уровень социальной поддержки. Стресс, связанный с городской жизнью, также может негативно сказаться на здоровье, приводя к различным заболеваниям и психическим расстройствам. Ввиду высокой загруженности многие люди просто не уделяют должного внимания правильному питанию, забывают о необходимости регулярных физических нагрузок и не следят за качеством сна.

Для решения проблемы ведения здорового образа жизни предлагается реализация следующих функций Телеграм-бота:

- трекер питания;
- трекер сна;
- трекер физической активности;
- календарь прививок.

Стоит отметить, что на российском рынке присутствуют похожие продукты, такие как:

1. «Gymboss» – бот-тренер по фитнесу и здоровому питанию [11]. Позволяет подобрать индивидуальный план тренировок и питания в зависимости от целей пользователя. Однако бот не позволяет использовать данные с внешних устройств (например, фитнес-браслетов).

2. «Gym Ru Bot» – бот для похожих целей, однако с ограниченным функционалом: отсутствие аналитики не предполагает динамическое изменение плана тренировок, а отсутствие советов по контролю рациона может снизить их эффективность [12].

3. «KRV Sport» – бот, предоставляющий советы по питанию и ведению тренировок без возможности сохранения статистики, что не позволяет отслеживать динамику прогресса, а из-за отсутствия регулярных напоминаний пользователь может вовсе перестать использовать бота [13].

4. «Sleepy» – бот для улучшения качества сна [14]. Имеет внушительный функционал по контролю и улучшению сна, не содержит дополнительных функций.

Как видно из сказанного выше, преимущества данного бота над аналогичными продуктами заключается в индивидуальном подходе к каждому пользователю, а также предоставлении клиентам целой экосистемы для поддержания здоровья (таблица 1).

Трекер питания – это приложение функция, позволяющая пользователю контролировать свое питание.

Таблица 1. Краткий сравнительный анализ ботов

| | HelthBot | Gymboss | Gym Ru Bot | Слипи | KRV Sport |
|------------------------|----------|---------|------------|-------|-----------|
| Ежедневные напоминания | + | + | + | - | - |
| Ведение статистики | + | + | + | + | - |
| Аналитика | + | + | - | + | - |
| Трекер питания | + | + | - | - | + |
| Трекер сна | + | - | - | + | - |
| Трекер активности | + | + | + | - | + |
| Календарь прививок | + | - | - | - | - |

Таблица 2. Дневные нормы калорий в зависимости от физической активности [8]

| Группы населения по интенсивности труда | Возраст | Мужчины | | | | Женщины | | | |
|--|---------|----------------|------------------|------------------------------------|------------------|----------------|------------------|------------------------------------|------------------|
| | | Большие города | Небольшие города | Дополнительная физическая нагрузка | | Большие города | Небольшие города | Дополнительная физическая нагрузка | |
| | | | | Большие города | Небольшие города | | | Большие города | Небольшие города |
| Лица, работа которых не связана с затратами физического труда (работники умственного труда, служащие) | 18–40 | 2800 | 3000 | 3100 | 3300 | 2400 | 2600 | 2650 | 2850 |
| | 40–60 | 2600 | 2800 | 2800 | 3000 | 2200 | 2400 | 2350 | 2550 |
| Работники механизированного труда и сферы обслуживания (телефонисты, продавцы и т.п.) | 18–40 | 3000 | 3200 | 3300 | 3500 | 2550 | 2750 | 2800 | 3000 |
| | 40–60 | 2800 | 3000 | 3000 | 3200 | 2350 | 2550 | 2500 | 2700 |
| Работники механизированного труда, прикладывающие значительные физические усилия (станочники, текстильщики и т.п.) | 18–40 | 3200 | 3400 | 3500 | 3700 | 2700 | 2900 | 2950 | 3150 |
| | 40–60 | 2900 | 3100 | 3100 | 3300 | 2500 | 2700 | 2650 | 2850 |
| Работники частично механизированного труда (шахтёры, механизаторы, металлурги и т.п.) | 18–40 | 3700 | 3900 | 4000 | 4200 | 3150 | 3350 | 3400 | 3600 |
| | 40–60 | 3400 | 3600 | 3600 | 3800 | 2900 | 3100 | 3050 | 3250 |
| Люди пожилого возраста | 60–70 | 2350 | 2500 | 2500 | 2650 | 2100 | 2200 | 2200 | 2500 |
| | >70 | 2200 | | | | 2000 | | | |

Ученые дают следующие советы для улучшения ежедневного рациона [8]:

1. Сбалансированное питание. Питание должно быть сбалансированным, то есть включать в себя достаточное количество белков, жиров, углеводов, витаминов и минералов. Также необходимо соблюдать нормы потребления калорий. В зависимости от пола, возраста, места проживания и других факторов которые нормы для каждого человека рассчитываются индивидуально (таблица 2).

2. Разнообразие. Для получения всех необходимых питательных веществ следует потреблять продукты из разных групп. Для белков это мясо, рыба, яйца, орехи, семена и бобовые. Для жиров – мононенасыщенные жиры (оливковое масло, авокадо, некоторые орехи), стоит избегать потребления насыщенных жиров и трансжиров. Для углеводов – цельнозерновые продукты (овес, гречка, овсянка, коричневый рис, хлеб из цельного зерна). Для витаминов и минералов – фрукты, овощи, орехи.

3. Соблюдение пропорций. Голодание так же вредно для организма, как и переедание. Ученые рекомендуют потреблять примерно 0,8 грамма белка на килограмм веса, 20–35% калорий из жиров, а также 45–65% калорий из углеводов [7]. Регулярность приема пищи также способствует улучшению пищеварения.

4. Ограничение употребления сахара и насыщенных жиров. Такие продукты как шоколад, десерты, цельномолочные продукты, обработанное мясо и т.п. могут привести к развитию сердечно-сосудистых заболеваний и диабета.

5. Контроль потребления воды. Вода по-

могает поддерживать гидратацию организма, улучшает пищеварение, способствует выведению вредных веществ, регулирует температуру тела. Обезвоживание может привести к головным болям, усталости, проблемам с сердцем и почками [4].

Бот будет руководствоваться вышеописанными правилами, что позволит скорректировать рацион пользователя или предложить новый на основе его предпочтений и образа жизни. Также бот будет напоминать о своевременном приеме пищи с помощью уведомлений на устройство пользователя.

Трекер сна – это функция бота для отслеживания качества сна пользователя. Она состоит из двух частей: браслета, который носится на запястье, и алгоритма анализа полученных с устройства данных. Браслет оснащен акселерометром, пульсометром и звукозаписывающим датчиком, предназначенных для определения времени засыпания и пробуждения, фазы сна, а также отслеживания движений во сне [6].

Акселерометр – это устройство, измеряющее ускорение. Он состоит из пружины, которая может перемещаться внутри корпуса. Когда пользователь засыпает, пружина остается неподвижной. Когда пользователь просыпается и начинает двигаться, пружина также начинает двигаться. Это движение регистрируется акселерометром и передается на процессор, который обрабатывает данные и передает их на смартфон. На основе этих данных бот может определить, когда пользователь заснул и проснулся, а также сколько времени он провел во сне.

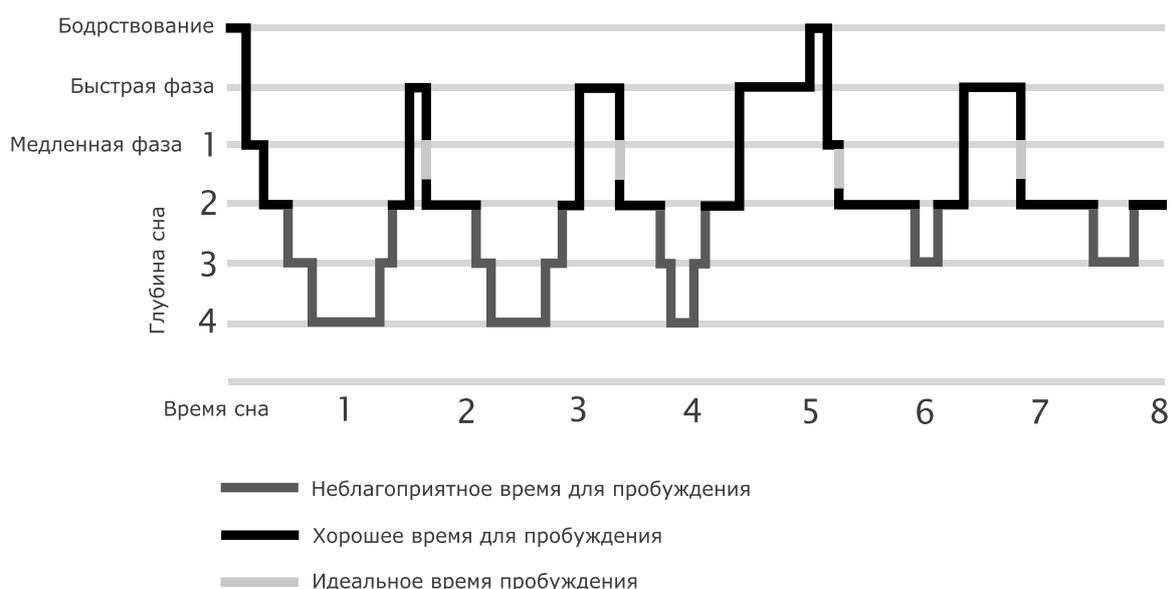


Рисунок 2. Фазы сна человека [5]

Пульсометр – это устройство, предназначенное для измерения частоты сердечных сокращений. Прибор использует метод оптической плетизмографии: метод измерения кровяного давления без прямого контакта с артерией [1]. Измерения происходят при помощи световых датчиков, расположенных на обратной стороне устройства так, чтобы осуществлялся контакт с кожей. Светодиоды излучают свет на кожный покров, а световой сенсор улавливает уровень его поглощения. По этим параметрам можно определить частоту сердечного ритма, что в свою очередь позволяет фиксировать переходы из одной фазы сна в другую.

Чувствительный звукозаписывающий датчик способен улавливать издаваемые человеком во время сна звуки, что позволяет не только фиксировать межфазовые переходы, но и анализировать храп пользователя, диагностируя апноэ сна, и регистрировать остановки дыхания.

Информация с этих датчиков передается на смартфон пользователя, а затем Телеграм-боту, который анализирует полученные данные и делает выводы о качестве сна, его продолжительности и легкости пробуждения. Бот также способен выбрать оптимальное время для пробуждения пользователя, когда организм физически готов проснуться (рисунок 2). Если человеку нужно встать в конкретное время, то умный будильник предлагает установить временной интервал, в

пределах которого бот автоматически определит наилучший момент для пробуждения.

Трекер физической активности – это функция бота для отслеживания регулярной физической активности пользователя. Введение данной функции обусловлено тем, что большое количество людей, особенно в урбанизированных странах, ведут сидячий образ жизни. Это может привести к ряду проблем, включая ожирение, сердечно-сосудистые заболевания, диабет и даже некоторые виды рака [3].

Бот позволяет определять количество пройденных пользователем шагов несколькими способами:

Использование данных о местоположении телефона. Бот использует GPS-данные телефона пользователя, чтобы определить, где он находится и сколько шагов прошел. Это простое решение, так как не требует дополнительных устройств, однако этот метод может снизить время работы батареи телефона. Такое решение будет полезно для тех, кто хочет отслеживать свою физическую активность во время прогулок или пробежек.

Запрос у пользователя информации о количестве пройденных шагов или затраченном на активность времени. Данный метод предполагает ручной ввод данных и может быть полезным для тех, у кого низкая физическая активность, однако для большинства пользователей такой вариант, вероятно, будет неудобным.



Рисунок 3. Частота занятий различными видами физической активности [10]

Интеграция с фитнес-трекерами. В отличие от первых двух вариантов, где данные об активности являются оценочными, а использование дополнительных устройств не предполагается, здесь предлагается считывать данные с умных часов, фитнес-браслетов и других устройств, оснащенных различными датчиками отслеживания состояния человека. Этот метод отлично подойдет тем, кто уделяет много времени физическим нагрузкам, избавит от необходимости ручного ввода, а также предоставит более точную аналитику.

Вне зависимости от метода сбора данных, бот предоставляет следующие возможности:

1. Отслеживание прогресса пользователя в достижении своих целей по физической активности.
2. Предоставление рекомендаций по улучшению физической активности и здоровья, согласно общепринятым рекомендациям (рисунок 3).
3. Анализ данных для определения лучших практик и стратегий в области здоровья и фитнеса.
4. Отправка уведомлений и напоминаний о необходимости физической активности или здоровом образе жизни.
5. Создание отчетов и аналитики для изучения тенденций в области здоровья и физической активности.

Календарь прививок – это функция бота, предоставляющая пользователю информацию о рекомендуемом графике прививок. График индивидуален для каждого пользователя и составляется с учетом его возраста, потребностей и прочей необходимой информации (список сделанных прививок, дата последней прививки и т.д.).

Календарь прививок имеет следующие функции:

1. Ведение базы данных по прививкам. Бот хранит информацию о сделанных пользователем прививках с указанием даты, типа прививки и реакции организма.
2. Расчет интервала между прививками. Бот определяет оптимальное время между прививками на основе рекомендаций Минздрава и ВОЗ.
3. Напоминание о прививке. Бот отправляет уведомление пользователю о необходимости сделать следующую прививку.
4. Интеграция с другими сервисами. Календарь связан другими функциями бота, такими как отслеживание физической активности и контроль питания, что помогает пользователю видеть, как его привычки в области здоровья влияют на график прививок.

Заключение

В данной статье был рассмотрен Телеграм-бот для поддержания физического здо-

ровья, его основные функции и возможности. Бот способен отслеживать физическую активность пользователей, помогать контролировать питание и соблюдать режим сна. Данные инструменты будут способствовать улучшению здоровья человека и выработке полезных привычек. Бот также имеет функцию «Календарь прививок», с помощью которой пользователи смогут следить за графиком проведения вакцинаций и получать своевременные уведомления о нужных прививках.

Результаты этого исследования могут помочь в создании эффективных мер для поддержания здоровья, улучшения общего самочувствия людей и повышения их уровня жизни. В дальнейшем планируется расширение функционала бота и доработка его возможностей.

Литература

1. Галкин М.А., Семикин Г.И. Аппаратно-программное обеспечение диагностики сердечно-сосудистой системы человека с применением метода оптической плетизмографии высокого разрешения // Гуманитарный вестник. 2013. № 12 (14). С. 2.
2. Вялков А.И. Информационные технологии в сфере здравоохранения // Управление и экономика здравоохранения. М.: ГЭОТАР-Медиа, 2009. С. 8–9.
3. Рекомендации ВОЗ по вопросам физической активности и малоподвижного образа жизни: краткий обзор. Женева: Всемирная организация здравоохранения, 2020. URL: <https://chocmp.ru/wp-content/uploads/2022/07/9789240014909-rus.pdf> (дата обращения: 15.11.2023).
4. Ермолаева Е.Л., Грибина Г.А. О значении воды для человека // Международный студенческий научный вестник. 2018. № 6. С. 53.
5. Ковальзон В.М., Долгих В.В. Регуляция цикла бодрствование-сон // Неврологический журнал. 2016. № 6. С. 316–322.
6. Что такое трекер сна и как он работает? URL: <https://www.askona.ru/blog/suggestions/chto-takoe-treker-sna-i-kak-on-rabotaet/> (дата обращения: 11.11.2023).
7. Сбалансированное питание: сколько калорий нужно есть в сутки? URL: <https://profilaktica.ru/for-population/profilaktika-zabolevaniy/vse-opravilnom-pitanii/skolko-kaloriy-nuzhno-est-v-den/> (дата обращения: 12.11.2023).
8. Здоровое питание. URL: <https://www.who.int/ru/news-room/fact-sheets/detail/healthy-diet> (дата обращения: 12.11.2023)
9. Физическая активность. URL: <https://www.who>

- int/ru/news-room/fact-sheets/detail/physical-activity (дата обращения: 13.11.2023).
10. Неделя информирования о важности физической активности. URL: <https://volosovocrb.ru/novosti/404-nedelya-informirovaniya-o-vazhnosti-fizicheskoy-aktivnosti.html> (дата обращения: 13.11.2023).
11. Бот-тренер по фитнесу и здоровому питанию. URL: <https://t.me/gymbosbot> (дата обращения: 15.10.2023).
12. Бот для проведения тренировок. URL: https://t.me/gym_ru_bot (дата обращения: 15.10.2023).
13. Бот, предоставляющий советы по питанию и ведению тренировок. URL: https://t.me/Sport_nutrition_Bot (дата обращения: 15.10.2023).
14. Чат бот для улучшения качества сна. URL: <https://t.me/sleepyrobot> (дата обращения: 15.10.2023).

Получено 22.11.2023

Малина Алиа Бахтагалиевна, к.э.н., доцент кафедры информационных технологий (ИТ) Самарского государственного технического университета (СамГТУ). 443100, Российская Федерация, г. Самара, ул. Молодогвардейская, 244; магистрант кафедры информатики, прикладной математики и методики их преподавания Самарского государственного социально-педагогического университета. 443099, Российская Федерация, г. Самара, ул. М. Горького, 65/67. Тел. +7 846 278-43-11. E-mail: kuzdavletova_ab@mail.ru

Голубкин Дмитрий Александрович, студент кафедры ИТ СамГТУ. 443100, Российская Федерация, г. Самара, ул. Молодогвардейская, 244. Тел. +7 846 278-43-11. E-mail: cookie lord.public@gmail.com

Вехова Александра Данииловна, студент кафедры ИТ СамГТУ. 443100, Российская Федерация, г. Самара, ул. Молодогвардейская, 244. Тел. +7 846 278-43-11. E-mail: aleksav2014@yandex.ru

THE USE OF INFORMATION AND COMMUNICATION TECHNOLOGIES TO MAINTAIN HUMAN PHYSICAL HEALTH

Malina A.B.^{1,2}, Golubkin D.A.¹, Vekhova A.D.¹

¹Samara State Technical University, Samara, Russian Federation

²Samara State Social and Pedagogical University, Samara, Russian Federation

An important factor in the successful functioning of a person is the healthy state of their body. The article discusses the use of information and communication technologies to maintain human physical health through the use of a Telegram bot. This technology can improve the quality of life of people who need to maintain their physical well-being. Due to the ease of use and high prevalence of the Telegram messenger, the bot is available to a wide range of users, including the elderly. The bot will provide users with personalized recommendations based on an analysis of their physical activity, sleep quality, diet and stress levels. In addition, it will help in planning and tracking progress towards achieving the set goals for improving health. The article presents the results of preliminary studies confirming the effectiveness of the proposed approach, which can significantly increase the level of health care in the digital age.

Keywords: *telegram bot, development, health, healthy lifestyle, physical activity, sleep, vaccination calendar*

DOI: 10.18469/ikt.2023.21.3.10

Malina Aliya Bakhtagalievna, Samara State Technical University, 244, Molodogvardeyskaya Street, Samara, 443100, Russian Federation; Associate Professor of Information Systems Department, PhD in Economics. Samara State Social and Pedagogical University, 65/67, M. Gorky Street, Samara, 443099, Russian Federation; Master's Degree Student of Informatics, Applied Mathematics and Methods of their Teaching Department. Tel. +7 846 278-43-11. E-mail: kuzdavletova_ab@mail.ru

Golubkin Dmitry Aleksandrovich, Samara State Technical University, 244, Molodogvardeyskaya Street, Samara, 443100, Russian Federation; Student of Information Systems Department. Tel. +7 846 278-43-11. E-mail: cookie lord.public@gmail.com

Vekhova Alexandra Danilovna, Samara State Technical University, 244, Molodogvardeyskaya Street, Samara, 443100, Russian Federation; Student of Information Systems Department. Tel. +7 846 278-43-11. E-mail: aleksav2014@yandex.ru

References

1. Galkin M.A., Semikin G.J. Hardware and software for diagnostics of human cardiovascular system using method of optical plethysmography of high resolution. *Gumanitarnyj vestnik*, 2013, no. 12 (14), pp. 2. (In Russ.)
2. Vyalkov A.I. Information technologies in healthcare. *Upravlenie i ekonomika zdavoohraneniya*. Moscow: GEOTAR-Media, 2009, pp. 8–9. (In Russ.)
3. *World Health Organization Recommendations on Physical Activity and Sedentary Lifestyle: A Brief Review*. Geneva: World Health Organization, 2020. URL: <https://chocmp.ru/wp-content/uploads/2022/07/9789240014909-rus.pdf> (accessed: 15.11.2023). (In Russ.)
4. Ermolaeva E.L., Gribina G.A. The Importance of water for humans. *Mezhdunarodnyj studencheskij nauchnyj vestnik*, 2018, no. 6, pp. 53. (In Russ.)
5. Kovalzon V.M., Dolgikh V.V. Regulation of the wakefulness-sleep cycle. *Nevrologicheskij zhurnal*, 2016, no. 6, pp. 316–322. (In Russ.)
6. What is a sleep tracker and how does it work? URL: <https://www.askona.ru/blog/suggestions/chto-takoe-treker-sna-i-kak-on-rabotaet/> (accessed: 11.11.2023). (In Russ.)
7. Balanced diet: how many calories should i eat per day? URL: <https://profilaktica.ru/for-population/profilaktika-zabolevaniy/vse-o-pravilnom-pitanii/skolko-kaloriy-nuzhno-est-v-den/> (accessed: 12.11.2023). (In Russ.)
8. Healthy eating. URL: <https://www.who.int/ru/news-room/fact-sheets/detail/healthy-diet> (accessed: 12.11.2023). (In Russ.)
9. Physical activity. URL: <https://www.who.int/ru/news-room/fact-sheets/detail/physical-activity> (accessed: 13.11.2023). (In Russ.)
10. Physical activity awareness week. URL: <https://volosovocrb.ru/novosti/404-nedelya-informirovaniya-o-vazhnosti-fizicheskoy-aktivnosti.html> (accessed: 13.11.2023). (In Russ.)
11. Fitness and healthy eating trainer bot. URL: <https://t.me/gymbosbot> (accessed: 15.10.2023). (In Russ.)
12. Bot for conducting workouts. URL: https://t.me/gym_ru_bot (accessed: 15.10.2023). (In Russ.)
13. Bot providing advice on nutrition and workouts. URL: https://t.me/Sport_nutrition_Bot (accessed: 15.10.2023). (In Russ.)
14. Sleep quality improvement chatbot. URL: <https://t.me/sleepyrobot> (accessed: 15.10.2023). (In Russ.)

Received 22.11.2023

УДК 372.811.1

РАЗРАБОТКА ПРИЛОЖЕНИЯ ДЛЯ АНАЛИЗА КОММЕНТАРИЕВ АНГЛОЯЗЫЧНОГО YOUTUBE

Дукальская И.В., Киргизов А.Г.

Поволжский государственный университет телекоммуникаций и информатики, Самара, РФ

E-mail: i.dukalskaya@psuti.ru, kirgizoffand@yandex.ru

Данная статья представляет разработку инновационного приложения, цель которого – сбор и анализ комментариев на англоязычных видеороликах YouTube с целью выявления и систематизации наиболее часто употребляемых англоязычных аббревиатур. Это приложение не только облегчает процесс изучения английского языка, но и предоставляет глубокое понимание использования аббревиатур в контексте реальных диалогов, способствуя более эффективному их использованию в повседневной коммуникации. Позволяет взглянуть на английский язык со стороны носителя. Подробно рассматриваются особенности разработки приложения, такие как использование YouTube Data API, анализ текста комментариев с помощью регулярных выражений и сохранение данных для

дальнейшего изучения. Данная работа уделяет внимание вызовам при работе с API (Application Programming Interface), методам обработки текста и показывает возможности исследования английского языка в онлайн-среде.

Ключевые слова: методы извлечения данных, сохранение данных, анализ комментариев, изучение английского языка, обработка текста, онлайн платформы, англоязычные аббревиатуры

Введение

В условиях современного мира, где культурные преграды сокращаются благодаря глобализации и интернету, изучение иностранного языка становится ключевой целью. Чтобы обогатить свой словарный запас и научиться понимать язык, мы ищем новые методы изучения. Один из перспективных подходов – анализ реальных диалогов и комментариев на англоязычных платформах, таких как, например, YouTube. Это не только открывает нам доступ к разнообразным языковым проявлениям, но и позволяет изучать язык в контексте реальных общений. В данной статье речь пойдет о разработке инновационного приложения, специализирующегося на анализе комментариев к видеороликам на YouTube. Это приложение предлагает возможность выявления и выделения наиболее часто используемых англоязычных сокращений. Оно не только облегчает процесс изучения английского языка, но и дает более глубокое понимание использования сокращений в реальных диалогах, что помогает более эффективно освоить их для повседневного использования.

Это приложение является не только полезным инструментом для изучения языка, но также открывает новые возможности для исследования и использования английского языка. Позволяет посмотреть на его некоторые аспекты новым взглядом [4; 5].

Особенности разработки приложения

YouTube Data API (Application Programming Interface) является основой для функционирования приложения. Это API предоставляет доступ к базе данных комментариев, оставленных на англоязычных видеороликах YouTube. Приложение использует ключ разработчика (`developer_key`), получаемый каждым пользователем индивидуально, и уникальный идентификатор видеоролика (`video_id`), чтобы понимать, с каким конкретным видеороликом идет работа. Благодаря всему этому осуществляются запросы к API с помощью библиотеки `googleApiClient` и обеспечивается эффективный и быстрый сбор информации о комментариях (рисунок 1) [7].

Одной из главных особенностей являются ограничения, установленные самим YouTube на частоту запросов. Необходимо контролировать

количество запросов к API, чтобы избежать блокировок или ограничений доступа к данным. Также важно учитывать, что некоторые данные могут быть ограничены, что может повлиять на полноту или доступность информации при использовании API.

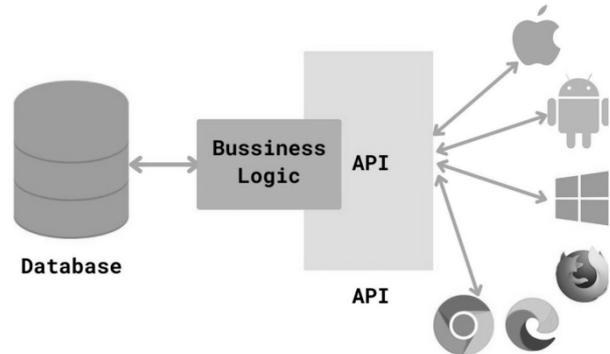


Рисунок 1. Схема API

Еще одной трудностью при работе с API является обработка большого объема данных, особенно при загрузке комментариев и ответов к ним. Необходимо эффективно организовывать код и процесс сбора данных, чтобы избежать возможных ошибок и обеспечить оптимальную производительность приложения.

Кроме того, важно учитывать обновления и изменения в API, так как YouTube периодически вносит изменения в свои правила и функционал. Это может потребовать соответствующих обновлений в коде приложения для поддержания его работоспособности и соответствия актуальным требованиям платформы. Для решения данной проблемы был использован такой способ создания приложения как версионирование. При каждом обновлении приложения выпускается новая версия. Устаревшие методы помечаются как `deprecated`, а новые вносятся в документацию. Тем самым мы получаем новый функционал, не нарушая работу старого.

Таким образом, использование YouTube Data API предоставляет важные возможности для приложения, но требует внимательного подхода к управлению запросами, обработке данных и поддержанию соответствия изменениям в API, чтобы обеспечить надежную и эффективную работу приложения.

Одним из ключевых направлений деятельности приложения является глубокий анализ текста комментариев, оставленных на англоязычных видео-

роликах на платформе YouTube. Основной целью является извлечение и систематизация англоязычных аббревиатур и сокращений из этих комментариев, с целью выявления наиболее часто встречающихся языковых элементов. Для достижения этой цели прибегаем к применению методов обработки текста и регулярных выражений, что позволяет нам точно выделять и анализировать аббревиатуры, используемые в комментариях.

Процесс анализа текста комментариев включает в себя создание специализированных алгоритмов, способных эффективно выявлять и извлекать аббревиатуры из текстового контента. Уделяется внимание разработке и оптимизации этих алгоритмов, чтобы обеспечить точность и полноту извлечения аббревиатур, а также эффективность обработки больших объемов текста.

Основными инструментами для извлечения аббревиатур и сокращений являются регулярные выражения, позволяющие проводить гибкий и точный поиск заданных шаблонов текста. Используются разнообразные шаблоны регулярных выражений с учетом различных вариантов написания и использования аббревиатур, что позволяет нам оптимально охватить разнообразие аббревиатур, применяемых в комментариях на YouTube.

Однако анализ текста комментариев не лишен сложностей. Разнообразие стилей и контекстов использования аббревиатур, а также их многообразие и динамичность могут представлять вызовы при точной классификации и интерпретации. Это требует постоянного совершенствования методов обработки текста и алго-

ритмов извлечения данных для более точного и всестороннего анализа.

Таким образом, приложение активно исследует и анализирует текстовый контент комментариев, используя методы обработки текста и регулярные выражения для точного выделения англоязычных аббревиатур и сокращений, несмотря на некоторые сложности и особенности данного процесса.

Проанализированные данные сохраняются в файл в формате CSV (формат Microsoft Excel). В этом файле содержится информация о комментариях, включая `etag`, `id`, текст комментария, автор и другие сведения. Эти данные могут быть использованы для дальнейшего изучения и анализа.

Процесс разработки

Процесс сбора комментариев с англоязычных видеороликов на YouTube начинается с использования YouTube Data API в рамках разработанного приложения на языке программирования Python.

Инициализация API и отправка запросов: первым шагом в процессе сбора комментариев является инициализация YouTube Data API с помощью ключа разработчика (DEVELOPER_KEY). Посредством вызова метода `youtube.commentThreads().list()` с указанием параметров, таких как `part`, `maxResults`, `pageToken`, и `videoId` для определения идентификатора видеоролика (VIDEO_ID), осуществляется запрос к API для получения корневых комментариев.

Циклический процесс загрузки комментариев: после отправки первого запроса на получение комментариев для указанного видеоролика про-

| | |
|----|--|
| 1 | textOriginal |
| 2 | Book an English class on LiveXP for \$0.99: https://bit.ly/livexp_antonio Use my discount code ANTONIO30 to get 30% off. 🇺🇸 |
| 3 | I am lowkey watching your all videos . Is that good Anton? |
| 4 | Your explanation was legit! 🤩 |
| 5 | Thanks |
| 6 | Nice ☑️ |
| 7 | Love it ❤️ |
| 8 | The origin of to bail is the act of jumping out of an airplane, with a parachute we hope. |
| 9 | You are excellent |
| 10 | I really like all your videos. You should be a teacher If you are not already one |
| 11 | ❤️❤️❤️ |
| 12 | Thanks 😊 |
| 13 | I like your short hair! |
| 14 | I super thanks you. |
| 15 | Your videos are absolutely amazing - congrats! |
| 16 | Thanks for sharing these useful slang words your voice is so nice and clear .I really love your videos. |

Рисунок 2. Пример сохранения комментариев

исходит циклический процесс, в котором используется параметр `nextPageToken` для получения последующих страниц комментариев. Для этого повторяем запросы, указывая `pageToken` равным значению `nextPageToken`, чтобы получить следующую порцию комментариев. Это продолжается до тех пор, пока `nextPageToken` не станет равным `None`, что означает достижение конца списка комментариев.

Сохранение комментариев: полученные комментарии сохраняются в переменную `items`. Каждый комментарий представлен в формате JSON и содержит различные атрибуты, такие как идентификаторы, текст комментария, информацию об авторе и времени публикации.

Дополнительная загрузка ответов на комментарии (далее – «реплаев»): помимо извлечения корневых комментариев, приложение проверяет наличие реплаев к каждому комментарию. Если у корневого комментария есть реплаи, используется аналогичный процесс для загрузки реплаев на основании их `parentId`, чтобы получить дополнительные комментарии, относящиеся к ним.

Сохранение данных в файл CSV: по завершении сбора комментариев и их реплаев приложение сохраняет полученные данные в файл формата CSV. Для этого используется модуль `csv` для создания файла `youtuberresults.csv` и записи всех собранных комментариев и реплаев в этот файл с определенной структурой и разделителями между данными (рисунок 2) [9].

Таким образом, процесс сбора комментариев с англоязычных YouTube-видеороликов осуществляется путем систематического использования YouTube Data API, опираясь на различные параметры запросов и обработку полученных данных для сохранения их в файле для последующего анализа и обработки.

Пример использования

Для того чтобы протестировать приложение, было выбрано 3 видео с платформы YouTube по различной тематике. Для процесса анализа текста был написан скрипт, который скачивает список аббревиатур и сленга с OpenApi NLTK, ищет количество вхождений слов во всех комментариях, считает их соотношение к общему тексту [8; 10]. Пример работы кода с тестовым видео:

По результату первого анализа:

Slang Words, their Percentage, and Counts:

Omg: 0.03% (1 occurrences)

Lol: 0.03% (1 occurrences)

bro: 0.03% (1 occurrences)

Dude: 0.03% (1 occurrences)

Total Words: 3,095

По результату второго анализа:

lol: 0.11% (4 occurrences)

Bro: 0.03% (1 occurrences)

idk: 0.03% (1 occurrences)

Lol: 0.06% (2 occurrences)

mate: 0.03% (1 occurrences)

bro: 0.06% (2 occurrences)

Brb: 0.03% (1 occurrences)

dude: 0.03% (1 occurrences)

lmao: 0.03% (1 occurrences)

OMG: 0.03% (1 occurrences)

Idk: 0.03% (1 occurrences)

Dude: 0.03% (1 occurrences)

Total Words: 3,556

По результату третьего анализа:

Slang Words, their Percentage, and Counts:

lol: 0.01% (1 occurrences)

CHILL: 0.01% (1 occurrences)

dude: 0.02% (2 occurrences)

chill: 0.05% (5 occurrences)

Lol: 0.01% (1 occurrences)

OMG: 0.02% (2 occurrences)

Omg: 0.01% (1 occurrences)

Props: 0.01% (1 occurrences)

Total Words: 10,530

В данной статистике видно, что большей популярностью пользуются сокращения для неформальной речи. Далее мы можем узнать их значение, посмотреть примеры использования и начать употреблять их в своей речи, тем самым улучшая свой уровень английского. Таким же образом можно анализировать технические и профессиональные видео, собирая тематическую лексику [1; 2; 3].

Заключение

Это приложение, совмещающее в себе технологии обработки текста и доступ к обширной базе данных комментариев на YouTube, открывает перед пользователями уникальную возможность погружения в англоязычную среду. Анализ комментариев на основе современных методов обработки текста и регулярных выражений позволяет выделить, систематизировать и изучить разнообразные аббревиатуры, активно используемые в реальных диалогах пользователей на этой платформе [6].

Это приложение не только обогащает словарный запас изучающих английский язык, но и стимулирует интерес к языку через интерактивное и практическое изучение. Анализ аббревиатур, наиболее часто встречающихся

в реальных комментариях, предоставляет не только знания о сленге и сокращениях, но и глубокое понимание контекста и специфики языка, используемого в повседневной коммуникации.

Эффективное использование современных методов анализа текста и доступ к обширной базе данных комментариев на YouTube делает это приложение значимым инструментом в контексте образования и изучения языка. Объединение технологии и образования позволяет приложению предложить новый и инновационный подход к изучению английского языка, делая этот процесс более интересным, доступным и эффективным для всех желающих.

Литература

1. Peters P. The Cambridge Dictionary of English Grammar. Cambridge: Cambridge University, 2004. 512 p.
2. Adams M. Slang: The People's Poetry. Oxford: Oxford University, 2009. 240 p.
3. Joshi M. Abbreviations in English. Independently published, 2019. 120 p.
4. 79 Incredible English Slang Words to Help you Understand Native Speakers. URL: <https://storylearning.com/blog/english-slang-words> (дата обращения: 20.11.23).
5. Slang Words: List of 100 Common Slang Words & Phrases you Need to Know! URL: <https://eslforums.com/slang-words/> (дата обращения: 20.11.23).
6. Re – Regular Expression Operations. URL: <https://docs.python.org/3/library/re.html> (дата обращения: 20.11.23).
7. Добавьте функциональность YouTube в свое приложение. URL: <https://developers.google.com/youtube/v3?hl=ru> (дата обращения: 22.11.23).
8. NLTK Documentation. Natural Language Toolkit. URL: <https://www.nltk.org> (дата обращения: 22.11.23).
9. CSV File Reading and Writing. URL: <https://docs.python.org/3/library/csv.html> (дата обращения: 23.11.23).
10. How to Learn Big Data? Step by Step Roadmap. URL: <https://www.mltut.com/how-to-learn-big-data-step-by-step/> (дата обращения: 24.11.23).

Получено 15.01.2024

Дукальская Ирина Владимировна, к.филол.н., доцент, доцент кафедры иностранных языков Поволжского государственного университета телекоммуникаций и информатики (ПГУТИ). 443090, Российская Федерация, г. Самара, Московское шоссе, 77. Тел. +7 927 204-77-93. E-mail: i.dukalskaya@psuti.ru

Киргизов Андрей Геннадьевич, студент кафедры программной инженерии ПГУТИ. 443090, Российская Федерация, г. Самара, Московское шоссе, 77. Тел. +7 932 861-85-21. E-mail: kirgizoffand@yandex.ru

DEVELOPMENT OF AN APPLICATION FOR ANALYZING COMMENTS ON ENGLISH-LANGUAGE YOUTUBE

Dukalskaya I.V., Kirgizov A.G.

Povolzhskiy State University of Telecommunications and Informatics, Samara, Russian Federation

E-mail: i.dukalskaya@psuti.ru, kirgizoffand@yandex.ru

This article introduces the development of an innovative application aimed at analyzing comments on English-language YouTube videos to identify and systematize the most frequently used English abbreviations. This application not only facilitates the process of learning English, but also provides a deep understanding of the use of abbreviations in the context of real dialogues, contributing to a more effective use of them in everyday communication. It provides a native speaker's perspective on the English language. Details the features of developing an application, such as using the YouTube Data API, analyzing comment text using regular expressions and saving data for further study. This paper focuses on the challenges of working with the API (Application Programming Interface), text processing techniques and shows the possibilities of exploring English in an online environment.

Keywords: *data extraction methods, data storage, comment analysis, English language learning, text processing, online platforms, English-language abbreviations*

DOI: 10.18469/ikt.2023.21.3.11

Dukalskaya Irina Vladimirovna, Povolzhskiy State University of Telecommunications and Informatics, 77, Moskovskoe shosse, Samara, 443090, Russian Federation; Associate Professor of Foreign Languages Department, PhD in Philology. Tel. +7 927 204-77-93. E-mail: i.dukalskaya@psuti.ru

Kirgizov Andrey Gennadievich, Povolzhskiy State University of Telecommunications and Informatics, 77, Moskovskoe shosse, Samara, 443090, Russian Federation; Student of Software Engineering Department. Tel. +7 932 861-85-21. E-mail: kirgizoffand@yandex.ru

References

1. Peters P. *The Cambridge Dictionary of English Grammar*. Cambridge: Cambridge University, 2004, 512 p.
2. Adams M. *Slang: The People's Poetry*. Oxford: Oxford University, 2009, 240 p.
3. Joshi M. *Abbreviations in English*. North Charleston: Independently published, 2019, 120 p.
4. 79 Incredible English Slang Words to Help you Understand Native Speakers. URL: <https://storylearning.com/blog/english-slang-words> (accessed: 20.11.23).
5. Slang Words: List of 100 Common Slang Words & Phrases you Need to Know! URL: <https://eslforums.com/slang-words/> (accessed: 20.11.23).
6. Re – Regular Expression Operations. URL: <https://docs.python.org/3/library/re.html> (accessed: 20.11.23).
7. Add YouTube functionality to your app. URL: <https://developers.google.com/youtube/v3?hl=ru> (accessed: 22.11.23) (In Russ).
8. NLTK Documentation. Natural Language Toolkit. URL: <https://www.nltk.org> (accessed: 22.11.23).
9. CSV File Reading and Writing. URL: <https://docs.python.org/3/library/csv.html> (accessed: 23.11.23).
10. How to Learn Big Data. URL: <https://www.mltut.com/how-to-learn-big-data-step-by-step/> (accessed: 24.11.23).

Received 15.01.2024

ТЕХНОЛОГИИ РАДИОСВЯЗИ, РАДИОВЕЩАНИЯ И ТЕЛЕВИДЕНИЯ

УДК 004.722.45

ИНТЕГРИРОВАННАЯ КОСМИЧЕСКАЯ СЕТЬ: РЕЗУЛЬТАТЫ ПЕРВИЧНОГО МОДЕЛИРОВАНИЯ МАРШРУТИЗИРУЕМОЙ МЕЖСПУТНИКОВОЙ СЕТИ

Кустов Н.Д.

Сибирский государственный университет науки и технологий имени академика М. Ф. Решетнева,

Красноярск, РФ

E-mail: kustovnd@sibsau.ru

В работе приведены первые результаты, полученные при моделировании маршрутизируемой межспутниковой сети на основе входных данных космической миссии ReshUCube-3. В рамках миссии планируется запуск группировки из четырех малых космических аппаратов типа CubeSat. Для обеспечения концепции интегрированной космической сети на физическом уровне планируется использование технологии LoRa, а в качестве протокола маршрутизации на сетевом уровне – Routing Protocol for Low-Power and Lossy Networks. Модель учитывает орбитальные параметры спутниковой группировки, конфигурацию антенных систем космических аппаратов и наземной станции, параметры приемо-передающих радиомодулей. На основе данных параметров проведен анализ межспутниковых каналов связи. Определены максимальная достигаемая дистанция между спутниками, энергетика каналов, среднесуточное время покрытия с учетом существующих маршрутов и т.д. Планируется, что предложенная модель послужит основой в процессе проектирования и эксплуатации межспутниковой сети.

Ключевые слова: *интегрированная космическая сеть, межспутниковая связь, ReshUCube, CubeSat, LoRa, Routing Protocol for Low-Power and Lossy Networks*

Введение

В настоящее время приоритетным направлением развития миссий малых космических аппаратов (МКА) является формирование низкоорбитальных группировок, выполняющих функций связи. Как правило, такие группировки должны обеспечивать глобальное покрытие и сервисы передачи данных космос-Земля через наземные базовые станции (НС). Перспективными проектами в России, представляющими такую концепцию, на текущий момент, являются: глобальная низкоорбитальная многоспутниковая система передачи данных «Марафон IoT» [1], а также низкоорбитальная спутниковая группировка связи от «Бюро 1440» [2].

Следующим закономерным этапом развития подобных систем, сопряженным с внедрением радиосистем 5G и 6G, представляется обеспечение непосредственного маршрутизируемого межспутникового взаимодействия внутри группировки [3; 4].

В свою очередь, СибГУ им. М.Ф. Решетнева ведет активную деятельность по разработке и эксплуатации МКА в рамках проекта «Space-п». На сегодняшний день запущены и успешно эксплуатируются два МКА типа CubeSat: ReshUCube-1 [5] и ReshUCube-2. А в 2024 году планируется одновременный запуск группировки МКА, состоящей из одного аппарата формата 3U и трех аппаратов формата 1U. Основной целью планируемой миссии является отработка технологии интегрированной космической сети: сетевых функций на линиях космос-Земля и космос-космос (межспутник).

Ранее были описаны архитектурные и технические решения, принятые в рамках миссии аппарата-демонстратора ReshUCube-2 [3]. В общем смысле, был предложен сетевой стек, основанный на технологиях LoRa (Long Range), стандартах 6LoWPAN (IPv6 over Low-Power Wireless Personal Area Networks) и TCP/IP (Transmission Control Protocol / Internet Protocol), который, в перспективе, позволит организовать маршрутизируемую IPv6 сеть с использованием, к примеру, протокола RPL (Routing Protocol for Low-Power and Lossy Networks). Планируется, что данные решения лягут в основу предложенной технологии и будущей космической миссии ReshUCube-3.

Целью данной работы является демонстрация первых результатов, полученных в рамках имитационного моделирования системы межспутниковой связи с учетом орбитальных параметров и параметров каналов связи, приближенных к исходным данным планируемой космической миссии ReshUCube-3.

Постановка задач моделирования

Ключевыми задачами работы являются:

- определение максимальной дистанции между МКА и наземной станцией (центром управления полетами, ЦУП);
- определение максимальной дистанции между МКА, на которой возможно поддержание непрерывного канала связи;
- определение среднесуточной продолжительности сеансов связи (активности канала связи) между отдельным МКА и НС;
- определение среднесуточной продолжительности сеансов связи между группировкой МКА и НС с учетом всех возможных маршрутов с промежуточными узлами-МКА в зависимости от межспутниковой дистанции.

Глобальной задачей при этом является верификация модели. В том случае, если модель является репрезентативной, дальнейшие результаты моделирования могут быть использованы в процессе проектирования и эксплуатации группировки, а также смежных задачах.

Входные данные

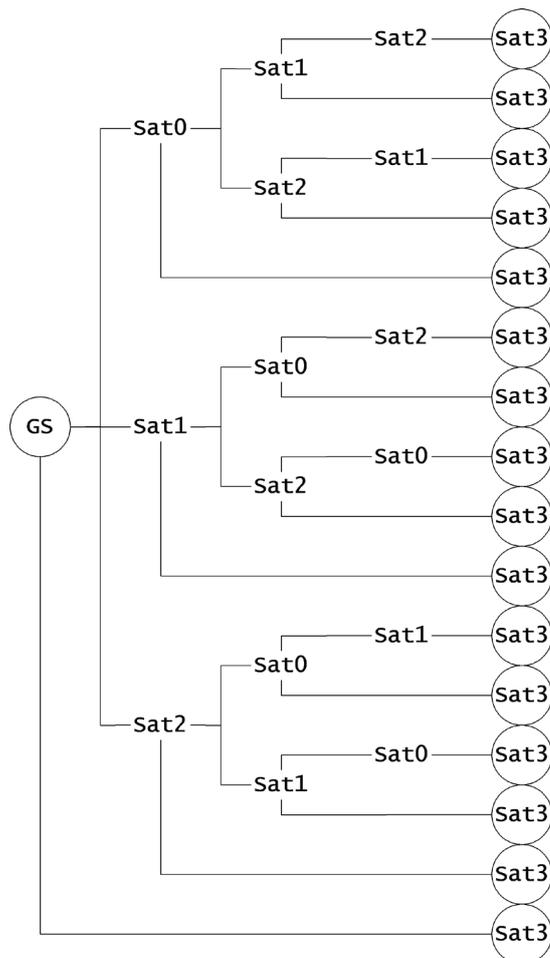


Рисунок 1. Моделируемые маршруты

Достоверно определить орбитальные параметры спутниковой группировки до момента запуска не представляется возможным. Точными параметрами орбиты, коэффициентами ускорения, торможения и т.д., в данном рассмотрении, можно пренебречь. В связи с этим, за основу были взяты параметры TLE (Two-Line Element set) четырех уже запущенных аппаратов «СтратоСат ТК-1-А», «СтратоСат ТК-1-Б», «СтратоСат ТК-1-В» и «СтратоСат ТК-1-Г». 11 июля 2023 г. данные спутники одновременно вышли из транспортного контейнера на орбите, что наиболее близко отражает принцип разделения, планируемый в рамках миссии ReshUCube-3.

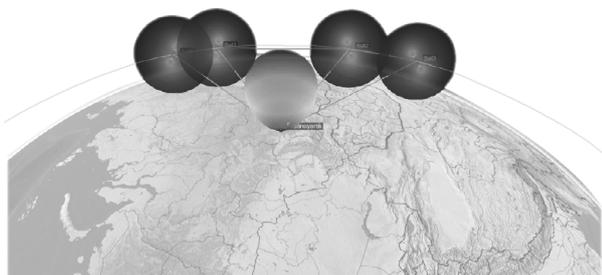


Рисунок 2. Визуальная симуляция

Расположение ЦУП было задано следующими координатами: 56.0121° с.ш., 92.97465° в.д. (что соответствует расположению ЦУП СибГУ им. М.Ф. Решетнева).

В качестве антенной системы каждого из четырех моделей МКА определена модель дипольной антенны, рассчитанной на 435 МГц. Подобная система предусмотрена антенным модулем, который был разработан для аппаратов миссии ReshUCube-3.

В качестве приемно-передающей системы модели ЦУП определена модель Гауссовской антенны с диаметром 0,5 м и коэффициентом эффективности диафрагмы 0,65. Модель похожа на параболическую антенну с погрешностью около -6 дБ относительно точки обзора.

Параметры приемников и передатчиков заданы на основе значений, приведенных в работах [7; 8], а также исходных данных планируемой миссии (часть из которых приведена в работе [6]). Отметим также, что приведенные параметры заданы пессимистично.

Так, для приемников МКА установлены следующие параметры: системные потери – 5 дБ; потери перед приемником – 3 дБ; отношение коэффициента усиления антенны к температуре шума – 25 дБ/К; минимальное отношение энергии сигнала на бит к энергетической спектральной плотности шума – 6,5 дБ.

Для приемника ЦУП: системные потери – 5 дБ; потери перед приемником – 3 дБ; отношение коэффициента усиления антенны к температуре шума – 20 дБ/К; минимальное отношение энергии сигнала на бит к энергетической спектральной

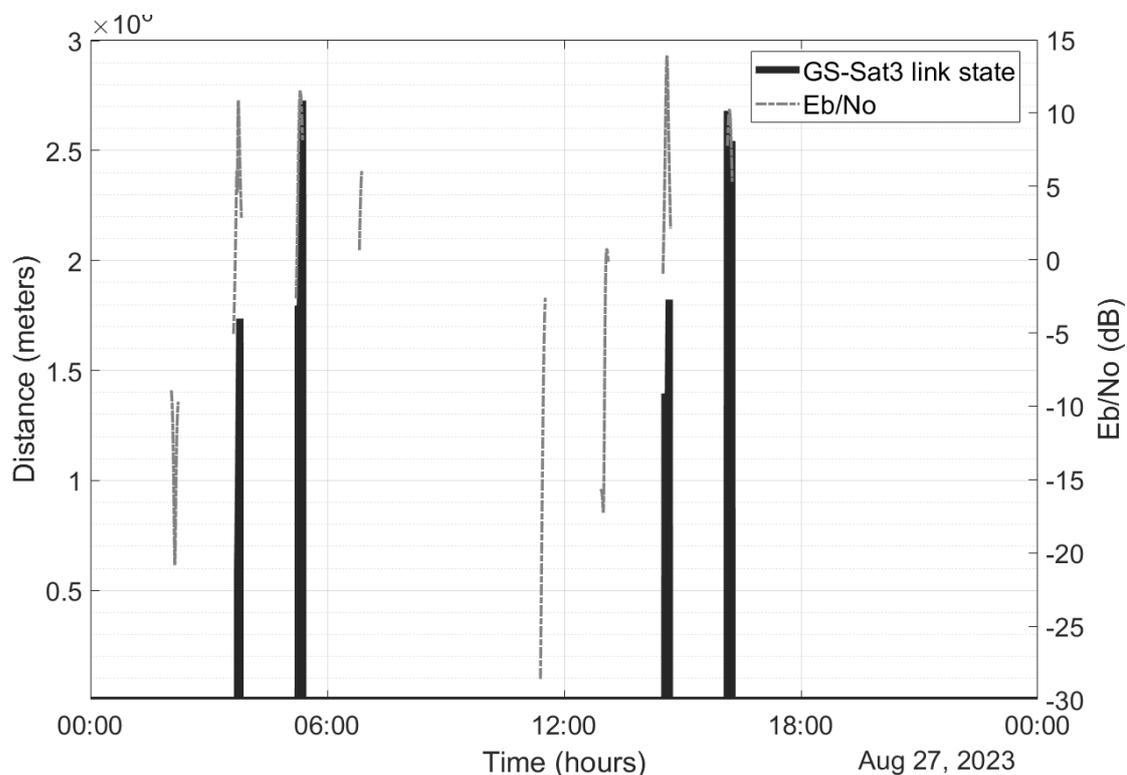


Рисунок 3. Дистанция между Sat3 и GS по отношению к E_b/N_0

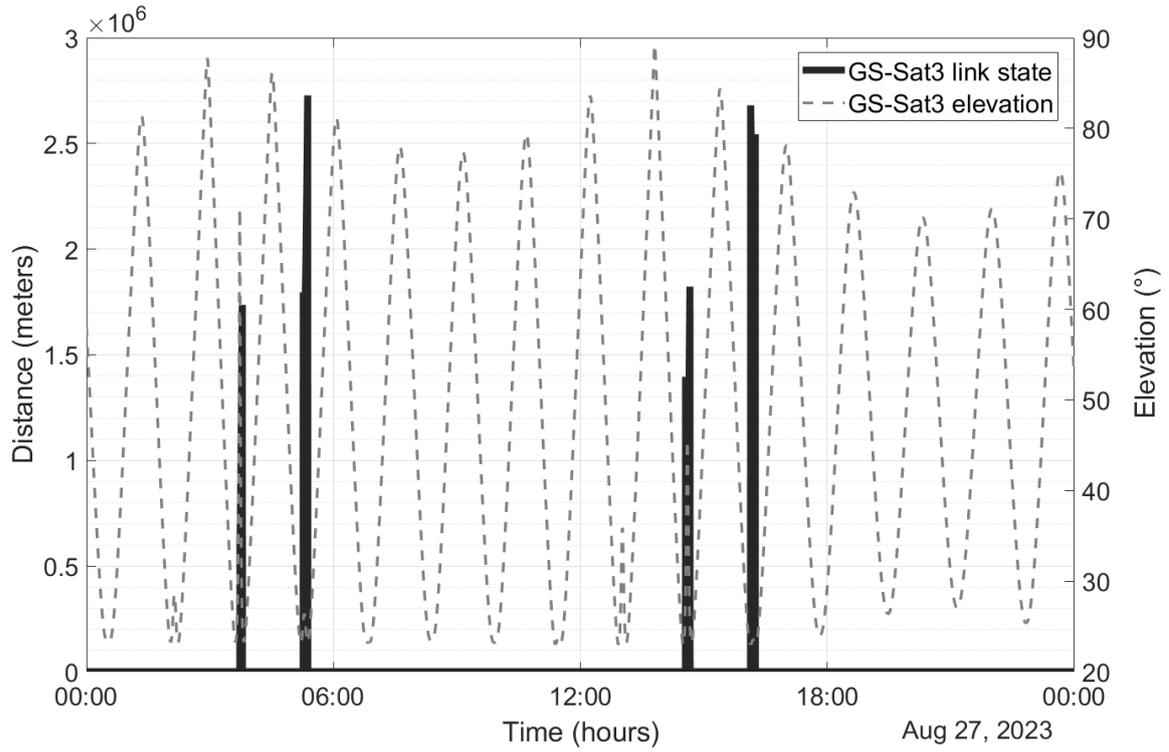


Рисунок 4. Дистанция между Sat3 и GS по отношению к углу возвышения

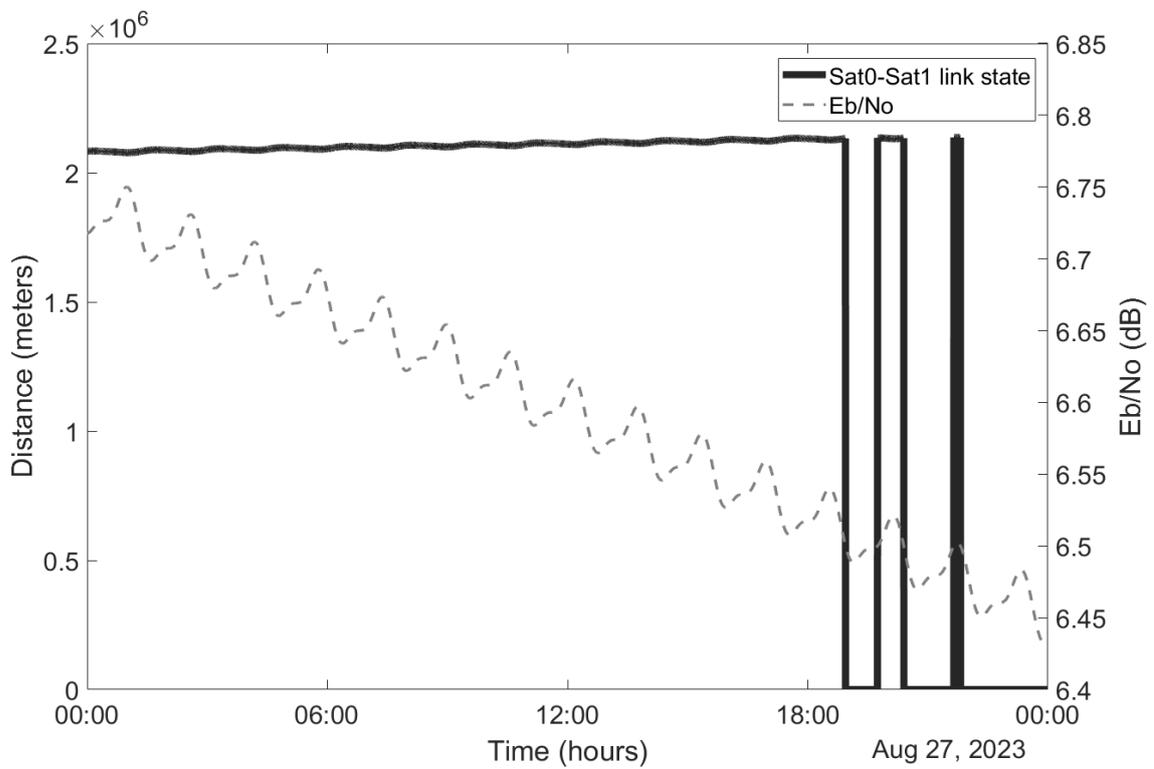


Рисунок 5. Дистанция между Sat₀ и Sat₁ по отношению к E_b/N_0

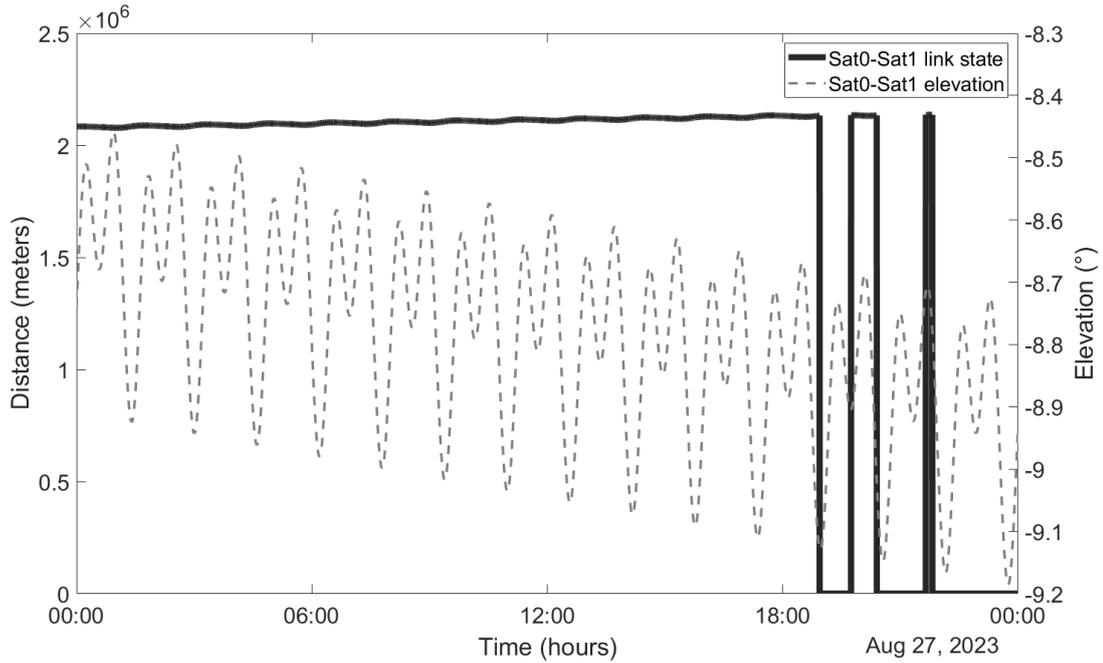


Рисунок 6. Дистанция между Sat0 и Sat1 по отношению к углу возвышения

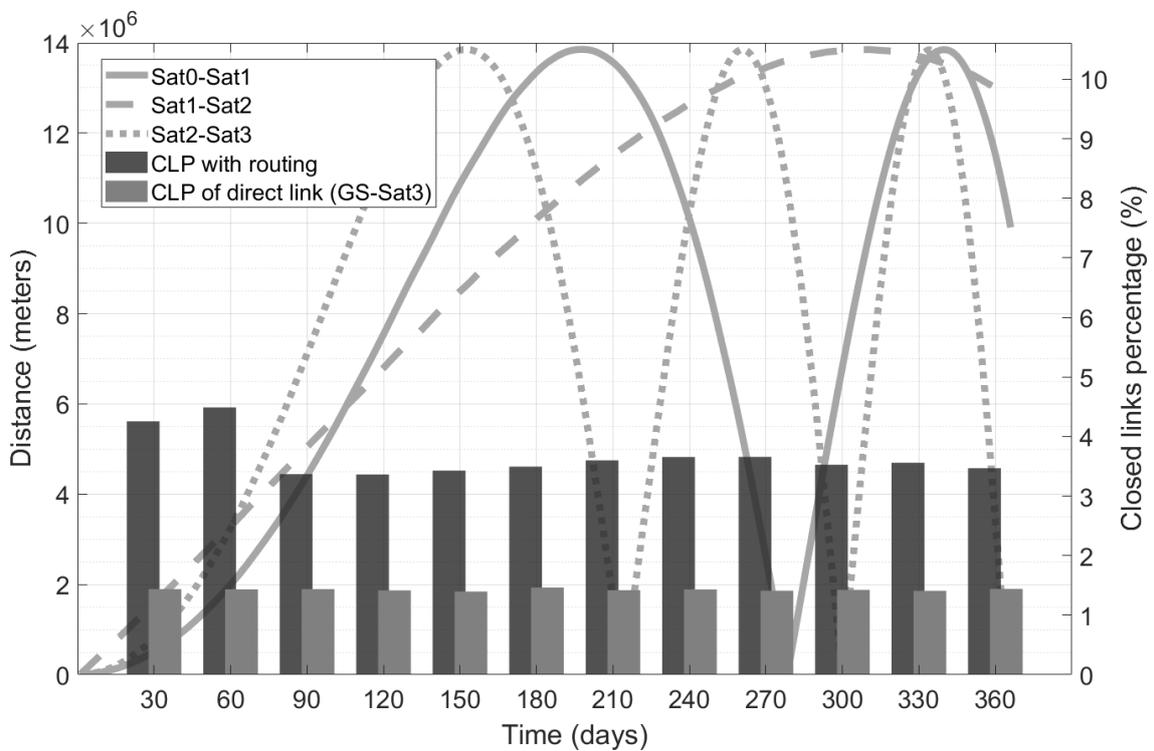


Рисунок 7. Процент активного использования каналов в зависимости от межспутниковой дистанции

ной плотности шума – 6,5 дБ.

Для передатчиков МКА: системные потери – 5 дБ; частота – 435 МГц; битрейт – 5,47 кбит/с; мощность – 0 дБВт.

Для передатчика ЦУП: системные потери – 5 дБ; частота – 435 МГц; битрейт – 5,47 кбит/с; мощность – 10 дБВт.

Важно отметить, что МКА ориентированы в надир для получения оптимального позиционирования диаграммы направленности антенной системы.

Как было упомянуто ранее, некоторые виды моделирования включают в себя моделирование маршрутов. Под «маршрутом», в данном случае, подразумевается прямой канал связи между НС (Ground Station, GS) и целевым узлом (Satellite 3, Sat3) или составной – через промежуточные узлы (Satellite 0-3, Sat0-3). Маршрут считается действительным, если каждый из составных каналов удовлетворяет условию моделируемой радиовидимости. Все возможные вариации маршрутов показаны на рисунке 1.

Результаты моделирования

Первым важным результатом имитационного моделирования является динамическая визуальная симуляция, представленная на рисунке 2. Здесь, в зависимости от текущего модельного времени, можно наблюдать траекторию орбиты МКА, диаграммы направленности антенных систем МКА и ЦУП и отображение каналов связи в случае соблюдения условия радиовидимости. Все это позволяет дать визуальную оценку релевантности модели.

Переходя к основным задачам моделирования, приведем результаты суточного моделирования канала связи между Sat3 и GS.

На рисунке 3 представлен график дистанции между Sat3 и GS в периоды времени, когда канал между ними активен, а также сопутствующие значения энергии сигнала на бит к энергетической спектральной плотности шума (E_b/N_0). Можно заметить, что канал активен только в тот момент, когда значение E_b/N_0 превышает пороговое (6,5 дБ). При этом значение максимальной достигнутой дистанции составило 2 727 290 м.

На рисунке 4 представлен тот же график дистанции, но по отношению к углу возвышения между Sat3 и GS. Стоит отметить, что значение угла принимает абсолютные значения: за сутки угол принимает пиковые значения 15 раз, половину из которых Sat3 находится на обратной стороне Земли. Также наблюдаются случаи, когда Sat3 и GS находятся в зоне прямой радиовидимости под острыми углами, однако энергетики

сигнала недостаточно для того, чтобы образовать действительный канал связи.

Следующим этапом моделирования является определение максимальной дистанции между Sat0 и Sat1 в момент времени, когда канал между ними прерывается.

На рисунке 5 представлен график дистанции между Sat0 и Sat1, а также значение E_b/N_0 (по аналогии с рисунком 3). Значение максимальной достигнутой дистанции до прерывания канала составило 2 132 929 м. Значение E_b/N_0 при этом убывает линейно по отношению к увеличению межспутниковой дистанции.

На рисунке 6 представлен тот же график дистанции, но по отношению к углу возвышения между Sat0 и Sat1 (по аналогии с рисунком 4). Аналогичным образом наблюдается линейная зависимость между дистанцией и углом.

Финальным этапом моделирования, в текущем рассмотрении, является определение процента активного использования каналов (среднесуточной продолжительности сеансов связи) в случае прямого канала между Sat3 и GS, а также в случае использования всех описанных ранее маршрутов.

Для того, чтобы определить влияние межспутниковой дистанции на эффективность использования каналов, было рассчитано 365 дней модельного времени. За это время МКА несколько раз «разлетаются» и «сближаются» в случае, если отсутствуют активные механизмы удержания в группировке.

Кроме того, было взято 12 месячных отсчетов, которые показывают процент активности канала от общего времени моделирования. Эти отсчеты были взяты для случая, когда используются все возможные маршруты, а также случая с единственным прямым маршрутом Sat3-GS. Результат моделирования приведен на рисунке 7.

Моделирование показало, что в случае с прямым маршрутом процент активности канала в месяц остается близок к среднему значению 1,4267% в течение года, что составляет, в среднем, около 20 мин/сутки.

В случае же, когда рассматриваются все возможные маршруты, процент активности зависит от межспутникового расстояния. Среднее значение в год составляет 3,6551% (т.е. около 52 мин/сутки). Пиковое же значение достигается тогда, когда межспутниковое расстояние близко к максимальному (около 2000 км) и составляет 4,4883% (около 64 мин/сутки).

Заключение

В рамках работы была разработана модель межспутниковой сети, основанная на входных данных планируемой миссии ReshUCube-3. Ключевые результаты моделирования: определение максимальной дистанции космос-космос и космос-Земля, а также эффективности покрытия с учетом активных маршрутов.

Стоит отметить, что некоторые результаты моделирования сопоставимы с объективными показателями, полученными в рамках активных миссий со схожей спецификой (например, [9; 10]). Тем не менее, миссия ReshUCube-3 послужит инструментом для предметной верификации и уточнения модели.

Предложенная модель может быть полезна при проектировании антенных систем и приемопередающих систем МКА и НС, отладки протоколов маршрутизации и других сетевых функций, а также в сопутствующих задачах управления группировкой и др. На этапе эксплуатации уточненная модель может служить в качестве цифрового двойника для реальной системы.

В будущих работах планируется провести более подробную оценку влияния переменных параметров, эффекта Доплера, задержек распространения и т.д. На следующем этапе планируется ввести детализацию модели в разрезе физических технологий передачи (модуляции, кодирования и др.), а также используемых сетевых протоколов и технологий.

Литература

1. Ань Н.Д., Ершов А. Система «Марафон IoT» и новый рынок спутникового IoT // Технологии и средства связи. 2021. № S1. С. 75–80.
2. Ходаков А.М. Обзор состояния и перспектив развития частных российских космических компаний // Финансово-экономическая аналитика в системе управления современными социально-экономическими процессами: материалы Международной научно-практической конференции. Самара: САМАРАМА, 2023. С. 168–171.
3. Space-air-ground integrated network: a survey / J. Liu [et al.] // IEEE Communications Surveys and Tutorials. 2018. Vol. 20, no. 4. P. 2714–2741.

DOI: 10.1109/COMST.2018.2841996

4. Space-air-ground integrated network (SAGIN) for 6G: requirements, architecture and challenges / H. Cui [et al.] // China Communications. 2022. Vol. 19, no. 2. P. 90–108. DOI: 10.23919/JCC.2022.02.008
5. Предварительные результаты космической миссии ReshUCube-1 / В.Х. Ханов [и др.] // Решетневские чтения: материалы XXVI Международной научно-практической конференции, посвященной памяти генерального конструктора ракетно-космических систем академика М.Ф. Решетнева. Красноярск, 2022. С. 452–454.
6. Кустов Н.Д., Евдокимов К.С., Шахматов А.В. Интегрированная космическая сеть: обоснование архитектурных и технических решений космической миссии ReshUCube-2 // Сибирский аэрокосмический журнал. 2023. Т. 24, № 2. С. 260–272. DOI 10.31772/2712-8970-2023-24-2-260-272
7. Нгуен Д.А. Оценка энергетики абонентских радиолиний и информационной емкости спутниковой низкоорбитальной системы интернета вещей // T-Comm: Телекоммуникации и транспорт. 2021. Т. 15, № 11. С. 32–39. DOI 10.36724/2072-8735-2021-15-11-18-31
8. Шорин О.А., Бокк Г.О., Щепнов С.Г. Сравнение эффективности модуляции с расширением спектра сети Lora и традиционной техники дискретной фазовой модуляции. Часть 1 // Экономика и качество систем связи. 2023. № 3 (29). С. 37–53.
9. First flight-testing of LORA modulation in satellite radio communications in Low-Earth orbit / A. Zadorozhny [et al.] // IEEE Access. 2022. Vol. 10. P. 100006–100023. DOI: 10.1109/ACCESS.2022.3207762
10. NORBY cubesat nanosatellite: design challenges and the first flight data / V.Y. Prokopyev [et al.] // Journal of Physics: Conference Series: 16. Novosibirsk, 2021. P. 012–038. DOI: 10.1088/1742-6596/1867/1/012038.

Получено 11.12.2023

Кустов Никита Дмитриевич, ассистент кафедры безопасности информационных технологий Сибирского государственного университета науки и технологий имени академика М.Ф. Решетнева. 660037, Российская Федерация, г. Красноярск, проспект им. газеты «Красноярский рабочий», 31. Тел. +7 902 222-82-07. E-mail: kustovnd@sibsau.ru

SPACE INTEGRATED NETWORK: ROUTED INTER-SATELLITE NETWORK PRIMARY MODELING RESULTS

Kustov N.D.

*Reshetnev Siberian State University of Science and Technology, Krasnoyarsk, Russian Federation
E-mail: kustovnd@sibsau.ru*

The paper presents the first results obtained by modeling a routed inter-satellite network based on the parameters of the ReshUCube-3 space mission. ReshUCube-3 is a mission in which it is planned to launch a satellite constellation of four CubeSats. A network stack was proposed to provide the concept of a space integrated network. It is planned to use LoRa technology on the physical layer, and RPL as a routing protocol on the network layer. The model takes into account the orbital parameters of the satellite constellation, the antenna systems of spacecraft and ground station, and the parameters of radio modules. Based on this information, the analysis of inter-satellite communication channels was carried out. The maximum distance between the satellites, the energy properties of the channels, the average daily coverage time, etc. are determined. An important feature is the consideration of existing routes between satellites. It is planned that the proposed model will serve as a basis in the process of designing and operating an inter-satellite network.

Keywords: *space integrated network, inter-satellite communication, ReshUCube, CubeSat, LoRa, RPL*

DOI: 10.18469/ikt.2023.21.3.12

Kustov Nikita Dmitrievich, Reshetnev Siberian State University of Science and Technology, 31, Krasnoyarsky Rabochy Avenue, Krasnoyarsk, 660014, Russian Federation; Assistant of Information Technology Security Department. Tel. +7 908 222-82-07. E-mail: kustovnd@sibsau.ru

References

1. Anh N.D., Ershov A. Marathon IOT system and new satellite IOT market. *Texnologii i sredstva Svyazi*, 2021, no. S1, pp. 75–80. (In Russ.)
2. Xodakov A. M. Overview of the state and prospects of development of private Russian space companies. *Finansovo-ekonomicheskaya analitika v sisteme upravleniya sovremennymi social'no-ekonomicheskimi processami: materialy Mezhdunarodnoj nauchno-prakticheskoy konferencii*. Samara: SAMARAMA, 2023, pp. 168–171. (In Russ.)
3. Liu J. et al. Space-air-ground integrated network: a survey. *IEEE Communications Surveys & Tutorials*, 2018, vol. 20, no. 4, pp. 2714–2741. DOI: 10.1109/COMST.2018.2841996
4. Cui H. et al. Space-air-ground integrated network (SAGIN) for 6G: requirements, architecture and challenges. *China Communications*, 2022, vol. 19, no. 2, pp. 90–108. DOI: 10.23919/JCC.2022.02.008.
5. Xanov V.X. et al. Preliminary results of the reshucube-1 space mission. *Reshetnevskie chteniya: materialy XXVI Mezhdunarodnoj nauchno-prakticheskoy konferencii, posvyashchennoj pamyati general'nogo konstruktora raketno-kosmicheskikh sistem akademika M. F. Reshetneva*. Krasnoyarsk, 2022, pp. 452–454. (In Russ.)
6. Kustov N.D., Evdokimov K.S., Shaxmatov A.V. Space integrated network: architectural and technical solutions justification of the reshucube-2 space mission. *Sibirskij aerokosmicheskij zhurnal*, 2023, vol. 24, no. 2, pp. 260–272. DOI 10.31772/2712-8970-2023-24-2-260-272 (In Russ.)
7. Nguyen D.A. Estimation of energy intensity of subscriber radio lines and information capacity of the leo satellite system of the internet of things. *T-Comm: telekommunikacii i transport*, 2021, vol. 15, no. 11, pp. 32–39. DOI 10.36724/2072-8735-2021-15-11-18-31 (In Russ.)
8. Shorin O.A., Bokk G.O., Shhepnov S.G. Comparison of the modulation efficiency with the expansion of the LORA network spectrum and the traditional technique of discrete phase modulation. *Ekonomika i kachestvo sistem svyazi*, 2023, no. 3(29), pp. 37–53 (In Russ.)

9. Zadorozhny A. et al. first flight-testing of LORA modulation in satellite radio communications in Low-Earth orbit. *IEEE Access*, 2022, vol. 10, pp. 100006–100023. DOI: 10.1109/ACCESS.2022.3207762
10. Prokopyev V.Y. et al. NORBY cubesat nanosatellite: design challenges and the first flight data. *Journal of Physics: Conference Series*, 2021, pp. 012–038. DOI: 10.1088/1742-6596/1867/1/012038

Received 11.12.2023

УПРАВЛЕНИЕ И ПОДГОТОВКА КАДРОВ ДЛЯ ОТРАСЛИ ТЕЛЕКОММУНИКАЦИЙ

УДК 004.855.5

ИСПОЛЬЗОВАНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ТРУДОУСТРОЙСТВА ВЫПУСКНИКОВ

Селиванова Е.С.¹, Газизов Т.Т.^{1,3}, Долганова Н.Ф.¹, Пираков Ф.Д.²

¹Томский государственный педагогический университет, Томск, РФ

²Томский государственный университет систем управления и радиоэлектроники, Томск, РФ

³Сахалинский государственный университет, Южно-Сахалинск, РФ

E-mail: slyvva@tspu.edu.ru

В современном мире технологии продолжают играть все более весомую роль во всех сферах жизни человека. Образование не является исключением и идет в ногу со временем. Одно из более перспективных направлений в этой области является использование методов машинного обучения для анализа данных о выпускниках и прогнозировании их дальнейшего трудоустройства, на основе накопленной информации об обучающихся студентах. Такая информация не только помогает определиться с будущей профессией, но и позволит учебным заведениям более эффективно координировать свою деятельность и повышать качество образования. Кроме того, использование машинного обучения в образовании может привести к созданию новых, более эффективных методов обучения, учитывающих индивидуальные особенности каждого студента. Результатом данной работы является интеллектуальная система обработки больших данных, способная адаптироваться к текущему состоянию рынка труда и помогающая выпускникам в более раннем определении в своей будущей профессии. Анализ данных проводился на выборке обучающихся Томского государственного педагогического университета 2021–2023 учебных годов.

Ключевые слова: анализ данных, трудоустройство, машинное обучение, обработка информации, искусственный интеллект, модель, прогнозирование

Введение и постановка задачи

Ценнейшим товаром на современном рынке услуг является информация о пользователе, анализ которой посредством алгоритмов машинного обучения позволяет обеспечить ее целенаправленное использование в дальнейшем. Информация о человеке, его достижениях и увлечениях считается одним из ценнейших продуктов не только на рынке рекламы, но и в образовательной среде [1; 2]. Система образования является одной из ключевых сфер жизнедеятельности любого человека и активно внедряет и использует тренды в развитии информационных технологий [3; 4]. В образовании целесообразно применять методы машинного обучения. Можно предположить, что в сфере образования возможно прогнозировать будущие места работы студентов педагогического вуза, используя данные выпускников

[5; 6]. Это позволит выявить траекторию профессиональной подготовки молодых специалистов и оценить качество предоставляемых образовательных услуг вузом [7].

В связи с этим целью работы является проведение обработки данных о трудоустройстве и прогнозирование места работы студентов на примере выпускников, используя модели искусственного интеллекта.

Для достижения поставленной цели необходимо решить следующие основные задачи:

1. Определить параметры, влияющие на трудоустройство выпускников.
2. Обработать данные о трудоустройстве выпускников.
3. Разработать интеллектуальную систему для прогнозирования трудоустройства выпускников.

Машинное обучение в прикладных задачах

Одним из наиболее распространенных направлений использования машинного обучения является оптимизация подборов туров в турагентствах [8]. Именно методы автоматического поиска закономерностей в больших объемах информации (данных) являются фундаментом задач прогнозирования и одним из наиболее перспективных направлений использования современных информационных технологий в различных отраслях экономики, в том числе и в туризме. В туристическом бизнесе важно понимать и оперативно предлагать клиенту определенные направления проведения досуга, в зависимости от его предпочтений, образа жизни, социального статуса и прочего. Минимизировать трудовые затраты на удовлетворение потребностей конкретного клиента, предоставив лучший вариант отдыха для него, зачастую помогают инструменты аналитики, механизмы моделирования и разработка алгоритмов. Анализ пользовательских предпочтений осуществляется на фиксации данных и их изучении. Отзывы посетителей о выбранных туристических объектах и их сегментирование по каким-либо схожим характеристикам (как со стороны объектов, так и клиентов) лежат в основе «рекомендательных систем, предлагающих пользователю наиболее подходящие для него точки притяжения, отели, рестораны и прочее». Система спрашивает у гостя при посещении сайта о его предпочтениях и возможностях и на основании отзывов пользователей предоставляет наиболее подходящие варианты отдыха данному клиенту. Из чего следует, что уже на этапе посещения сайта клиенту сразу «предсказывается» его «идеальный» тур, и на услугах менеджера по данному фронту работы можно экономить.

Data Science и Machine Learning

Чтобы научиться извлекать полезную информацию из полученных данных, необходимо их правильно обрабатывать. Для этого используется Data Science – наука о данных, которая включает в себя все инструменты, методы и технологии и позволяет обрабатывать данные и использовать их для собственной выгоды, выверяя закономерности в больших массивах данных и на их основе прогнозировать вероятность получения того или иного исхода [9].

Три основных составляющих Data Science:

1. Организация информации – это процесс ее хранения и форматирования.

2. Агрегация данных – это объединение начальных сведений в новый вид и (или) представление.

3. Доставка данных – это действие, обеспечивающее доступ к коллекциям агрегированных данных.

Существует большое количество областей для работы с данными, использующих искусственный интеллект, и одна из них – это машинное обучение (machine learning, ML) [10; 11], позволяющее создавать программы, улучшающиеся в процессе обучения.

Основные задачи ML:

1. Регрессия (предсказание числовых значений на основе заданных признаков).

2. Классификация (на основе набора признаков происходит разделение объектов).

3. Кластеризация (распределение данных на группы).

Большинство задач, которое решается посредством ML, использует определенные методы, то есть алгоритмы, позволяющие компьютеру обучаться на основе опыта и предоставленных данных, и сделать предсказание или принять решение на основе этого обучения.

Способы машинного обучения [12; 13]:

1. *Машинное обучение с учителем.*

В этом способе человек обучает машину с помощью огромной выборки данных, параметры которой варьируются до тех пор, пока не получится нужного результата.

2. *Машинное обучение без учителя.*

Машинное обучение без учителя предполагает исследование компьютером определенного набора данных, где выявляются скрытые закономерности корреляции между переменными.

3. *Машинное обучение с частичным привлечением учителя.*

Гибридный способ, в основе которого лежит обучение с учителем и без. Учитель, отметив небольшую часть данных, дает возможность понять, каким образом сгруппировать остальные.

4. *Обучение с подкреплением.*

При данном способе обучения машине позволяет взаимодействовать с окружением и в зависимости от настроек получать вознаграждение при правильном выполнении задания.

Для анализа данных выпускников педагогического вуза и реализации прогноза трудоустройства студентов выбран способ машинного обучения с учителем.

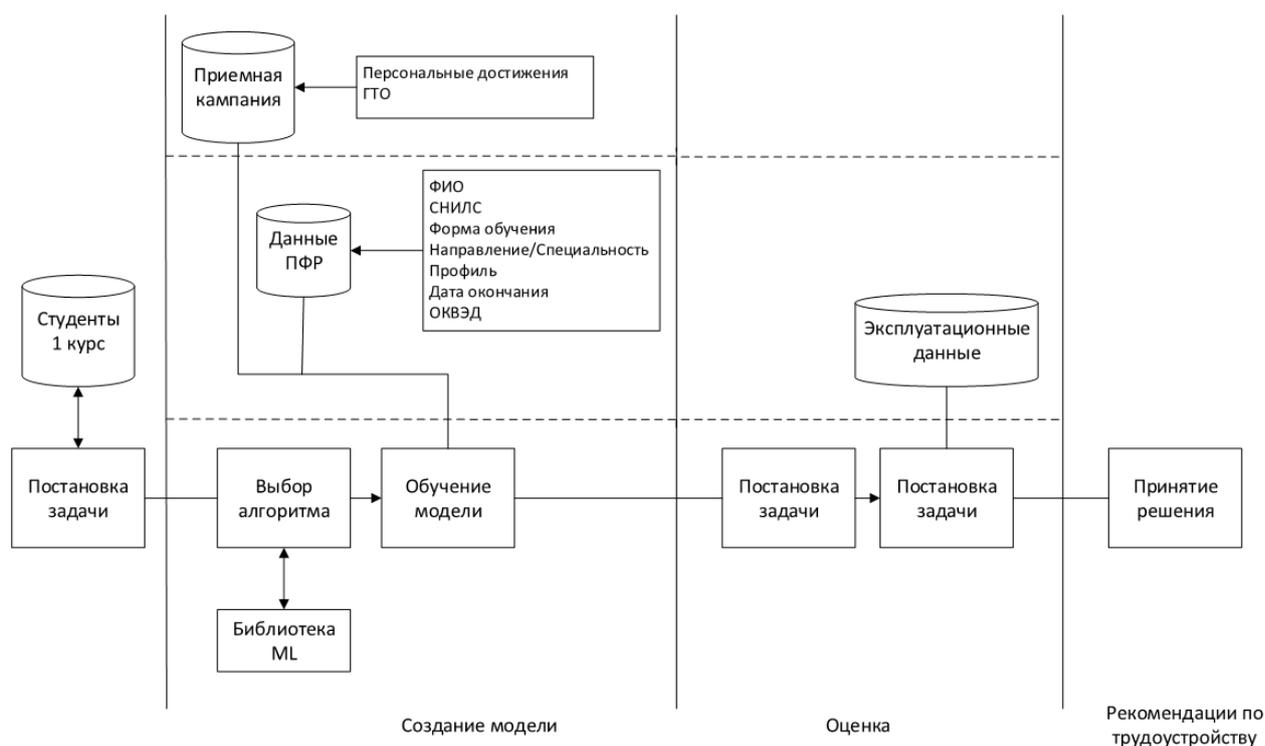


Рисунок 1. Схема работы модели машинного обучения

Таблица 1. Пример данных выпускников из ПФР

| Форма | Направление/Специальность | Профиль 1 | Дата | Работодатель | Наименование ОКВЭД |
|---------|---|------------------------|------------|--|--|
| заочная | 44.03.03: Специальное (дефектологическое) образование | Логопедия | 24.10.2022 | ОГАУЗ «ОПЦ Им. И.Д. Евтушенко» | Деятельность больничных организаций |
| очная | 44.03.05: Педагогическое образование (с двумя профилями подготовки) | Начальное образование | 09.01.2022 | Рогожкин Валерий Андреевич | Предоставление услуг по дневному уходу за детьми |
| очная | 44.03.05: Педагогическое образование (с двумя профилями подготовки) | Математика | 09.01.2022 | МАОУ СОШ № 58 г. Томска | Образование среднее общее |
| очная | 44.03.05: Педагогическое образование (с двумя профилями подготовки) | Математика | 09.01.2022 | МАОУ Гимназия № 26 г. Томска | Образование среднее общее |
| заочная | 44.03.01: Педагогическое образование | Начальное образование | 10.08.2022 | МАОУ «ОШ №5 г. Асино» | Образование основное общее |
| очная | 44.03.05: Педагогическое образование (с двумя профилями подготовки) | Технология | 09.01.2022 | МАОУ Гимназия № 55 им. Е.Г. Версткиной г. Томска | Образование среднее общее |
| заочная | 44.03.01: Педагогическое образование | Дошкольное образование | 11.05.2020 | МБДОУ Детский сад «Светлячок» | Образование дошкольное |
| очная | 44.03.05: Педагогическое образование (с двумя профилями подготовки) | Русский язык | 26.06.2021 | МАОУ ДО ДДТ «Созвездие» | Образование дополнительное детей и взрослых |

Реализация моделей машинного обучения

Реализовывать алгоритмы машинного обучения было решено на веб-платформе Google Colaboratory (Colab), которая позволяет создать и запускать код на языке Python, не устанавливая на компьютер дополнительных программ.

Python один из наиболее популярных языков для машинного обучения. Основными причинами его использования для написания алгоритмов машинного обучения являются простота использования, большой объем доступных библиотек и широкое пользование [14].

Pandas предоставляет набор инструментов для манипулирования данными (чтение, запись, обработка и анализ). Эта библиотека позволяет работать с данными как со структурой, состоящей из столбцов и строк, что облегчает выполнение различных операций над данными [14].

Sklearn написана на языке Python и использует библиотеки NumPy и SciPy для работы с массивами и линейной алгеброй. Библиотека является простым и эффективным инструментом для предиктивного анализа данных [15].

Для задачи регрессии были выбраны два алгоритма машинного обучения [16]:

1. *Логистическая регрессия.* Логистическая регрессия (Logistic Regression) позволяет делать прогнозы для точек в двоичной системе: 0 или 1. Если значение равно или больше 0,5, объект относится к категории «1». Если значение меньше 0,5 – к катего-

рии «0». Каждый признак имеет свою метку, которая равна либо 0, либо 1. Алгоритм решает задачи бинарной классификации, так как алгоритм применяет сигмоидальную функцию. В данном методе выполняется условие, где $0 \leq Y \leq 1$, что достигается применением сигмоидальной (логистической) функции:

$$Y = \frac{1}{1 + e^{-F(x)}},$$

где $F(x)$ – стандартное уравнение регрессии. При этом нужно учесть, что, если значение не равно 0 или 1, значение аппроксимируется.

2. *Случайный лесной регрессор.* Случайный лесной регрессор (Random Forest Regressor) – это модель машинного обучения, которая может решать различные задачи, такие как классификация, регрессия, кластеризация и другие, используя алгоритм случайного леса. Random Forest состоит из множества базовых деревьев решений, которые работают вместе. Каждое дерево предсказывает класс, и класс, за который «проголосует» большинство деревьев, становится ответом классификатора. Одним из преимуществ модели Random Forest Regressor является ее способность обрабатывать большие объемы данных и работать с большим количеством переменных. Она также устойчива к выбросам в данных и способна выявлять сложные нелинейные зависимости между переменными. Теоретическая часть алгоритма,

Таблица 2. Результаты прогноза с помощью метода «Случайный лес»

| № | Место работы | Вероятность трудоустройства |
|---|---|-----------------------------|
| 1 | Образование среднее общее | 0,6919 |
| 2 | Образование основное общее | 0,0685 |
| 3 | Деятельность в области спорта прочая | 0,0630 |
| 4 | Образование дополнительное детей и взрослых | 0,0624 |
| 5 | Образование высшее | 0,0263 |
| 6 | Деятельность по обработке данных, предоставление услуг по размещению информации и связанная с этим деятельность | 0,0249 |
| 7 | Деятельность по предоставлению прочих вспомогательных услуг для бизнеса, не включенная в другие группировки | 0,0236 |
| 8 | Образование профессиональное среднее | 0,0219 |
| 9 | Управление эксплуатацией жилого фонда за вознаграждение или на договорной основе | 0,0175 |

по сравнению с другими методами, достаточно проста, необходима только формула итогового классификатора:

$$a(x) = \frac{1}{N} \sum_{i=1}^N b_i(x),$$

где N – количество деревьев;

i – счетчик для деревьев;

b – решающее дерево;

x – сгенерированная на основе данных выборка.

Конвейер машинного обучения разделяется на несколько основных этапов: определение задачи, данные, оценка, моделирование и применение. Одно не следует за другим, так как от одного пункта можно вернуться к другому. Сбор данных предполагает обработку сырых данных, то есть информацию из ПФР и данные студентов 1 курса. Под «моделированием» понимается использование алгоритма ML для выявления закономерностей в собранных данных. И последним этапом является применение этой модели уже на эксплуатационных данных. На первоначальном этапе прогнозирования рекомендаций по дальнейшему трудоустройству были указаны основные параметры: направление подготовки, направленность (профили), форма обучения. В дальнейшем для более корректного и точного предсказания есть возможность расширения передаваемых в модель параметров из приемной кампании вуза, где указаны персональные достижения поступающего абитуриента (рисунок 1).

Для обучения модели прогнозирования будущего трудоустройства студентов использовались данные о выпускниках, полученные из ПФР, пример которых представлен в таблице 1. Информация, полученная о выпускниках 2021–2022 года, представляет собой Excel-таблицу с данными бывших студентов, которая содержит: ФИО, СНИЛС, форму обучения, направление/специальность, профиль, дату окончания, работодателя и ОКВЭД (общероссийский классификатор видов экономической деятельности) (таблица 1).

Таким образом, выбрав язык программирования Python, файл с данными обработан программной библиотекой pandas и применен в работе с библиотекой, предназначенной для машинного обучения, scikit-learn (Алгоритм 1 и 2).

Алгоритм 1. Реализация модели машинного обучения LogisticRegression на определенном наборе данных.

```
job = pd.read_excel("job_2021.xlsx")
df = pd.get_dummies(job,
columns=["form", "spec_napravl", "profile"])
y = df.position
x = df.drop('position', axis=1)
model = LogisticRegression(solver='lbfgs',
max_iter=1000)
model.fit(x,y)
val_of = pd.DataFrame(val)
model.predict(val_of)
```

Алгоритм 2. Реализация модели машинного обучения RandomForestRegressor на определенном наборе данных.

```
job = pd.read_excel("job_2021.xlsx")
df = pd.get_dummies(job,
columns=["form", "spec_napravl", "profile"])
y = pd.get_dummies(job.position,
columns=["position"])
x = df.drop(['position'], axis=1)
model = RandomForestRegressor(random_
state=0)
model.fit(x,y)
val_of = pandas.DataFrame(val)
a = model.predict(val_of)
```

Результаты прогнозирования

Модели, обученные на конкретном примере, могут спрогнозировать, кем студент педагогического вуза будет работать в будущем. Для проверки «предсказания» выбраны такие метки для параметров.

1. Направление подготовки: 44.03.05 Педагогическое образование (с двумя профилями подготовки).

2. Направленность (профили): русский язык и литература.

3. Форма обучения: очная.

В результате проделанной работы была разработана интеллектуальная система для прогнозирования трудоустройства выпускников.

После проведенного анализа и вывода прогноза был получен результат: «Студент с такими параметрами, вероятнее всего, станет учителем (оказание услуг по предоставлению общего среднего образования)» (таблица 2).

В результате использования данного алгоритма с учетом входных параметров, описывающих студента педагогического вуза, определяется наиболее вероятный результат его трудоустройства. Путем применения логистической регрессии был получен прогноз, согласно которому студент станет учителем в будущем (рисунок 2).

```
array(['Образование среднее общее'], dtype=object)
```

Рисунок 2. Результат прогноза с помощью метода логистической регрессии

Рассмотренный алгоритм «Случайный лес» представляет собой метаоценщик, который состоит из ряда классификационных деревьев принятия решений, построенных на различных подвыборках набора данных, и использует технику усреднения для повышения точности прогнозов и предотвращения переобучения. Этот алгоритм позволяет получить вероятностное распределение потенциального места работы студента в будущем, которое представлено в виде таблицы 2.

Заключение

В заключение следует отметить, что в ходе работы была обработана информация о трудоустройстве выпускников педагогического вуза, а также разработана интеллектуальная система анализа больших данных, которая позволяет прогнозировать возможное место работы студента в будущем и оценивать качество образовательных услуг. Использование машинного обучения значительно улучшает точность таких прогнозов, а также помогает определить востребованные навыки и знания для работодателей. Это, в свою очередь, позволяет адаптировать учебные программы и повысить конкурентоспособность образовательного учреждения. Кроме того, такая система может быть использована для определения наиболее перспективных направлений для будущих специалистов и оценки качества образовательных услуг.

Исследование выполнено за счет гранта Российского научного фонда № 23-28-00577.

Литература

1. Наговицын Р.С. Прогнозирование трудоустройства выпускников педагогического института на основе технологий искусственного интеллекта // Образовательное пространство в информационную эпоху: материалы международной научно-практической конференции. Москва, 2022. С. 244–253.
2. Наговицын Р.С. Искусственный интеллект для реализации прогноза профессиональной занятости будущих учителей физической культуры // Российское государство, право, экономика и общество: проблемы и пути развития: материалы III Национальной научно-практической конференции. Казань, 2021. С. 86–91.
3. Гладкова И.А., Щанина Е.В. Прогнозирование трудоустройства выпускников вузов // Теоретические и практические аспекты раз-

вития современной науки: теория, методология, практика: материалы международной научно-практической конференции. Уфа, 2019. С. 151–155.

4. Наговицын Р.С. Трудоустройство выпускников педагогических профилей на основе технологий искусственного интеллекта и анализа данных // Мир образования – образование в мире. 2023. № 1 (89). С. 165–176.
5. Фадеев А.С., Змеев О.А., Газизов Т.Т. Модель университета 4.0 // Научно-педагогическое обозрение. 2020. № 2 (30). С. 172–178. DOI: 10.23951/2307-6127-2020-2-172-178
6. Основные направления информатизации деятельности томского государственного педагогического университета / А.Н. Клишин [и др.] // Вестник Томского государственного педагогического университета. 2015. № 3 (156). С. 110–118.
7. Наговицын Р.С. Искусственный интеллект по обработке данных абитуриентов для прогнозирования их дальнейшего трудоустройства после обучения // Региональные вузы-драйверы пространственного развития России: материалы Всероссийской с международным участием научно-практической конференции, посвященной 90-летию Удмуртского государственного университета. Ижевск, 2021. С. 92–104.
8. Изучение опыта прогнозирования туристских потоков с применением алгоритмов машинного обучения / С.А. Лочан [и др.] // Известия высших учебных заведений. Серия: Экономика, финансы и управление производством. 2021. № 4 (50). С. 145–155.
9. Наука о данных. URL: [https://www.tadviser.ru/index.php/Статья:Наука_о_данных_\(Data_Science\)](https://www.tadviser.ru/index.php/Статья:Наука_о_данных_(Data_Science)) (дата обращения: 10.09.2023).
10. Лебедев И.С. Адаптивное применение моделей машинного обучения на отдельных сегментах выборки в задачах регрессии и классификации // Информационно-управляющие системы. URL: <https://cyberleninka.ru/article/n/adaptivnoe-primenenie-modeley-mashinnogo-obucheniya-na-otdelnyh-segmentah-vybornki-v-zadachah-regressii-i-klassifikatsii> (дата обращения: 15.09.2023).
11. Сейдаметова З.С. Задачи и алгоритмы машинного обучения: вероятностные графические модели // Информационно-компьютерные технологии в экономике, образовании и социальной сфере. 2019. № 1 (23). С. 180–187.
12. Згонникова А.О., Прокопенко А.А. Машинное обучение и обучение на протяжении всей

- жизни // Новые научные исследования: материалы VIII Международной научно-практической конференции. Пенза, 2022. С. 22–24.
13. Бородин И.Д. Рефлексия в машинном обучении на примере обучения деревьев решений // Аллея науки. 2017. Т. 4, № 9. С. 857–865.
14. Аналитикам: большая шпаргалка по Pandas. URL: <https://smysl.io/blog/pandas/> (дата обращения: 19.09.2023).
15. Scikit-learn. Машинное обучение в Python. URL: <https://scikit-learn.ru/> (дата обращения: 19.09.2023).
16. Микшина В.С., Павлов С.И. Принятие решений с использованием ансамбля классификаторов // Информационные технологии в науке, образовании и управлении. 2019. № 2 (12). С. 50–54.

Получено 07.11.2023

Селиванова Елизавета Сергеевна, магистрант кафедры информатики Томского государственного педагогического университета (ТГПУ). 634061, Российская Федерация, г. Томск, ул. Герцена, 49. Тел. +7 961 096-90-90. E-mail: slvnva@tspu.edu.ru

Газизов Тимур Тальгатович, д.т.н., профессор кафедры информатики ТГПУ. 634061, Российская Федерация, г. Томск, ул. Герцена, 49; научный сотрудник центра экспертизы и испытаний Сахалинского государственного университета. 693008, Российская Федерация, г. Южно-Сахалинск, пр. Коммунистический, 33. Тел. +7 382-2 31-13-66. E-mail: gtt@tspu.edu.ru

Долганова Надежда Филипповна, старший преподаватель кафедры информатики ТГПУ. 634061, Российская Федерация, г. Томск, ул. Герцена, 49. Тел. +7 923 422-47-45. E-mail: dolganovaNF@tspu.edu.ru

Пираков Фаррух Джамshedович, аспирант кафедры технологий электронного обучения Томского государственного университета систем управления и радиоэлектроники. 634050, Российская Федерация, г. Томск, пр-т Ленина, 40. Тел. +7 952 157-21-24. E-mail: pfd@tspu.edu.ru

USING MACHINE LEARNING METHODS FOR PREDICTING THE EMPLOYMENT OF GRADUATES

Selivanova E.S.¹, Gazizov T.T.^{1,3}, Dolganova N.F.¹, Pirakov F.D.²

¹Tomsk State Pedagogical University, Tomsk, Russian Federation

²Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russian Federation

³Sakhalin State University, Yuzhno-Sakhalinsk, Russian Federation

E-mail: slvnva@tspu.edu.ru

In today's world, technology continues to play an increasingly important role in all spheres of human life. Education is no exception and keeps pace with the times. One of the more promising directions in this area is the use of machine learning methods to analyze data on graduates and predict their future employment, based on the accumulated information about students. Such information not only helps to determine the future profession, but also will allow educational institutions to co-ordinate their activities more effectively and improve the quality of education. In addition, the use of machine learning in education can lead to the creation of new, more effective teaching methods that take into account the individual characteristics of each student. The result of this work is an intelligent system of big data processing, capable of adapting to the current state of the labor market and helping graduates in an earlier determination of their future profession. Data analysis was conducted on a sample of students of Tomsk State Pedagogical University of 2021-2023 academic years.

Keywords: *Data analysis, employment, machine learning, information processing, artificial intelligence, model, prediction*

DOI: 10.18469/ikt.2023.21.3.13

Selivanova Elizabeth Sergeevna, Tomsk State Pedagogical University, 49, Herzen Street, Tomsk, 634061, Russian Federation; Master's Degree Student of Informatics Department. Tel. +7 961 096-90-90. E-mail: slvnva@tspu.edu.ru

Gazizov Timur Talgatovich, Tomsk State Pedagogical University, 49, Herzen Street, Tomsk, 634061, Russian Federation; Professor of Computer Science Department, Doctor of Technical Science. Sakhalin State University, 33, Kommunisticheskoy Avenue, Yuzhno-Sakhalinsk, 693008, Russian Federation; Researcher at the Center for Expertise and Testing. Tel. +7 382-2 31-13-66. E-mail: gtt@tspu.edu.ru

Dolganova Nadezhda Pilippovna, Tomsk State Pedagogical University, 49, Herzen Street, Tomsk, 634061, Russian Federation; Senior lecturer of Informatics Department. Tel. +7 923 422-47-45. E-mail: dolganovaNF@tspu.edu.ru

Pirakov Farrukh Jamshedovich, Tomsk State University of Control Systems and Radioelectronics, 40, Lenina Prospect, 634050, Tomsk, Russian Federation; Department of E-Learning Technologies, Automation, PhD Student. Tel. +7 952 157-21-24. E-mail: pfd@tspu.edu.ru

References

1. Nagovitsyn R.S. Forecasting the employment of graduates of a pedagogical institute based on artificial intelligence technologies. *Obrazovatel'noe prostranstvo v informacionnuyu epokhu: materialy mezhdunarodnoj nauchno-prakticheskoy konferencii*. Moscow, 2022, pp. 244–253. (In Russ.)
2. Nagovitsyn R.S. Artificial intelligence for implementing the forecast of professional employment of future physical education teachers. *Rossiyskoe gosudarstvo, pravo, ekonomika i obshchestvo: problemy i puti razvitiya: materialy III Nacional'noj nauchno-prakticheskoy konferencii*. Kazan, 2021. pp. 86–91. (In Russ.)
3. Gladkova I.A., Shchanina E.V. Forecasting job placement of university graduates. *Teoreticheskie i prakticheskie aspekty razvitiya sovremennoj nauki: teoriya, metodologiya, praktika: materialy Mezhdunarodnoj nauchno-prakticheskoy konferencii*. Ufa, 2019. pp. 151–155. (In Russ.)
4. Nagovitsyn R.S. Employment of graduates of pedagogical profiles based on artificial intelligence technologies and data analysis. *Mir obrazovaniya – obrazovanie v mire*, 2023, no. 1 (89), pp. 165–176. (In Russ.)
5. Fadeev A.S., Zmeev O.A., Gazizov T.T. University Model 4.0. *Nauchno-pedagogicheskoe obozrenie*, 2020, no. 2, pp. 172–178. (In Russ.)
6. Klishin A.N. et al. Main directions for applying information technologies to the automation of tspu activities. *Vestnik Tomskogo gosudarstvennogo pedagogicheskogo universiteta*, 2015, vol. 3 (156), pp. 110–118. (In Russ.)
7. Nagovitsyn R.S. Artificial intelligence for processing the data of applicants to predict their further employment after training. *Regional'nye vuzy-drajvery prostranstvennogo razvitiya Rossii: materialy Vserossiyskoy s mezhdunarodnym uchastiem nauchno-prakticheskoy konferencii, posvyashchennoj 90-letiyu Udmurtskogo gosudarstvennogo universiteta*. Izhevsk, 2021, pp. 92–104. (In Russ.)
8. Lochan S.A. et al. Study of experience in forecasting tourist flows using machine learning algorithms. *Izvestiya vysshih uchebnyh zavedeniy. Seriya: Ekonomika, financy i upravlenie proizvodstvom*, 2021, vol. 4 (50), pp. 145–155. (In Russ.)
9. Data science. URL: [https://www.tadviser.ru/index.php/Статья:Наука_о_данных_\(Data_Science\)](https://www.tadviser.ru/index.php/Статья:Наука_о_данных_(Data_Science)) (accessed: 10.09.2023). (In Russ.)
10. Lebedev I.S. Adaptive application of machine learning models on separate sample segments in regression and classification tasks. *Informacionno-upravlyayushchie sistemy*. URL: <https://cyberleninka.ru/article/n/adaptivnoe-primenenie-modeley-mashinnogo-obucheniya-na-otdelnyh-segmentah-vyboriki-v-zadachah-regressii-i-klassifikatsii> (accessed: 15.09.2023). (In Russ.)
11. Seidametova Z.S. Machine learning problems and algorithms: probabilistic graphical models. *Informacionno-komp'yuternye tekhnologii v ekonomike, obrazovanii i social'noj sfere*, 2019, no. 1 (23), pp. 180–187. (In Russ.)
12. Zgonnikova A.O., Prokopenko A.A. Machine learning and lifelong learning. *New scientific*

research: Novye nauchnye issledovaniya: materialy VIII Mezhdunarodnoj nauchno-prakticheskoy konferencii, Penza, 2022, pp. 22–24. (In Russ.)

13. Borodin I.D. Reflexion in machine learning on the example of decision tree learning. *Alleya nauki*, 2017, vol. 4, no. 9, pp. 857–865. (In Russ.)
14. For Analysts: the big cheat sheet on Pandas. URL: <https://smysl.io/blog/pandas/> (accessed: 19.09.2023). (In Russ.)
15. Scikit-learn. Machine learning in Python. URL: <https://scikit-learn.ru/> (accessed: 19.09.2023). (In Russ.).
16. Mikshina V.S., Pavlov S.I. Decision-making using an ensemble of classifiers. *Informacionnye tekhnologii v nauke, obrazovanii i upravlenii*, 2019, no. 2 (12). pp. 50–54. (In Russ.)

Received 07.11.2023