УДК 681.3.016
DOI: 10.46548/21vek-2021-1055-0012

СИСТЕМА КРИТЕРИЕВ И АЛГОРИТМ ОБРАБОТКИ ИНФОРМАЦИИ И ПРИНЯТИЯ РЕШЕНИЙ ДЛЯ ПРОГРАММНОГО МОДУЛЯ ОТОБРАЖЕНИЯ НАИБОЛЕЕ ЗНАЧИМЫХ СОБЫТИЙ МОНИТОРИНГА В ИНФОРМАЦИОННОЙ СИСТЕМЕ

Щемелинин Дмитрий Александрович, кандидат технических наук,
старший научный сотрудник Центра научных исследований «Три Би»
Санкт-Петербургский политехнический университет Петра Великого
(194064, Россия, г. Санкт-Петербург, ул. Политехническая, 29, e-mail:dshchmel@gmail.ru)

Аннотация. В данной работе представлена созданная система критериев обработки информации и принятия решений для отображения наиболее значимых событий мониторинга в информационных вычислительных средах (ИС) и разработанный алгоритм анализа и обработки информации по предложенным критериям с целью сокращения количества информационного шума, отображающегося на графическом интерфейсе системы непрерывного мониторинга *NMC* (англ. *Network Monitoring Console*). Для анализа процесса мониторинга ИС, была собрана производственная статистика из исследуемых глобально-распределенных вычислительных комплексов (ГРВК) с использованием непрерывной системы мониторинга *Zabbix*, которая используется для получения параметрических данных и метрик от нескольких десятков тысяч виртуальных электронно-вычислительных машин (*VM*, англ. *Virtual Machine*) и сохранения данных о производительности в системах управления базами данных (СУБД) *SQL* (англ. *Structured Query Language*) для вывода информации на *NMC*. Исследуемые производственные процессы мониторинга и выявления отказов в ИС основаны на методе экспертных оценок и не всегда эффективны в случае наступления масштабного отказа работоспособности ГРВК. Целью научного исследования стала разработка критериев и моделей интеграции, позволяющих создать программный модуль корреляции событий для уменьшения ложных экспертных оценок при определении первопричин отказа ГРВК в системе мониторинга и сократить время восстановления сервисов, сократить большой объем передачи служебных данных и автоматически определять наиболее значимые мониторинговые события в системе принятия решений.

Ключевые слова: мониторинг, большие данные, информационные технологии, принятие решений, обработка информации, критерии обработки информации, метрики мониторинга, корреляция событий, облачные технологии.

SYSTEM OF CRITERIA AND ALGORITHM OF INFORMATION PROCESSING AND DECISION-MAKING FOR THE SOFTWARE MODULE FOR DISPLAYING THE MOST SIGNIFICANT MONITORING EVENTS IN THE INFORMATION SYSTEM

Shchemelinin Dmitry Aleksandrovich, candidate of technical sciences,
senior researcher at the Tri B Research Center
Peter the Great St. Petersburg Polytechnic University
(194021, Russia, Saint-Petersburg, Polytechnic Street, 29, e-mail: dshchmel@gmail.ru)

Abstract. This paper presents the created system of information processing and decision-making criteria for displaying the most significant monitoring events in information computing environments (*IS*) and the developed algorithm for analyzing and processing information according to the proposed criteria in order to reduce the amount of information noise displayed on the graphical interface of the *NMC* continuous monitoring system (*English Network Monitoring Console*). To analyze the IS monitoring process, production statistics were collected from the studied globally distributed computing complexes (*GDVK*) using the continuous monitoring system *Zabbix*, which is used to obtain parametric data and metrics from several tens of thousands of virtual electronic computers (*VM*, *English Virtual Machine*) and storing performance data in database management systems (*DBMS*) *SQL* (*English Structured Query Language*) for displaying information on *NMC*. The investigated production processes for monitoring and detecting failures in the IS are based on the method of expert assessments and are not always effective in the event of a large-scale failure of the *GDVK* operability. The purpose of the research was to develop integration criteria and models that allow creating a software module for event correlation to reduce false expert assessments when determining the root causes of failure of the *GDVK* in the monitoring system and reduce the recovery time of services, reduce the large volume of service data transmission and automatically determine the most significant monitoring events in the system. decision making.

Keywords: monitoring, big data, information technology, decision making, information processing, information processing criteria, monitoring metrics, event correlation, cloud technologies.

Введение. Большинство *IT* компаний (англ. I*nformation Technology*), обслуживающих ИС, построенные по принципу информационного облака, сталкиваются с проблемой эффективности средств мониторинга с точки зрения визуализации состояний ГРВК. Пытаясь привлечь новых клиентов, *IT* компа-

нии инвестируют больше ресурсов для повышения надежности и емкости своих ГРВК. Это вызывает значительный рост инфраструктуры ИС, что увеличивают процент генерируемых метрик и сценариев приложений, которые находятся под мониторингом и могут дать промежуточный результат самого мониторинга - мониторинг событий. Обзор источников [1-7] показал, что существующие средства мониторинга не способны полностью удовлетворить требования обслуживающего персонала по группировке сообщений мониторинга и подавлению, так называемого, шума событий, состоящего из множества триггеров, происходящих в ГРВК. В то же время обслуживающему персоналу требуется четкое представление о состоянии элементов системы, чтобы своевременно и надлежащим образом поддерживать работоспособность ГРВК. Персонал разработчиков программного обеспечения (ПО), также нуждается в эффективной системе непрерывного мониторинга с возможностью корреляции событий в ГРВК для отображения возникшей первопричинной проблемы, влияющей на работоспособность ИС.

В системе облачного мониторинга международной телекоммуникационной компании *RingCentral* (*USA*), являющейся, согласно аналитическим отчетам Gartner 2017, 2018, 2019 и 2020, мировым Топ 20 лидером в секторе *UCaaS* (англ. *Unified Communication as a Service*), генерируется около 2 миллионов событий в минуту в среднем на 50 000 компонент в работающих в ГРВК [8], который построен с использованием технологии облачных вычислений и использует вычислительные компоненты развернутые на *VM*. Статистический анализ частоты событий (рис. 1), отображающихся на графическом интерфейсе системы непрерывного мониторинга *NMC* (англ. *Network Monitoring Console*) для ГРВК, показал, что перебой в работоспособности только одного сетевого узла вычислений может привести к значительным всплескам частоты событий мониторинга, генерируемых другими вычислительными компонентами, работающими в логической взаимосвязи с отказавшим сетевым узлом [9].

Также, представленный статистический анализ показал, что подобный информационный шум на *NMC* возникает практически каждый день. Это в свою очередь, делает невозможным упорядочивание событий мониторинга, с точки зрения отображения причинно-следственных связей и приводит к невозможности оперативной обработки таких событий вручную силами дежурной смены инженеров обслуживания ГРВК. Более того, неверно истолкованные события и, как следствие, неправильные упреждающие действия дежурной смены могут спровоцировать глобальный отказ предоставления информационных услуг для пользователей *UCaaS*.

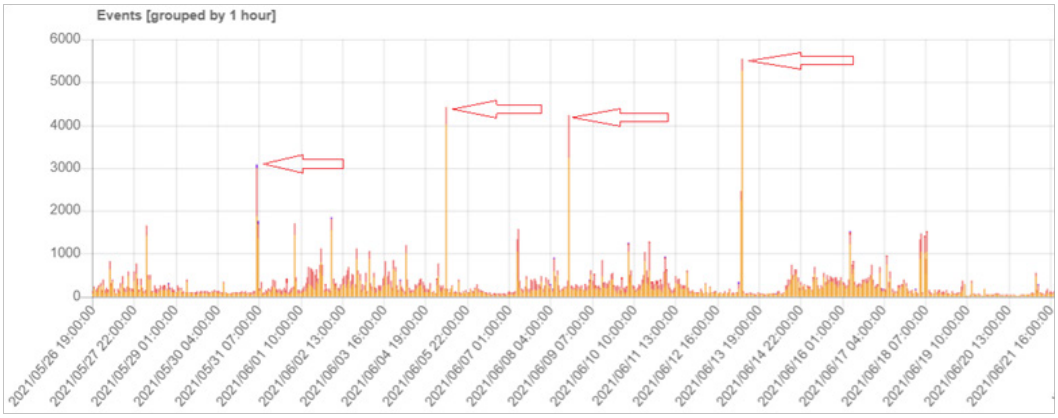

Рисунок 1 – Количество событий мониторинга в час на NMC при отказе в ГРВК

Целью данной работы является создание системы критериев обработки информации и принятия решений для отображения наиболее значимых событий мониторинга в ИС и разработка алгоритма анализа и обработки информации по предложенным критериям с целью сокращения количества информационного шума отображающегося на *NMC*.

Методы и материалы исследования. Разработка системы критериев обработки информации и принятия решений для отображения наиболее значимых событий мониторинга в ИС базировалась на основе объективных данных мониторинга, полученных в исследуемых ГРВК с использованием системы непрерывной мониторинга *Zabbix* [10, 11]. Для решения поставленных задач была разработана информационная модель интеграции (рис. 2) разрозненных производственных процессов в *VM* с использованием инструментальных средств моделирования СУБД *Gliffy* [12-15].

В отличие от существующих решений, предложенная новая система критериев и структура данных описывает облачную ИС покомпонентно с использованием критерия *System Unit* (перевод с англ. Системная Единица), которое объединяет вычислительные компоненты и узлы ИС по набору уникальных параметров, свойственных только данному типу. Кроме того, для каждого вычислительного узла, было предложено включить дополнительный критерий *System Relation* (перевод с англ. Системная Связь), что позволило объединить любые два вычислительных узла, с точки зрения их системных связей и далее реализовать классы реляционных связей с логикой «родитель-потомок», учитывающих все виды физических и логических сетевых интерфейсов и соединений, а также транзит сетевого трафика.

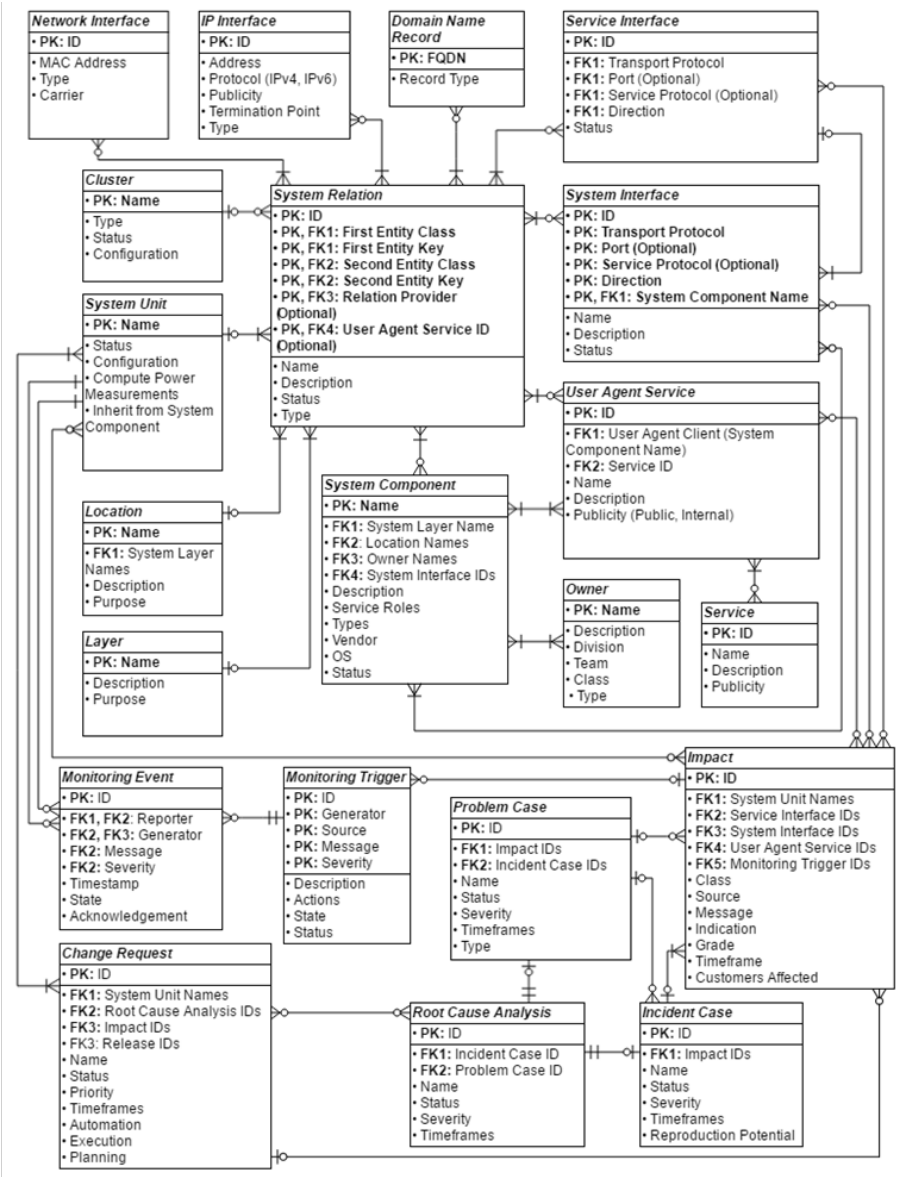

Рисунок 2 – Предложенная модель корреляции событий мониторинга ГРВК

Для описания «родителя» всех вычислительных компонент и сетевых узлов был введен критерий *System Component* (перевод с англ. Системный Компонент), который включает в себя верхне-уровневый набор параметров вычислительных компонент и сетевых узлов. Причем, Системный Компонент может быть представлен в нескольких географических *Locations* (перевод с англ. Локация) и представлять уникальный системно-вычислительный *Layer* (перевод с англ. Уровень), которые в ГРВК могут быть использованы для выполнения различающихся вычислений, таких как: вычислительные окружения для предоставления IT услуг клиентам компании либо лабораторные окружения для разработчиков ПО и тестировщиков.

Для каждой системной единице были введены следующие идентификаторы позволяющие объединять *Monitoring Events* (перевод с англ. Мониторинговые События) по принадлежностям:

• *User Agent Service* (перевод с англ. Пользовательский Сервис);

• *User Agent Client* (перевод с англ. Пользовательский Клиент);

• *Owner* (перевод с англ. Владелец);

• *System Interface* (перевод с англ. Системный Интерфейс);

• *Network Interface* (перевод с англ. Сетевой Интерфейс);

• *Domain Name Record* (перевод с англ. Запись Доменного Имени);

• *Cluster* (перевод с англ. Кластер);

• *Service* (перевод с англ. Сервис).

В случае возникновения отказа в ГРВК, Мониторинговые События могут быть также сгруппированы по результатам оперативного технического расследования причин его возникновения (англ. *Root Cause Analysis*). Для такого сценария были предложены и реализованы следующие идентификаторы:

• *Monitoring Trigger* (перевод с англ. Мониторинговый Триггер);

• *Problem Case* (перевод с англ. Проблемный Слу-

чай);

• *Impact* (перевод с англ. Негативное Влияние);

• *Incident Case* (перевод с англ. Случай Инцидента);

• *Change Request* (перевод с англ. Запрос на Изменение).

Результаты исследования. Был разработан алгоритм анализа и обработки информации по предложенным критериям с целью сокращения количества информационного шума отображающегося на *NMC*. Он выглядит следующим образом:

Шаг 1. Сбор данных: сервер мониторинга *Zabbix* собирает все параметры работоспособности компонент ГРВК и выдает предупреждение в случае обнаружения аномалии в работе ИС;

Шаг 2. Проверка: на *Zabbix*-сервере по прибытии события в систему корреляции событий мониторинга, оно должно либо соответствовать предложенной системе критериев обработки информации, указанным в правилах обработки событий, либо быть отброшенным;

Шаг 3. Анализ: если событие принято к обработке, далее набор выборок опрашивается на предмет наличия конкретного идентификатора;

Шаг 4. Событие соответствующим образом дополняется служебной информацией;

Шаг 5. Корреляция событий с использованием предложенной системы критериев обработки информации и принятия решений событий мониторинга;

Шаг 6. Анализ вариантов организации иерархии в базах данных с использованием модели данных для описания иерархических объектов с произвольными атрибутами и сравнительного [8, 9], что позволяет исключить отображение информационного шума и выводить на графический интерфейс только иерархическое древо событий мониторинга, сгруппированных под родительским перво-причинным триггером;

Шаг 7. Отображение результата вычислений на *NMC* (рис. 3), на котором отображаются только ключевые наиболее значимые аномальные событий, когда как вторичные события, создающие информационный шум, скрываются, с возможностью их просмотра.

При открытии оператором *NMC*, вторичные события выделяются серым цветом.

В ходе исследования, было экспериментально установлено (рис. 4), что программная реализация алгоритма анализа и обработки событий мониторинга по предложенным критериям позволило значительно сократить частоту отображения событий мониторинга на *NMC* и обеспечить непрерывный процесс обеспечения различных телекоммуникационных сервисов клиентам через Интернет с минимальными операционными затратами.

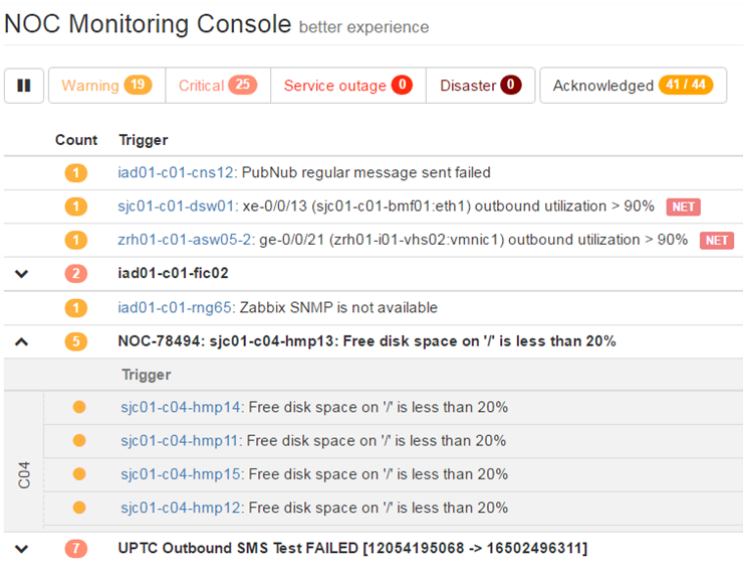

Рисунок 3 – Интегральный интерфейс визуализации (NMC) состояния ГРВК

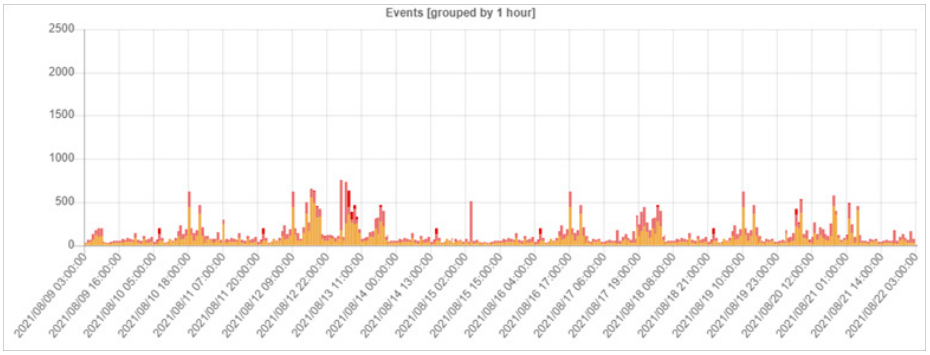

Рисунок 4 – Количество событий мониторинга в час на NMC при отказе в ГРВК с применением алгоритма анализа и обработки событий мониторинга по предложенным критериям

Заключение. Основным результатом работы стала сформулированная система критериев корреляции и обработки информации для визуализации только ключевых наиболее значимых аномальных событий в больших информационных потоках ГРВК на *NMC* облачной инфраструктуры, которая позволила реализовать алгоритм подавления информационного шума в программном модуле автоматической группировки событий мониторинга [16], что значительным образом отличается от существующих систем непрерывного мониторинга и обеспечивает визуализацию происходящих в ИС событий по новому. Практическая реализация и внедрение нового алгоритма анализа и обработки информации по предложенным критериям с целью сокращения количества информационного шума отображающегося на *NMC* в крупной телекоммуникационной компании *RingCentral* позволили уменьшить ложные экспертные оценки при определении первопричин отказа ГРВК в системе мониторинга и сократить время восстановления сервисов с целью их достижения мирового уровня доступности 99,999% в режиме 24/7 [8].

СПИСОК ЛИТЕРАТУРЫ:

1. Батура Т., Мурзин Ф., Семич Д. Облачные технологии: основные понятия, задачи и тенденции развития // Программные продукты, системы. – 2014. – №1. URL: http://swsys-web.ru/cloud-computing-basic-concepts-problems.html (дата обращения: 08.21.2021).

2. Кучерова К.Н., Мещеряков С.В., Щемелинин Д.А. Сравнительный анализ систем мониторинга глобально распределенных вычислительных комплексов // Сравнительный анализ в проектировании и управлении: сб. науч. тр. XX Международной науч.-практич. конф., СПб, СПбПУ, – 2016. - С. 303-309.

3. Лавлинский Н.Е. Мониторинг сайта – отказоустойчивость и качество обслуживания // «Метод Лаб», – 2015. URL: https://www.methodlab.ru/articles/monitoring_saita (дата обращения: 21.08.2021).

4. Рудницкий П. Сравнение систем мониторинга // Prudnitskiy.PRO, 2013. URL: https://prudnitskiy.pro/2013/11/14/monitoring-comparsion/ (дата обращения: 21.08.2021).

5. F. Lanubile, C. Ebert, R. Prikladnicki, A. Vizcaino. Collaboration Tools for Global Software Engineering // IEEE Software, Vol. 27, Issue 2, – 2010. URL: http://ieeexplore.ieee.org/abstract/document/5420797/ (дата обращения: 21.08.2021).

6. A Dell Technical Whitepaper. Sizing and Best Practices for Deploying Virtual Desktops with Dell EqualLogic Virtual Desktop Deployment Utility in a VMware Environment, 2012. URL: http://en.community.dell.com/dell-groups/dtcmedia/m/ (дата обращения: 21.08.2021).

7. Bernstein D. Containers and Cloud: From LXC to Docker to Kubernetes // IEEE Cloud Computing, Vol. 1, Issue 3, 2014. URL: http://ieeexplore.ieee.org/document/7036275/ (дата обращения: 21.08.2021).

8. Официальный Интернет-сайт RingCentral. URL: http://www.ringcentral.com/ (дата обращения: 21.08.2021).

9. Ефимов В.В., Щемелинин Д.А., Яковлев К.А. Интеграционная модель данных для управления непрерывным обслуживанием глобально распределенных вычислительных систем // Труды междунар. науч.-техн. конф. КОМОД-2017,-СПб, Изд-во Политехн. ун-та. – 2017. URL: http://dcn.icc.spbstu.ru/fileadmin/userfiles/Documents/Erasmus/Sbornik_Comod_2017/COMOD-2017_paper_14.pdf (дата обращения: 21.08.2021).

10. Щемелинин Д.А. Программные модели и методы мониторинга состояния процес-синговых узлов в облачной инфокоммуникационной системе с использованием Zabbix // Программные системы и вычислительные методы. – 2021. – № 2. DOI: 10.7256/2454-0714.2021.2.35617 URL: https://nbpublish.com/library_read_article.php?id=35617 (дата обращения: 21.08.2021).

11. Zabbix Enterprise-class Monitoring System // URL: http://www.zabbix.com (дата обращения: 21.08.2021)

12. Иванов В.М., Мещеряков С.В. Методы оптимального проектирования баз данных производственного оборудования / // – СПб.: СПбПУ, – 2012.

13. Иванов В.М., Мещеряков С.В. Реализация модели данных для описания иерархических объектов с произвольными атрибутами // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление.– 2009.– № 1 (72).– С. 139-143.

14. Мещеряков С.В. Сравнительный анализ вариантов организации иерархии в базах данных // Системы управления и информационные технологии.– 2009.– № 1 (35).– С. 34-37.

15. S.V. Mescheryakov, V.V. Efimov, A.N. Volkov, D.A. Shchemelinin Integrated Data Model for Managing a Multi-Service Dynamic Infrastructure // Computer Modeling and Simulation: труды междунар. науч.-техн. конф., СПб, Изд-во Политехн. ун-та, 2014. URL: http://dcn.icc.spbstu.ru/index.php?id=344 (дата обращения: 21.08.2021).

16. Кучерова К.Н., Мещеряков С.В., Щемелинин Д.А. Прогностическое моделирование и визуализация в облачной системе мониторинга // Распределенные компьютерные и теле-коммуникационные сети: управление, вычисление, связь (DCCN-2016): Материалы 19 междунар. науч. конф., Т. 1, М: РУДН, – 2016. URL: http://dccn.ru (дата обращения: 21.08.2021).

Статья поступила в редакцию 12.07.2021
Статья принята к публикации 15.09.2021