ISSN 2518-1092

НАУЧНЫЙ РЕЗУЛЬТАТ

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

RESEARCH RESULT. INFORMATION TECHNOLOGY

6(2) 2021

16+

Сайт журнала:
rrinformation.ru
сетевой научный рецензируемый журнал
online scholarly peer-reviewed journal

НАУЧНЫЙ РЕЗУЛЬТАТ
RESEARCH RESULT

НАУЧНЫЙ РЕЗУЛЬТАТ. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ
RESEARCH RESULT. INFORMATION TECHNOLOGIES

Журнал зарегистрирован в Федеральной службе по надзору в сфере связи, информационных технологий и массовых коммуникаций (Роскомнадзор)
Свидетельство о регистрации средства массовой информации Эл. № ФС77-69101 от 14 марта 2017 г.

The journal has been registered at the Federal service for supervision of communications information technology and mass media (Roskomnadzor)
Mass media registration certificate El. № FS 77-69101 of March 14, 2017

Том 6, № 2. 2021

Volume 6, № 2. 2021

СЕТЕВОЙ НАУЧНО-ПРАКТИЧЕСКИЙ ЖУРНАЛ
Издается с 2016 г.
ISSN 2518-1092

ONLINESCHOLARLYPEER-REVIEWEDJOURNAL
First published online: 2016
ISSN 2518-1092

Учредитель: Федеральное государственное автономное образовательное учреждение высшего образования «Белгородский государственный национальный исследовательский университет»
Издатель: НИУ «БелГУ». Адрес издателя: 308015 г. Белгород, ул. Победы, 85.
Журнал выходит 4 раза в год

Founder: Federal state autonomous educational establishment of higher education «Belgorod State National Research University»
Publisher: Belgorod State National Research University
Address of publisher: 85 Pobeda St., Belgorod, 308015, Russia
Publication frequency: 4 /year

СОДЕРЖАНИЕ

CONTENTS

ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ
INFORMATION SYSTEM AND TECHNOLOGIES

УДК 004.056 **DOI: 10.18413/2518-1092-2021-6-2-0-1**

Нестеренко В.Р.
Маслова М.А.

ИСПОЛЬЗОВАНИЕ ТЕХНОЛОГИИ BLOCKCHAIN ДЛЯ ОБЕСПЕЧЕНИЯ БЕЗОПАСНОСТИ В РАСПРЕДЕЛЕННОМ ИНТЕРНЕТЕ ВЕЩЕЙ

Севастопольский государственный университет, ул. Университетская, д. 33, г. Севастополь, 299053, Россия

e-mail: vladimir.nesterenko.workmail@gmail.com, mashechka-81@mail.ru

Аннотация
Применение блокчейна в сетях интернета вещей – новаторский подход, который способен сделать коммуникации между устройствами такой сети распределенными, автономными и безопасными. Блокчейн в данном контексте представляет из себя совокупность криптографически связанных блоков. Транзакции в сети исполняют роль основных носителей информации о состоянии узлов, а также выходной информации самих узлов для автономного функционирования сети. Узлом является “умное” устройство, датчик или же микроконтроллер, который связывает группу датчиков. Блокчейн применятся для обеспечения защищенной передачи и обработки данных устройств в сети интернета вещей. В данной статье рассмотрены основные возможности и вызовы при применении технологии в распределенных сетях.
Ключевые слова: internet of things (интернет вещей), blockchain (блокчейн), peer-to-peer network, proof of authentication (доказательство аутентификации), децентрализованная сеть.
Для цитирования: Нестеренко В.Р., Маслова М.А. Использование технологии blockchain для обеспечения безопасности в распределенном интернете вещей // Научный результат. Информационные технологии. – Т.6, №2, 2021. – С. 3-8. DOI: 10.18413/2518-1092-2021-6-2-0-1

Nesterenko R.V.
Maslova M.A

USING BLOCKCHAIN TECHNOLOGY TO ENSURE SECURITY IN THE DISTRIBUTED INTERNET OF THINGS

Sevastopol state University, 33 Universitetskaya St., Sevastopol, 299053, Russia

e-mail: vladimir.nesterenko.workmail@gmail.com, mashechka-81@mail.ru

Abstract
The use of Blockchain in the Internet of Things networks is an innovative approach that can make communication between devices of such a network distributed, autonomous and secure. The blockchain in this context is a set of cryptographically connected blocks. Transactions in the network act as the main carriers of information about the state of the nodes, as well as the output information of the nodes themselves for the autonomous functioning of the network. A node is a "smart" device, a sensor, or a microcontroller that connects a group of sensors. Blockchain will be used to provide secure data transmission and processing of devices in the Internet of Things network. This article discusses the main opportunities and challenges in the application of technology in distributed networks.
Keywords: internet of things, blockchain, peer-to-peer network, proof of authentication, a decentralized network.
For citation: Nesterenko R.V., Maslova M.A. Using blockchain technology to ensure security in the distributed internet of things // Research result. Information technologies. – T.6, №2, 2021. – P. 3-8. DOI: 10.18413/2518-1092-2021-6-2-0-1

ВВЕДЕНИЕ

С развитием коммуникационных технологий и с повсеместным введением 5G-сетей, технология “Интернет вещей” начала развиваться с экспоненциальной скоростью. Умный дом, умные города, система e-Health, интернет вещей для промышленных предприятий, распределенный интеллект и другие системы – эффективное и привычное для общества средство улучшения многих процессов, например, процессов орошения урожая на основе датчиков и других процессов, которые могут быть автоматизированы. Подобный подход к процессам уменьшает влияние человеческого фактора и способствует повышению эффективности предприятия, где есть все предпосылки для применения технологии IoT.

При всей своей эффективности и распространенности, технология Internet of things (интернет вещей) имеет множество вызовов и проблем, связанных с безопасностью и безопасной конфигурацией устройств Интернета вещей.

Существование огромного количества подобных устройств несет в себе опасность, так как злоумышленник может взять их под контроль и организовывать с помощью устройств интернета вещей ДДОС-атаки и другие манипуляции с трафиком, которые отсылают данные устройства на сервер.

Одним из примеров согласованных атак множества устройств интернета вещей является ботнет. Ботнет – это совокупность взломанных устройств, находящихся под контролем злоумышленника. Mirai – червь и ботнет, образованный взломанными (скомпрометированными) устройствами типа «интернет вещей» (видеопроигрыватели, «умные» веб-камеры, прочее). Данный ботнет взламывает устройства, подбирая пароль к 23 порту (telnet). В централизованных системах интернета вещей иногда достаточно бывает взломать сервер или микроконтроллер, отвечающий за коммуникации между большой группой устройств, чтобы получить возможность контроля всех устройств, коммуницирующих по централизованному протоколу с взломанным сервером [1, 3, 8].

ОСНОВНАЯ ЧАСТЬ

ДЕЦЕНТРАЛИЗОВАННЫЙ ПОДХОД В INTERNET OF THINGS (Интернете вещей). Централизация системы управления интернетом вещей может являться уязвимостью, так как такая архитектура значительно сокращает время, за которое все устройства в такой сети могут быть взяты под контроль злоумышленником.

Выходом служит использование децентрализованного подхода к организации такой сети, где каждое устройство выступает самостоятельным узлом. В случае такой коммуникации, злоумышленнику придется взломать каждое устройство, а не только центральный сервер. Использование протоколов централизованной коммуникации в децентрализованной сети не является достаточно безопасным и эффективным. Использование технологии Блокчейн для организации коммуникации между устройствами в такой сети наиболее оправданное решение, так как информация будет передаваться в виде безопасных, подписанных транзакций, которые должны быть записаны в распределенном реестре каждого узла.

Подобный подход обеспечивает следующие преимущества и свойства взаимодействия устройств в распределенной сети [5, с. 3]:

1) децентрализация;
2) безопасность;
3) идентификация;
4) гибкость сети;
5) автономность работы сети;
6) надежность информации.

Децентрализация предполагает устранение проблем безопасности централизованного подхода к организации интернета вещей, увеличивая погрешность, однако также увеличивая

эффективность такой сети и безопасность. Транзакции между узлами безопасны, подписаны секретным ключом узла отправителя и проверенные узлом получателем, таким образом обеспечивается безопасность и идентификация. В любой момент в сеть может быть подключено любое количество устройств, которые получат актуальную копию распределенного реестра – таким образом обеспечивается гибкость сети. Автономность работы заключается в невозможности приостановить работу всей сети, выведя из строя какой-либо ее компонент, как это может происходить в централизованной сети при выводе из строя сервера. Надежность информации в сети заключается в том, что в блоках распределенного реестра будут находиться только верифицированные майнерами или иным способом транзакции, содержащие выходную информацию устройств [2, 4].

Децентрализация и peer-to-peer организация сети показывают высокий уровень безопасности, надежности, гибкости сети и возможность автономной работы ее частей.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ

Рассмотрим эффективный консенсус и хранение распределенного реестра. При всех преимуществах децентрализованной сети все еще остаются актуальными следующие вызовы: как хранить распределенный реестр на узле и какой алгоритм консенсуса использовать для эффективной работы сети.

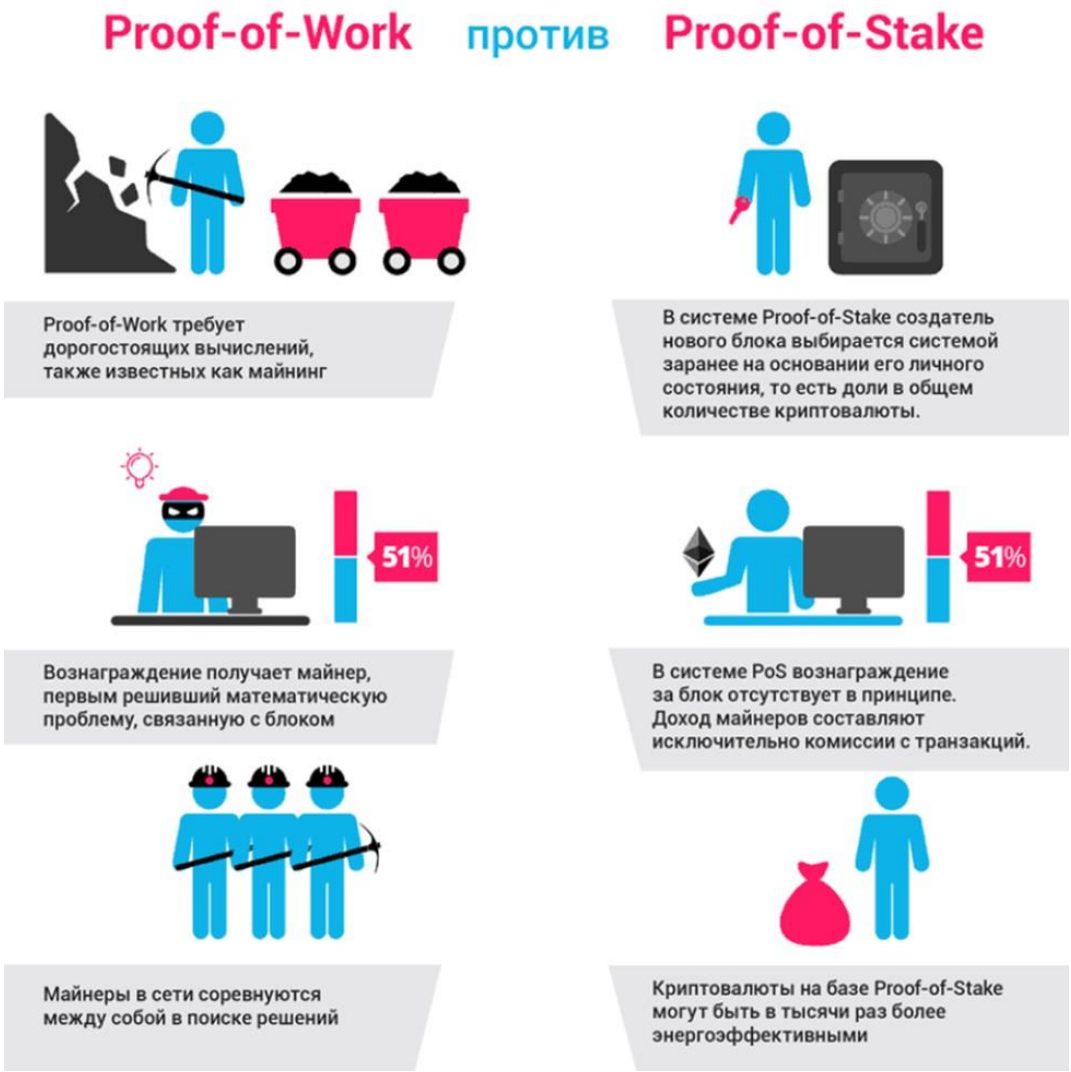


Рис. Доказательство работы и доказательство доли
Fig. Proof-of-work and Proof-of-Stake

Proof-of-work (доказательство работы) – стандартный алгоритм сети Биткоин, который позволяет на основе доказательства каких-то сложных вычислений, доказать проделанную работу для верификации транзакций и криптографического закрытия блока. В огромной сети данный алгоритм консенсуса оказывается очень затратным с точки зрения расходуемой энергии на вычисления для верификации и закрытия блоков.

Сеть интернета вещей должна обеспечивать коммуникацию и принятие решения в режиме реального времени. Данное требование делает доказательство работы неэффективным для решения поставленной задачи. Так как в закрытой сети интернета вещей доказательством работы должны заниматься сами устройства – работа такой сети может быть нарушена из-за высокой нагрузки на устройства, при вычислении доказательства работы.

Майнером считается участник сети, заинтересованный в поддержании работоспособности такой сети за некоторую награду. Возможно организовать распределенную сеть, основанную на доказательстве работы, только сделав интернет вещей открытым для внешних майнеров. В таком случае придется обеспечить достаточно высокую вовлеченность майнеров, чтобы не было задержки в генерации новых блоков и не было дополнительной нагрузки на узлы в сети.

Другим выходом может являться выбор более легковесного алгоритма консенсуса. Например, Proof-of-Stake или доказательство части. Данный алгоритм консенсуса менее требователен к ресурсам, чем доказательство работы (см. рис.1). Однако в соответствии с [7, с. 2], предпочтительным и легковесным алгоритмом консенсуса для распределенного интернета вещей является Proof-of-Authentication (доказательство аутентификации).

Для реализации данного алгоритма необходимо в общей таблице коммуникаций хранить соответствия открытого ключа и MAC-адреса устройства.

Доказательство аутентификации может быть реализовано следующим образом [6, с.9]:

- Выбираются доверенные узлы.
- Недоверенный узел собирает транзакции в блок.
- Недоверенный узел подписывает блок и отправляет всем доверенным узлам.
- Доверенные узлы сопоставляют открытый ключ узла и его MAC-адрес.
- Если все доверенные узлы успешно аутентифицировали узел, отправивший блок, то происходит рассылка этого блока всем узлам сети.
- При получении блока остальные узлы находят хэш заголовка и открывают новый блок с этим хэш-значением в поле “Предыдущий хэш”.

Если доверенный узел не смог аутентифицировать блок и приславший этот блок узел, тогда рейтинг доверия падает на 1. При низком рейтинге доверия происходит переназначение доверенных узлов в сети. Такой алгоритм позволяет очень сильно сократить нагрузку на устройства, а также обеспечить проверку подлинности присланной информации с помощью механизма ЭЦП. Использование приведенного выше алгоритма помимо надежного алгоритма консенсуса также способствует улучшению безопасности хранимых данных.

Хранение распределенного реестра возможно реализовать в облаке, чтобы каждый узел имел возможность доступа к своему участку облака. Таким образом, данные не будут занимать место на самом устройстве.

Также возможно хранить не все, но только самые актуальные блоки с данными на устройствах. Такой способ позволит отказаться от взаимодействия с облаком, а также будет экономить память самих узлов.

ЗАКЛЮЧЕНИЕ

Применение технологии Блокчейн в организации защищенной распределенной сети интернета вещей - очень перспективная и новаторская технология. Данный подход в организации распределенной сети позволяет обеспечить прежде всего автономность работы узлов, высокий уровень безопасности инфраструктуры и элементов интернета вещей, а также идентификацию с помощью ЭЦП. Однако, у применения данной технологии есть некоторые ограничения:

стандартные алгоритмы консенсуса не подходят из-за их привязанности к внешним майнерам или из-за высоких энергозатрат, однако, алгоритм доказательства аутентификации является надежным алгоритмом консенсуса для сети, где майнерами выступают узлы самой сети. Данный алгоритм способен обеспечивать необходимое быстродействие устройств и коммуникацию в реальном времени. Хранение данных возможно реализовать двумя, предложенными в этой статье способами:

1) обеспечить доступ узла к участку облака, где будет храниться копия реестра,
2) хранить только наиболее актуальные блоки в памяти самого узла, обеспечив такому реестру легковесность.

Список литературы

1. Афонькин А.Ю., Ноздрина Н.А. Перспективы развития технологии блокчейн в ближайшем будущем // Научные тенденции: Вопросы точных и технических наук./Сборник научных трудов по материалам XVI международной научной конференции. 2018. С. 20-21.
2. Гончаренко Ю.Ю., Арзамасцев Д.А. Программный модуль для контроля и ведения электронного документооборота на основе технологии блокчейн // Научный результат. Информационные технологии. – Т.5, №3, 2020.
3. Гончаренко Ю. Ю., Паво Ф. Н. Разработка децентрализованного приложения для реализации цифровой идентичности с использованием технологии блокчейн // Вестник УрФО № 3(29) / 2018, С. 23–28.
4. Михаленко Ю.А., Крюкова А.А. Блокчейн как один из элементов цифровизации государства // Вестник Евразийской науки, 2018 №1, https://esj.today/PDF/10ECVN118.pdf
5. Alam, Tanweer. (2019). Blockchain and its Role in the Internet of Things (IoT). International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 151-157. 10.32628/CSEIT195137.
6. Deepak Puthal and Saraju P. Mohanty and Venkata P. Yanambaka and Elias Kougianos (2020). PoAh: A Novel Consensus Algorithm for Fast Scalable Private Blockchain for Large-scale IoT Frameworks
7. Puthal, Deepak & Mohanty, Saraju. (2019). Proof of Authentication: IoT-Friendly Blockchains. IEEE Potentials. 38. 26-29. 10.1109/MPOT.2018.2850541.
8. Горшкова С. Новые технологии на службе интеллектуального права: блокчейн, искусственный интеллект, виртуальная реальность // Сборник научных трудов IX Международного юридического форума (IP форум) // Правовая защита интеллектуальной собственности: проблемы теории и практики Москва, 12–13 февраля 2021 года.

References

1. Afonkin A.Yu., Nozdrina N.A. Prospects for the development of blockchain technology in the near future // Scientific trends: Questions of exact and technical sciences / Collection of scientific papers based on the materials of the XVI International Scientific Conference. 2018. p. 20-21.
2. Goncharenko Yu.Yu., Arzamassev D.A. Program module for monitoring and maintaining electronic document management based on blockchain technology. Research Result. Information Technologies. – Vol. 5, No. 3, 2020
3. Goncharenko Yu.Yu., Pavo F.N. Development of a decentralized application for implementing digital identity using blockchain technology // Bulletin of the Ural Federal District No. 3 (29), 2018, pp. 23-28.
4. Mikhalenko Yu.A., Kryukova A.A. Blockchain as one of the elements of state digitalization // Bulletin of Eurasian Science, 2018 No. 1, https://esj.today/PDF/10ECVN118.pdf
5. Alam, Tanweer. (2019). Blockchain and its Role in the Internet of Things (IoT). International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 151-157. 10.32628/CSEIT195137.
6. Deepak Puthal and Saraju P. Mohanty and Venkata P. Yanambaka and Elias Kougianos (2020). PoAh: A Novel Consensus Algorithm for Fast Scalable Private Blockchain for Large-scale IoT Frameworks
7. Puthal, Deepak & Mohanty, Saraju. (2019). Proof of Authentication: IoT-Friendly Blockchains. IEEE Potentials. 38. 26-29. 10.1109/MPOT.2018.2850541.
8. Gorshkova S. New technologies in the service of intellectual property law: blockchain, artificial intelligence, virtual reality. // Collection of scientific papers of the IX International Legal Forum (IP Forum). // Legal protection of intellectual property: Problems of Theory and Practice Moscow, February 12-13, 2021.

Нестеренко Владимир Романович, студент второго курса магистратуры кафедры Информационная безопасность Института радиоэлектроники и информационной безопасности
Маслова Мария Александровна, старший преподаватель кафедры «Информационная безопасность» Института радиоэлектроники и информационной безопасности

Nesterenko Vladimir Romanovich, second-year master's student of the Department Information security, Institute of Radioelectronics and Information security
Maslova Maria Alexandrovna, senior lecturer of the Department «Information security», Institute of Radioelectronics and Information security

УДК 004.056.55 **DOI: 10.18413/2518-1092-2021-6-2-0-2**

Буханцов А.Д.[1]
Саджид А.Ю.[1]
Устинов А.Н.[1]
Родионов С.В.[2]

ИССЛЕДОВАНИЕ НАДЁЖНОСТИ ШИФРОВАНИЯ РЕЧИ В ТЕХНОЛОГИИ МОБИЛЬНОЙ СВЯЗИ GSM

[1)] Белгородский государственный национальный исследовательский университет,
ул. Победы д. 85, г. Белгород, 308015, Россия
[2)] Украинский государственный университет железнодорожного транспорта,
пл. Фейербаха, д. 7, г. Харьков, 61050, Украина

e-mail: bukhantsov@bsu.edu.ru, 1275117@bsu.edu.ru, 1319385@bsu.edu.ru, rodionov.serhii@kart.edu.ua

Аннотация
В статье проводится анализ стойкости алгоритма поточного шифра A5/1 в системе сотовой связи GSM на основе разработанной программы моделирования процесса формирования ПСП комбинированной схемой из линейных регистров и реализованного в программной среде MatLab статистического теста FIPS 140 – 1 / FIPS 140 – 2. Результаты анализа позволяют установить, что последовательность, формируемая на основе комбинированной схемы из ЛРР A5/1, не является псевдослучайной, что может позволить выработать рекомендации по дальнейшему совершенствованию механизма защиты в данной системе мобильной связи.
Ключевые слова: шифрование, алгоритм A5/1, сеть GSM, аутентификация.
Для цитирования: Буханцов А.Д., Саджид А.Ю., Устинов А.Н., Родионов С.В. Исследование надёжности шифрования речи в технологии мобильной связи GSM // Научный результат. Информационные технологии. – Т.6, №2, 2021 – С. 9-17. DOI: 10.18413/2518-1092-2021-6-2-0-2

Buhantsov A.D.[1]
Sadjiid A.Yu.[1]
Ustinov A.N.[1]
Rodionov C.V.[2]

RESEARCH OF SPEECH ENCRYPTION RELIABILITY IN GSM MOBILE COMMUNICATION TECHNOLOGY

[1)] Belgorod State National Research University,
85 Pobedy St., Belgorod, 308015, Russia
[2)] Ukrainian State University of Railway Transport
7 Feuerbach Square, Kharkiv, 61050, Ukraine

e-mail: bukhantsov@bsu.edu.ru, 1275117@bsu.edu.ru, 1319385@bsu.edu.ru, rodionov.serhii@kart.edu.ua

Abstract
The article analyzes the strength of the A5/1 stream cipher algorithm in the GSM cellular communication system based on the developed program for modeling the PSP process with a combined circuit of linear registers and the FIPS 140-1 / FIPS 140-2 statistical test implemented in the MatLab software environment. The results of the analysis make it possible to establish that the sequence formed on the basis of the combined scheme from the LRR A5 / 1 is not pseudo-random, which can make it possible to develop recommendations for further improving the protection mechanism in this mobile communication system.
Keywords: encryption, algorithm A5/1, network GSM, authentication.
For citation: Buhantsov A.D., Sadjiid A.Yu., Ustinov A.N., Rodionov C.V. Research of speech encryption reliability in GSM mobile communication technology // Research result. Information technologies. – T.6, №2, 2021. – P. 9-17. DOI: 10.18413/2518-1092-2021-6-2-0-2

ВВЕДЕНИЕ

В настоящее время повсеместно используются беспроводные технологии связи, которые с одной стороны являются удобными в использовании, но с другой стороны увеличивают риски утечки данных, что требует дальнейшего повышения степени защищенности таких технологий. Одной из гарантий надежности внедряемых криптографических алгоритмов является их открытость, которая позволяет экспертному сообществу проводить их анализ и находить слабые места с целью дальнейшего совершенствования их защищенности. Одной из таких систем является стандарт сотовой связи GSM, механизмы защиты которого хорошо известны и широко опубликованы в открытой печати [1, 3-6].

В данной статье рассмотрены принципы организации безопасной передачи информации в системе сотовой связи GSM, проведен процесс моделирования некоторых элементов системы защиты и сформулированы предложения по их совершенствованию.

Одним из алгоритмов шифрования, применяемых в данном стандарте, является поточный шифр A5/1. В свое время он имел ряд преимуществ в обеспечении конфиденциальности беспроводной связи. Для того чтобы понять, насколько данный алгоритм эффективен в наше время, будет проведена оценка стойкости использования данного метода шифрования при помощи моделирования схемы A5/1 и статического теста для генераторов случайных чисел FIPS 140–1 / FIPS 140–2.

ОРГАНИЗАЦИЯ ЗАЩИТЫ СЕТИ GSM

В технологи GSM (Global System for Mobile Communications) элементы безопасности реализуется в трех объектах: SIM-карта, GSM-телефон и сеть. Модуль идентификации абонента (SIM) содержит [3]:

- IMSI – TMSI – PIN;
- Ключ аутентификации *Ki* (64-битный);
- Алгоритм A8 генерации ключа шифрования *Kc*;
- Алгоритм аутентификации *A3*;
- SIM-карта защищена PIN-кодом и принадлежит оператору.

Из этого следует, что техническая безопасность GSM обеспечивается набором алгоритмов, используемых для организации соединения сотового телефона с сетью оператора GSM.

АУТЕНТИФИКАЦИЯ МОБИЛЬНОЙ СТАНЦИИ

Каждый абонент мобильного телефона получает стандартный модуль аутентификации (SIM-карту), который содержит следующие данные на период использования системы связи:

- IMSI (International Mobile Station Indentifer) – международный идентификационный номер мобильного абонента,
- свой индивидуальный ключ аутентификации K_i,
- алгоритм аутентификации *A3*.

Также каждому абоненту системы связи присваивается «временное удостоверение личности» или временный международный идентификационный номер пользователя TMSI (Temporary Mobile Station Indentifer). После завершения процесса аутентификации и запуска режима шифрования временный идентификационный номер – TMSI передается на мобильную станцию, только в зашифрованной форме. Этот номер TMSI используется для всего последующего доступа к системе. Необходимая информация об участниках хранится в базах данных оператора сети участника [4].

С помощью заложенной в SIM информации в результате взаимного обмена данными между мобильной станцией и сетью осуществляется полный цикл аутентификации и разрешается доступ абонента к сети. На рисунке 1 представлена схема аутентификации.

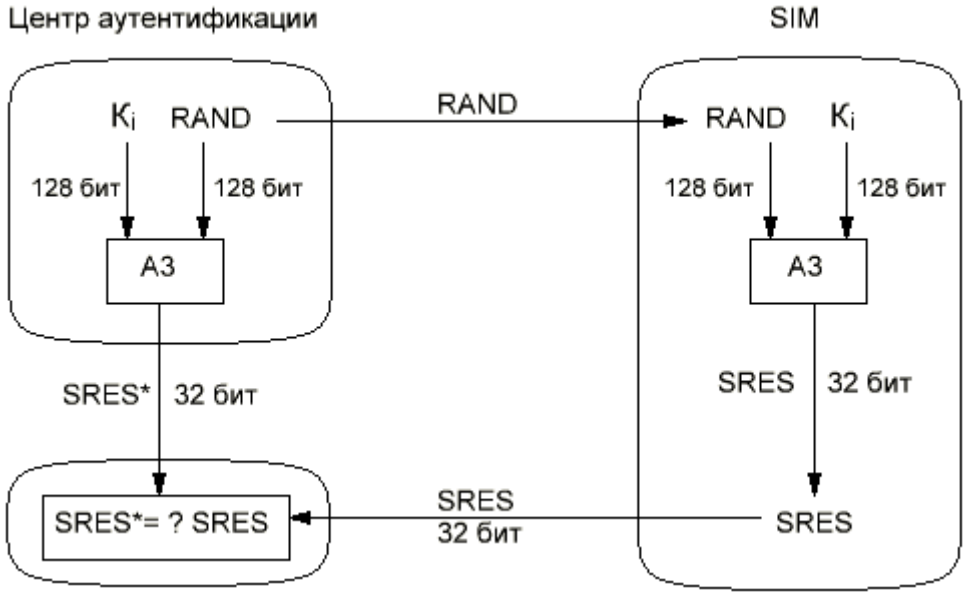


Рис. 1. Схема процесса аутентификации GSM
Fig. 1. GSM authentication Process Diagram

Процесс аутентификации выглядит следующим образом [4]:

- На каждой SIM-карте запрограммирован уникальный ключ аутентификации абонента. Центр аутентификации (ЦА) имеет список, который сопоставляет номер K_i с SIM-картой.
- Когда SIM-карта запрашивает вызов, 128-битное случайное число мгновенно генерируется ЦА и передается на SIM-карту.
- Алгоритм A3, который запрограммирован внутри SIM-карты, обрабатывает число RAND и число K_i, и генерирует 32-битный вывод, называемый подписанным номером ответа (SRES).
- Тот же процесс выполняется на стороне ЦА.
- SIM-карта передает этот номер SRES в ЦА.
- ЦА сравнивает полученный SRES с SRES, сгенерированным на стороне сети.
- SIM-карта аутентифицируется тогда и только тогда, когда два SRES одинаковы.

ШИФРОВАНИЕ В СЕТИ GSM

Сеть GSM использует информацию, хранящуюся на SIM-карте и в телефоне, для обеспечения зашифрованной связи и аутентификации. Шифрование GSM применяется только к связи между мобильным телефоном и базовой станцией. Остальная часть передачи по обычной фиксированной сети или радиорелейной сети не защищена, и ее можно легко перехватить или изменить. Шифрование GSM достигается за счет использования общего секретного ключа. Ключ 64-бит разделен для обеспечения конфиденциальности данных [5]. Невозможно зашифровать все данные; например, некоторая информация о маршруте должна быть отправлена в виде открытого текста. Подробный процесс шифрования данных показан на рисунке 2.

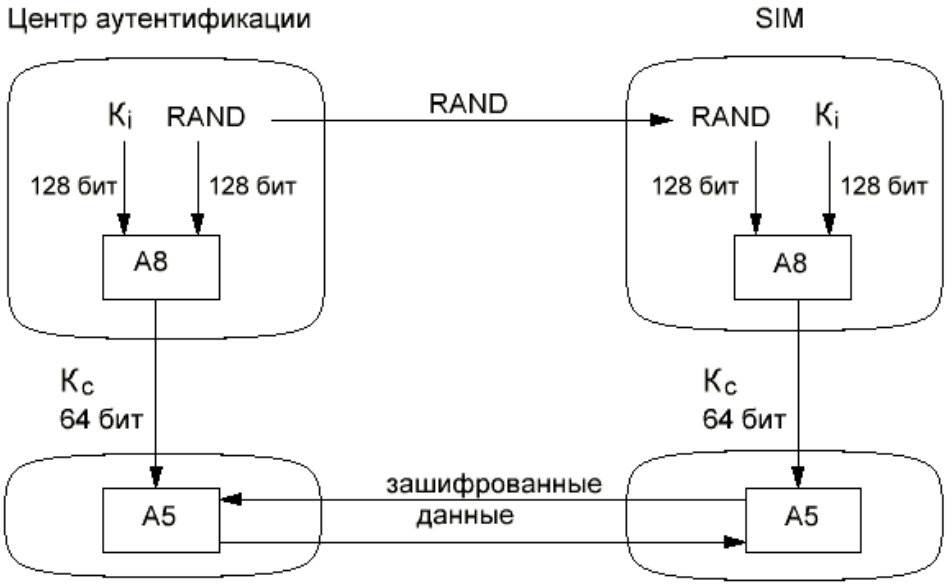


Рис. 2. Схема процесса шифрования GSM
Fig. 2. GSM encryption Process Diagram

Алгоритм выполнения шифрования выглядит следующим образом [4]:

- ЦА генерирует случайное число (RAND) из 128 бит и отправляет его в МС.
- RAND и число *Ki* обрабатываются алгоритмом A8 с обеих сторон. Алгоритм A8 создает 64-битный ключ шифрования *Кс*.
- Алгоритм A5 использует ключ *Кс* при поточном шифровании данных.

В системе мобильной связи стандарта GSM алгоритм шифрования речи A5 использует сложение по модулю-2 данных (после соответствующего преобразования речи в двоичную последовательность) и ПСП (псевдослучайная последовательность), которая вырабатывается комбинированной схемой из ЛЛР с псевдослучайным тактированием [3].

Алгоритм состоит из 3-х линейных регистров сдвига с обратной связью (ЛРОС) длиной 19, 22 и 23. Алгоритм A5 существует в двух модификациях A5 / 2 и A5 / 1.

Рассмотрим подробнее алгоритм A5 / 1.

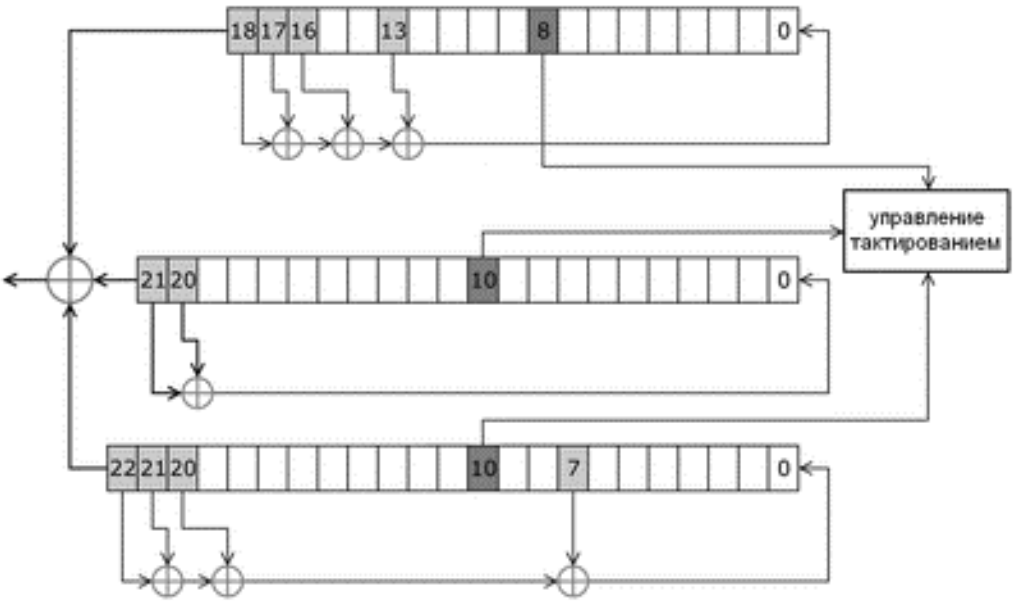


Рис. 3. Алгоритм шифрования A5/1 стандарта GSM
Fig. 3. A5/1 GSM encryption algorithm

Из рисунка 3 видно, что генератор гаммы потокового шифра A5/1 состоит из трех ЛРР, соответствующих примитивным полиномам: ($x^{19}+x^{18}+x^{17}+x^{14}+1$; $x^{22}+x^{21}+1$; $x^{23}+x^{22}+x^{21}+x^{8}+1$).

Темным цветом выделены биты, от которых существенно зависят функции обратной связи. Все три регистра используют псевдослучайное тактирование, которое работает по следующему правилу: биты с отвода 8 первого ЛРР, а также с отводов 10 второго и третьего ЛРР, подаются на так называемый *мажоритарный элемент*. Последний выдает на выходе значение 0 или 1, в зависимости от того, появляется ли на его входах больше нулей или единиц. Далее выход этого мажоритарного элемента сравнивается со значениями выходов на трех отводах ЛРР с номерами 8, 10, 10 (которые подавались ранее на входы этого мажоритарного элемента), и каждый ЛРР продвигается на один такт тогда и только тогда, когда сравниваемые биты оказываются одинаковыми. Шифрующая гамма формируется как сумма по модулю 2 выходов всех трех ЛРР. Ключом являются начальные заполнения всех ЛРР, которые вводятся на начальном этапе без псевдослучайного тактирования. Общая длинна ключа составляет 64 бита. [3]

ИССЛЕДОВАНИЕ НАДЕЖНОСТИ ШИФРА A5/1

Проведем небольшое исследование шифра A5/1 на предмет анализа стойкости и проверим возможности её повышения.

Тестировать полученную последовательность будем с помощью статического теста для генераторов случайных чисел FIPS 140 - 1 и FIPS 140 – 2.

Рассматриваемый стандарт FIPS 140 – 1 рекомендуется для оперативного тестирования последовательностей, формируемых некоторым ГПСП (генератором псевдослучайных последовательностей) и использует 4 теста. При этом для проведения тестирования требуется битовая строка длиной 20 тыс. бит. При этой длине исследуемой последовательности допустимые интервалы для статистик, вычисляемых по каждому из тестов, задаются в явном виде (то есть нет необходимости предварительно выбирать соответствующие уровни значимости).

В стандарте FIPS 140 – 2 для уменьшения вероятности принятия ошибочного решения были пересмотрены (сделаны более жесткими) допустимые интервалы для каждого из статистических тестов.

Рассмотрим кратко предлагаемые стандартом тесты [6]:

1. Монобитный тест (частотный тест).

В исследуемой последовательности количество единичных бит N_1 должно находиться в следующем интервале:

FIPS 140 -1 $9654 < N_1 < 10346$

FIPS 140 -2 $9725 < N_1 < 10275$

2. Покер – тест (блочный тест).

По исследуемой последовательности подсчитывается следующая статистика:

$$X_3 = \frac{2^m}{k}\left(\sum_{i=1}^{2^m} n_i^2\right) - k \qquad (1)$$

где m – длина подсчитываемых неперекрывающихся подпоследовательностей для данного стандарта (принято m=4);

n_i – количество появлений подпоследовательности i – того типа длины m (для m = 4 существует $2^m = 2^4 = 16$ типов подпоследовательностей);

k – общее количество неперекрывающихся подпоследовательностей длины m (для данного стандарта k = 20000 / 4 = 5000).

Значение полученной статистики X_3 должно находиться в следующем интервале.

FIPS 140 -1 $1.03 < X_3 < 57.4$

FIPS 140 -2 $1.16 < X_3 < 46.17$

3. Тест серий.

Серией считают подпоследовательность исходной последовательности, которая состоит из битов одного типа (либо “0”, либо “1”), которым не предшествует и за которыми не следует бит того же типа (“0” или “1” соответственно).

В данном тесте для исследуемой последовательности подсчитывается количество единичных S_i и количество нулевых Z_i серий длины i (**$1 \le i \le 6$**, серии большей длины в данном тесте рассматриваются как серии длины **6**).

Тест серий считается успешно пройденным, если все **12** подсчитанных значения (S_i и Z_i **,** **$1 \le i \le 6$)** принадлежат соответствующим интервалам, приведенным в следующей таблице.

Таблица 1

Разрешенные интервалы

Table 1

Allowed intervals

Длина серии		1	2	3	4	5	6
Допустимые диапазона	*FIPS 140 -1*	2267-2733	1079-1421	502-748	223-402	90-223	90-223
	FIPS 140 - 2	2343-2657	1135-1365	542-708	251-373	111-201	111-201

4. Тест максимальной длины серии.

Тест считается успешно пройденным, если в исследуемой последовательности не существуют серии длиной 34 (FIPS 140 - 1), 26 (FIPS 140 - 1) и более.

ЭКСПЕРИМЕНТ

Алгоритмы рассмотренных тестов были реализованы в программной среде MatLab. Если хоть один из тестовых результатов не будет удовлетворять требованиям, то тестируемая последовательность не соответствует необходимой степени защиты стандарта FIPS.

Также для моделирования процесса генерации ПСП с помощью комбинированной схемы из ЛРР использовалась программа формирования ПСП на основе комбинированной схемы из линейных регистров сдвига [2]. Программа обеспечивает выполнение следующих функций:

– формирование трех линейных рекуррентных регистров на основе произвольных полиномов не более тридцать первой степени;

– произвольный выбор номера отвода на мажоритарный элемент от каждого из регистров, а также инверсию схемы мажоритарного элемента;

– произвольное начальное битовое заполнение сформированных линейных регистров сдвига;

– выбор шага генерации результирующей псевдослучайной последовательности и ее визуальное отображение;

– визуальное отображение псевдослучайной последовательности, формируемой отдельным регистром на очередном шаге генерации.

Начнем тестирование с получения ПСП. Программа позволяет задать: примитивные полиномы до максимальной степени 31; номер отвода от ЛРР; количество выполняемых шагов (число бит получаемой ПСП) и начальное заполнение регистров. Окно программы показано на рисунке 4. Изначально поля полиномов заполнены согласно стандарту шифра A5/1.

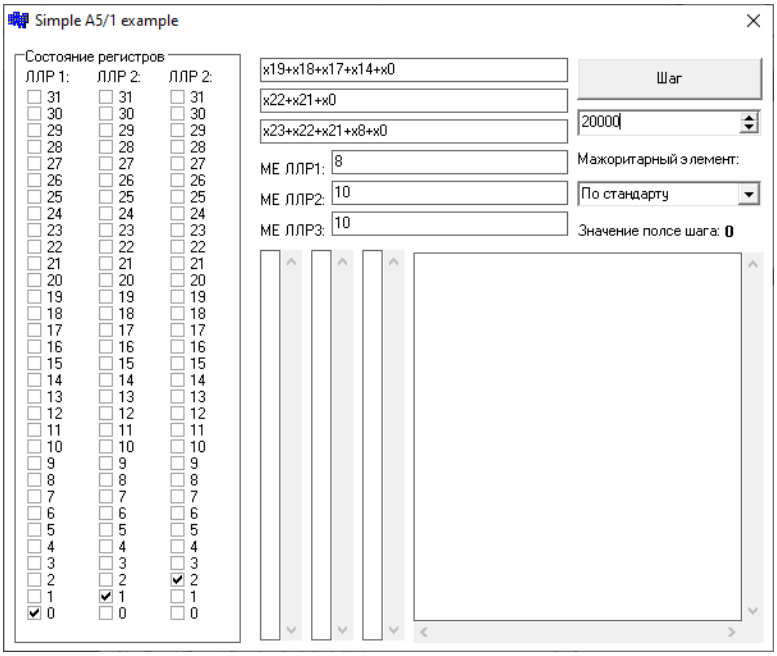


Рис. 4. Окно программы "Генератор А5-1", начальный интерфейс
Fig. 4. "Generator A5-1" program window, initial interface

После запуска программы на рисунке 5, можно увидеть сгенерированную последовательность.

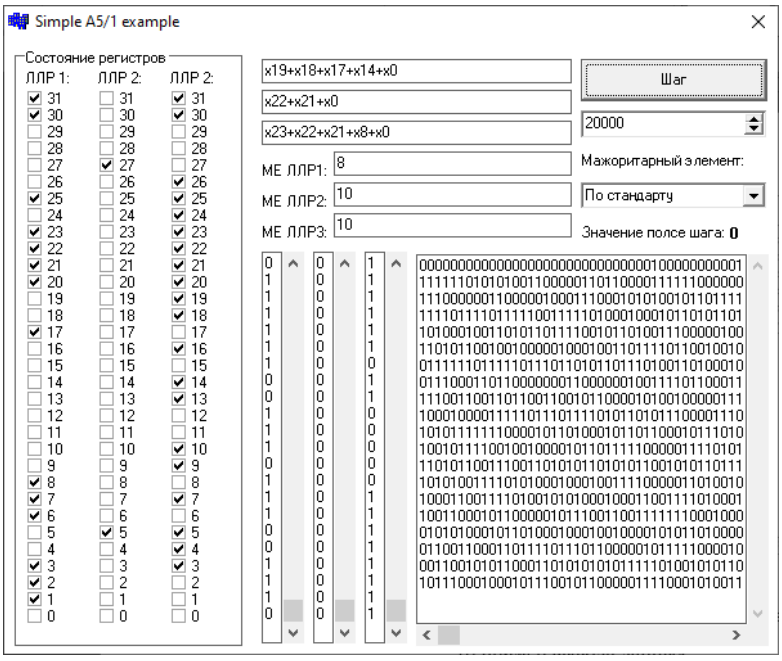


Рис. 5. Окно программы "Генератор А5-1", вывод данных
Fig. 5. Program window "Generator A5-1", data output

Полученную последовательность тестируем по стандартам FIPS 140 – 1 и FIPS 140 – 2, и получаем следующие результаты:

Таблица 2

Результат монобитного и покер теста

Table 2

The result for monobit and poker

Тесты	X	FIPS 140 – 1	FIPS 140 – 2
Монобитный	9929	9654 < X<10346	9725< X < 10275
Покер	12.25	1.03 <X<57.4	2.16 < X < 46.17

Таблица 3

Результат теста серий

Table 3

Series Test result

Серии	X_0	X_1	FIPS 140 – 1	FIPS 140 – 2
1	2629	2522	$2267<X_{0,1}<2733$	$2343 <X_{0,1}<2657$
2	1237	1197	$1079<X_{0,1}<1421$	$1135 <X_{0,1}<1365$
3	597	686	$502<X_{0,1}<748$	$542 <X_{0,1}<708$
4	300	321	$223<X_{0,1}<402$	$251 <X_{0,1}< 373$
5	157	139	$90<X_{0,1}<223$	$111 <X_{0,1}<201$
6	86	77	$90<X_{0,1}<223$	$111 <X_{0,1}< 201$

Таблица 4

Результат теста максимальной длины серий

Table 4

Test result of the maximum length of the series

Тест	X	FIPS 140 – 1	FIPS 140 – 2
Максимальная серия длинны	30	34	26

По результатам тестирования можно увидеть, что формируемая последовательность не проходит тест из 6-ти серий и тест максимальной длины серий. Из этого следует, что исследуемая последовательность на выходе комбинированной схемы из ЛЛР, соответствующей генератору A5/1, не проходит тестирование.

Таким образом, можно сделать вывод, что последовательность, формируемую комбинированной схемой A5/1, нельзя считать псевдослучайной, следовательно, ее предсказуемость значительно снижает защищенность исследуемой системы сотовой связи.

ЗАКЛЮЧЕНИЕ

В данной работе был проведен анализ стойкости алгоритма поточного шифра A5/1 с помощью статистического теста FIPS 140–1 / FIPS 140–2, реализованного в программной среде MatLab, и моделирующей программы формирования ПСП на основе комбинированной схемы из линейных рекуррентных регистров.

По результатам тестирования удалось установить, что последовательность, формируемая на основе комбинированной схемы из ЛРР A5/1, не проходит тест, и, следовательно, не является псевдослучайной. Поэтому остается актуальной задача поиска полиномов для комбинированной схемы поточного шифра в технологии GSM, позволяющей формировать последовательность с необходимыми свойствами для достижения требуемого уровня защищенности голосовой связи и данных.

Список литературы

1. Защита информации в системах мобильной связи. Учебное пособие для вузов. Под ред. А.В. Заряева и С.В. Скрыля. – 2-е изд. испр. и доп. М: Горячая линия Телеком, 2015. – 171 с.

2. Буханцов А.Д., Черноморец А.А., Болгова Е.В. "Программная система формирования псевдослучайной последовательности на основе комбинированной схемы из линейных регистров сдвига" / Свидетельство о государственной регистрации программы для ЭВМ № 2015619256 от 27.08.2015.

3. Ветров Ю.В. Криптографические методы защиты информации в телекоммуникационных системах: учеб. пособие / Ю.В. Ветров, С.Б. Макаров. – СПб.: Изд-во Политехн. ун-та, 2011. – 174 с.

4. Andreas Bubla (2004). Kryptographie in Mobilfunknetze (GSM, UMTS) [Online], availed at: https://www.bubla.info/informatik/files/krypto_gsm_umts_paper.pdf (Accessed 20 April 2021)

5. Асосков А.В., Иванов М.А., Мирский А.А., Рузин А.В., Сланин А.В., Тютвин А.Н. Поточные шифры. – М.: КУДИЦ-ОБРАЗ, 2003. – 336 с.

6. Попов В.И. Основы сотовой связи стандарта GSM / В.И. Попов. – М.: Эко-Трендз, 2005. – 29 с.

References

1. Zashchita informatsii v sistemakh mobilnoy svyazi. Uchebnoye posobiye dlya vuzov. Pod red. A.V. Zaryayeva i S.V. Skrylya. – 2-e izd. ispr. i dop. M: Goryachaya liniya Telekom. 2015. – 171 s.

2. Bukhantsov A.D., Chernomorets A.A., Bolgova E.V. "Programmnaya sistema formirovaniya psevdosluchaynoy posledovatelnosti na osnove kombinirovannoy skhemy iz lineynykh registrov sdviga" / Svidetelstvo o gosudarstvennoy registratsii programmy dlya EVM № 2015619256 ot 27.08.2015

3. Vetrov Yu.V. Kriptograficheskiye metody zashchity informatsii v telekommunikatsionnykh sistemakh: ucheb. posobiye / Yu.V. Vetrov. S.B. Makarov. – SPb.: Izd-vo Politekhn. un-ta. 2011. – 174 s.

4. Andreas Bubla (2004). Kryptographie in Mobilfunknetze (GSM, UMTS) [Online], availed at: https://www.bubla.info/informatik/files/krypto_gsm_umts_paper.pdf (Accessed 20 April 2021)

5. Acoskov A.V., Ivanov M.A., Mirskiy A.A., Ruzin A.V., Slanin A.V. Tyutvin A.N. Potochnyye shifry. – M.: KUDITs-OBRAZ. 2003. – 336 s.

6. Popov V.I. Osnovy sotovoy svyazi standarta GSM / V.I. Popov. – M.: Eko-Trendz. 2005. – 296 s.

Буханцов Андрей Дмитриевич, кандидат технических наук, старший научный сотрудник, доцент кафедры информационно-телекоммуникационных систем и технологий
Саджид Александр Юрьевич, студент кафедры информационно-телекоммуникационных систем и технологий
Устинов Алексей Николаевич, студент кафедры информационно-телекоммуникационных систем и технологий
Родионов Сергей Викторович, кандидат технических наук, доцент, доцент кафедры транспортной связи

Buhantsov Andrey Dmitrievich, Candidate of Technical Sciences, Senior Researcher, Associate Professor of the Department of Information and Telecommunications Systems and Technologies
Sajiid Alexander Yuryevich, student of the Department of Information and Telecommunications Systems and Technologies
Ustinov Alexey Nikolaevich, student of the Department of Information and Telecommunications Systems and Technologies
Rodionov Sergey Viktorovich, Candidate of Technical Sciences, Associate Professor, Associate Professor of the Department of Transport Communications

UDC 004.492.3 **DOI: 10.18413/2518-1092-2021-6-2-0-3**

Goncharenko Yu.Yu.
Devitsyna S.N.
Shapovalov P.A.

AUTOMATIC DETECTION OF JAVASCRIPT SNIFFERS

Sevastopol State University, 33 Universitetskaya St., Sevastopol, 299053, Russia

e-mail:iuliay1985@mail.ru, sndevitsyna@sevsu.ru, actek2009@gmail.com

Abstract
The method of automatic detection of javascript sniffers in the code of online stores is considered. As well as its software implementation, which allows you to detect embedded javascript code with a high probability, notify the resource administrator and the user who is on the payment page, thereby protecting the online store from reputational threat.
Keywords: javascript sniffer, online store protection, javascript sniffer detection program.
For citation: Goncharenko Yu.Yu., Devitsyna S.N., Shapovalov P.A. Automatic detection of javascript sniffers // Research result. Information technologies – T.6, №2, 2021. – P. 18-24. DOI: 10.18413/2518-1092-2021-6-2-0-3

Гончаренко Ю.Ю.
Девицына С.Н.
Шаповалов П.А.

АВТОМАТИЧЕСКОЕ ОБНАРУЖЕНИЕ JAVASCRIPT-СНИФФЕРОВ

Севастопольский государственный университет, ул. Университетская, д. 33, г. Севастополь, 299053, Россия

e-mail:iuliay1985@mail.ru, sndevitsyna@sevsu.ru, actek2009@gmail.com

Аннотация
Рассмотрен метод автоматического обнаружения javascript-снифферов в коде интернет-магазинов, а также его программная реализация, которая позволяет с высокой вероятностью обнаруживать встроенный javascript-код, уведомлять администратора ресурса и пользователя, находящегося на странице оплаты, тем самым защищая интернет-магазин от репутационных угроз.
Ключевые слова: javascript-сниффер, защита интернет-магазина, программа обнаружения javascript-сниффера.
Для цитирования: Гончаренко Ю.Ю., Девицына С.Н., Шаповалов П.А. Автоматическое обнаружение javascript-снифферов // Научный результат. Информационные технологии. – Т.6, №2, 2021. – С. 18-24. DOI: 10.18413/2518-1092-2021-6-2-0-3

INTRODUCTION

Online stores are a dynamically evolving online commerce tool. According to forecasts of the Russian research agency Data Insight [Data Insight, 2021], the annual volume of revenue from Internet commerce by 2023 will more than double and will amount to 2.4 trillion. rubles.

The convenience of purchasing goods on the Internet has a downside: buyers who use bank cards to pay online face a variety of cyber threats, one of which is JavaScript sniffers [Technology and Media, 2021; Central Bank of the Russian Federation, 2018; Group-IB, 2021].

Sniffer is a type of malicious code injected by cybercriminals into a victim's website script to intercept user-entered data: bank card numbers, names, addresses, logins, passwords, and so on. When a site is infected, all parties are involved in the chain of victims - end users, payment systems, banks and large companies that sell their goods and services via the Internet.

Having analyzed the existing methods for detecting vulnerabilities in web ap-plications, as well as attack methods [Lukatsky A., 2008; Mark Dowd, John McDonald and Justin Schuh, 2007], a new method for detecting JavaScript sniffers is proposed.

MAIN PART

Analysis of the javascript sniffer detection technique

The technique of automatic detection of javascript sniffers consists in comparing the states of the same page at different intervals according to a certain algorithm. At the initial stage, it is assumed that the web application is "clean", that is, it is not infected with javascript sniffer and other malicious code. The detection system saves a perfect "snapshot" of the page that needs to be monitored for in-fection and puts it in a store of the original state.

An impression represents the following states:

- DOM-tree of the document;
- the counted number of internal and external url-links;
- the weight of each link that has the extension js, gif, png, jpg.

At the necessary time intervals, the system accesses the url page, which is reg-istered, makes its "impression" and compares the current "impression" with the original one. Figure 1 shows the main nodes of the warning system and the inter-action between them.

In case of detection of differences in the states of "snaps" - the notification system about the threat of infection of the url-page being monitored is triggered. A notification about this fact is sent to the resource administrator, and also, if the client of interaction with the browser is connected, to the user's browser window, which is located on the monitored page.

Software implementation of a javascript sniffer detection system

The javascript sniffer detection system consists of the following nodes:

- web-based management interface;
- parsing module;
- module for processing and saving results;
- module for comparison and decision making;
- threat notification module;
- repositories of the initial and current state.

To implement the detection system, the author has chosen the php programming language. Intermediate data will be stored in text files. The advantage of this configuration is that it is cross-platform and easy to deploy.

The web interface is implemented on the bootstrap 4 framework. It consists of two sections - snapshot and monitoring. The monitoring section has an advanced and simple interface. Advanced - for system testing and detailed study of detect-ed threats. Simple interface - combined with the notification module i.e. when accessing the simple interface, information about the presence or absence of a threat is returned without the details that are present in the extended interface.

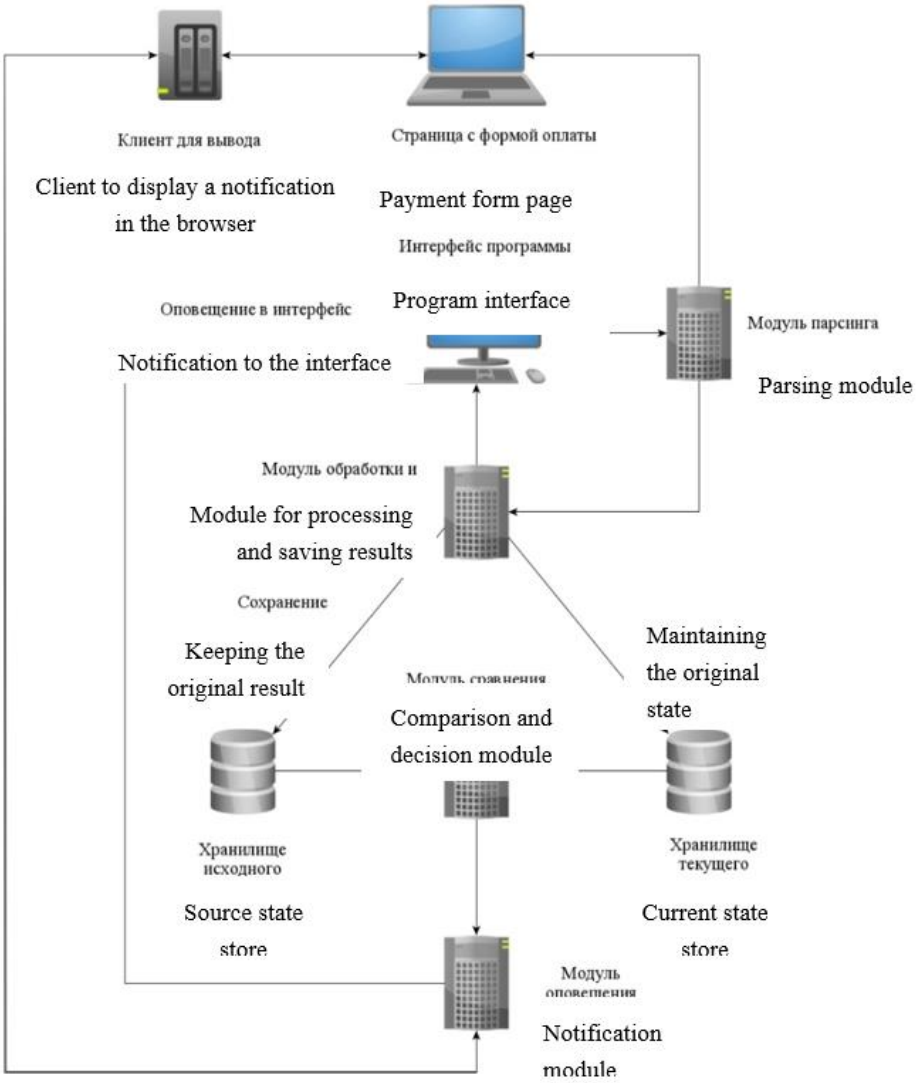


Fig. 1. Nodes of the sniffer notification system

Testing program code

A test page with a form of payment is used to test the scenario of work (see Fig. 2).

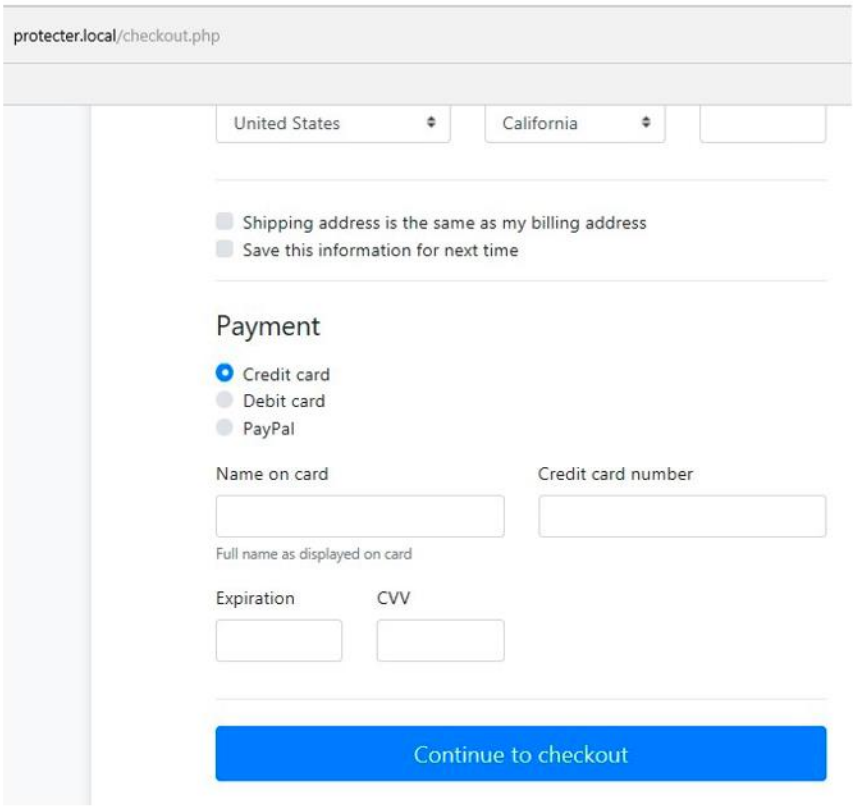


Fig. 2. Test page with payment form

The formed "snapshot" of this page allowed adding it to monitoring (see Fig. 3).

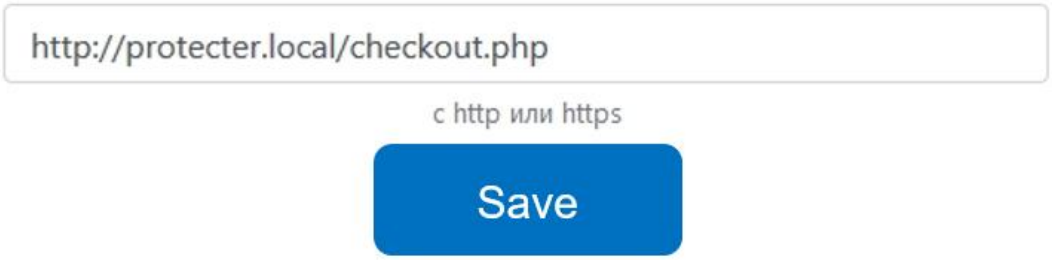


Fig. 3. Procedure for adding a page to monitoring

The monitoring page is a screen divided into two parts (see Fig. 4). On the left, the initial state is displayed, and on the right, the current state. Until threats are detected, the system is in equilibrium.

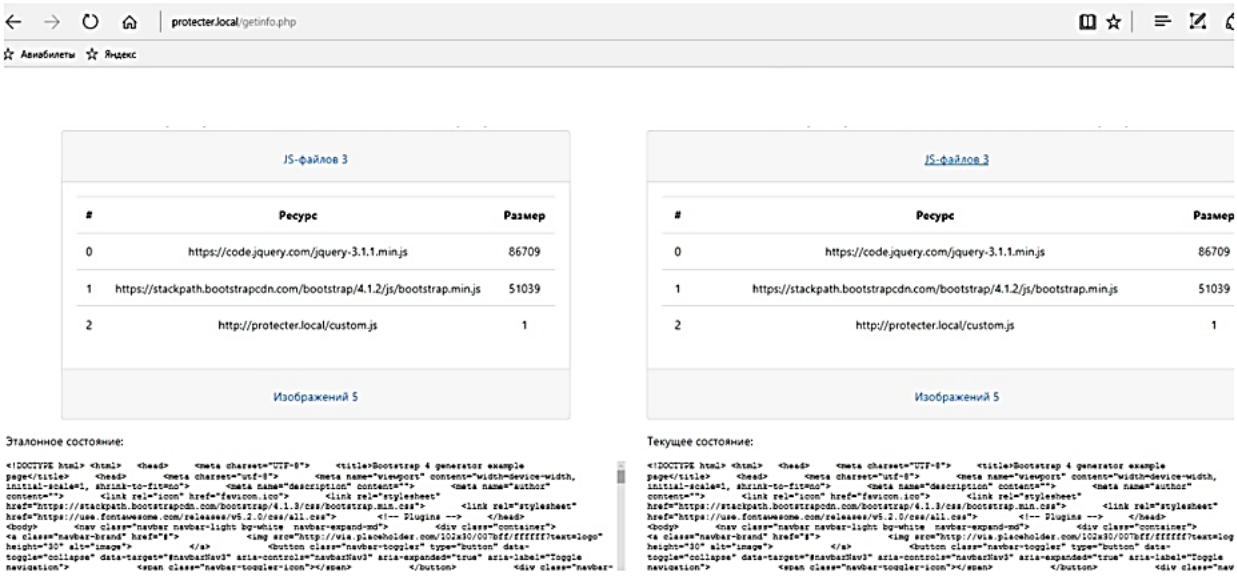


Fig. 4. Advanced interface of the monitoring results page

If you put arbitrary javascript code into a page with a form, a corresponding warning will be displayed in the advanced interface of the detection system (see Fig. 5).

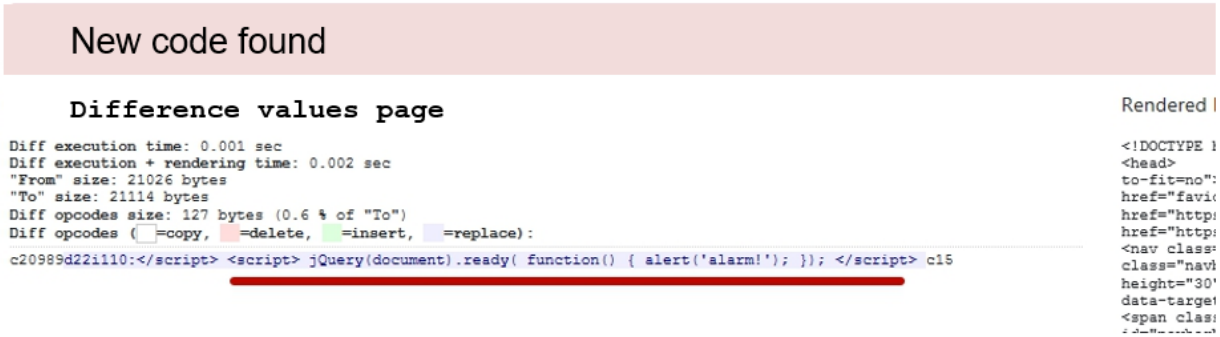


Fig. 5. Notification in the extended interface

And in the browser, the user who is on the page with the payment form will receive a notification about the threat (see Fig. 6).

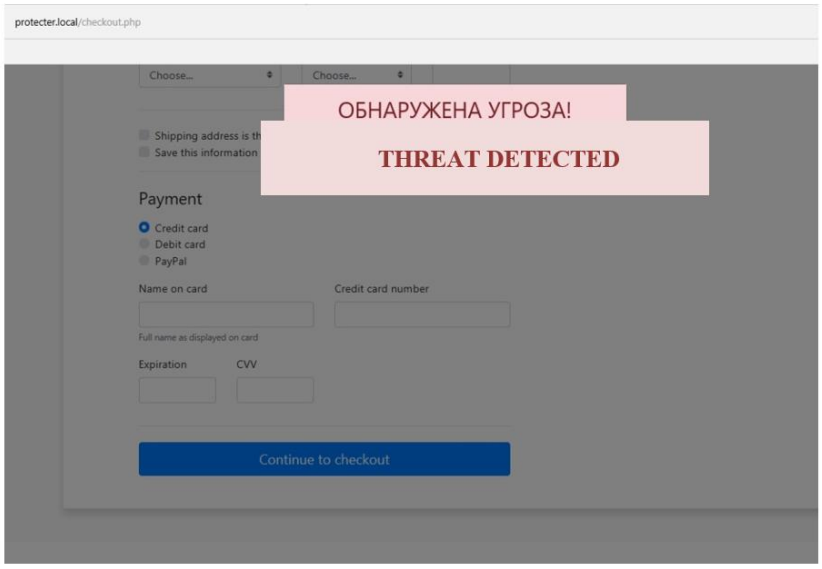


Fig. 6. Notification of a threat in the user's browser

If an attacker tries to place malicious code not directly on the page, but through a file already connected to this page, the system will also notify about this, since the file size will change.

CONCLUSIONS

Thus, using the proposed method for comparing the initial and current state, it is possible to detect the embedded sniffers code with 100% probability, since any change in the code or size of an external file, even by 1 byte, triggers the notification system.

The developed program for detecting changes in the source code can be used:

- to warn about the injection of malicious code into a web application;
- in analytical systems, where it is necessary to track changes in the current state from the initial one;
- an additional module for existing WAFs;
- to compare two texts and find differences.

References

1. Data Insight, 2021. URL: http://www.datainsight.ru/public (date of circulation: 07.05.2021).

2. Technology and Media, 2021. URL: https://www.rbc.ru/technology_and_media, (date of circulation: 07.05.2021).

3. Central Bank of the Russian Federation, 2018. Overview of the main types of computer attacks in the credit and financial sector in 2018. URL: https://cbr.ru/Content/Document/File/72724/DIB_2018_20190704.pdf (date of circulation: 08.05.2021).

4. Group-IB, 2021. Getting to know sniffers: the ReactGet family. URL: https://www.group-ib.ru/blog/reactget, (date of circulation: 09.05.2021).

5. Group-IB. Getting to know sniffers-2: G-Analytics, 2021. URL: https://www.group-ib.ru/blog/g-analytics, (date of circulation: 08.05.2021).

6. Group-IB. Getting to know sniffers-3: Illum, 2021. URL: https://www.group-ib.ru/blog/illum last accessed 2019/06/06 (date of circulation: 08.05.2021).

7. Group-IB. Getting to know sniffers-4: CoffeMokko, 2021. URL: https://www.group-ib.ru/blog/coffemokko (date of circulation: 08.05.2021).

8. Lukatsky A., 2008. Detection of attacks. Textbook. BHV. ISBN: 5-94157-246-8. (In Russian).

9. Mark Dowd, John McDonald, and Justin Schuh, 2007. The art of software security assessment: identifying and preventing software vulnerabilities. Indianapolis, Ind: Addison-Wesley.

Гончаренко Юлия Юрьевна, доктор технических наук, доцент, профессор кафедры «Информационная безопасность» Института радиоэлектроники и информационной безопасности
Девицына Светлана Николаевна, кандидат технических наук, доцент, доцент кафедры «Информационная безопасность» Института радиоэлектроники и информационной безопасности
Шаповалов Павел Анатольевич, главный специалист одела информационной безопасности Аппарата Законодательного Собрания города Севастополя

Goncharenko Yuliya Yur'evna, Doctor of Technical Sciences, Associate Professor, Professor of the Department of Information Security of Institute of Radio-Electronics and Information Security of Sevastopol State University
Devitsyna Svetlana Nikolaevna, Candidate of Technical Sciences, Associate Professor, Associated Professor of the Department of Information Security of Institute of Radio-Electronics and Information Security of Sevastopol State University
Shapovalov Pavel Anatol'evich, Chief Specialist of the Information Security Department of the Legislative Assembly of the City of Sevastopol

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ПРИНЯТИЕ РЕШЕНИЙ
ARTIFICIAL INTELLIGENCE AND DECISION MAKING

УДК 004.94 **DOI: 10.18413/2518-1092-2021-6-2-0-4**

Жихарев А.Г.[1]
Маматов Р.А.[2]

СИСТЕМНО-ОБЪЕКТНЫЙ ПОДХОД В КОНТЕКСТЕ АГЕНТНОГО МОДЕЛИРОВАНИЯ

[1)] Белгородский государственный технологический университет им. В.Г. Шухова, ул. Костюкова, 46, Белгород, 308012, Россия
[2)] Управление Росгвардии по Белгородской области, ул. Преображенская, 60А, Белгород, 308009, Россия

e-mail: zhikharev@bsu.edu.ru

Аннотация
В статье рассматривается соотношение системно-объектного подхода и метода агентного имитационного моделирования. Проводится анализ основных структурных элементов и механизмов агентного подхода, на основании которого авторы предлагают новый способ формального описания агента средствами исчисления систем как функциональных объектов. Рассматривается структура агента в терминах вышеупомянутого исчисления и механизмы описания его поведения. Также рассмотрен модифицированный алгоритм системно-объектного имитационного моделирования с точки зрения построения агентной модели.
Ключевые слова: система, организационная система, метод системного анализа, агентный подход, имитационная модель.
Для цитирования: Жихарев А.Г., Маматов Р.А. Системно-объектный подход в контексте агентного моделирования // Научный результат. Информационные технологии. – Т.6, №2, 2021. – С. 25-31. DOI: 10.18413/2518-1092-2021-6-2-0-4

Zhikharev A.G.[1]
Mamatov R.A.[2]

THE RATIO OF THE SYSTEM-OBJECT AND AGENT-BASED APPROACHES TO THE CONSTRUCTION OF SIMULATION MODELS

[1)] Belgorod state technological university named after V.G. Shukhov, 46 Kostyukova street, Belgorod, 308012, Russia
[2)] Rosgvardia Directorate for the Belgorod Region, st. Preobrazhenskaya, 60A, Belgorod, 308009, Russia

e-mail: zhikharev@bsu.edu.ru

Abstract
The article discusses the relationship between the system-object approach and the method of agent-based simulation. The analysis of the main structural elements and mechanisms of the agent-based approach is carried out, on the basis of which the authors propose a new way of formal description of the agent by means of calculating systems as functional objects. The structure of an agent in terms of the above-mentioned calculus and mechanisms for describing its behavior are considered. A modified algorithm of system-object simulation modeling from the point of view of building an agent-based model is also considered.
Keywords: system, organizational system, system analysis method, agent-based approach, simulation model.
For citation: Zhikharev A.G., Mamatov R.A. The ratio of the system-object and agent-based approaches to the construction of simulation models // Research result. Information technologies. – T.6, №2, 2021. – P. 25-31. DOI: 10.18413/2518-1092-2021-6-2-0-4

ВВЕДЕНИЕ

Сегодня, практически во всем мире можно наблюдать беспрецедентное развитие производственно-технологических секторов национальных экономик. Это связано с бурным развитием научно-технических отраслей, с возрастающими запросами со стороны потребления, да и пандемия, вызванная новой коронавирусной инфекцией, также внесла весомый вклад в роботизацию многих производственное-технологических систем. В тоже время, несправедливо говорить о том, что исключительно пандемия 2020-2021 года привела к сплошной роботизации производственных и организационно-деловых процессов. Естественно, это не так, этот процесс начался достаточно давно, еще в период появления первых технических средств автоматизации, однако пандемия послужила своего рода, катализатором процессов вытеснения человека из производственно-технологических и организационно-деловых процессов. Комплексная автоматизация и роботизация любых процессов, в том числе и производственно-технологических требует высокого уровня информационно-аналитического обеспечения, причем не только на этапах реорганизации процессов, но и на этапах, связанных с их эксплуатацией. Для создания такого информационно-аналитического обеспечения используются различные методы и технологии. К отдельной категории информационно-аналитического обеспечения процессов можно отнести методы построения имитационных моделей, позволяющих прогнозировать поведение той или иной системы при прочих равных условиях. Одним из, относительно, современных подходов является агентный подход [1], который показал себя эффективным инструментарием во многих областях народного и мирового хозяйства. В литературе [2-4] отмечается, что появление агентного моделирования связано, в первую очередь, с нарастающей сложностью моделируемых систем. Агентный подход позволяет моделировать поведение системы в целом за счет описания ее участников – агентов. Ключевой задачей агентного моделирования при исследовании систем является прогнозирование поведения системы за счет моделирования децентрализованных агентов, участвующих в реализации функционального запроса надсистемы. Здесь можно выявить противоречие с точки зрения общей теории систем. Противоречие заключается в том, что система – есть целостный объект, в то время как агентное моделирование представляет моделируемую область в виде децентрализованных агентов. В этом, по мнению авторов, заключаются ключевые проблемы, связанные с использованием агентного подхода, касающиеся, в свою очередь, методов и подходов к построению агентных систем. В связи с вышесказанным, актуальной является задача создания методологии агентного моделирования, позволяющей ввести конкретные формальные правило описания агентов как составных частей моделируемой системы, при этом методология должна обеспечивать учет общесистемных принципов и закономерностей.

МЕТОДЫ

Рассмотрим метод системно-объектного имитационного моделирования в контексте агентного подхода.

Ранее авторским коллективом был разработан системно-объектный метод представления организационных знаний [5], который в дальнейшем получил развитие в качестве метода системно-объектного имитационного моделирования [6].

С формальной точки зрения системно-объектная модель любой предметной области может быть представлена комбинацией УФО-элементов, базовой иерархией связей [7] предметной области и множеством связей УФО-элементов [8], таким образом, системно-объектная модель, с формальной точки зрения, может быть определена тремя компонентами:

- (потоковые объекты системы) иерархия связей системы;
- узловые объекты системы;
- связи системы.

Потоковый объект системно-объектной модели – частный случай объекта в терминах теории объектов [9], представляющий собой именованный набор свойств реального объекта моделируемой предметной области. Например, потоковый объект, описывающий «жидкость», скорее всего, будет иметь такие поля (свойства), как: плотность, температура, цвет и т.п. Набор существенных свойств моделируемого объекта реального мира будет зависеть от целей моделирования.

Узловой объект системно-объектной модели – именованный объект в терминах теории объектов [9], представляющий собой формальное определение системы моделируемой предметной области. Узловой объект содержит две группы полей: поля, описывающие интерфейсные характеристики (узел); поля, описывающие объектные характеристики (объект). Кроме того, в качестве метода узлового объекта выступает функция системы (функция).

Связь системы – именованное ребро, связывающее два узловых объекта предметной области.

Далее рассмотрим формальное определение компонентов системно-объектной модели. Системно-объектная модель M имеет вид, как показано в выражении 1.

$$M =< L, S, C >, \quad (1)$$

где:

- L – множество потоковых объектов модели M (иерархия связей системно-объектной модели);
- S – множество узловых объектов модели M;
- C – множество связей системно-объектной модели M.

Множество потоковых объектов содержит, во-первых, базовую иерархию связей – статическая часть множества (неизменна для любой системно-объектной модели), во-вторых, потоковые объекты моделируемой предметной области (динамическая часть множества).

В общем виде множество потоковых объектов имеет следующее определение:

$$L = \{l | l = [r_1, \dots, r_n]\}, \quad (2)$$

где: l – представляет собой именованный потоковый объект, который, в свою очередь содержит множество пар r_n следующего формата – «идентификатор: значение».

Следует отметить, что множество L содержит, фактически, именованные множества, представляющие собой потоковые объекты, причем, размеры потоковых объектов (мощности множеств l) разные, в зависимости от моделируемого объекта предметной области и количества выделенных существенных его свойств.

Таким образом, множество L складывается из двух подмножеств: множество базовых потоковых объектов (статическая часть иерархии потоковых объектов) и множество потоковых объектов предметной области (динамическая часть иерархии потоковых объектов). Таким образом любое множество потоковых объектов имеет следующую структуру:

$$L = \{l_v, l_e, l_d, l_c, l_1, \dots, l_n\}, \quad (3)$$

где: l_v – потоковый объект родитель, представляющий собой класс вещественных объектов; l_e – потоковый объект родитель, представляющий собой класс энергетических объектов; l_d – потоковый объект родитель, представляющий собой класс информационных объектов; l_c – потоковый объект родитель, представляющий собой класс информационных управляющий объектов. n – количество потоковых объектов системно-объектной модели.

Далее в рассматриваемых формализмах первая часть множества потоковых объектов будет опущена, так как она неизменна для любой системно-объектной модели.

Множество узловых объектов S соответствует множеству систем как УФО-элементов. Система как УФО-элемент представляет собой триединую конструкцию, где должны быть учтены структурные характеристики системы, определяющиеся перекрестком входящих и исходящих связей системы. Таким образом, структурные характеристики системы, представляют собой интерфейс системы, за счет которого она может рассматриваться в контексте некоторой

надсистемы. Функциональные характеристики системы определяются запросом надсистемы – внешняя детерминанта системы, которая формально представлена как раз в виде интерфейса УФО-элемента (далее – узловой объект). Также функциональная характеристика определяется конкретными процедурами преобразования входных связей узлового объекта в выходные. Объектные характеристики узлового объекта представляют собой набор параметров той субстанции, которая реализует функцию узлового объекта. Таким образом, множество узловых объектов системно-объектной модели представлено в следующем формальном виде:

$$S = \{s_1, \dots, s_n\}, \quad (4)$$

где n – количество узловых объектов (систем); s_n – узловой объект.

Для учета в формальном аппарате исчисления систем как функциональных объектов контекста системы, введем во множество узловых объектов специальный элемент, представляющий собой контекст моделируемой системы, далее такой узловой объект будем обозначать с нижним индексом «*kontext*» – $s_{context}$. Такой узловой объект является «черным ящиком» с точки зрения моделирования, однако его необходимо учитывать, так как именно контекст системы определяет ее внешнюю детерминанту, которая в модели представлена входящими и исходящими потоковыми объектами первого уровня. таким образом, множество узловых объектов системно-объектной модели всегда будет включать контекстный узловой объект.

Каждый n-й элемент множества S представляет собой специальный узловой объект (соответствующий конкретной системе/УФО-элементу), который в соответствии с исчислением объектов Абади-Кардели состоит из полей и метода и имеет следующий вид:

$$s_n = [U, f, O], \quad (4)$$

где: U – представляет собой множество полей для описания интерфейсных потоковых объектов узлового объекта s_n (порты узлового объекта), соответствующих множеству функциональных связей моделируемой системы. Множество U, в свою очередь можно разделить на два подмножества: подмножество потоковых объектов, которые выступают в качестве входных портов и подмножество потоковых объектов, выступающих в качестве выходных портов узлового объекта, таким образом:

$$U = L? \cup L!, \quad (5)$$

где $L?$ – представляет собою множество входящих интерфейсных потоковых объектов, соответствующих входящим связям системы, $L!$ – представляет собою множество исходящих интерфейсных потоковых объектов, соответствующих выходящим связям системы.

Функция в выражении (4) f – представляет собой метод узлового объекта s_n, описывающий процесс преобразования ресурсов, поступающих посредством входящих интерфейсных потоковых объектов (входящих связей системы) $L?$ в выходящие – $L!$. Далее метод узлового объекта будем представлять в виде функции от множества входящих потоковых объектов:

$$f(L?)L! \quad (6)$$

O – представляет собою множество полей для описания объектных характеристик узлового объекта (системы) s_n, элементы которого имеют формат «идентификатор: значение».

Т.е. множество O можно понимать, как набор полей узлового объекта. Причем, поля узлового объекта можно условно поделить на три группы: первая группа – характеризует входные порты узлового объекта (например, пропускная способность входящего порта), вторая группа характеризует выходные порты узлового объекта (например, пропускная способность исходящего порта), третья группа – характеризует объект, участвующий в реализации метода узлового объекта. Таким образом, множество полей для описания объектных характеристик системы состоит из трех подмножеств:

$$O = O? \cup O! \cup Of, \quad (7)$$

где: множество полей *O?* содержит параметры, характеризующие входы системы, *O!* – множество параметров, характеризующих выходы системы и *Of* – множество параметров, характеризующих объект, участвующий в реализации функции системы. С формальной точки зрения, перечисленные множества объектных характеристик содержат постоянные величины, задействованные в описании функции узла.

Принимая во внимание описанные выше компоненты узлового объекта, выражение (4) можно записать в следующем уточненном виде:

$$s_n = [L?, L!; f(L?)L!; O?, O!, Of] \quad (8)$$

Для учета внутренних связей системы в определение системно-объектной модели вводится множество *C*, где содержатся все связи между узловыми объектами. Каждая связь, характеризуется, как минимум тремя компонентами: источник связи – экземпляр узлового объекта, для которого связь выступает в качестве исходящей; получатель связи – экземпляр узлового объекта, для которого связь выступает в качестве входящей; тип связи – экземпляр потокового объекта. Таким образом, множество связей системно-объектной модели определим следующим образом:

$$C = \{(s_{out}, s_{in}, l) | s_{out} \in S, s_{in} \in S, l \in L\}, \quad (9)$$

где: s_{out} – экземпляр узлового объекта – источника связи; s_{in} – экземпляр узлового объекта – получателя связи; *l* – тип связи между указанными выше узловыми объектами.

РЕЗУЛЬТАТЫ

Ключевым понятие в рамках агентного подхода является агент – представляющий собой децентрализованную модель объекта реального мира. Агент, в свою очередь, имеет следующие характеристики [10]:

- идентификатор;
- характеристики, определяющие поведение агента;
- правила поведения агента;
- связь с другими агентами системы;
- механизмы принятия решения по взаимодействию с другими агентами;
- внешняя среда агента;
- цель существования и функционирования.

Из перечисленных характеристик можно увидеть взаимосвязь терминов «агент» и «система». В терминах системно-объектного подхода, любого агента представим в виде узлового объекта, как показано в выражении 8. Тогда системно-объектная агентная модель может быть представлена как обычная системно-объектная модель (выражение 1). Структура такой модели может быть следующей:

- *L* – множество потоковых объектов модели *M* (виды возможных связей между агентами);
- *S* – множество агентов модели *M*;
- *C* – множество связей между агентами модели *M*.

Таким образом, каждый агент системно-объектной модели может быть однозначно идентифицирован через соответствующее множество. В качестве характеристик, определяющих поведение агента выступают объектные характеристики узлового объекта, определяющие его функциональность. Правила поведения агента могут быть описаны в рамках метода узлового объекта. Механизмы принятия решений по взаимодействию с другими агентами могут быть представлены в виде интерфейса узлового объекта и его метода. Связи с другими агентами соответствуют обычным связям между узловыми объектами. Внешняя среда агента может быть представлена специальным контекстным узловым объектом, определяющим общие правила поведения агентов. Цель существования и функционирования агента можно представить как в

некотором числовом виде (например, конкретные значения определенных его характеристик), либо как внешняя детерминанта системы.

ЗАКЛЮЧЕНИЕ

Проведенный выше анализ показывает, что средства системно-объектного подхода к построению имитационных моделей содержат все необходимое для реализации агентных моделей. Таким образом, становится возможным применение метода и алгоритмов системно-объектного моделирования в описании агентов. Это позволит сформулировать конкретный метод описания агентов предметной области, учитывая при этом общесистемные принципы и закономерности.

БЛАГОДАРНОСТИ

Исследования выполнены при финансовой поддержке проектов Российского фонда фундаментальных исследований № 19-07-00290, 19-07-00111.

Список литературы

1. Bonabeau Eric Agent-based modeling: methods and techniques for simulating human systems. Proc. National Academy of Sciences. 2002. № 99(3). P 7280-7287.
2. Gilbert N., Pyka A., Ahrweiler P. Innovation Networks – A Simulation Approach // Journal of Artificial Societies and Social Simulation. 2001. Vol. 4. № 3.
3. Macal C., North M. Tutorial on Agent-Based Modelling and Simulation // Journal of Simulation. 2010. Vol. 4. P. 151–162.
4. Cederman L. Emergent Actors in World Politics: How States and Nations Develop and Dissolve. Princeton University Press. 1997.
5. Matorin S., Zhikharev A., Zimovets O. Object Calculus in the System–Object Method of Knowledge Representation. Scientific and Technical Information Processing. 2018. 45(5). P 307–316.
6. Matorin S., Zhikharev A. Calculation of the function objects as the systems formal theory basis. Advances in Intelligent Systems and Computing. 2018. 679. P. 182–191.
7. Zhikharev A., Matorin S., Tinyakov O., Shcherbinina N., Migal L. Formalization of system-object method of knowledge representation by calculation of systems as functional objects. Journal of Physics: Conference Series. 2021. 1801(1). 012025.
8. Zhikharev A., Matorin S., Egorov I. Formal principles of system-object simulation modeling of technological and production processes. Journal of Advanced Research in Dynamical and Control Systems. 2018. 10(10). P. 1806–1812.
9. Abadi M., Cardelli L. A Theory of Objects. New York: Springer-Verlag. 1996.

References

1. Bonabeau Eric Agent-based modeling: methods and techniques for simulating human systems. Proc. National Academy of Sciences. 2002. № 99(3). P 7280-7287.
2. Gilbert N., Pyka A., Ahrweiler P. Innovation Networks – A Simulation Approach // Journal of Artificial Societies and Social Simulation. 2001. Vol. 4. № 3.
3. Macal C., North M. Tutorial on Agent-Based Modelling and Simulation // Journal of Simulation. 2010. Vol. 4. P. 151–162.
4. Cederman L. Emergent Actors in World Politics: How States and Nations Develop and Dissolve. Princeton University Press. 1997.
5. Matorin S., Zhikharev A., Zimovets O. Object Calculus in the System–Object Method of Knowledge Representation. Scientific and Technical Information Processing. 2018. 45(5). P 307–316.
6. Matorin S., Zhikharev A. Calculation of the function objects as the systems formal theory basis. Advances in Intelligent Systems and Computing. 2018. 679. P. 182–191.
7. Zhikharev A., Matorin S., Tinyakov O., Shcherbinina N., Migal L. Formalization of system-object method of knowledge representation by calculation of systems as functional objects. Journal of Physics: Conference Series. 2021. 1801(1). 012025.

8. Zhikharev A., Matorin S., Egorov I. Formal principles of system-object simulation modeling of technological and production processes. Journal of Advanced Research in Dynamical and Control Systems. 2018. 10(10). P. 1806–1812.
9. Abadi M., Cardelli L. A Theory of Objects. New York: Springer-Verlag. 1996.

Жихарев Александр Геннадиевич, кандидат технических наук, доцент, доцент кафедры программного обеспечения вычислительной техники и автоматизированных систем
Маматов Роман Александрович, старший инспектор отделения организации службы ОМОН, Управление Росгвардии по Белгородской области

Zhikharev Alexander Gennadievich, Candidate of Technical Sciences, Associate Professor, Associate Professor of the Department of Computer Engineering and Automated Systems Software
Mamatov Roman Aleksandrovich, Senior Inspector of the Department of Organization of the OMON Service, Rosgvardia Directorate for the Belgorod Region

УДК 004.04 **DOI: 10.18413/2518-1092-2021-6-2-0-5**

Наумов Р.К.[1]
Самылкин М.С.[1]
Копейкин М.В.[2]

СПОСОБЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ СРЕДСТВАМИ СУБД

[1)] Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики», Кронверкский пр., д. 49, г. Санкт-Петербург, 197101, Россия
[2)] Федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский горный университет», 21-я лин. В.О., д. 2, Санкт-Петербург, 199106

e-mail: ruslan.naumow.dake@gmail.com, maksamylkin@gmail.com

Аннотация
Рост объема неструктурированных данных, генерируемых различными приложениями и сервисами, и принятие компаниями факта ценности таких данных привели к востребованности систем, способных анализировать большие объемы данных без участия человека. Удовлетворить данную потребность могут системы интеллектуального анализа данных, однако повышение эффективности таких систем является актуальной задачей. Несмотря на растущую популярность NoSQL решений, основными системами управления базами данных все еще являются реляционные СУБД. В статье особое внимание уделено тому, что современные РСУБД могут использоваться не только в качестве надежных хранилищ данных. В настоящее время первоочередной задачей развития РСУБД является интеграция в них интеллектуального анализа данных. Благодаря тому, что данные остаются в хранилище, система не тратит ресурсы на выгрузку анализируемого набора данных из базы данных и загрузку результатов анализа обратно. Данный подход повысит как скорость разработки, за счет использования сервисов, вшитых в СУБД, так и производительность всей системы. В статье рассматриваются основные задачи интеллектуального анализа данных и существующие алгоритмы их решения. Описываются основные методики внедрения интеллектуального анализа данных в СУБД. Особое внимание уделено подходу, в котором система анализа данных рассматривается как внутренний сервис СУБД. В работе представлены известные системы и библиотеки анализа данных, разработанные для РСУБД, а также варианты расширений языка запросов SQL.
Ключевые слова: интеллектуальный анализ данных, реляционные СУБД, кластеризация, поиск шаблонов, классификация.
Для цитирования: Наумов Р.К., Самылкин М.С., Копейкин М.В. Способы интеллектуального анализа данных средствами СУБД // Научный результат. Информационные технологии. – Т.6, №2, 2021. – С. 32-40. DOI: 10.18413/2518-1092-2021-6-2-0-5

Naumov R.K.[1]
Samylkin M.S.[1]
Kopeikin M.V.[2]

DATA MINING METHODS USING DBMS TOOLS

[1)] Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, 49 Kronverkskiy prospekt, St. Petersburg, 197101, Russia
[2)] Saint Petersburg Mining University, 2 21-st line V.O., St. Petersburg, 199106, Russia

e-mail: ruslan.naumow.dake@gmail.com, maksamylkin@gmail.com

Abstract
The growth in the volume of unstructured data generated by various applications and services, and the acceptance by companies of the fact that such data is valuable, have led to the demand for systems that can analyze large amounts of data without human intervention. Data mining systems

can satisfy this need, but improving the efficiency of such systems is an urgent task. Despite the growing popularity of NoSQL solutions, the main database management systems are still relational databases. In the article, special attention is paid to the fact that modern RDBMS can be used not only as reliable data stores. Currently, the primary task of RDBMS development is to integrate data mining into them. Due to the fact that the data remains in the storage, the system does not waste resources on unloading the analyzed data set from the database and loading the analysis results back. This approach will increase both the speed of development, due to the use of services embedded in the DBMS, and the performance of the entire system. The article discusses the main problems of data mining and the existing algorithms for solving them. The main methods of implementing data mining in a DBMS are described. Special attention is paid to the approach in which the data analysis system is considered as an internal DBMS service. The paper presents well-known data analysis systems and libraries developed for RDBMS, as well as variants of SQL query language extensions.
Keywords: data mining, relational DBMS, clustering, pattern mining, classification.
For citation: Naumov R.K., Samylkin M.S., Kopeikin M.V. Data mining methods using DBMS tools // Research result. Information technologies. – T.6, №2, 2021. – P. 32-40. DOI: 10.18413/2518-1092-2021-6-2-0-5

ВВЕДЕНИЕ

На сегодняшний день спрос на эффективные и мощные инструменты для работы с данными вызван многими факторами. К ним можно отнести потребность в беспрерывных и высоконагруженных системах для анализа данных. Такая потребность объясняется ростом числа приложений и сервисов, постоянно генерирующих большие объемы неструктурированных данных. Скорость прироста таких данных для одного ресурса может превышать 1 Тб в день. Большие данные могут включать в себя документы, электронную почту, информацию социальных сетей, аудио, видеофайлы, и т.д. [1] Важными факторами в анализе данных являются их объем и скорость накопления, однако, критичным является наличие эффективных методов для их обработки.

Под интеллектуальным анализом данных понимают совокупность алгоритмов, методов и программного обеспечения для обнаружения в данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия стратегически важных решений в различных сферах человеческой деятельности. [2]

Для извлечения данных, т.е. полезных знаний, из Больших данных используются ETL-системы, которые включают в себя процессы трансформации и очистки данных. Такие процессы приводят к образованию массивных хранилищ данных. Один из разработчиков СУБД Ingres и PostgreSQL пишет [3], что для решения данной проблемы необходимо пользоваться технологиями, предоставляемыми СУБД.

По данным авторитетного портала DB-engines.com [4], собирающего данные о реляционных и NoSQL СУБД, на апрель 2021 года самыми популярными системами являются Oracle, MySQL, Microsoft SQL Server, PostgreSQL и MongoDB (рис. 1).

Из этой статистики можно сделать вывод, что, несмотря на растущую популярность NoSQL решений, разработчики все еще отдают большее предпочтение реляционным базам данных.

На конгрессе [5] специалистами в сфере обработки и анализа данных был поднят вопрос генерирования большого объема данных в процессе цифровой трансформации общества. Основным решением данной проблемы стал полный контроль всех процессов от получения данных до извлечения из них полезных знаний.

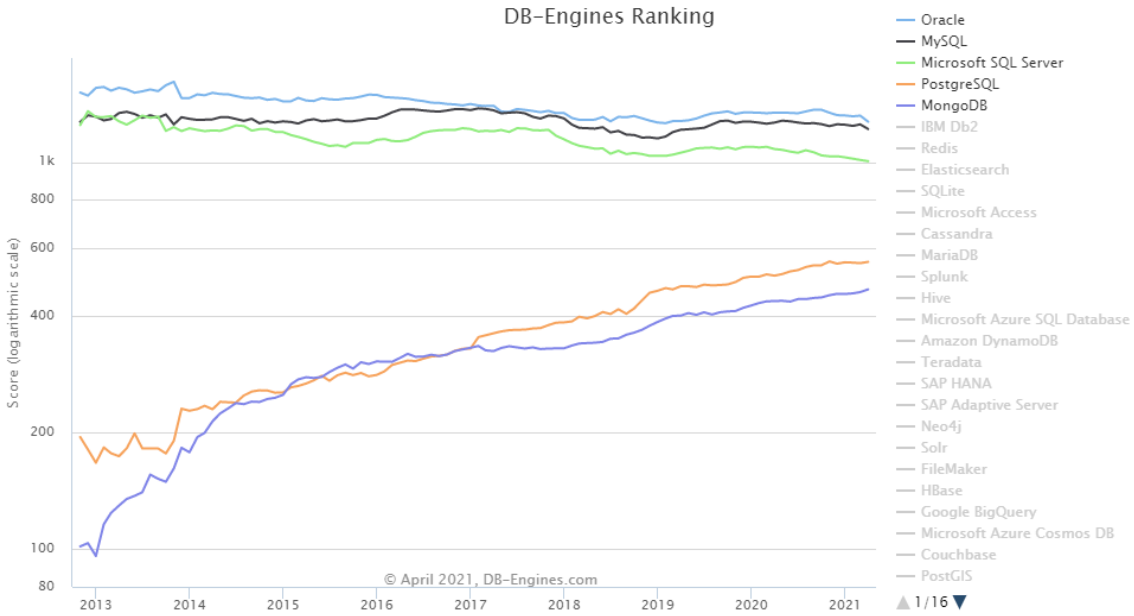


Рис. 1. Рейтинг СУБД портала DB-engines.com
Fig. 1. DB-Engines Ranking

На сегодняшний день внедрение методов интеллектуального анализа данных во внутренние процессы систем управления базами данных становится основной траекторией развития РСУБД. По расчетам автора работы [6] размещение верно подобранных алгоритмов обработки данных в непосредственной близости к анализируемым данным способствует увеличению производительности системы. Такая интеграция позволяет достичь минимизации расходов по выгрузке анализируемых данных из базы данных и загрузке результатов анализа обратно. Помимо этого, при обработке данных внутри хранилища, программист может воспользоваться внутренними сервисами СУБД, например, механизмом репликации, целостности и отказоустойчивости. Ускорить обработку данных можно с помощью индексации полей, алгоритмов оптимизации и других средств, которые заложены в архитектуру РСУБД.

Целью статьи является исследование способов интеграции интеллектуального анализа данных в реляционные СУБД. Основными задачами исследования являются:

1. Обзор существующих алгоритмов решения наиболее часто встречающихся задач интеллектуального анализа данных, таких как, кластеризация и поиск шаблонов;
2. Поиск подходов к внедрению интеллектуального анализа данных в РСУБД;
3. Определение и анализ существующих методов интеллектуального анализа данных в РСУБД.

ОСНОВНАЯ ЧАСТЬ

Задачи кластеризации и поиска шаблонов

Под задачей кластеризации подразумевается разбиение множества объектов, имеющих похожую структуру, на неопределенные заранее группы (кластеры) в зависимости от схожести их свойств. Задача кластеризации применяется повсеместно: сегментирование изображений, анализ социальных исследований и т.д. В алгоритмах кластеризации используются следующие статистические данные: количество точек в ячейке, а также результаты агрегатных функций (min, max, avg и т.д.).

Основные алгоритмы кластеризации перечислены ниже:

- Алгоритм четкой кластеризации;
- Алгоритм нечеткой кластеризации;
- Алгоритм раздельной (partitioning) кластеризации;

- Алгоритм k-средних (k-means);
- Алгоритмы k-medoids и PAM;
- Агломеративный иерархический алгоритм;
- Дивизимный иерархический алгоритм;
- Плотностная (density-based) кластеризация;
- Решетчатая (grid-based) кластеризация;

Примером агломеративного подхода кластеризации является алгоритм AGNES, в то время как для дивизимного подхода реализован алгоритм DIANA. Одним из примеров плотностной кластеризации является алгоритм DBSCAN, а для решетчатой кластеризации можно выделить алгоритм STING.

Задача поиска шаблонов заключается в нахождении явных закономерностей на имеющемся наборе объектов. Поиск шаблонов часто применяется в разнообразных областях человеческой деятельности, начиная с медицины, заканчивая анализом звуковых дорожек.

Традиционно для решения задачи поиска часто встречающихся зависимостей между данными применяют алгоритм Apriori. Суть действия данного алгоритма заключается в последовательном преобразовании наборов кандидатов в независимые множества с последующим отбором этих кандидатов по удовлетворяющему числу поддержки. Однако, у данного алгоритма есть недостаток: при большом значении таких наборов и низком пороговом значении поддержки появляются значительные расходы во времени. Исходя из этого были разработаны улучшенные алгоритмы Apriori: AprioriTid, DHP, DIC, Eclat, Partition.

Еще одним способом поиска шаблонов является алгоритм FP-Growth, который основывается на построении FP-дерева. В такой структуре данных наборы и значения их поддержки хранятся более компактно. Данный алгоритм также был доработан – появились алгоритмы OpportuneProject и AFOPT.

Подходы к интеграции интеллектуального анализа данных в РСУБД

Становление интеллектуального анализа данных, как отдельной дисциплины, дало начало научным изысканиям в области внедрения интеллектуального анализа данных в РСУБД, но на сегодняшний день данное направление не перестает быть актуальным. В научных трудах [7, 8] выделяют понятие «связывание». Данный термин подразумевает под собой интеграцию между интеллектуальным анализом данных и системой управления базами данных Понятия «связывание» принято делить на слабое, среднее и сильное связывание.

Слабое связывание представляет собой распределенную систему, в которой подсистема интеллектуального анализа данных существует автономно и не зависит от СУБД. Анализирующая система пользуется средствами СУБД для выгрузки данных из базы данных и обратной загрузки результатов анализа в базу данных. Данный подход используют такие open-source системы анализа данных как RapidMiner, Pentaho и т.д.

Принцип среднего связывания основывается на том, что система интеллектуального анализа данных имеет возможность выполнять примитивные операции, которые часто используются на стадии преданализа данных, с помощью средств СУБД. Такими операциями можно считать операции индексации, соединения отношений, а также исполнение агрегатных функций. Однако при данном подходе все еще подразумевается отделение анализирующей системы от системы управления базами данных. Здесь СУБД используется в качестве хранилища заранее вычисленных промежуточных результатов интеллектуального анализа, используемых наиболее часто.

Идея того, что система интеллектуального анализа данных закладывается в архитектуру СУБД и рассматривается как ее внутренний сервис, основывает принцип сильного связывания. Он заключается в использовании средств, обеспечивающих выполнение запросов на анализ данных на самом сервере базы данных. Оптимизация и выполнение функций анализа данных основывается на встроенных в СУБД сервисах и методах обработки запросов. Использование методики

сильного связывания способствует повышению скорости разработки и производительности эксплуатации всей информационной системы, однако в то же время является тяжело реализуемым.

Принцип сильного связывания системы интеллектуального анализа данных и СУБД может быть реализован с помощью двух подходов: интеллектуальный анализа данных, как сервис СУБД, и интеллектуальный анализа данных, как функция, написанная на языке SQL или его расширениях. В свою очередь каждый подход делят на различные способы реализации, которые схематично представлены на рисунке 2.

Внедренная в СУБД подсистема представляет собой механизм, поддерживающий сторонний язык анализа данных или расширяющий язык SQL добавлением необходимых для анализа функций, операторов и т.д.

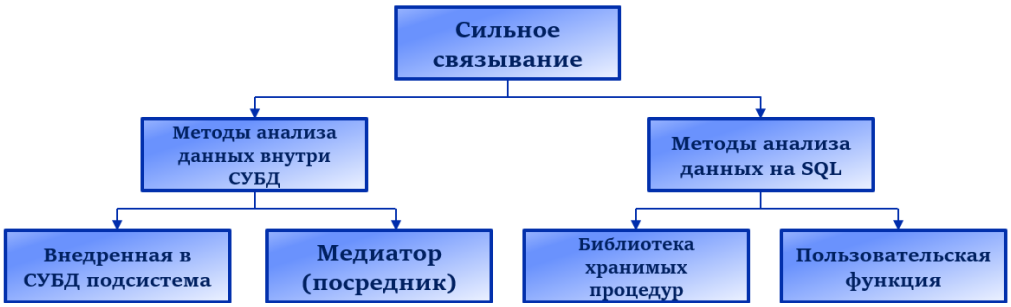


Рис. 2. Подходы к реализации «сильного связывания»
Fig. 2. Approaches to the implementation of "tight coupling" approach

Медиатор реализуется в виде посредника между архитектором баз данных и системой управления базами данных. Функцией медиатора является предоставление некоторого интерфейса или языка запросов для программиста, то есть обеспечивает преобразование запросов интеллектуального анализа данных в запросы на SQL.

Библиотека хранимых процедур разрабатывается программистом и представляет собой набор хранимых процедур, хранящихся в виде подпрограмм и компилирующихся однократно. После компиляция процедура постоянно хранится на сервере базы данных. При подключении библиотеки хранимых процедур к приложению базы данных процесс интеллектуального анализа выполняется внутри ядра СУБД.

Разработка пользовательских функций предполагает написание подпрограммы-функции, которая затем хранится на сервере СУБД. Пользовательская функция вызывается на исполнение с помощью вставки выражения в оператор SQL. Результат рассчитывается в процессе выполнения конкретного запроса и может быть представлен в виде скалярного, либо табличного типа. Пользовательская функция обычно реализовывается на SQL или его расширении, однако в большинстве современных СУБД возможна реализация на языках высокого уровня.

Методы интеллектуального анализа данных в РСУБД

Перейдем к обзору известных систем и библиотек для интеллектуального анализа данных, которые были разработаны для реляционных СУБД, а также вариантов расширений языка запросов SQL.

Занимающая одно из лидирующих мест среди инструментов для работы с базами данных СУБД Microsoft SQL Server (рис. 1) поддерживает стандарт OLE DB for Data Mining и специализированный язык запросов Data Mining Extensions (DMX). Язык DMX базируется на языке SQL, однако поддерживает только часть возможностей стандарта SQL:2009. Особенностью языка DMX является представление его операндов. Ими являются не традиционные реляционные таблицы (отношения), а сочетания данных, параметров, алгоритмов, применяющихся для анализа, и фильтров, задающих ход обработки данных. На рисунке 3 представлен пример запроса на языке DMX, а именно запрос кластеризации данных по заданным значениям.

```
SELECT PredHist(Cluster())
FROM [TM Clustreing]
    NATURAL PREDICTION JOIN
     (SELECT 1998 AS [Birth_Year],
   'Kingisepp' AS [City],
   1 AS [Childrens]) AS t
```

Рис. 3. Пример кластеризации данных на языке DMX
Fig. 3. Example of data clustering using the DMX language

Похожая реализация присутствует в СУБД Oracle, разработанной одноименной компанией, в виде модуля Oracle Data Mining [10]. При построении запроса к базе данных на интеллектуальный анализ данных используется PL/SQL API, реализованный пакетом DMBS_DATA_MINING. Пример запроса классификации данных на языке Oracle Data Mining представлен на рисунке 4. В первую очередь происходит создание модели "covid_risk", после чего происходит выборка по полученному в модели предсказанию (PREDICTION).

```
DBMS_DATA_MINING.CREATE_MODEL (
model_name => 'covid_risk_model',
    function => DBMS_DATA_MINING.classification,
data_table_name => 'covid_country_data',
case_id_column_name => 'country_id',
target_column_name => 'covid_risk',
    settings_table_name => 'credit_risk_model_settings');

SELECT country_name
FROM covid_country_data
WHERE PREDICTION (covid_risk_model USING *) = 'LOW'
```

Рис. 4. Пример классификации данных на языке Oracle Data Mining
Fig. 4. Example of data classification using the Oracle Data Mining language

Самым заметным средством интеллектуального анализа данных в реляционных СУБД PostgresSQL и разработанной на ее основе Greenplum является open-source библиотека MADlib. Широкий набор механизмов, предоставляемый данной подсистемой, дает возможность проводить кластеризацию и классификацию данных, осуществлять регрессионный анализ и пользоваться другими методами для анализа свойств данных. Особенностью библиотеки является адаптированность этих алгоритмов к реляционной составляющей системы без участия сторонних аналитических приложений. Обращение к базе данных происходит за счет исполнения заранее написанных на языке программирования Python пользовательских функций, которые выступают в роли коннектора и формируют корректную структуру таблиц. Пример вызова функции библиотеки MADlib, классифицирующую данные, представлен на рисунке 5.

```
SELECT madlib.create_nb_probs_view (
'example_feature_probs',        -- таблица выходных вероятностей
'example_priors',               -- таблица выходных классов
'class_example_topredict',      -- таблица с данными для классификации
'id',                           -- имя ключевого столбца
'attributes',                   -- имя столбца атрибутов
3,                              -- количество атрибутов
'example_classified'            -- название нового представления (view)
);
```

Рис. 5. Пример классификации данных с помощью библиотеки MADlib
Fig. 5. Example of data classification using the MADlib library

Одним из вариантов расширения языка SQL для кластеризации данных является использование оператора CLUSTER BY, который был предложен в статье [11]. Работа оператора заключается в группировке результирующих строк с помощью встроенного алгоритма кластеризации. Данный механизм группировки отличается от традиционного, предусмотренного стандартом SQL, оператора GROUP BY, который выполняет группировку по точному совпадению значений в строках результирующей выборки. Одной из систем, в которой используется оператор CLUSTER BY, является PosgGIS – расширение СУБД PostgreSQL для работы с геоданными. Пример использования оператора приведен на рисунке 6. Существуют аналоги для различных СУБД, реализующие алгоритм кластеризации с помощью расширения языка SQL. К ним можно отнести SIMILAR GROUP BY для PostgreSQL, DISTRIBUTE BY для SPARK и др.

```
SELECT country_name
FROM counties
    WHERE mainland = 'Europe'
    CLUSTER BY population
```

Рис. 6. Пример кластеризации данных с помощью оператора CLUSTER BY
Fig. 6. Example of clustering data using the CLUSTER BY statement

Помимо приведенных ранее способов интеллектуального анализа данных существует исследовательское направление реализации алгоритмов анализа данных в РСУБД. Данные реализации позволят без дополнительных манипуляций над кодом переносить алгоритмы между различными СУБД. В таблице 1 приведены наиболее заметные в научных трудах SQL-реализации задачи поиска шаблонов.

Таблица 1

SQL-реализации задач поиска шаблонов

Table 1

SQL implementations of pattern mining tasks

Используемый алгоритм	**SQL-реализация**
Apriori	K-Way-Join
	Three-Way-Join
	Subquery
	Two-Group-Bys
	Set-oriented Apriori
	RDB-MINER
Universal quantification	Quiver
FP-Growth	Propad
	FP-TDG

Помимо представленных ранее задач, с помощью SQL также предлагается решать задачи классификации [12]. Классификация также, как и кластеризация, является задачей разделения конечного числа объектов на группы (классы), однако в отличии от задачи кластеризации имеет заранее определенную структуру и семантику классов. Одним из основных подходов к классификации является построение дерева решений. Несмотря на то, что данные, построенные на графах, имеют не реляционную природу, использование интеллектуального анализа таких данных в РСУБД является актуальным направлением. Так, например, в работе [13] был предложен алгоритм анализа структур деревьев с помощью SQL, а в статье [14] авторы описали алгоритм поиска полного подграфа, который основывается на применении средств РСУБД.

ЗАКЛЮЧЕНИЕ

На сегодняшний день наблюдается тенденция ускоренного роста объема данных. Неструктурированные данные генерируются различными сервисами и приложениями, которые являются частью жизни большинства людей. Все больше компаний уделяют внимание не только привлечению клиентов, но и сбору данных о них, а для их обработки без участия человека используют средства интеллектуального анализа данных. Реляционные СУБД по оценкам специалистов и мнению сообщества занимают лидирующую позицию среди инструментов управления данными. Перспективной траекторией развития РСУБД является внедрение в них средств интеллектуального анализа данных. Интеграция повысит скорость обработки данных, за счет минимизации ресурсных расходов по выгрузке анализируемых выборок из базы данных и загрузке результатов анализа обратно. Помимо этого, программист сможет воспользоваться внутренними сервисами СУБД, заложенными в ее архитектуру.

В статье были рассмотрены существующие алгоритмы решения задачи кластеризации и задачи поиска шаблонов. Определены основные подходы к интеграции интеллектуального анализа данных: подход слабого, среднего и сильного связывания. Наибольшее внимание уделено последнему подходу, более удобному с точки зрения прикладного программиста, но требующему для интеграции больших усилий. Приведены примеры такой интеграции, реализованной на основе библиотек хранимых процедур и пользовательских функций. Также в работе представлены некоторые примеры расширения языка SQL с помощью добавления специальных операторов, а также описаны известные SQL-реализации алгоритмов анализа данных.

Список литературы

1. Соловьев А.И. Хранение и Обработка Больших Данных // Тенденции Развития Науки и Образования. 2018. № 6 С. 47–51.
2. Цымблер М.Л. Обзор методов интеграции интеллектуального анализа данных в СУБД // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2019. № 2. С. 32–62.
3. Stonebraker M., Madden S., Dubey P. Intel “Big Data” Science and Technology Center Vision and Execution Plan. SIGMOD Record. 2013. № 1. С. 44–49.
4. DB-Engines Ranking, 2021 г. URL: https://db-engines.com/en/ranking (дата обращения: 03.04.2021).
5. Abadi D., Agrawal R., Ailamaki A. The Beckman Report on Database Research // Commun. ACM. 2016. № 2. С. 92–99.
6. Ordonez C. Statistical Model Computation with UDFs // IEEE Trans. Knowl. Data Eng. 2010. № 12. С. 1752–1765.
7. Han J., Kamber M. Data Mining: Concepts and Techniques // Morgan Kaufmann, 2006. С. 743.
8. Sarawagi S., Thomas S., Agrawal R. Integrating Mining with Relational Database Systems: Alternatives and Implications // ACM SIGMOD International Conference on Management of Data. 1998. С. 343–354.
9. Salal Y.K., Abdullaev S.M. Using of Data Mining Techniques to Predict of Student’s Performance in Industrial Institute of Al-Diwaniyah, Iraq // Bulletin of the South Ural State University Series: Computer Technologies, Automatic Control, Radio Electronics. 2019. № 19. С. 121-130.
10. Соболева А.Д., Сабинин О.Ю. Разработка метода композиции алгоритмов машинного обучения для решения задачи прогнозирования на примере технологии Oracle Data Mining // Theoretical & Applied Science. 2018. № 3. С. 147-154.
11. Sun P., Huang Y., Zhang C. Cluster-By: An Efficient Clustering Operator in Emergency Management Database Systems // Web-Age Information Management – WAIM. 2013. С. 152–164.
12. Li J., Cheng T., Zhao Z. High Efficient Classification Mining Method for Regional Large Data Features under Complex Attribute Environment // Academic Journal of Manufacturing Engineering. 2020. № 18(3). С. 113-119.
13. Padmanabhan S., Chakravarthy S. HDB-Subdue: A Scalable Approach to Graph Mining // Data Warehousing and Knowledge Discovery, 11th International Conference. 2009. С. 325–338.
14. Srihari S., Chandrashekar S., Parthasarathy S. A Framework for SQL Based Mining of Large Graphs on Relational Databases // Advances in Knowledge Discovery and Data Mining, 14th Pacific-Asia Conference. 2010. С. 160–167.

15. Integration of Data Mining Techniques to PostgreSQL Database Manager System / Viloria A., Acuña, G.C., Alcázar F., D.J., Hernández-Palma, H., Fuentes, J.P., Rambal, E.P. // Procedia Computer Science. 2019. С. 575–580.

16. Аверьянова Е.В., Малышева Е.Ю. Алгоритмы интеллектуального анализа данных в Microsoft SQL Server // Вестник Поволжского государственного университета сервиса. Серия: Экономика. 2017. № 1. С. 115-120.

References

1. Soloviev A.I. Storage and Processing of Big Data // Trends in the Development of Science and Education. 2018. No. 6 pp. 47–51.

2. Zymbler M.L. Overview of Methods for Integrating Data Mining into DBMS // Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering. 2019. No. 2. pp. 32–62.

3. Stonebraker M., Madden S., Dubey P. Intel “Big Data” Science and Technology Center Vision and Execution Plan. SIGMOD Record. 2013. No. 1. pp. 44–49.

4. DB-Engines Ranking, 2021. URL: https://db-engines.com/en/ranking (date of the request: 03.04.2021).

5. Abadi D., Agrawal R., Ailamaki A. The Beckman Report on Database Research // Commun. ACM. 2016. No. 2. pp. 92–99.

6. Ordonez C. Statistical Model Computation with UDFs // IEEE Trans. Knowl. Data Eng. 2010. No. 12. pp. 1752–1765.

7. Han J., Kamber M. Data Mining: Concepts and Techniques // Morgan Kaufmann, 2006. pp. 743.

8. Sarawagi S., Thomas S., Agrawal R. Integrating Mining with Relational Database Systems: Alternatives and Implications // ACM SIGMOD International Conference on Management of Data. 1998. pp. 343–354.

9. Salal Y.K., Abdullaev S.M. Using of Data Mining Techniques to Predict of Student’s Performance in Industrial Institute of Al-Diwaniyah, Iraq // Bulletin of the South Ural State University Series: Computer Technologies, Automatic Control, Radio Electronics. 2019. No. 19. pp. 121-130.

10. Soboleva A.D., Sabinin O.Yu. Development of a Method for Composition of Machine Learning Algorithms for Solving a Forecasting Problem using the Example of Oracle Data Mining Technology // Theoretical & Applied Science. 2018. No. 3. pp. 147-154.

11. Sun P., Huang Y., Zhang C. Cluster-By: An Efficient Clustering Operator in Emergency Management Database Systems // Web-Age Information Management – WAIM. 2013. pp. 152–164.

12. Li J., Cheng T., Zhao Z. High Efficient Classification Mining Method for Regional Large Data Features under Complex Attribute Environment // Academic Journal of Manufacturing Engineering. 2020. No. 18(3). pp. 113-119.

13. Padmanabhan S., Chakravarthy S. HDB-Subdue: A Scalable Approach to Graph Mining // Data Warehousing and Knowledge Discovery, 11th International Conference. 2009. pp. 325–338.

14. Srihari S., Chandrashekar S., Parthasarathy S. A Framework for SQL Based Mining of Large Graphs on Relational Databases // Advances in Knowledge Discovery and Data Mining, 14th Pacific-Asia Conference. 2010. pp. 160–167.

15. Integration of Data Mining Techniques to PostgreSQL Database Manager System / Viloria A., Acuña, G.C., Alcázar F., D.J., Hernández-Palma, H., Fuentes, J.P., Rambal, E.P. // Procedia Computer Science. 2019. pp. 575–580.

16. Averyanova E.V., Malysheva E.Y. Data Mining Algorithms in Microsoft SQL Server // Bulletin of the Volga State University of Service. Series: Economics. 2017. No. 1. pp. 115-120.

Наумов Руслан Кириллович, инженер мегафакультета трансляционных информационных технологий, студент факультета инфокоммуникационных технологий
Самылкин Максим Сергеевич, студент факультета безопасности информационных технологий
Копейкин Михаил Васильевич, кандидат технических наук, доцент факультета фундаментальных и гуманитарных дисциплин, кафедра информационных систем и вычислительной техники

Naumov Ruslan Kirillovich, engineer, student, Faculty of Translational Information Technologies
Samylkin Maxim Sergeevich, student of the Faculty of Information Technology Security
Kopeikin Mikhail Vasilievich, Candidate of Technical Sciences, Associate Professor of the Faculty of Fundamental and Humanitarian Disciplines, Department of Information Systems and Computer Engineering

УДК 004.9 **DOI: 10.18413/2518-1092-2021-6-2-0-6**

Скрипина И.И.[1]
Зайцева Т.В.[2]
Путивцева Н.П.[2]

АНАЛИЗ И ВЫБОР МАТЕМАТИЧЕСКОЙ МОДЕЛИ С ПОМОЩЬЮ МЕТОДА АНАЛИЗА ИЕРАРХИЙ

[1)] Белгородский государственный аграрный университет имени В.Я. Горина, ул. Вавилова, д.1, п. Майский, Белгородский р-н, Белгородская обл., 308503, Россия
[2)] Белгородский государственный национальный исследовательский университет, ул. Победы д. 85, г. Белгород, 308015, Россия

e-mail: skripina@bsu.edu.ru, zaitseva@bsu.edu.ru, putivzeva@bsu.edu.ru

Аннотация
В настоящее время существует множество различных моделей, используемых для прогнозирования, классификация которых рассмотрена в данной статье. Так же в статье было рассмотрено использование такого инструмента, как метод анализа иерархий для определения наиболее подходящей для прогнозирования модели.
Ключевые слова: модель, прогнозирование, классификация, метод анализа иерархий.
Для цитирования: Скрипина И.И., Зайцева Т.В., Путивцева Н.П. Анализ и выбор математической модели с помощью метода анализа иерархий // Научный результат. Информационные технологии. – Т.6, №2, 2021. – С. 41-46. DOI: 10.18413/2518-1092-2021-6-2-0-6

Skripina I.I.[1]
Zaitseva T.V.[2]
Putivtseva N.P.[2]

ANALYSIS AND SELECTION OF A MATHEMATICAL MODEL USING THE HIERARCHY ANALYSIS METHOD

[1)] Belgorod State Agrarian University named after V.Ya. Gorin, 1 Vavilova St., Maysky, Belgorod district, Belgorod region, 308503, Russia
[2)] Belgorod State National Research University, 85 Pobedy St., Belgorod, 308015, Russia

e-mail: skripina@bsu.edu.ru, zaitseva@bsu.edu.ru, putivzeva@bsu.edu.ru

Abstract
Currently, there are many different models used for forecasting, the classification of which we will consider in this article. The article also discussed the use of such a tool as the method of analyzing hierarchies to determine the most suitable model for forecasting.
Keywords: model, forecasting, classification, hierarchy analysis method.
For citation: Skripina I.I., Zaitseva T.V., Putivtseva N.P. Analysis and selection of a mathematical model using the hierarchy analysis method // Research result. Information technologies. – T.6, №2, 2021. – P. 41-46. DOI: 10.18413/2518-1092-2021-6-2-0-6

Для успешного функционирования любого производства необходимо принимать решения, в условиях, которые бы соответствовали условиям функционирования реального объекта. Одним из инструментов, помогающим в принятии решений в таких условиях, являются системы поддержки принятия решений, использующие имитационное или математическое моделирование. Рассмотрим классификацию математических моделей на рисунке 1.

В настоящее время все большую популярность набирают комбинированные модели, которые объединяют в себе несколько моделей и методов. Это дает возможность нивелировать недостатки одних моделей при помощи использования достоинства других. Данный подход помогает повысить точность прогнозов, и, как следствие, эффективность. Однако, при разработке таких моделей необходимо учитывать их сложность и ресурсоемкость.

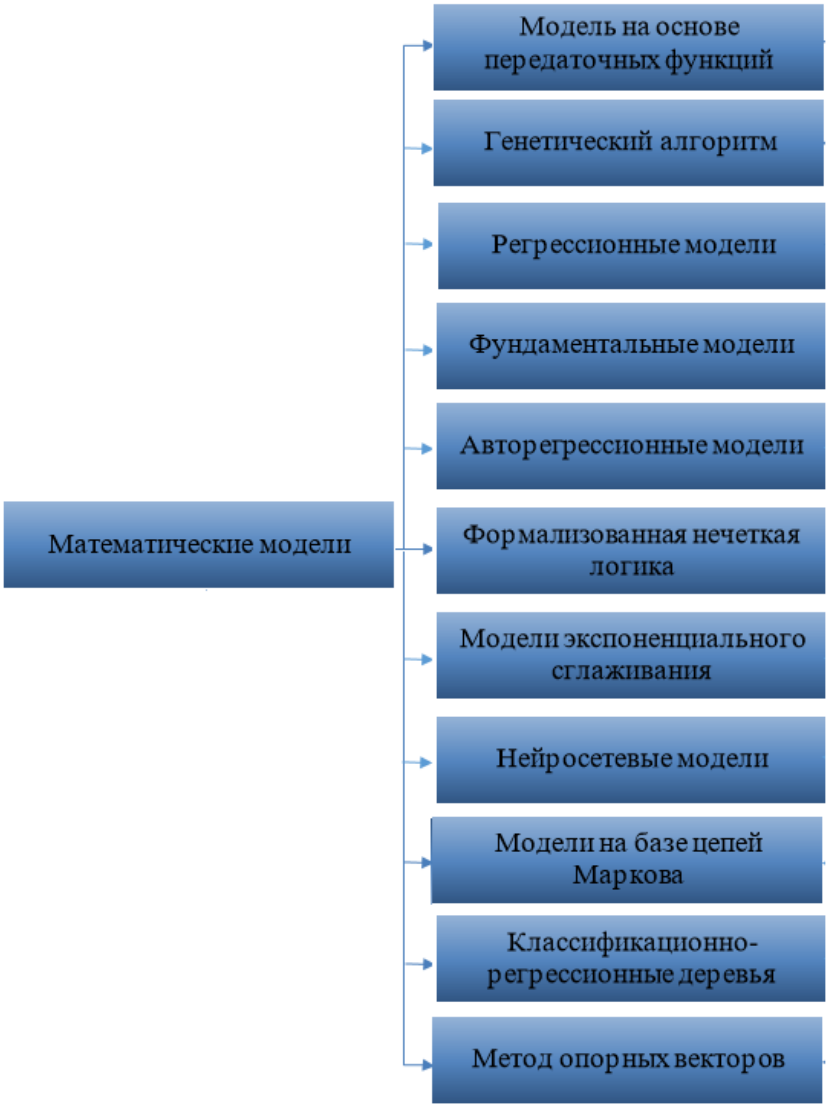


Рис. 1. Классификация математических моделей
Fig. 1. Classification of mathematical models

Рассмотренные математические модели (некоторые модели являются разновидностями моделей одного класса, поэтому были объединены в укрупненный класс) были рассмотрены в качестве альтернатив:

- регрессионные модели;
- авторегрессионные модели;
- модель группового учета элементов;
- адаптивные модели временных рядов;
- нейросетевые модели;
- модели прогнозирования на основе цепей Маркова;
- модели на базе классификационно-регрессионных деревьев CART.

В качестве критериев были выбраны характерные особенности рассмотренных моделей:

- возможность анализа промежуточных значений;
- возможность учета нелинейных зависимостей;
- учет ретроспективных данных;
- универсальность;
- возможность масштабирования;
- адаптивный подбор горизонта планирования;
- временные затраты;

- адаптируемость к изменению исходных данных;
- количество исходных данных;
- гибкость.

Для выбора наиболее предпочтительной математической модели был использован метод многокритериального оценивания, позволяющий задавать степень предпочтительности одного объекта над другим с использованием порядковой шкалы со значениями от 1 до 9 и обратными величинами путем заполнения обратносимметричных матриц парных сравнений. Данный метод реализован в ряде компьютерных систем поддержки принятия решений. Для решения задачи была использована СППР «Решение».

Иерархия выбора математической модели прогнозирования представлена на рисунке 2.

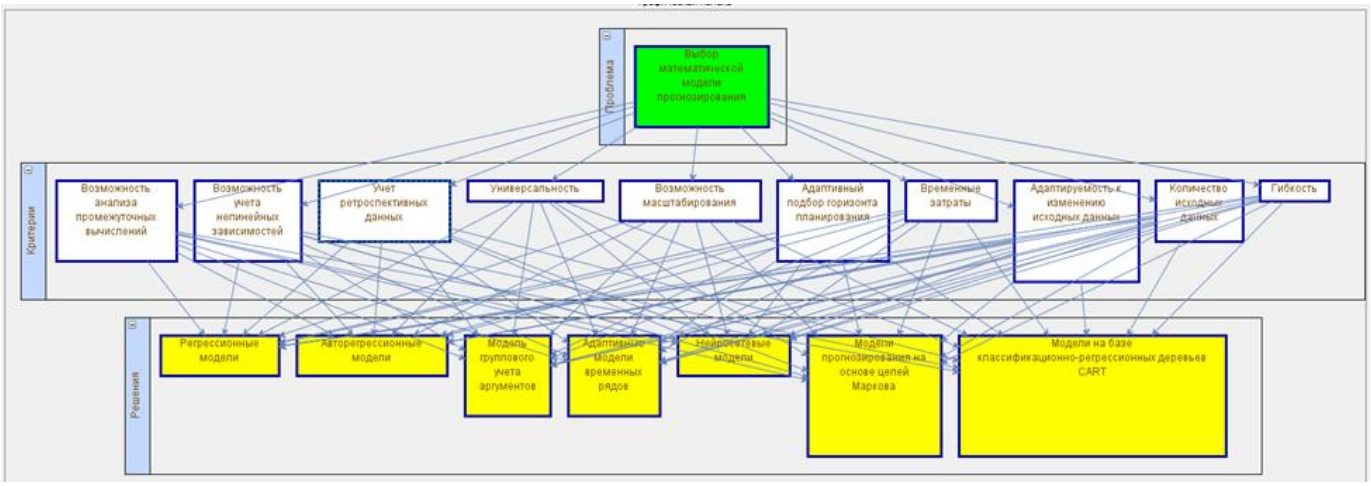
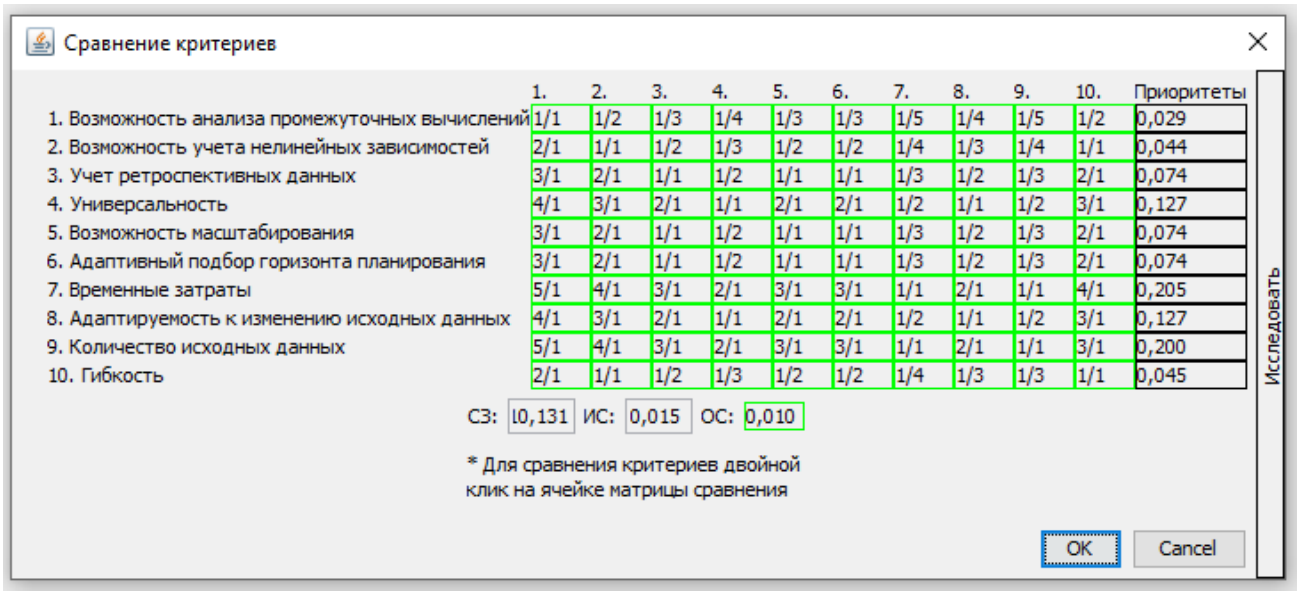


Рис. 2. Иерархия выбора математической модели прогнозирования
Fig. 2. Hierarchy of choosing a mathematical forecasting model

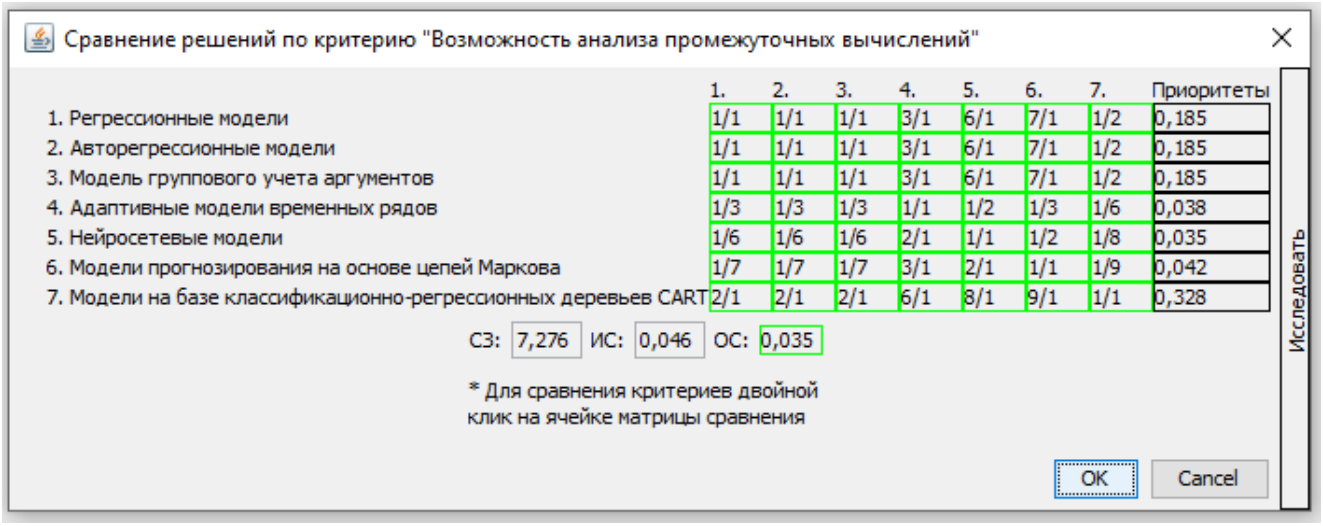
На основе представленных показателей с помощью МАИ выберем модель прогнозирования. Для этого построим матрицы парных сравнений критериев и альтернатив по предложенным критерием.


Сравнение критериев

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	Приоритеты
1. Возможность анализа промежуточных вычислений	1/1	1/2	1/3	1/4	1/3	1/3	1/5	1/4	1/5	1/2	0,029
2. Возможность учета нелинейных зависимостей	2/1	1/1	1/2	1/3	1/2	1/2	1/4	1/3	1/4	1/1	0,044
3. Учет ретроспективных данных	3/1	2/1	1/1	1/2	1/1	1/1	1/3	1/2	1/3	2/1	0,074
4. Универсальность	4/1	3/1	2/1	1/1	2/1	2/1	1/2	1/1	1/2	3/1	0,127
5. Возможность масштабирования	3/1	2/1	1/1	1/2	1/1	1/1	1/3	1/2	1/3	2/1	0,074
6. Адаптивный подбор горизонта планирования	3/1	2/1	1/1	1/2	1/1	1/1	1/3	1/2	1/3	2/1	0,074
7. Временные затраты	5/1	4/1	3/1	2/1	3/1	3/1	1/1	2/1	1/1	4/1	0,205
8. Адаптируемость к изменению исходных данных	4/1	3/1	2/1	1/1	2/1	2/1	1/2	1/1	1/2	3/1	0,127
9. Количество исходных данных	5/1	4/1	3/1	2/1	3/1	3/1	1/1	2/1	1/1	3/1	0,200
10. Гибкость	2/1	1/1	1/2	1/3	1/2	1/2	1/4	1/3	1/3	1/1	0,045

Рис. 3. Матрица парных сравнений критериев
Fig. 3. Matrix of paired comparisons of criteria

Следующим этапом было вычисление локальных приоритетов альтернатив по каждому из критериев путем заполнения соответствующих матриц парных сравнений (рис. 4).



Рс.4. МПС альтернатив по критерию «Возможность анализа промежуточных значений»
Rs. 4. MPS of alternatives according to the criterion "The possibility of analyzing intermediate values"

Наилучшими альтернативами с точки зрения критерия «Возможность анализа промежуточных значений» являются модели на базе классификационно-регрессионных деревьев CART. На втором месте по предпочтительности следующие модели: регрессионные модели, авторегрессионные модели, модель группового учета элементов.

Наилучшими альтернативами с точки зрения критерия «Возможность учета нелинейных зависимостей» являются нейросетевые модели и модели на базе классификационно-регрессионных деревьев CART.

Наилучшими альтернативами с точки зрения критерия «Учет ретроспективных данных» являются авторегрессионные модели и адаптивные модели временных рядов.

Наилучшими альтернативами с точки зрения критерия «Универсальность» являются нейросетевые модели и адаптивные модели временных рядов.

Наилучшими альтернативами по критерию «Возможность масштабирования» являются нейросетевые модели и модели прогнозирования на основе цепей Маркова.

Наилучшими альтернативами по критерию «Адаптивный подбор горизонта прогнозирования» являются адаптивные модели временных рядов и авторегрессионные модели.

Наилучшими альтернативами по критерию «Временные затраты» являются – модели на базе классификационно-регрессионных деревьев CART и модели прогнозирования на основе цепей Маркова.

Наилучшей альтернативой по критерию «Адаптируемость к изменению исходных данных» являются модели прогнозирования на основе цепей Маркова. На втором месте по предпочтительности – регрессионные, авторегрессионные и нейросетевые модели.

Наиболее предпочтительными альтернативами по критерию «Количество исходных данных» являются модель группового учета элементов и нейросетевые модели. На первом месте по предпочтительности нейросетевые модели, на втором – регрессионные модели.

Проведя все сравнения для иерархии, можно перейти к результатам ранжирования моделей. Как видно из рисунка 5, наилучшей оценки заслуживают нейросетевые модели.

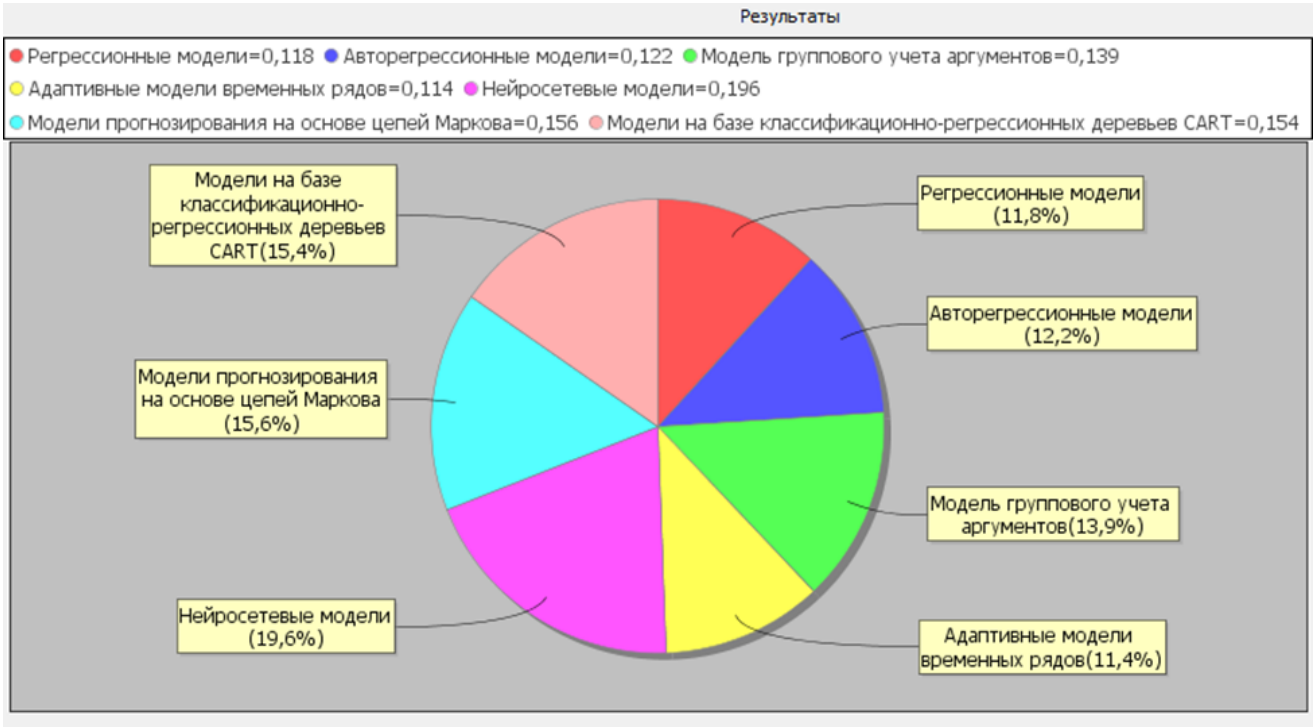


Рис. 5. Окно вывода результатов оценки критериев в СППР «Решение»
Fig. 5. The window for displaying the results of the evaluation of criteria in the DSS "Solution"

Таким образом, анализ результатов показал, что наиболее предпочтительной моделью для прогнозирования являются нейросетевыми. Модели прогнозирования на основе цепей Маркова и модели на базе классификационно-регрессионных деревьев CART находятся соответственно на втором и третьем месте по предпочтительности.

Список литературы

1. Блюмин, С.Д. Модели и методы принятия решений в условиях неопределенности [Текст] / С.Л. Блюмин, И.А. Шуйкова. – Липецк: ЛЭГИ, 2001. – 138 с.

2. Ногин, В.Д. Принятие решений в многокритериальной среде: количественный подход [Текст] / В.Д. Ногин. – Москва: ФИЗМАТЛИТ, 2002. – 176 с.

3. Огурцов, А.Н. Алгоритм повышения согласованности экспертных оценок в методе анализа иерархий [Текст] / А.Н. Огурцов, Н.А. Староверова // Вестник Ивановского государственного энергетического университета. – Иваново, 2013. – №5. – С.81-84.

4. Саати, Т. Принятие решений. Метод анализа иерархий [Текст] / Т. Саати. – Москва: Радио и связь, 1993. – 278 с.

5. Скрипина И.И. Экспертная оценка приоритетности объектов инвестирования на основе метода анализа иерархий. [Текст] / И.И. Скрипина, Е.С. Сорокина // Научные ведомости Белгородского государственного университета. Серия Экономика. Информатика. – 2017. – №9(258) – С. 133-141.

6. Путивцева Н.П. Разработка программной поддержки иерархической многокритериальной процедуры оценки качества экспертов. [Текст] / Н.П. Путивцева, Т.В. Зайцева, О.П. Пусная, С.В. Игрунова, Е.В. Нестерова, Е.В. Калюжная, Е.А. Зайцева. // Научные ведомости Белгородского государственного университета. Серия Экономика. Информатика. – 2016. – №16(237) – С. 172-179.

References

1. Blyumin, S.D. Models and methods of decision-making in conditions of uncertainty [Text] / S.L. Blyumin, I.A. Shuikova. – Lipetsk: LEGI, 2001 – 138 p.

2. Nogin, V.D. Decision-making in a multi-criteria environment: a quantitative approach [Text] / V.D. Nogin. – Moscow: FIZMATLIT, 2002. – 176 p.

3. Ogurtsov, A.N. Algorithm for improving the consistency of expert assessments in the method of hierarchy analysis [Text] / A.N. Ogurtsov, N.A. Staroverova // Bulletin of the Ivanovo State Energy University. – Ivanovo, 2013. – No. 5. – pp. 81-84.

4. Saati, T. Decision-making. Method of hierarchy analysis [Text] / T. Saati. – Moscow: Radio and Communications, 1993. – 278 p.

5. Skripina I.I. Expert assessment of the priority of investment objects based on the method of hierarchy analysis [Text] / I.I. Skripina, E.S. Sorokina // Belgorod State University Scientific Bulletin Economics Information technologies. – 2017. – №9(258) – P. 133-141.

6. Putivtseva N.P. Implementation of the program support of the hierarchical multicriteria procedure of the evaluation of experts' quality [Text] / N.P. Putivceva, T.V. Zajceva, O.P. Pusnaja, S.V. Igrunova, E.V. Nesterova, E.V. Kaljuzhnaja, E.A. Zajceva // Belgorod State University Scientific Bulletin Economics Information technologies. – 2016. – №16(237) – P. 172-179.

Скрипина Ирина Ивановна, старший преподаватель кафедры математики, физики, химии и информационных технологий
Зайцева Татьяна Валентиновна, кандидат технических наук, доцент, доцент кафедры прикладной информатики и информационных технологий
Путивцева Наталья Павловна, кандидат технических наук, доцент кафедры прикладной информатики и информационных технологий

Skripina Irina Ivanovna, Senior Lecturer, Department of Mathematics, Physics, Chemistry and Information Technology
Zaitseva Tatyana Valentinovna, Candidate of Technical Sciences, Associate Professor, Associate Professor of the Department of Applied Informatics and Information Technologies
Putivtseva Natalia Pavlovna, Candidate of Technical Sciences, Associate Professor of the Department of Applied Informatics and Information Technologies

КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ
COMPUTER SIMULATION

УДК 004.838.2 **DOI: 10.18413/2518-1092-2021-6-2-0-7**

Черноморец А.А.
Болгова Е.В.
Черноморец Д.А.

О СКРЫТНОМ ВНЕДРЕНИИ ДАННЫХ В ВИДЕОПОТОК НА ОСНОВЕ ТРЕХМЕРНОГО СУБПОЛОСНОГО АНАЛИЗА

Белгородский государственный национальный исследовательский университет,
ул. Победы д. 85, г. Белгород, 308015, Россия

e-mail: chernomorets@bsu.edu.ru, bolgova_e@bsu.edu.ru, chernomorets_d@bsu.edu.ru

Аннотация
Работа посвящена разработке основных теоретических положений метода скрытного внедрения данных на основе трехмерного субполосного анализа. Рассмотрена задача скрытного внедрения данных в отдельные блоки кадров видеопотока. Приведено описание подобласти пространственных частот при трехмерном косинус преобразовании. Предложены понятия части и доли квадрата нормы блока кадров при косинус преобразовании, соответствующие заданной подобласти пространственных частот, а также соотношения для их вычисления на основании значений элементов соответствующих субполосных матриц. Приведены соотношения, являющиеся основой построения проекций блока кадров на собственные векторы субполосных матриц, образующих трехмерный ортонормированный базис, а также соотношения, которые являются основой для проведения субполосного синтеза блока кадров на основе измененных в соответствии с внедряемыми данными проекциями блока кадров. Сформулированы основные этапы метода скрытного субполосного внедрения данных в блоки кадров видеопотока. Рассмотренный метод скрытного субполосного внедрения данных в заданный блок кадров на основе трехмерного субполосного анализа может быть применен для повышения скрытности внедренных данных.

Ключевые слова: скрытное внедрение, блок кадров, субполосный анализ, субполосные матрицы, собственные векторы, косинус преобразование, проекции блока кадров.
Для цитирования: Черноморец А.А., Болгова Е.В., Черноморец Д.А. О скрытном внедрении данных в видеопоток на основе трехмерного субполосного анализа // Научный результат. Информационные технологии. – Т.6, №2, 2021. – С. 47-55. DOI: 10.18413/2518-1092-2021-6-2-0-7

Chernomorets A.A.
Bolgova E.V.
Chernomorets D.A.

ON HIDDEN DATA EMBEDDING INTO THE VIDEO STREAM BASED ON THREE-DIMENSIONAL SUBBAND ANALYSIS

Belgorod State National Research University, 85 Pobedy St., Belgorod, 308015, Russia

e-mail: chernomorets@bsu.edu.ru, bolgova_e@bsu.edu.ru, chernomorets_d@bsu.edu.ru

Abstract
The paper is devoted to the development of the main theoretical provisions of the method of hidden data embedding based on three-dimensional subband analysis. The problem of hidden data embedding in separate blocks of video stream frames is considered. The description of the spatial frequencies subdomain of the three-dimensional cosine transform is given. The concepts of the part and the fraction of the square of the norm of a frames block with a cosine transform

corresponding to a given spatial frequencies subdomain, as well as formulas for their calculation based on the corresponding subband matrices elements values, are proposed. The formulas that are the basis for constructing the frames block projections on the eigenvectors of subband matrices forming a three-dimensional orthonormal basis are given, as well as the formulas that are the basis for conducting sub-band synthesis of a frames block based on the frames block projections modified in accordance with the embedding data. The main stages of the method of hidden subband data embedding in video stream frames blocks are formulated. The considered method of hidden subband data embedding into a given frames block based on three-dimensional subband analysis can be used to increase the secrecy of embedded data.
Keywords: hidden embedding, frames block, subband analysis, subband matrices, eigenvectors, cosine transform, frames block projections.
For citation: Chernomorets A.A., Bolgova E.V., Chernomorets D.A. On hidden data embedding into the video stream based on three-dimensional subband analysis // Research result. Information technologies. – T.6, №2, 2021. – P. 47-55. DOI: 10.18413/2518-1092-2021-6-2-0-7

ВВЕДЕНИЕ

В настоящее время в условиях высокого уровня развития средств создания, передачи, хранения мультимедийных данных достаточно актуальным является создание информационных систем, обеспечивающих защиту авторских прав на созданные звуковые записи, цифровые изображения, видеозаписи.

В существующих системах защиты авторских прав на мультимедийную продукцию, зачастую, для идентификации автора применяют скрытное внедрение данных в звуковые записи, изображения, а также в кадры видеопотока. При внедрении в видеопоток в большинстве случаев скрытное внедрение данных осуществляется в отдельные кадры видеопотока (рисунок 1) на основании методов внедрения в изображения [1-5].

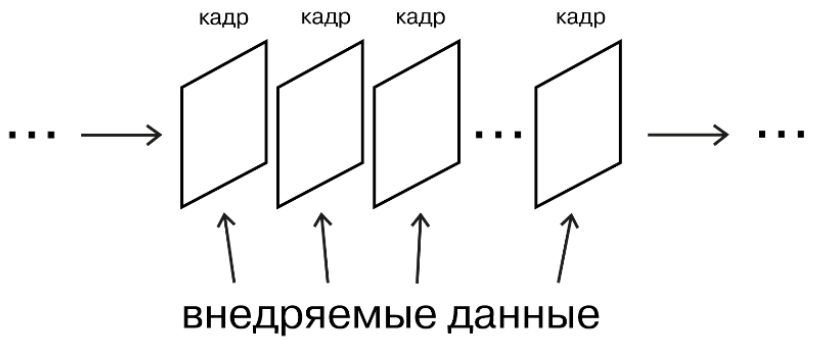


Рис. 1. Скрытное внедрение данных в отдельные кадры видеопотока
Fig. 1. Hidden data embedding into individual frames of a video stream

Для повышения скрытности внедряемых данных одним из подходов является скрытное внедрение данных в блоки кадров видеопотока (рисунок 2). В данном случае видеопоток представляется в виде последовательности блоков кадров, при этом каждый блок кадров рассматривается как единый контейнер, в который осуществляется скрытное внедрение данных.

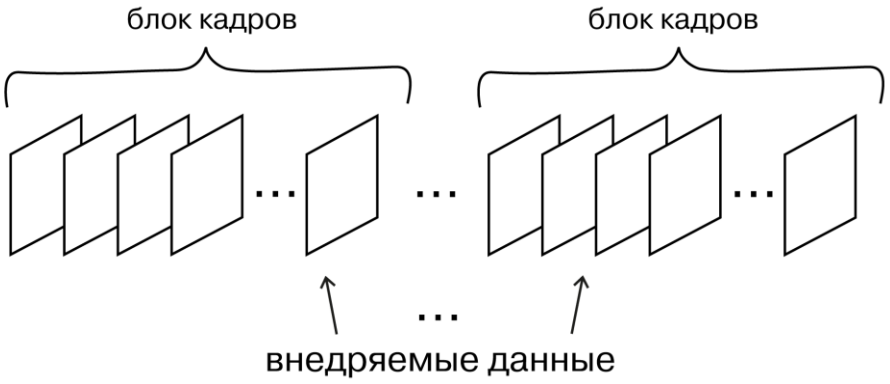


Рис. 2. Скрытное внедрение данных в блоки кадров видеопотока
Fig.2. Hidden data embedding into video stream frames blocks

Представим отдельный блок кадров в виде трехмерной матрицы $\Phi = (f_{ijk})$, $i = 1,2,...,N_1$, $j = 1,2,...,N_2$, $k = 1,2,...,N_3$, значений пикселей на кадрах блока, где $N_1 \times N_2$ – размерность кадра, N_3 – количество кадров в блоке. Адекватной теоретической основой многих задач обработки трехмерных данных, в частности значений пикселей блока кадров видеопотока, является их частотное представление на основе результатов косинус преобразования по дискретным данным [6-9].

$$f_{ijk} = \frac{8}{\pi^3} \int_0^{\pi} \int_0^{\pi} \int_0^{\pi} F^{\Phi}(u,v,w) \cos(u(i-\frac{1}{2})) \cos(v(j-\frac{1}{2})) \cos(w(k-\frac{1}{2})) du dv dw, \quad (1)$$

где $F^{\Phi}(u,v,w)$ – частотная характеристика, получаемая в результате косинус преобразования [7-9]:

$$F^{\Phi}(u,v,w) = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sum_{k=1}^{N_3} f_{ijk} \cos(u(i-\frac{1}{2})) \cos(v(j-\frac{1}{2})) \cos(w(k-\frac{1}{2})), \quad (2)$$

u, v, w – нормированные пространственные круговые частоты, принимающие значения в области определения косинус преобразования D_π (рисунок 3) [7, 9, 10],

$$(u,v,w) \in D_\pi, \quad (3)$$

$$D_\pi = \{(u,v,w) \mid 0 \le u, v, w < \pi\}. \quad (4)$$

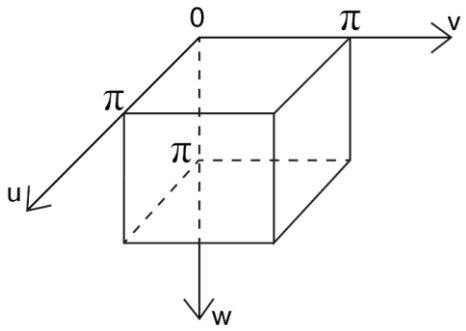


Рис. 3. Область D_π определения трехмерного косинус преобразования
Fig. 3. The definition domain D_π of the three-dimensional cosine transform

Для применения субполосного анализа/синтеза [11-13] в задаче скрытного внедрения данных в блок кадров Ф, основанного на анализе распределения квадрата нормы его косинус-преобразования по подобластям пространственных частот области определения косинус-преобразования, разобьем область D_π на $R_1 \times R_2 \times R_3$ подобластей $V_{r_1 r_2 r_3}$, $r_1 = 1,2,...,R_1$, $r_2 = 1,2,...,R_2$, $r_3 = 1,2,...,R_3$, следующего вида (рисунок 4):

$$V_{r_1 r_2 r_3} = D^u_{r_1} \cap D^v_{r_2} \cap D^w_{r_3}, \tag{5}$$

где трехмерные субполосы $D^u_{r_1}$, $D^v_{r_2}$ и $D^w_{r_3}$ трехмерного пространства пространственных частот имеют вид:

$$D^u_{r_1} = [u_{r_1,1}, u_{r_1,2}), \quad 0 \le u_{r_1,1} < u_{r_1,2} < \pi, \quad r_1 = 1,2,...,R_1, \tag{6}$$

$$D^v_{r_2} = [v_{r_2,1}, v_{r_2,2}), \quad 0 \le v_{r_2,1} < v_{r_2,2} < \pi, \quad r_2 = 1,2,...,R_2, \tag{7}$$

$$D^w_{r_3} = [w_{r_3,1}, w_{r_3,2}), \quad 0 \le w_{r_3,1} < w_{r_3,2} < \pi, \quad r_3 = 1,2,...,R_3, \tag{8}$$

$$u_{r_1,1} = (r_1 - 1)\frac{\pi}{R_1}, \quad u_{r_1,2} = r_1 \frac{\pi}{R_1}, \tag{9}$$

$$v_{r_2,1} = (r_2 - 1)\frac{\pi}{R_2}, \quad v_{r_2,2} = r_2 \frac{\pi}{R_2}. \tag{10}$$

$$w_{r_3,1} = (r_3 - 1)\frac{\pi}{R_3}, \quad w_{r_3,2} = r_3 \frac{\pi}{R_3}. \tag{11}$$

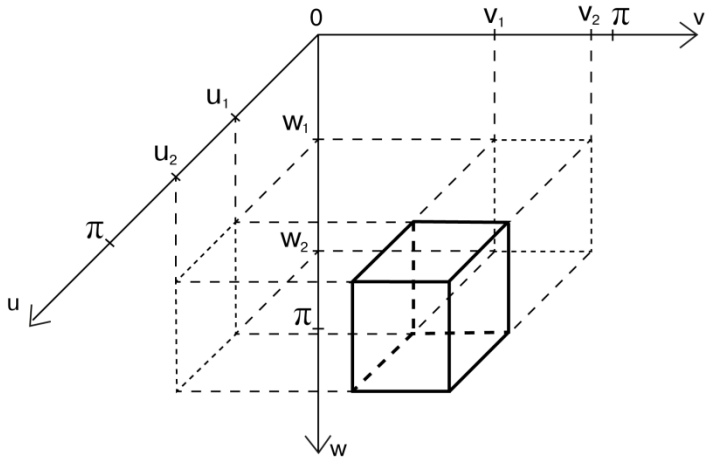


Рис. 4. Подобласть пространственных частот при трехмерном косинус преобразовании
Fig. 4. The spatial frequencies subdomain of the three-dimensional cosine transform

Для анализа информативности подобластей пространственных частот с позиции значимости соответствующих им данных для представления блока кадров рассмотрим понятие доли квадрата нормы блока кадров Ф.

На основании равенства Парсеваля для косинус-преобразования справедливо следующее равенство для квадрата нормы блока кадров Ф:

$$\|\Phi\|^2 = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sum_{k=1}^{N_3} f_{ijk}^2 = \frac{8}{\pi^3} \int_0^\pi \int_0^\pi \int_0^\pi (F^\Phi(u,v,w))^2 \, du dv dw. \tag{12}$$

Часть $E_{r_1 r_2 r_3}(\Phi)$ квадрата нормы блока кадров Ф при косинус-преобразовании, соответствующей заданной подобласти пространственных частот $V_{r_1 r_2 r_3}$, вычислим на основании следующего соотношения:

$$E_{r_1r_2r_3}(\Phi)=\frac{8}{\pi^3}\iiint\limits_{(u,v,w)\in V_{r_1r_2r_3}}(F^{\Phi}(u,v,w))^2\,dudvdw. \tag{13}$$

Очевидно, что для квадрата нормы блока кадров Ф справедливо следующее равенство:

$$\|\Phi\|^2=\sum_{r_1=1}^{R_1}\sum_{r_2=1}^{R_2}\sum_{r_3=1}^{R_3}E_{r_1r_2r_3}(\Phi)\,. \tag{14}$$

На основании соотношений (12) и (13) долю $P_{r_1r_2r_3}(\Phi)$ квадрата нормы блока кадров Ф, соответствующую заданной подобласти пространственных частот $V_{r_1r_2r_3}$, предлагается вычислять следующим образом:

$$P_{r_1r_2r_3}(\Phi)=\frac{E_{r_1r_2r_3}(\Phi)}{\|\Phi\|^2}\,. \tag{15}$$

Для вычисления значений части $E_{r_1r_2r_3}(\Phi)$ квадрата нормы блока кадров Ф при косинус-преобразовании, соответствующей заданной подобласти пространственных частот $V_{r_1r_2r_3}$, преобразуем выражение (13) следующим образом – подставим (2) в (13) и выполним преобразования:

$$E_{r_1r_2}(\Phi)=\frac{8}{\pi^3}\iiint\limits_{(u,v,w)\in V_{r_1r_2r_3}}\sum_{i_1=1}^{N_1}\sum_{j_1=1}^{N_2}\sum_{k_1=1}^{N_3}f_{i_1j_1k_1}\cos(u(i_1-\frac{1}{2}))\cos(v(j_1-\frac{1}{2}))\cos(w(k_1-\frac{1}{2}))\cdot$$

$$\cdot\sum_{i_2=1}^{N_1}\sum_{j_2=1}^{N_2}\sum_{k_2=1}^{N_3}f_{i_2j_2k_2}\cos(u(i_2-\frac{1}{2}))\cos(v(j_2-\frac{1}{2}))\cos(w(k_2-\frac{1}{2}))dudvdw=$$

$$=\sum_{i_1=1}^{N_1}\sum_{j_1=1}^{N_2}\sum_{k_1=1}^{N_3}\sum_{i_2=1}^{N_1}\sum_{k_2=1}^{N_2}\sum_{k_2=1}^{N_3}f_{i_1j_1k_1}f_{i_2j_2k_2}g^{r_1}_{i_1i_2}h^{r_2}_{j_1j_2}z^{r_3}_{k_1k_2}\,, \tag{16}$$

где

$$g^{r_1}_{i_1i_2}=\frac{2}{\pi}\int\limits_{u\in D^u_{r_1}}\cos(u(i_1-\frac{1}{2}))\cos(u(i_2-\frac{1}{2}))du\,, \tag{17}$$

$$h^{r_2}_{j_1j_2}=\frac{2}{\pi}\int\limits_{v\in D^v_{r_2}}\cos(v(j_1-\frac{1}{2}))\cos(v(j_2-\frac{1}{2}))dv\,, \tag{18}$$

$$z^{r_3}_{k_1k_2}=\frac{2}{\pi}\int\limits_{w\in D^w_{r_3}}\cos(w(k_1-\frac{1}{2}))\cos(w(k_2-\frac{1}{2}))dw\,. \tag{19}$$

В работах [12-15] показано, что элементы (17)-(19) могут быть вычислены следующим образом: значения элементов $g^{r_1}_{i_1i_2}$, $i_1,i_2=1,2,...,N_1$, могут быть представлены, применяя значения (9), в виде:

$$g^{r_1}_{i_1i_2}=a^{r_1}_{i_1i_2}+\tilde{g}^{r_1}_{i_1i_2}\,, \tag{20}$$

где

$$a^{r_1}_{i_1i_2}=\begin{cases}\dfrac{\sin(u_{r_1,2}(i_1-i_2))-\sin(u_{r_1,1}(i_1-i_2))}{\pi(i_1-i_2)}, & i_1\neq i_2,\\ \dfrac{u_{r_1,2}-u_{r_1,1}}{\pi}, & i_1=i_2,\end{cases} \tag{21}$$

$$\tilde{g}^{r_1}_{i_1i_2}=\frac{\sin(u_{r_1,2}(i_1+i_2-1))-\sin(u_{r_1,1}(i_1+i_2-1))}{\pi(i_1+i_2-1)}\,. \tag{22}$$

Значения элементов $h^{r_2}_{j_1 j_2}$, $j_1, j_2 = 1,2,...,N_2$, и элементов $z^{r_3}_{k_1 k_2}$, $k_1, k_2 = 1,2,...,N_3$, также могут быть вычислены на основании соотношений (20) и (21) при подстановке значений частот $v_{r_2,1}$, $v_{r_2,2}$ (10) и $w_{r_3,1}$, $w_{r_3,2}$ (10) соответственно.

Элементы $g^{r_1}_{i_1 i_2}$, $i_1, i_2 = 1,2,...,N_1$; $h^{r_2}_{j_1 j_2}$, $j_1, j_2 = 1,2,...,N_2$, $z^{r_3}_{k_1 k_2}$; $k_1, k_2 = 1,2,...,N_3$, (17)-(19) образуют субполосные матрицы косинус преобразования $G_{r_1} = (g^{r_1}_{i_1 i_2})$, $i_1, i_2 = 1,2,...,N_1$; $H_{r_2} = (h^{r_2}_{j_1 j_2})$, $j_1, j_2 = 1,2,...,N_2$, и $Z_{r_3} = (z^{r_3}_{k_1 k_2})$, $k_1, k_2 = 1,2,...,N_3$, соответствующих заданной подобласти пространственных частот $V_{r_1 r_2 r_3}$ вида (5).

Субполосные матрицы косинус-преобразования $G_{r_1} = (g^{r_1}_{i_1 i_2})$, $i_1, i_2 = 1,2,...,N_1$; $H_{r_2} = (h^{r_2}_{j_1 j_2})$, $j_1, j_2 = 1,2,...,N_2$, и $Z_{r_3} = (z^{r_3}_{k_1 k_2})$, $k_1, k_2 = 1,2,...,N_3$, а также их собственные числа и собственные векторы представляют собой математический инструмент субполосного анализа/синтеза блока кадров видеопотока.

В работах [14-15] показано, что субполосные матрицы G_{r_1}, H_{r_2} и Z_{r_3} являются вещественными, симметрическими матрицами, следовательно, они имеют полные наборы ортонормированных собственных векторов и соответствующие собственные числа.

Обозначим, $\vec{q}^{\,r_1}_i = (q^{r_1}_{i_1 i})$, $i, i_1 = 1,2,...,N_1$; $\vec{u}^{\,r_2}_j = (u^{r_2}_{j_1 j})$, $j, j_1 = 1,2,...,N_2$; $\vec{w}^{\,r_3}_k = (w^{r_3}_{k_1 k})$, $k, k_1 = 1,2,...,N_3$, – наборы собственных векторов субполосных матриц G_{r_1}, H_{r_2} и Z_{r_3}.

Элементы $\gamma^{r_1 r_2 r_3}_{ijk}$, $i = 1,2,...,N_1$, $j = 1,2,...,N_2$, $k = 1,2,...,N_3$, значения которых предлагается вычислять на основании следующего соотношения,

$$\gamma^{r_1 r_2 r_3}_{ijk} = \sum_{i_1=1}^{N_1} \sum_{j_1=1}^{N_2} \sum_{k_1=1}^{N_3} f_{i_1 j_1 k_1} q^{r_1}_{i_1 i} u^{r_2}_{j_1 j} w^{r_3}_{k_1 k}, \quad (23)$$

можно считать значениями проекций блока кадров Ф в трехмерном ортонормированном базисе, составленном из собственных векторов $\vec{q}^{\,r_1}_i$, $i = 1,2,...,N_1$; $\vec{u}^{\,r_2}_j$, $j = 1,2,...,N_2$, и $\vec{w}^{\,r_3}_k$, $k = 1,2,...,N_3$, субполосных матриц G_{r_1}, H_{r_2} и Z_{r_3}, соответствующих заданной подобласти пространственных частот $V_{r_1 r_2 r_3}$.

Для восстановления значений элементов $f_{i_1 j_1 k_1}$, $i_1 = 1,2,...,N_1$, $j_1 = 1,2,...,N_2$, $k_1 = 1,2,...,N_3$, блока кадров Ф предлагается использовать следующее соотношение:

$$f_{i_1 j_1 k_1} = \sum_{i_1=1}^{N_1} \sum_{j_1=1}^{N_2} \sum_{k_1=1}^{N_3} \gamma^{r_1 r_2 r_3}_{ijk} q^{r_1}_{i_1 i} u^{r_2}_{j_1 j} w^{r_3}_{k_1 k}. \quad (24)$$

Соотношение (23) является основой представления используемого для скрытного внедрения блока кадров Ф в виде множества его проекций в трехмерном ортонормированном базисе, составленном из собственных векторов субполосных матриц.

Соотношение (24) является основой для проведения субполосного синтеза блока кадров на основе измененных в соответствии с внедряемыми данными проекциями блока кадров в трехмерном ортонормированном базисе.

Метод скрытного субполосного внедрения данных в заданный блок кадров на основе трехмерного субполосного анализа-синтеза разработан в соответствии с основными положениями метода скрытного субполосного внедрения данных в изображения [10, 16]:

- выбор подобласти пространственных частот, в которую рекомендуется скрытно внедрять данные, предлагается осуществлять на основе анализа значений долей $P_{r_1 r_2 r_3}(\Phi)$ (15) квадрата

нормы блока кадров Ф, соответствующих различным подобластям пространственных частот $V_{r_1r_2r_3}$ вида (5);

- выбор неинформативных проекций блока кадров в трехмерном ортонормированном базисе, составленном из собственных векторов субполосных матриц, которые соответствуют выбранной подобласти пространственных частот, предлагается осуществлять на основании анализа распределения квадратов значений соответствующих проекций (23);

- изменение значений выбранных для скрытного внедрения проекций предлагается осуществлять на основании внедряемых данных, представленных в двоичном виде, в соответствии с требованиями метода скрытного субполосного внедрения данных в изображения [10, 16];

- синтез блока кадров, содержащего внедренные данные, предлагается осуществлять в соответствии с соотношением (24),

- извлечение внедренных данных также осуществляется в соответствии с методом скрытного субполосного внедрения данных в изображения.

Предлагаемый метод скрытного субполосного внедрения данных в заданный блок кадров на основе трехмерного субполосного анализа-синтеза позволяет повысить скрытность внедрения.

ВЫВОДЫ

В работе рассмотрены основные положения метода скрытного внедрения данных в видеопоток на основе трехмерного субполосного анализа. Для повышения скрытности внедряемых данных предложено осуществлять скрытное внедрение данных в отдельные блоки кадров видеопотока. Сформулированы основные теоретические положения субполосного анализа при трехмерном косинус преобразовании с позиций решения задачи скрытного внедрения. Приведены соотношения, являющиеся основой построения проекций блока кадров Ф на собственные векторы субполосных матриц, образующих трехмерный ортонормированный базис, а также соотношения, которые являются основой для проведения субполосного синтеза блока кадров на основе измененных в соответствии с внедряемыми данными проекциями блока кадров. Сформулированы основные этапы метода скрытного субполосного внедрения данных в блоки кадров видеопотока.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-07-00657.

Список литературы

1. Грибунин В.Г. Цифровая стеганография / В.Г. Грибунин, И.Н. Оков, И.В. Туринцев. – М.: Солон-пресс, 2016. – 262 с.

2. Конахович Г.Ф. Компьютерная стеганография. Теория и практика / Г.Ф Конахович., А.Ю Пузыренко. – Киев: «МК–Пресс», 2006. – 288 с.

3. Аграновский А.В. Стеганография, цифровые водяные знаки и стеганоанализ / А.В. Аграновский. – М.: Вузовская книга, 2009. – 220 с.: ил.

4. Cox I.J. Digital watermarking and steganography / I.J Cox., M. Miller, J Bloom., J. Fridrich, T Kalker. – Morgan Kaufmann, 2007. – 593 p.

5. Hartung F. Multimedia. Watermarking Techniques / F Hartung., M. Kutter // Proceedings IEEE, Special Issue on Identification and Protection of Multimedia Information. – 1999. – 87(7). – Pp. 1079-1107.

6. Залманзон Л.А. Преобразования Фурье, Уолша, Хаара и их применение в управлении, связи и других областях / Л.А. Залманзон. – М.: Наука. Гл. ред. физ.-мат. лит., 1989. – 496 с.

7. Болгова Е.В. О сосредоточенности энергии косинусного преобразования [Текст] / Е.В. Болгова // Научные ведомости БелГУ. Сер. Экономика. Информатика. – 2017. – № 9(258). – Вып. 42. – С. 111-121.

8. Черноморец А.А. О равенстве Парсеваля для косинусного преобразования Фурье изображений [Текст] / Черноморец А.А., Болгова Е.В., Ходырева А.А. IV международная научно-практическая конференция: Фундаментальная наука и технологии - перспективные разработки. – North Charleston, 2014. – С. 160-163.

9. Черноморец А.А. Об анализе данных на основе косинусного преобразования [Текст] / А.А. Черноморец, Е.В. Болгова // Научные ведомости БелГУ. Сер. Экономика. Информатика. – 2015. – № 1(198). – Вып. 33/1. – С. 68-73.

10. Болгова Е.В. О методе субинтервального скрытного внедрения данных в изображения [Текст] / Е.В. Болгова, А.А. Черноморец // Научные ведомости БелГУ. Сер. Экономика. Информатика. – 2018. – Т. 45. – № 1.– С. 192-201.

11. Черноморец Д.А. Представление изображений на основе базиса собственных векторов субполосных матриц косинус-преобразования / Д.А. Черноморец, Е.В. Болгова, А.А. Черноморец, А.А. Барсук // Научный результат. Информационные технологии. – 2019. Т. 4. –№ 1. – С. 3-8.

12. Жиляков, Е.Г. Вариационные алгоритмы анализа и обработки изображений на основе частотных представлений: Монография [Текст] / Е.Г. Жиляков, А.А. Черноморец. – Белгород: Изд-во ГИК, 2009. – 146 с.

13. Жиляков, Е.Г. Вариационные методы анализа и построения функций по эмпирическим данным: моногр. [Текст] / Е.Г. Жиляков. – Белгород: Изд-во БелГУ, 2007. – 160 с.

14. Жиляков, Е.Г. О субполосном анализе изображений [Текст] / Е.Г. Жиляков, А.А. Черноморец // ГрафиКон’2013: 23-я Международная конференция по компьютерной графике и зрению: 16–20 сентября, 2013 г., Владивосток, Институт автоматики и процессов управления ДВО РАН: Труды конференции. – С. 230-233.

15. Жиляков, Е.Г. О частотном анализе изображений [Текст] / Е.Г. Жиляков, А.А. Черноморец // Вопросы радиоэлектроники. Сер. ЭВТ. – 2010. – Вып. 1. – С. 94-103.

16. Zhilyakov E.G. Hidden data embedding method based on the image projections onto the eigenvectors of subinterval matrices / E.G. Zhilyakov, A.A Chernomorets., E.V Bolgova., I.I Oleynik., D.A. Chernomorets // International Journal of Engineering & Technology. – 2018. – 7(3.19). – P. 72-80.

References

1. Gribunin V.G. Digital steganography / V.G. Gribunin, I.N. Okov, I.V. Turintsev. – M.: Solon-press, 2016, 262 p.

2. Konakhovich G.F. Computer steganography. Theory and practice / G.F. Konakhovich, A.Yu. Puzyrenko. – Kiev: «MK–Press», 2006, 288 p.

3. Agranovskiy A.V. Steganography, digital watermarks and steganoanalysis / A.V. Agranovskiy. – M.: Vuzovskaya kniga, 2009. 220 p.

4. Cox I.J. Digital watermarking and steganography / I.J Cox., M. Miller, J Bloom., J. Fridrich, T Kalker. – Morgan Kaufmann, 2007. – 593 p.

5. Hartung F. Multimedia. Watermarking Techniques / F Hartung., M. Kutter // Proceedings IEEE, Special Issue on Identification and Protection of Multimedia Information. – 1999. – 87(7). – Pp. 1079-1107.

6. Zalmanzon L.A. Fourier, Walsh, Haar transforms and their application in control, communication and other fields / L.A. Zalmanzon. – M.: Science. Ch. ed. physical-mat. lit., 1989. – 496 p.

7. Bolgova E.V. About cosine transform energy concentration / E.V. Bolgova // Belgorod State University Scientific Bulletin Economics Information technologies. – 2017. – № 9(258). – Issue. 42. – P. 111-121.

8. Chernomorets A.A On Parseval's equality for the cosine Fourier transform of images / A.A Chernomorets, E.V Bolgova, Khodyreva A.A. // IV international scientific and practical conference: Fundamental science and technology - promising. – North Charleston, 2014. – P. 160-163.

9. Chernomorets A.A. On the analysis of data based on the cosine transformation / A.A Chernomorets, E.V. Bolgova // Belgorod State University Scientific Bulletin Economics Information technologies. – 2015. – № 1(198). – Issue. 33/1. – P. 68-73.

10. Bolgova E.V. On the method of subinterval data hidden embedding in images / E.V Bolgova, A.A Chernomorets // Belgorod State University Scientific Bulletin Economics Information technologies. – 2018. – T. 45. – № 1. – P. 192-201.

11. Chernomorets D.A. Images presentation based on subband cosine transform matrix eigenvectors basis / D.A. Chernomorets, E.V. Bolgova, A.A. Chernomorets, A.A. Barsuk // Research result. Information technologies. – 2019. – T. 4. – № 1. – P. 3-8.

12. Zhilyakov E.G. Variational algorithms for image analysis and processing based on frequency representations: Monograph // E.G. Zhilyakov, A.A. Chernomorets. – Belgorod: GIK, 2009. – 146 p.

13. Zhilyakov E.G. Variational methods of analysis and construction of functions from empirical data: Monograph / E.G. Zhilyakov. – Belgorod: BelGU Publishing House – 160 p.

14. Zhilyakov E.G. About subband image analysis / E.G. Zhilyakov, A.A. Chernomorets // GrafiKon'2013: 23rd International Conference on Computer Graphics and Vision: September 16–20, 2013, Vladivostok, Institute of Automation and Control Processes FEB RAS: Proceedings of the conference. – P. 230-233.

15. Zhilyakov E.G. On frequency analysis of images/ E.G. Zhilyakov, A.A. Chernomorets // Questions of radio electronics. Ser. EVT. – 2010. – Issue. 1. – P. 94-103.

16. Zhilyakov E.G. Hidden data embedding method based on the image projections onto the eigenvectors of subinterval matrices / E.G. Zhilyakov, A.A Chernomorets. E.V Bolgova, I.I Oleynik, D.A. Chernomorets // International Journal of Engineering & Technology. – 2018. – 7(3.19). – P. 72-80.

Черноморец Андрей Алексеевич, доктор технических наук, доцент, профессор кафедры прикладной информатики и информационных технологий
Болгова Евгения Витальевна, кандидат технических наук, доцент кафедры прикладной информатики и информационных технологий
Черноморец Дарья Андреевна, аспирант кафедры информационно-телекоммуникационных систем и технологий

Chernomorets Andrey Alekseevich, Doctor of Technical Sciences, Associate Professor, Professor of the Department of Applied Informatics and Information Technologies
Bolgova Evgeniya Vitalievna, Candidate of Technical Sciences, Associate Professor of the Department of Applied Informatics and Information Technologies
Chernomorets Darya Andreevna, postgraduate student of the Department of Information and Telecommunications Systems and Technologies