АЗЕРБАЙДЖАНСКАЯ РЕСПУБЛИКА

На правах рукописи

РАЗРАБОТКА МЕТОДОВ И АЛГОРИТМОВ ОБЕСПЕЧЕНИЯ КИБЕРУСТОЙЧИВОСТИ КИБЕРФИЗИЧЕСКИХ СИСТЕМ

Специальность: 3338.01 – Системный анализ, управление и обработка информации (информационные технологии)

Отрасль науки: Технические науки

ДИССЕРТАЦИЯ

на соискание ученой степени
доктора наук

Соискатель: _____ доктор философии по технике, доцент
Сухостат Людмила Валентиновна

Научный консультант: _____ академик НАН Азербайджана, доктор технических наук, профессор
Алгулиев Расим Магамед оглы

Баку – 2025

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ 5

ГЛАВА I. Анализ и исследование существующих методов и алгоритмов обеспечения киберустойчивости и информационной безопасности киберфизических систем 17

1.1. Анализ особенностей становления и развития киберфизических систем 17

1.2. Анализ и исследование методов и алгоритмов интеллектуальной обработки больших данных киберфизических систем 36

1.3. Анализ и исследование методов и алгоритмов обеспечения кибербезопасности информационных технологий и защищенности операционных технологий киберфизических систем 38

1.4. Анализ и исследование методов и алгоритмов оценки киберустойчивости киберфизических систем 51

1.5. Постановка научной задачи исследования 54

1.6. Концептуальная модель обеспечения киберустойчивости киберфизических систем 56

Выводы к главе I 59

ГЛАВА II. Разработка методов и алгоритмов интеллектуальной обработки больших данных для обеспечения киберустойчивости киберфизических систем 60

2.1. Алгоритм обнаружения выбросов в больших данных на основе метода к-средних для снижения риска блокировки киберфизических систем 60

2.2. Алгоритм анализа больших данных киберфизических систем на основе взвешенной кластеризации 80

2.3. Метод параллельной обработки больших данных для снижения риска выхода из строя киберфизических систем 89

2.4. Метод кластерного анализа больших данных киберфизических систем на основе консенсусного ансамбля 103
Выводы к главе II 124
ГЛАВА III. Разработка методов и алгоритмов обеспечения кибербезопасности информационных технологий киберфизических систем 125
3.1. Метод классификации кибератак на киберфизические системы с применением экстремального машинного обучения 125
3.2. Метод обнаружения DoS атак на киберфизические системы с применением ансамбля классификаторов 141
3.3. Алгоритм обнаружения вредоносного программного обеспечения в киберфизических системах на основе изображений 154
Выводы к главе III 176
ГЛАВА IV. Разработка методов и алгоритмов обеспечения защищенности операционных технологий киберфизических систем 177
4.1. Метод обнаружения аномалий в операционных технологиях киберфизических систем на основе иерархических скрытых Марковских моделей 177
4.2. Метод обнаружения аномалий в акустических сигналах киберфизических систем с применением трансферного обучения . 185
4.3. Метод обнаружения атак на устройства киберфизических систем с применением глубокой гибридной модели 197
4.4. Метод классификации отказов киберфизических систем на основе изображений 216
Выводы к главе IV 233

ГЛАВА V. Разработка методов оценки киберустойчивости киберфизических систем 234
5.1. Метод определения критичности уязвимостей функциональных компонентов киберфизических систем на основе Байесовского графа атак 234
5.2. Метод оценки рисков киберфизических систем с применением нечеткого интеграла Сугено 251
Выводы к главе V 270
ЗАКЛЮЧЕНИЕ 271
СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ 273
Список сокращений и условных обозначений 317

ВВЕДЕНИЕ

Актуальность работы. Современные информационные технологии являются важным средством развития государства и общества. Киберфизические системы (КФС) являются базовой технологией в реализации концепции Индустрия 4.0 и ключевым компонентом интеллектуальных объектов, которые широко используются в различных областях человеческой деятельности.

Ключевые технологические тенденции, лежащие в основе КФС, включают: «Интернет вещей» (Internet of Things), электрические сети, использующие смарт-технологии [353, с. 127-140; 383, с. 2375-2385], производственные системы с интегрированным искусственным интеллектом [271, с. 70-75], облачные вычисления [4, с. 18-31] и т.д. КФС являются основой для развития следующих сфер: смарт-производство, смарт-медицина [280, с. 1-6], смарт-здания и инфраструктуры, интеллектуальные транспортные системы [401, с. 117-124], мобильные системы, системы обороны и системы метеонаблюдения [51, с. 212-223].

КФС представляют собой комбинацию двух технологий: информационные технологии (ИТ) киберуровня и операционные технологии (ОТ) физического уровня. Киберуровень состоит из серверов, систем баз данных, хостов и т. д. Физический уровень состоит из сенсоров, исполнительных механизмов, систем обратной связи и т. д., которые отвечают за управление производственными объектами [87, с. 161-166].

КФС требует подключения многих устройств и систем для сбора и обмена огромными объемами данных на платформах ОТ и ИТ. Для этого существуют различные протоколы автоматизации, которые необходимо соединить для достижения поставленных задач. Стандарт ISO 22301:2019 имеет решающее значение для повышения устойчивости организаций к различным непредвиденным сбоям, обеспечивая непрерывность операций и услуг [227].

Актуальность к обеспечению кибербезопасности КФС резко возросла в

последние годы после серии кибератак.

Кибератаки на КФС представляют значительный риск для здоровья и безопасности человеческой жизни, угрожают серьезным ущербом окружающей среде. Интерес к обеспечению информационной безопасности КФС резко возрос после серии атак посредством компьютерного вируса Stuxnet в 2010 году [138, с. 80-91]. Далее последовали атаки на SCADA системы водоснабжения. Умные счетчики были взломаны в Пуэрто-Рико. Еще одной атакой стала Industroyer в 2016 году, направленная на отключение украинской энергосистемы. В апреле 2020 года скоординированные кибератаки на программируемые логические контроллеры водного сектора Израиля привели к неправильной работе оборудования. В июне того же года компания Honda Motor подверглась кибератаке. Заводское производство было временно приостановлено из соображений безопасности. А в июне Symantec выявила 31 крупную корпорацию США, ставшую объектом атаки с упором на производственный сектор. Атака вредоносного ПО 7 мая 2021 года остановила работу американской трубопроводной системы Colonial Pipeline по поставкам нефти и газа. 4 августа 2023 года были атакованы больницы и клиники в нескольких штатах США, что нарушило работу их компьютерных систем. В том же году энергетическая инфраструктура Дании подверглась одной из крупнейших кибератак. Кибератаки на энергетический сектор Тайваня наблюдались в апреле 2024 года. Таким образом, кибератаки на КФС растут с каждым годом.

Из-за постоянно меняющихся киберугроз необходима разработка методологии более высокого уровня, которая требует обнаружения внедренных и скрывающихся кибератак, обеспечивая более киберустойчивые и защищенные киберинфраструктуры [167, с. 1674-1683].

12 марта 2024 года Европейский парламент одобрил предложенный Закон о киберустойчивости (Cyber Resilience Act) [149]. Он будет применяться ко всем продуктам с цифровыми элементами, представленным на европейском

рынке, использование которых включает прямое или косвенное логическое или физическое подключение данных к устройству или сети.

Нарушение функционирования КФС и вывод из строя ее компонентов может привести к серьезному ущербу, что ставит вопрос о необходимости рассмотрения кибербезопасности КФС как одного из приоритетных.

Переход к Индустрии 4.0 поставил перед Азербайджанской Республикой, как и перед всем миром, новые задачи и вызовы. В Азербайджане обеспечены все возможности для развития Четвертой промышленной революции [321, с. 81-89]. В стране постоянно ведутся разработки передовых технологий [7, с. 98-108; 220, с. 768-770].

На сегодняшний день в Республике все технологии Индустрии 4.0 применяются в различных критических инфраструктурах, таких как нефтегазовые платформы, транспортные системы и т.д. [321, с. 81-89; 161, с. 86-99].

Активное сотрудничество с Всемирным экономическим форумом привело к созданию в Баку Центра анализа и координации четвертой промышленной революции согласно указу Президента Азербайджанской Республики от 6 января 2021 года [377]. Деятельность центра призвана обеспечить развитие искусственного интеллекта и машинного обучения, а также служить гражданам, развитию цифровой экономики и дальнейшему совершенствованию различных областей исследований в стране [5, с. 52].

17 апреля 2021 года был подписан указ «О некоторых мерах по обеспечению безопасности критической информационной инфраструктуры» [11].

Согласно закону Азербайджанской Республики о внесении изменений в Закон Азербайджанской Республики «Об информации, информатизации и защите информации» от 27 мая 2022 года добавляются понятия «критическая информационная инфраструктура», ее объект и субъект, а также правовая основа правил обеспечения ее безопасности [9]. 28 августа 2023 утверждена

«Стратегия информационной и кибербезопасности Азербайджанской Республики на 2023-2027 годы» [3]. 29 ноября 2024 года кабинет министров Азербайджана утвердил «Перечень объектов критической информационной инфраструктуры» [10].

Указом Президента Азербайджанской Республики от 16 января 2025 года была утверждена «Концепция цифрового развития в Азербайджанской Республике» [2].

Выявление современных кибератак имеет решающее значение для устойчивой работы КФС. Традиционные механизмы защиты становятся все более неэффективными из-за методов, используемых злоумышленниками. Риски кибератак возникают из-за интеграции кибер- и физического доменов, другими словами, интеграции доменов ИТ и ОТ. В этом контексте разработка адаптивных, более эффективных методов обнаружения кибератак является насущной потребностью для защиты ИТ и ОТ инфраструктур от таких угроз [332, с. 1-12].

Критические ситуации идеальны для киберпреступников, поскольку они могут воспользоваться самым слабым звеном – человеческим фактором. Киберпреступники используют человеческий фактор для получения несанкционированного доступа, кражи учетных данных и заражения систем вредоносными программами.

Важно заранее разработать алгоритмы обнаружения и меры противодействия всем известным кибератакам, чтобы уменьшить их влияние и минимизировать ущерб системе. Однако исследования показывают, что требуется более глубокий анализ кибербезопасности КФС, особенно применимости подходов на основе искусственного интеллекта.

В первую пятерку исследовательских университетов, занимающихся этой проблемой, входят Калифорнийский университет в Беркли (США), Пекинский университет науки и технологий (Китай), Королевский технологический институт КТН (Швеция), Миланский политехнический университет (Италия) и

Гамбургский технологический университет (Германия) [51, с. 212-223].

По мере продвижения к Индустрии 4.0 КФС становится все более сложной, распределенной и управляемой данными, что постепенно превращает обнаружение вторжений и аномалий в проблему больших данных [331, с. 50-67]. Сложной задачей становится необходимость анализа большого объема данных, разнородных по своей природе, с минимальными вычислительными затратами [1, с. 96-99; 246, с. 60-74; 43, с. 10-20; 29, с. 360-372].

Использование сенсоров, исполнительных механизмов и других устройств для сбора данных из КФС предоставляет полезную информацию для их последующего анализа. Однако сегодня объем хранимых данных становится слишком большим для обработки традиционными алгоритмами. Исследователям приходится прибегать к различным подходам, таким как разбиение данных и использование априорных знаний. Таким образом, работа с большими данными вызывает необходимость формализации новых методов, применяемых исследователями, и создания новых алгоритмов, использующих возможности методов машинного обучения для работы с данными больших объемов и размерностей. Тем не менее, существующие алгоритмы не всегда эффективны из-за высокой вычислительной сложности.

Методы машинного обучения широко применяются для анализа данных КФС на предмет наличия кибератак [20, с. 128-131]. Теоретически алгоритмы машинного обучения могут получить высокую производительность, т.е. минимизировать уровень ложных тревог и максимизировать точность обнаружения. Однако обычно требуется бесконечное число обучающих образцов [20, с. 128-131]. На практике это условие невозможно в силу ограничения вычислительной мощности и требования ответа в режиме реального времени. В связи с этим злоумышленники часто ищут и нацеливаются на самое слабое звено – наиболее уязвимый функциональный компонент в КФС.

Самые уязвимые устройства могут стать хорошей отправной точкой для

эффективного исследования уязвимостей системы безопасности [168, с. 1-6]. Существующие методы не могут всесторонне анализировать распространенные многоэтапные кибератаки в среде КФС.

Решение исследуемых проблем является важной задачей, в связи с этим были разработаны и применяются на практике различные методы и алгоритмы. Большинство из них рассматривают либо кибербезопасность только в среде ИТ, либо защищенность в среде ОТ. Обеспечение защиты от киберфизических атак является серьезной проблемой, требующей методов оценки киберустойчивости, способных исследовать тесные взаимодействия и взаимозависимости между кибер- и физическими компонентами в КФС [27, с. 1-16]. Однако существующие методы не учитывают эту специфику [352, с. 157-173].

В свете вышеизложенного можно сделать вывод, что на данный момент не существует единого подхода к проектированию защищенных и кибербезопасных КФС. Данное диссертационное исследование направлено на разработку методов и алгоритмов для обеспечения кибербезопасности компонентов ИТ и защищенности компонентов ОТ.

Цель работы. Целью диссертационной работы является разработка методов и алгоритмов обеспечения киберустойчивости и информационной безопасности киберфизических систем на основе интеллектуального анализа больших данных о кибератаках и отказах систем.

Объектом исследования является кибербезопасность ИТ и защищенность ОТ киберфизических систем, **предметом исследования** являются модели и методы управления процессами обеспечения киберустойчивости киберфизических систем.

Для достижения поставленной цели в работе решаются следующие **задачи:**

– Анализ и исследование существующих методов и алгоритмов обеспечения киберустойчивости КФС;

- Разработка концептуальной модели обеспечения киберустойчивости КФС;
- Разработка методов и алгоритмов интеллектуальной обработки больших данных для обеспечения киберустойчивости КФС;
- Разработка методов и алгоритмов обеспечения кибербезопасности информационных технологий КФС;
- Разработка методов обеспечения защищенности операционных технологий КФС;
- Разработка метода определения критичности уязвимостей функциональных компонентов для обеспечения киберустойчивости КФС;
- Разработка метода оценки рисков КФС.

Методы исследования базируются на применении теории информационной безопасности, машинного обучения, теории вероятностей, теории нечетких чисел, теории распознавания образов, анализа рисков и принятия решений.

Основные положения выносимые на защиту

- Концептуальная модель обеспечения киберустойчивости КФС;
- Алгоритм обнаружения выбросов в больших данных на основе метода к-средних для снижения риска блокировки КФС;
- Алгоритм анализа больших данных КФС на основе взвешенной кластеризации;
- Метод параллельной обработки больших данных для снижения риска выхода КФС из строя;
- Метод кластерного анализа больших данных КФС на основе консенсусного ансамбля;
- Метод классификации кибератак на КФС с применением экстремального машинного обучения;

- Метод обнаружения DoS атак на КФС с применением ансамбля классификаторов;
- Алгоритм обнаружения вредоносного программного обеспечения в КФС на основе изображений;
- Метод обнаружения аномалий в операционных технологиях КФС на основе иерархических скрытых Марковских моделей для системы водоочистки;
- Метод обнаружения аномалий в акустических сигналах КФС с применением трансферного обучения;
- Метод обнаружения атак на устройства КФС с применением глубокой гибридной модели для системы водоочистки;
- Метод классификации отказов КФС на основе изображений;
- Метод определения критичности уязвимостей функциональных компонентов КФС на основе Байесовского графа атак для транспортировки природного газа по трубопроводу;
- Метод оценки рисков КФС с применением нечеткого интеграла Сугено для генерации энергии ветра.

Научная новизна диссертационной работы состоит в следующем:

- Разработана трехуровневая концептуальная модель для решения ряда задач по обеспечению киберустойчивости КФС;
- Разработан алгоритм обнаружения выбросов в больших данных на основе метода к-средних для снижения риска блокировки КФС;
- Разработан алгоритм анализа больших данных КФС на основе взвешенной кластеризации;
- Разработан метод параллельной обработки больших данных для снижения риска выхода КФС из строя;
- Разработан метод кластерного анализа больших данных КФС на основе консенсусного ансамбля;

- Разработан метод классификации кибератак на КФС с применением экстремального машинного обучения;
- Разработан метод обнаружения DoS атак на КФС с применением ансамбля классификаторов;
- Разработан алгоритм обнаружения вредоносного программного обеспечения в КФС на основе изображений;
- Разработан метод обнаружения аномалий в операционных технологиях КФС на основе иерархических скрытых Марковских моделей для системы водоочистки;
- Разработан метод обнаружения аномалий в акустических сигналах КФС с применением трансферного обучения;
- Разработан метод обнаружения атак на устройства КФС с применением глубокой гибридной модели для системы водоочистки;
- Разработан метод классификации отказов КФС на основе изображений;
- Разработан метод определения критичности уязвимостей функциональных компонентов КФС на основе Байесовского графа атак для транспортировки природного газа по трубопроводу;
- Разработан метод оценки рисков КФС с применением нечеткого интеграла Сугено для генерации энергии ветра.

Теоретическая и практическая ценность. Научная и практическая ценность рассматриваемых в диссертации задач заключается в том, что разработанные методы и алгоритмы могут быть использованы для обеспечения киберустойчивости киберфизических систем с целью повышения защищенности ОТ и кибербезопасности ИТ. Разработанные методы и алгоритмы опубликованы в научных журналах и представлены в виде докладов на международных научных конференциях и являются общедоступными для применения. Научная обоснованность и достоверность полученных результатов подтверждаются применением современных методов исследований, а также

сравнением с результатами ранее предложенных методов и алгоритмов в исследуемой области с целью подтверждения их эффективности.

Апробация работы. Основные научно-теоретические и практические результаты диссертации были представлены и обсуждались на: семинарах Института Информационных Технологий Министерства Науки и Образования АР, а также на следующих республиканских и международных научных конференциях: III Республиканской научно-практической конференции «Riyaziyyatın tətbiqi məsələləri və yeni informasiya texnologiyaları» (г. Сумгаит, 15-16 декабря 2016 г.); I Республиканской конференции «Proqram mühəndisliyinin aktual elmi-praktiki problemləri» (г. Баку, 17 мая 2017 г.); I Республиканской научно-практической конференции «Big data: imkanları, multidissiplinar problemləri və perspektivləri» (г. Баку, 25 февраля 2016 г.); XIII Международной научно-технической конференции «Распознавание-2017» (г. Курск, Россия, 16-19 мая 2017 г.); IV Республиканской конференции «İnformasiya təhlükəsizliyinin aktual multidissiplinar elmi-praktiki problemləri» (г. Баку, 14 декабря 2018 г.); X Международной конеренции «Application of Information and Communication Technologies» (AICT-2016) (г. Баку, 12-14 октября 2016 г.); III Республиканской научно-практической конференции «İnformasiya təhlükəsizliyinin aktual problemləri» (г. Баку, 8 декабря 2017 г.); XV Международной научно-технической конференции «Распознавание-2019» (г. Курск, Россия, 14-17 мая 2019 г.); XIII Международной молодежной научной конференции «Technical Sciences. Industrial Management» (г. Боровец, Болгария, 11-14 марта 2020 г.); «Global Cyber Security Forum 2019» (г. Харьков, Украина, 14-16 ноября 2019); XIV Международной научно-технической конференции «Распознавание-2018» (г. Курск, Россия, 25-28 сентября 2018 г.); XIII Международной конеренции «Application of Information and Communication Technologies» (AICT-2019) (г. Баку, 23-25 октября 2019 г.); V Республиканской конференции «İnformasiya təhlükəsizliyinin aktual multidissiplinar elmi-praktiki problemləri» (г. Баку, 29 ноября 2019 г.); Международной конференции

«Information Security: Problems and Prospects» (г. Баку, 29 октября 2021 г.); Международном научном семинаре NATO ARW «Cybersecurity of Industrial Control Systems (ICS)» (г. Баку, 27-29 октября 2021 г.); XI «Национальном Суперкомпьютерном Форуме (НСКФ-21)» (г. Переславль-Залесский, Россия, 30 ноября – 3 декабря 2021 г.); XVI Международной конференции «Application of Information and Communication Technologies» (AICT-2022) (г. Вашингтон, США, 12-14 октября 2022 г.); XVII Международной научно-технической конференции «Распознавание-2023» (г. Курск, Россия, 12-15 сентября 2023 г.).

Соискатель принимал участие в конкурсах грантов, проводимым Фондом Развития Науки при Президенте Азербайджанской Республики (№ EİF-11-1(3)-82/08/1, EİF-RİTN-MQM-2/İKT-2-2013-7(13)-29/18/1, EİF-KETPL-2-2015-1(25)-56/05/1, AEF-MCG-2023-1(43)-13/04/1-M-04), Научным фондом Государственной нефтяной компании Азербайджанской Республики (ГНКАР) и Национальной Академией Наук Азербайджана.

Название организации, в которой выполнена диссертационная работа. Диссертация была выполнена в Институте Информационных Технологий Министерства Науки и Образования Азербайджанской Республики.

Публикации. По теме диссертационной работы опубликованы 33 научные работы, из них 17 статей, которые опубликованы в зарубежных журналах, в том числе 4 из них входят в базу данных Scopus, 11 работ входят в международную базу данных Web of ScienceTM Core Collection агентства Thomson Reuters и 2 работы входят в базу РИНЦ, 16 статей опубликованы в материалах международных и республиканских конференций.

Объем и структура работы. Диссертационная работа состоит из введения, пяти глав, заключения и списка литературы (409 наименований, в том числе 11 на азербайджанском языке, 13 на русском и 385 на английском языке). Основное содержание работы изложено на 318 страницах, включая 54 рисунка и 80 таблиц.

Общий объем диссертации в символах: 322919 символов, не считая рисунков, таблиц, графиков и списка литературы.

Титульный лист - 461 символ

Оглавление – 2538 символов

Введение – 17178 символов

Основное содержание диссертации (раздел, параграф, пункт) - 303118 символов (глава I – 62035 символов, глава II – 72511 символов, глава III – 58332 символа, глава IV – 66552 символа, глава V – 43688 символов), заключение – 2623 символа, список использованной литературы – 71041 символ, список сокращений и условных обозначений - 2204 символа).

ГЛАВА I. АНАЛИЗ И ИССЛЕДОВАНИЕ СУЩЕСТВУЮЩИХ МЕТОДОВ И АЛГОРИТМОВ ОБЕСПЕЧЕНИЯ КИБЕРУСТОЙЧИВОСТИ И ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ КИБЕРФИЗИЧЕСКИХ СИСТЕМ

Создание КФС систем поставило перед людьми новые задачи. Обеспечение информационной безопасности КФС является одной из наиболее сложных проблем в широком спектре средств защиты от кибератак [120, с. 1-7]. В данной главе описывается принцип работы КФС. Рассматриваются основные виды кибератак и угроз против КФС. Проводится анализ существующих исследовательских работ по интеллектуальной обработке больших данных КФС, обеспечению кибербезопасности ИТ, обнаружению аномалий в ОТ и оценке киберустойчивости КФС.

1.1. Анализ особенностей становления и развития киберфизических систем

Термин «киберфизическая система» был предложен Хелен Гилл в 2006 году на научном семинаре CPS NSF (National Science Foundation), проводимым Национальным научным фондом США. КФС — это система, которая может эффективно интегрировать кибер- и физические компоненты с использованием современных сенсорных, вычислительных и сетевых технологий. Сейчас КФС системы входят в списки приоритетных инноваций США и ряда стран Европы.

- Новизна и принципиальное отличие КФС от существующих встраиваемых систем или автоматизированных систем управления технологическими процессами (АСУ ТП), хотя и схожих внешне, заключается в том, что КФС интегрирует в себя кибернетическое начало, компьютерную технику и программные технологии, качественно новые исполнительные механизмы, заложенные в их окружающей среды и способны воспринимать ее изменения, реагировать на них, учиться и приспосабливаться [14, с. 517–528].
- С точки зрения информатики [260, с. 363-369] КФС представляет собой

интеграцию вычислительных и физических процессов. К ним относятся встроенные компьютеры, сетевые мониторы и контроллеры, обычно с обратной связью, где физические процессы влияют на вычисления и наоборот.

- С точки зрения технологий автоматизации [234, с. 3120-3121] КФС представляют собой специализированные системы, деятельность которых управляется вычислительными и коммуникационными ядрами, встроенными в объекты и структуры физической среды.
- Согласно NSF, КФС будущего намного превзойдут существующие системы по производительности, адаптивности, отказоустойчивости, безопасности и простоте использования.

Широкое внедрение КФС связано с концепцией «Industry 4.0» [397, с. 1028-1042], объединяющей технологии и знания, обеспечивающие автономность, надежность, системность и контроль без участия человека. Ключевые технологические тенденции, лежащие в основе КФС, включают Интернет вещей (Internet of Things, IoT) [45, с. 24-36; 6, с. 112-115; 154], большие данные, «умные» технологии [8, с. 95-107], облачные вычисления и т. д.

Индустрия 4.0 выступает в качестве инструмента для нового подхода к интеллектуальному производству, предоставляя различные новые возможности. Эта конкретная концепция, ориентированная на производство с использованием новых технологий, является лишь небольшим подмножеством, охватывающим многие промышленные области.

КФС являются основой для развития следующих направлений: умное производство, умная медицина, умные здания и инфраструктуры, умный город, умный транспорт, носимые устройства, мобильные системы, оборонные системы, метеорология и др. (Рис. 1.1.1). Быстрый рост приложений КФС приводит к ряду проблем с кибербезопасностью и конфиденциальностью [14, с. 517–528].

Согласно ISO/IEC 27001:2013, информационная безопасность – это сохранение конфиденциальности, целостности и доступности информации [51, с. 212-223].

Конфиденциальность – свойство, при котором информация не предоставляется и не раскрывается неуполномоченным лицам, организациям или процессам. Целостность есть свойство точности и полноты. Доступность – свойство быть доступным и пригодным для использования по требованию уполномоченного лица.

Другие свойства, такие как подлинность (authenticity), подотчетность (accountability), неотказуемость (non-repudiation) и надежность (reliability), также могут быть включены [92, с. 27-32].

В связи с широким распространением беспроводных технологий сбора и передачи данных и команд управления, где используется беспроводная сенсорная сеть (wireless sensor network, WSN), в отрасли возрастает потребность в разработке систем защиты информации. Удаленное расположение устройств КФС и их автономность приводят к риску вторжений и атак.

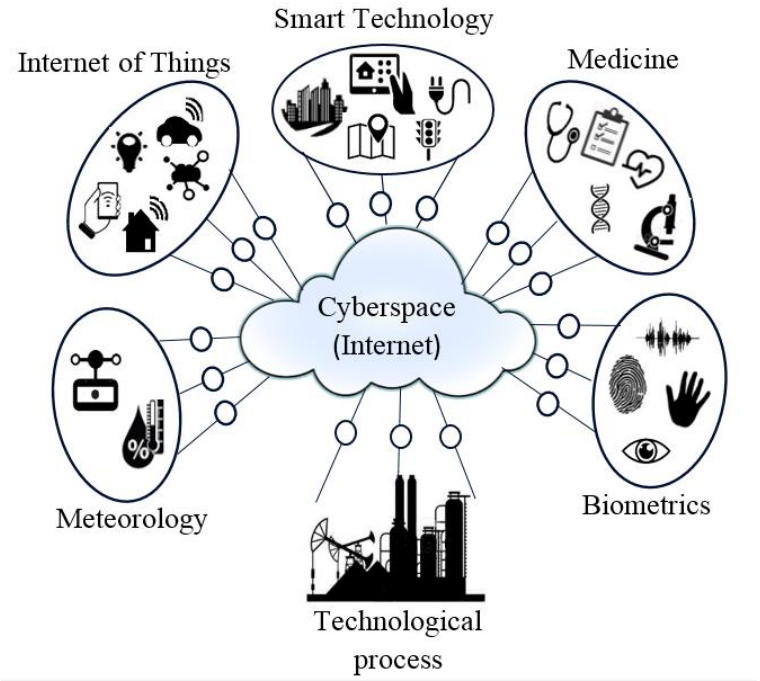


Рис. 1.1.1. Киберфизические системы [53, с. 213]

Работа с большими группами устройств может привести к компрометации некоторых из них. Безопасность КФС ставит ряд новых задач [272, с. 83-93]:

- рост количества IoT-устройств приводит к повышению уязвимости таких систем к кибератакам (например, DDoS (distributed denial-of-service));
- моделирование угроз безопасности;
- разработка формального подхода к оценке уязвимостей КФС;
- проектирование надежных и отказоустойчивых архитектур для обработки быстро развивающихся кибер- и физических угроз.

Поэтому необходимо разработать новые методологии и технологии (такие как сенсоры, ориентированные на людей, беспроводные и квантовые сенсоры, носимые биосенсоры, 2D/3D мультисенсорные системы и т. д.), чтобы удовлетворить требования КФС с точки зрения безопасности, надежности и защиты персональных данных [279, с. 678-682; 21, с. 92-96].

Новая вычислительная парадигма, известная как киберфизические системы социального назначения (Cyber-physical-social systems, CPSS) [341, с. 78-82], возникла из КФС и киберсоциальных систем (cyber-social systems, CSS). Эти системы расширяют КФС и включают в себя социальное пространство и признаки участия и взаимодействия людей [397, с. 1028-1042].

Особенности современных КФС систем включают [349, с. 466-469; 374, с. 127-155]:

- Большое количество устройств на базе встроенных процессоров и, как следствие, увеличение памяти для хранения данных.
- Встроенное и мобильное зондирование.
- Междоменные источники сенсоров и потоки данных.
- Взаимодействие кибер- и физической составляющих.
- Способность обучаться и адаптироваться.
- Взаимодействие через Интернет (например, IoT, «умная» среда и т.д.).
- Обеспечение надежной работы систем (таких как ATM и POS) с

централизованным автоматическим управлением. Качество алгоритма управления КФС может влиять на его сложность и надежность, повышая интенсивность вычислительной нагрузки.

- Наличие общего киберпространства обеспечивает обмен как внутри систем, так и с внешней средой, а также защиту информации в виде криптосистем, межсетевых экранов, антивирусов и т.д.
- Работа должна быть надежной и в некоторых случаях сертифицированной.
- Время отклика характеризует задержку обратной связи. Чем она выше, тем хуже контроль над качеством объекта.
- Надежность системы обеспечивается автоматизированным интеллектуальным управлением.
- Человек внутри/вне цикла [351, с. 3-9].

Принцип работы КФС. Архитектура КФС часто состоит из следующих основных уровней: киберуровня и физического уровня. Текущее состояние КФС включает переменные, которые представляют данные, полученные сенсорами, и управляющие переменные, представляющие управляющие сигналы [51, с. 212-223].

Нормальное значение определенного параметра процесса представляет собой целевое значение, которое необходимо достичь системе и поддерживать его. В КФС расстояние между значениями переменных процесса и соответствующими контрольными точками вычисляется контроллерами [51, с. 212-223]. После вычисления этого смещения контроллеры, используя сложный набор уравнений, разрабатывают локальное срабатывание и вычисляют новые переменные срабатывания и управления. Полученное управляющее значение отправляется на соответствующий исполнительный механизм, чтобы удерживать процесс ближе к определенному целевому значению.

Контроллеры также отправляют полученные измерения на главные серверы управления и выполняют с них выбранные команды. В КФС

операторы системы должны знать текущее состояние управляемых объектов. Таким образом, графический пользовательский интерфейс, называемый интерфейсом человек-машина (human-machine interface, HMI), предоставляет текущее состояние управляемого объекта человеку-оператору. В целом процесс работы КФС можно разделить на следующие этапы: 1) мониторинг; 2) сетевое взаимодействие; 3) вычислительная обработка; 4) приведение в действие. Киберуровень часто использует промышленные протоколы, такие как DNP3 и Modbus для связи с устройствами физического уровня.

Архитектуры КФС. На рис. 1.1.2 показана типичная базовая модель КФС, состоящая из нескольких уровней и зон, также известная как стандарт безопасности для таких систем (ISA-62443) [254, с. 38-62; 206, с. 1-15].

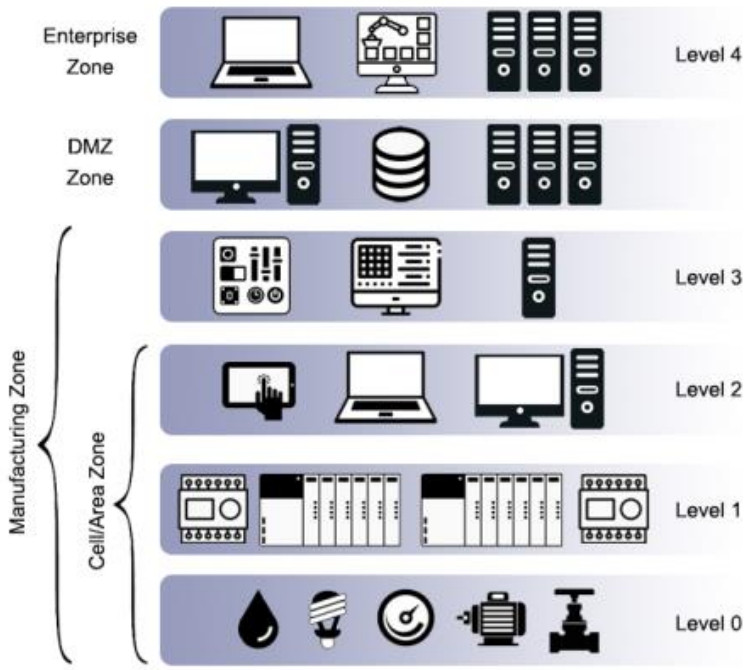


Рис. 1.1.2. Базовая модель по ISA-62443

Основные зоны можно охарактеризовать следующим образом [206, с. 1-15]:

- Зона предприятия представляет собой верхний уровень, также известный как ИТ-сеть.
- Демилитаризованная зона (Demilitarized Zone, DMZ). Все корпоративные

услуги (например, Интернет и электронная почта) находятся в этой зоне. Он включает в себя управление запасами, системы управления производством, контроллер домена, управление исправлениями, зеркало архиватора данных и другие.

- Зоны производства и управления. Первый применяется ко всей области ОТ, включая уровни с 0 по 3; последний находится на уровнях 0-2, поэтому он эквивалентен традиционной архитектуре КФС. Нулевой уровень содержит сенсоры для сбора данных и исполнительные механизмы. Первый уровень включает коммутационные узлы, которые подключают сенсоры к ПЛК. Второй и третий уровни – это ПЛК, рабочие места операторов, серверы хранения данных. Согласно [207, с. 1-6], уровень 3 включает в себя историю предприятия, производственное планирование и отчетность и другие. На продвинутом уровне реализованы VLAN, PVLAN, NIPS/HIPS, обнаружение аномалий и вторжений, интеллектуальные коммутаторы и т. д.

КФС традиционно используют систему аппаратных и программных компонентов SCADA для управления процессами, сбора данных, системного мониторинга, хранения данных лог-файлов и связи с промышленными устройствами [82, с. 81-97; 83, с. 481-489; 19, с. 257-259].

Самый нижний уровень типичной архитектуры SCADA обычно состоит из полевых элементов (также называемых терминалами или устройствами отключения звука), таких как сенсоры и исполнительные механизмы [107, с. 52-87; 123, с. 1-7]. Они управляются ПЛК и удаленными терминалами (Remote Terminal Unit, RTU). ПЛК и RTU – это микрокомпьютеры, которые отправляют управляющие сигналы на полевые элементы, собирают данные и передают их на главный терминал (master terminal unit, MTU).

Операторы контролируют отправку команд, загрузку новых конфигураций и мониторинг полевых элементов через HMI интерфейс, подключенный к MTU, что позволяет им собирать данные, отправлять команды на удаленные объекты

и изменять настройки и конфигурации [254, с. 38-62; 206, с. 1-15].

Политика и правила описывают безопасные конфигурации систем управления, охватывая такие вопросы (согласно ISO/IEC 62443):

- конфигурация сети;
- настройка брандмауэра;
- авторизация и аутентификация пользователей;
- связь между различными сетями управления технологическими процессами;
- использование беспроводной связи;
- домены и доверительные отношения;
- управление патчами и антивирусами;
- усиление защиты системы за счет закрытия программных портов, блокировки или отказа от неиспользуемых или опасных сервисов и отказа от съемных носителей данных;
- доступ к внешним сетям (например, к сети Интернет);
- правильное использование электронной почты.

Характеристики КФС включают [76, с. 71-119; 343, с. 64-73]: достоверность (fidelity), масштабируемость (scalability), воспроизводимость (reproducibility), адаптивность (adaptability), модульность (modularity), рентабельность (cost-effectiveness), измеримость (measurability), обособленность (isolation), удобство использования (usability), разнообразие (diversity), интероперабельность (interoperability), мониторинг (monitoring), открытость (openness), неоднородность (heterogeneity).

В КФС разработка стандартов, моделей и подходов очень важна. В связи с этим были предложены различные многоуровневые эталонные модели, чтобы лучше объяснить многоэлементную интеллектуальную производственную среду. Некоторые из самых известных моделей КФС систем [23, с. 14-23]:

- Модель эталонной архитектуры предприятия Purdue (Purdue Enterprise Reference Architecture Model, PERA);

- Модель архитектуры интеллектуальной сети (Smart Grid Architecture Model, SGAM);
- Модель эталонной архитектуры для «Индустрии 4.0» (Reference Architecture Model for Industrie 4.0, RAMI 4.0);
- Эталонная архитектура промышленного Интернета (Industrial Internet Reference Architecture, IIRA).

Наблюдение за состоянием и обеспечение безопасности для защиты критических инфраструктур представляет собой комплексный подход, который вводит процессы мониторинга на каждом из технических уровней моделей КФС (например, PERA [376, с. 281-291; 307], SGAM [366, с. 110-115], модель RAMI 4.0 [270] и IIRA [270]).

IIRA предоставляет архитектуру, которую можно использовать для построения систем промышленного Интернета вещей (Industrial IoT, IIoT) независимо от поля, в то время как модель RAMI 4.0 представляет собой архитектуру, которая обеспечивает эталон для применения основных частей Индустрии 4.0 [270; 165, с. 1-28]. Исследование организационных процессов, процедур и поведения человека [199, с. 196-202; 200, с. 43-52], а также мониторинг, обеспечивают повышение устойчивости критических инфраструктур к киберфизическим угрозам [177, с. 1-4; 199, с. 196-202; 277, с. 35-42; 327, с. 313-333].

Сравнительное исследование различных системных структур помогает понять преимущества и выбрать наилучший подход для проведения исследований по моделированию угроз.

Модель эталонной архитектуры Purdue Enterprise Reference Architecture Model. Эта логическая структура состоит из пяти зон [225]: зона предприятия, DMZ, производственная зона, зона Cell/Area и зона защищенности.

Зона Предприятия состоит из двух уровней (4 и 5 уровни). Уровень 5 состоит из системы и приложений корпоративной ИТ-инфраструктуры, а уровень 4 включает бизнес-планирование и логистику [141].

Доступ к среде КФС осуществляется через DMZ вместо прямой связи между системами в зоне Предприятия и средой КФС из-за высокого уровня риска [141].

Производственная зона (уровень 3) включает производственные операции и контроль на месте и отвечает за конечный продукт. Системы и приложения уровня взаимодействуют с системами в зоне предприятия через DMZ и с системами первых двух уровней [141].

В зону Cell/Area входят сенсоры и исполнительные механизмы уровней 0-2 [141]. Базовый уровень управления 1 обрабатывает информацию и отправляет данные конечному элементу. Управляющее оборудование, такое как ПЛК, работает под управлением операционных систем конкретного производителя [141].

Уровень 2 диспетчерского управления включает производственное оборудование, такое как HMI, системы сигнализации/предупреждения и т. д. Эти системы могут связываться с уровнями 1, а также с системами производственных и корпоративных областей через DMZ [141].

Системы зон безопасности отслеживают вторжения, восстанавливают сбои процессов в безопасное состояние, если они превышают определенный порог. Эти системы обычно изолированы от других КФС [141].

PERA — это функциональная модель большинства КФС. Каждый уровень может соответствовать зоне безопасности [157, с. 4196-4203]. Кроме того, PERA четко определяет типы протоколов, которые перемещаются между разными уровнями. Ограничения PERA заключаются в сложности проведения полного моделирования процесса КФС.

Модель архитектуры «умной» сети (Smart Grid Architecture Model). Европейская комиссия по стандартизации разработала общую инфраструктуру для получения информации о текущем состоянии электрической сети под названием SGAM [199, с. 196-202; 327, с. 313-333].

SGAM — это трехмерная модель, включающая три измерения: домены,

зоны и уровни. Он состоит из пяти основных слоев:

- бизнес-слой определяет бизнес-требования, бизнес-модели и учитывает нормативные ограничения;
- функциональный слой определяет необходимые функции и услуги системы, независимо от ее физической реализации;
- информационный уровень определяет модели данных, информацию и ее использование участниками энергосистемы;
- слой коммуникаций определяет протоколы и механизмы обмена данными;
- слой компонентов описывает физическое оборудование интеллектуальной сети. Эти физические устройства могут быть либо интеллектуальными электронными устройствами (intelligent electronic devices, IED) [175, с. 133-148], либо физическим коммутационным оборудованием. Происходит взаимодействие управления энергосистемой между доменами всех уровней в иерархических зонах.

Модель эталонной архитектуры Reference Architecture Model for Industrie 4.0. RAMI 4.0 представляет собой трехмерную архитектуру, состоящую из основных уровней, уровней иерархии и уровней управления жизненным циклом [270]. Модель имеет шесть основных уровней [270]: бизнес, функция, информация, коммуникация, интеграция и актив. Ось уровня очень близка к SGAM, с небольшой модификацией уровня интеграции в RAMI 4.0. Уровни иерархии представляют собой серию международных стандартов IEC62264/IEC61512, представляющих зоны ОТ от устройств управления, станций, рабочих центров, предприятия и подключенного «мира» [270].

Согласно IEC 62890, управление жизненным циклом включает два уровня: один для поставщиков во время проектирования, разработки продукта, а другой для клиентов во время внедрения, эксплуатации и обслуживания.

После создания продуктов в производственной среде на основе RAMI 4.0 и их установки их можно эксплуатировать и обслуживать в IIoT на основе IIRA.

Эталонная архитектура промышленного Интернета (Industrial Internet Reference Architecture). IIRA, представленная Консорциумом промышленного Интернета (Industrial Internet Consortium, IIC), направлена на создание комплексной модели IoT, независимой от конкретных областей и отраслей [269, 270]. Он основан на стандарте ISO/IEC 42010.

Базовые системные функции необходимы для поддержки универсального использования IIoT и реализации общих возможностей системы IIoT для бизнес-требований. IIRA состоит из пяти функциональных доменов [269, 270]: домен управления, домен операций, информационный домен, домен приложения и домен бизнеса. Домен управления включает в себя функции, выполняемые КФС. Домен операций включает удаленный мониторинг работоспособности, настройку и обновление, диагностику и профилактическое обслуживание.

Информационная область включает в себя функции сбора, преобразования и анализа данных. Домен приложения состоит из функций для применения логики, правил и моделей для конкретных вариантов использования. Бизнес-домен включает в себя функции для интеграции информации между бизнес-системами и приложениями для планирования работы, управления взаимоотношениями с клиентами, планирования ресурсов предприятия и системы управления производством [100, с. 1-12; 131, с. 68-91; 269; 270].

IIRA также определяет набор сквозных функций, которые включают подключение, управление распределенными данными, интеграцию и промышленную аналитику для поддержки этих основных системных функций [269, 270]. Основные характеристики системы, отмеченные в IIRA, включают безопасность, защищенность, надежность, отказоустойчивость, конфиденциальность и масштабируемость.

Угрозы кибербезопасности КФС. Поскольку КФС всегда должна быть высокодоступной, получение разрешения на ее отключение для технического обслуживания, связанного с кибербезопасностью, становится чрезвычайно

сложным. Таким образом, организации должны разработать и внедрить программу безопасности для защиты своей критической инфраструктуры.

Основные проблемы информационной безопасности КФС, выявленные экспертами, включают [100, с. 1-12; 131]:

— слабая защита от несанкционированного доступа (например, пароль по умолчанию);

— использование некриптографического шифрования;

— удаленное управление;

— использование Windows в качестве основной операционной системы;

— человеческий фактор;

и т. д.

Серверы или контроллеры связи обеспечивают связь между уровнями. Доступ к сенсорам осуществляется через протоколы и полевые шины (RS485, RS232, Fieldbus, Profibus, CAN, OPC и т. д.).

Подключение компонентов КФС к Интернету подвергает систему различным сценариям кибератак.

Согласно стандарту ISO/IEC 62443, кибератаки на КФС — это

— атака на связь;

— вторжение в базу данных;

— атака с повторным воспроизведением;

— поддельная авторизация и маскировка под законного пользователя;

— социальная инженерия;

— фишинговые атаки на менеджеров, инженеров SCADA и ИТ-администраторов;

— вредоносный код;

— отказ в обслуживании (DoS);

— расширение привилегий;

— физическое повреждение.

Необходимо повысить уровень безопасности КФС из-за растущей

вероятности сбоя системы, что может привести к экономическому ущербу для предприятия.

Угрозы информационной безопасности определяются на каждом уровне КФС. Определение включает в себя выявление источников угроз и оценку потенциала нарушителей, анализ уязвимостей, определение возможных сценариев реализации угроз КФС, оценку возможных последствий их реализации.

Среди угроз безопасности КФС были выявлены:

— угрозы отключения объекта управления или отказа в обслуживании [69, с. 389–405];

— угрозы нарушения целостности КФС как на программном уровне, так и на аппаратном уровне (вирусы класса Stuxnet, поражающие программируемые логические котроллеры Siemens, Triton);

— угрозы нарушения доступности из-за некорректной работы средств защиты информации;

— угрозы несанкционированного доступа к КФС.

Киберугрозы затрагивают: 1) конфиденциальность, которая необходима для обеспечения защиты персональных данных пользователя в КФС и предотвращения попытки злоумышленника изменить состояние физической системы путем «подслушивания» каналов связи между сенсорами и контроллером, и между контроллером и исполнительным механизмом; 2) целостность, когда данные или ресурсы могут быть изменены без разрешения; 3) достигаемость (availability) при сбоях в вычислительной технике, управлении, связи, оборудовании; 4) надежность, когда необходимо подтвердить, что обе вовлеченные стороны действительно являются теми, за кого себя выдают [344, с. 146-164; 371, с. 733-738]; 5) подлинность (authenticity), когда идентичность субъекта или ресурса может быть доказана как одно утверждение; 6) неотказуемость (non-repudiation), когда можно доказать, что действия или события имели место, так что они не могут быть

опровергнуты впоследствии; 7) подотчетность (accountability), когда действия субъекта можно однозначно проследить до самого субъекта.

Одной из основных характеристик киберугроз является то, что они масштабируемы, т. е. легко автоматизируются и тиражируются, и следует ожидать, что они свободно распространяются через ненадежные домены. Киберфизические угрозы – это угрозы, которые возникают в киберпространстве, но влияют на физическое пространство системы. Киберфизические угрозы исходят из киберпространства и воздействуют на физическое пространство КФС.

Классификация угроз КФС включает [385, с. 525-539]: подмену личности, фальсификацию данных, отказ от происхождения, раскрытие информации, повышение привилегий и отказ в обслуживании (Denial of service, DoS).

В работе [384, с. 157-160] были определены ключевые проблемы безопасности КФС: 1) понимание угроз и возможных последствий атак, 2) определение уникальных свойств КФС и их отличия от традиционной безопасности информационных технологий, 3) обсуждение механизмов безопасности применимых к КФС. С другой стороны, чтобы понять новые классы угроз КФС, например, в интеллектуальной сети и системах SCADA, полезно охарактеризовать взаимодействия между областью, которая является источником угроза и площадь, которая была затронута [300, с. 238-243].

Анализ современных публикаций по данной теме показал актуальность и перспективность КФС [371, с. 733-738]. Ученые разных стран (Германия, Китай, США и др.) посвятили этой технологии тысячи публикаций, исследующих создание таких систем.

В работе [276, с. 174-216] была представлена систематическая карта исследований по обеспечению безопасности КФС. В этой работе показаны ведущие университеты в некоторых развитых странах, которые занимаются исследованиями по этой проблеме.

Для анализа последних исследований в области обеспечения

кибербезопасности КФС были выделены четыре категории исследований [51, с. 212-223]: оценка последствий кибератак, моделирование атак на КФС, обнаружение атак на КФС и разработка архитектуры безопасности.

И согласно проведенному анализу, в первую пятерку исследовательских университетов, занимающихся этой проблемой, вошли Калифорнийский университет в Беркли (США), Пекинский университет науки и технологий (Китай), Королевский технологический институт КТН (Швеция), Миланский политехнический университет (Италия) и Гамбургский технический университет (Германия).

Развитие обнаружения атак для промышленных КФС рассматривается в соответствии с различными подходами обнаружения [167, с. 1674-1683]: 1) байесовское обнаружение с бинарной гипотезой; 2) подходы с использованием взвешенных наименьших квадратов; 3) $\chi 2$-детектор на основе фильтров Калмана; и 4) методы квази-FDI (обнаружение и изоляция неисправностей). Надежность, безопасность и устойчивость, а также стабильность обсуждались для управления способностью ослаблять различные кибератаки.

Согласно [82, с. 81-97], любая модель безопасности КФС должна включать уровни защиты безопасности со следующими характеристиками: сложное проникновение; надежный механизм аутентификации и контроля доступа; высокое время отклика; возможность обновления и возможности смягчения атак.

Следовательно, различные исследовательские сообщества очень активны в направлении безопасности КФС, что подтверждает важность этой проблемы. Однако все еще остается много нерешенных вопросов. В результате выделяются следующие направления исследований по кибербезопасности КФС:

1) Оценка последствий кибератак. Сложные и изощренные атаки предназначены для нанесения значительного ущерба кибер- и физическим компонентам КФС. Например, Stuxnet [138, с. 80-91], который является первым

вредоносным программным обеспечением, специально разработанным для нанесения физического ущерба промышленным инфраструктурам (перепрограммирование систем управления путем модификации кода ПЛК). Таким образом, необходимо оценить влияние кибератак на нормальное функционирование физических процессов. В таких ситуациях крайне важно не только продемонстрировать и оценить разрушительное воздействие кибератак, но и количественно оценить последствия и, в конечном итоге, обеспечить доступность конкретных киберактивностей.

2) Моделирование киберфизических атак. Модели атак и уязвимостей используются для выявления слабых мест в системах КФС для поддержки их стратегии поиска и понимания атак. Необходимо разработать модели атак для их оценки и принятия адекватных контрмер для обеспечения безопасности КФС. Злоумышленнику необходимо понимать условия отказа оборудования, принципы управления, поведение процесса и т. д. [252, с. 213-232].

3) Обнаружение кибератак на КФС. Важно заранее разработать алгоритмы обнаружения и контрмеры для всех известных атак, чтобы снизить влияние атак на ограниченное время и минимизировать ущерб системе.

4) Разработка архитектуры безопасности. Разработка КФС ограничена факторами безопасности. Основная задача проектирования сложных архитектур КФС — тестирование и проверка «безопасного дизайна» для обеспечения безопасности и надежности физических и киберкомпонентов. Необходимо разработать новые надежные алгоритмы контроля и оценки, которые учитывают более реалистичные модели атак с точки зрения безопасности.

Классифицируя рассмотренные публикации, их можно сгруппировать по четырем категориям, связанным с безопасностью SCADA, безопасностью интеллектуальных сетей (Smart Grid), мерами противодействия кибератакам и безопасностью коммуникаций.

Безопасность SCADA систем. Проблемы информационной безопасности

часто вызваны не столько внешними атаками, сколько несоблюдением персоналом положений и правил политики информационной безопасности предприятия [51, с. 212-223]. Это может привести к несанкционированному заражению компьютерными вирусами, троянами и червями.

Обнаружение заражения в системе SCADA может вызвать необходимость жесткой перезагрузки для очистки от вируса и остановит большинство процессов предприятия, но это не всегда осуществимо с экономической точки зрения.

В работе [252, с. 213-232] предложен метод оценки последствий распространения кибератак на КФС, оценивающий прямые и косвенные последствия атак на параметры управления, включая измерения сенсоров и сигналы контроллеров. Предложенный подход рассмотрен для электростанции с кипящей водой BWPP. Сравнивается «нормальное» поведение системы без каких-либо сбоев с аномальным поведением во время кибератак (DoS и атаки с отклонением).

Параметры системы разделены на два класса причинно-следственных параметров, которые могут быть одинаковыми или отличаться друг от друга. Предложены новые показатели, которые можно использовать для количественной оценки уровня важности каждого параметра в физическом процессе. Определены приоритеты в показаниях сенсоров и сигналов управления в зависимости от их чувствительности к атакам с использованием полученных количественных значений. В отличие от большинства предложенных методов, применимых к атакам, вызывающим остановку физического процесса, предложенный метод позволяет рассматривать кибератаки, которые не обязательно приводят к выходу из строя SCADA системы.

Меры противодействия кибератакам. В связи с растущим использованием IoT и автомобильных самоорганизующихся сетей (автомобильные сети ad hoc) в ближайшем будущем, они постоянно

развиваются и привлекают все большее внимание.

Злоумышленник может скомпрометировать некоторые транспортные средства и превратить их в автомобили-зомби, ожидающие приказов от сервера управления и контроля. Исследователи обсуждают проактивные и реактивные решения, которые можно использовать в качестве контрмер против кибератак.

В работе [286, с. 199-210] был предложен подход к моделированию и оценке атак и защитных действий для КФС. В этом подходе рассматривались отказы от истощения, проникновения и эксфильтрации. Определение оптимальной модели условий в КФС, таких как интервал обнаружения вторжения и моделирование уровня избыточности были исследованы.

Безопасность интеллектуальных сетей. Разработка методов смягчения последствий скоординированных киберфизических атак на интеллектуальные сети привлекла внимание исследователей в последние годы.

Результатом стало моделирование кибератак, используя уязвимость информации, связи и электрической сети, анализ уязвимостей электрической сети с неполной информацией, используя теорию графов и т.д.

Подход теории игр к оценке безопасности интеллектуальных сетей был предложен в работе [83, с. 481-489]. Прежде всего, авторы сосредоточились на киберфизической безопасности обширной территории (мониторинг, защита и контроль с точки зрения скоординированных кибератак).

Основное внимание было уделего изучению соответствующих вопросов киберфизической безопасности WAMPAC (Wide-Area Monitoring, Protection and Control).

Безопасность коммуникаций. Атаки на КФС могут изменить производственную систему, что приведет к ухудшению коммуникаций, функциональности или снижению производительности. Уязвимости КФС при различных кибератаках, мониторинга структурного состояния КФС, оценка последствий атак на коммуникации, вычисления и управление широко исследуются.

Уязвимость инфраструктуры КФС при преднамеренных атаках была исследована в работе [126, с. 337-348]. Были проанализированы надежность сети, ориентированной на Интернет, и сети, ориентированной на КФС. Аналитические и эмпирические результаты показали, что предложенный механизм повышает надежность КФС, даже в условиях слабой локальной возможности обнаружения и хрупкой сетевой структуры.

В работе [51, с. 212-223] были рассмотрены вышеуказанные категории с точки зрения вопросов безопасности, включая подлинность, конфиденциальность, надежность, отказоустойчивость и целостность. И было показано, что цели информационной безопасности затрагивались почти во всех рассмотренных работах.

1.2. Анализ и исследование методов и алгоритмов интеллектуальной обработки больших данных киберфизических систем

При анализе данных качество информации имеет первостепенное значение. Эта задача усложняется ростом больших объемов собираемой информации. Процессы сбора данных, их предварительной обработки, анализа, обмена, хранением и визуализацией взаимосвязаны в жизненном цикле больших данных КФС. Однако каждый из этих блоков подвержен различным угрозам кибербезопасности. В связи с этим исследователи уделяют особое внимание разработке эффективных методов обнаружения аномальной работы КФС. Работа с большими данными также требует больших вычислительных ресурсов.

Высокая степень важности решаемых задач привела к тому, что в этой области появилась целая плеяда различных методов. Методы отличаются друг от друга простотой реализации, пригодностью для обработки данных и лежащими в их основе базовыми принципами. Среди них выделяются методы кластеризации [152, с. 270-293].

Технология кластеризации используется во многих областях: медицине,

археологии, информационных системах и др. Кластеризация данных также часто используется в качестве начального шага для анализа данных [34, с. 85-99].

Использование таких подходов, как Big data в структурах безопасности КФС, помогает справиться с (1) объемом информации, производимой всеми взаимосвязанными устройствами, (2) скоростью обработки информации о физических процессах в режиме реального времени, (3) и разнообразием обрабатываемых данных (например, данные сенсоров, сетевой трафик, лог-файлы) [173; 391, с. 649-658].

Одной из таких особенностей является необходимость анализа большого объема данных, разнородных по своей природе, с минимальными вычислительными затратами [246, с. 60-74].

Предложенная авторами архитектура КФС основана на обработке потоков данных и параллельных вычислениях. Рассмотрены вопросы структуры системы, а также реализации сбора, хранения, нормализации, анализа и визуализации данных.

Новая система управления дорожным движением в «умном» городе, основанная на параллельной обработке больших данных, была описана [409, с. 653-664]. Он предоставляет масштабируемую и простую в использовании структуру для эффективной обработки сложных событий.

В [295, с. 369-380] представлена структура, которая может обрабатывать массивные данные интеллектуальной сети. В качестве примеров были рассмотрены «умный» дом и сеть из более чем 6000 умных счетчиков.

В работе [208, с. 1015-1021] были перечислены проблемы, связанные с большими данными. Несколько сенсоров и исполнительных механизмов на подключенных устройствах (включая сенсоры) извлекают огромные объемы данных. К задачам управления КФС в данном случае относятся: (1) аналитика информационных потоков, (2) вычисления и обработка информации, (3) безопасность и конфиденциальность.

Были рассмотрены решения по безопасности, связанные с практическими возможностями интеграции больших данных в КФС [231, с. 376-380]. Интеграция КФС с большими данными требует обработки информационных потоков в режиме реального времени и предполагает пакетную обработку для моделирования и изучения поведения системы.

Предложен подход к построению систем мониторинга безопасности сетей КФС, основанный на реализации параллельной потоковой обработки данных о событиях безопасности [246, с. 60-74]. Он реализован на платформе Hadoop с возможностью использования технологии комплексной обработки событий.

Kim и Yu [243, с. 93-102] предложили систему эффективного ведения и администрирования пациентов в медицинском учреждении. Они применили и внедрили системы планирования ресурсов предприятия (enterprise resource planning) в малых и средних больницах для анализа информации о пациентах и данных об оборудовании.

Системы аналитики в реальном времени позволяют использовать широкий спектр технологий больших данных.

Предложена интеллектуальная и безопасная гибридная четырехуровневая модель информационной системы с использованием машинного обучения для обработки больших данных в сфере здравоохранения [240, с. 1049-1059]. Уровень безопасности данных был включен для обеспечения высокой степени конфиденциальности.

Исследователи в [396, с. 649-658] сосредоточились на анализе временных рядов в промышленной среде и представили основу для сбора больших объемов данных в реальном времени с крупномасштабных сенсорных узлов. Он также включает передачу данных, обработку и визуализацию.

1.3. Анализ и исследование методов и алгоритмов обеспечения кибербезопасности информационных технологий и защищенности операционных технологий киберфизических систем

Различные типы интеграции компонентов КФС основаны на эффективной

связности. КФС включает в себя различные комбинации ключевых функций и зависит от их приложения. КФС рассматривает вычислительные компоненты, которые используют общие знания и информацию из физических процессов.

ИТ являются одной из составляющих современных КФС. Ее архитектуру можно рассматривать на различных уровнях модели ISO/OSI [265, с. 1097-1107; 249, с. 1-6; 51, с. 212-223]: от физического уровня до уровня приложений.

Физический уровень закладывает основу для архитектуры КФС. Физический уровень состоит из сенсоров, исполнительных механизмов, которые связаны друг с другом через беспроводные или проводные сети. Например, 2G/3G/4G, Wi-Fi, ZigBee, Bluetooth, WiMAX, считыватели и метки RFID, а также проводные технологии (PLC, NC и т. д.). 6LoWPAN (IPv6 по маломощным беспроводным персональным сетям, (IPv6 over Low power Wireless Personal Area Networks) — это протокол сетевого уровня, который может использоваться на любом физическом уровне и уровне канала передачи данных.

Этот уровень используется для подключения ZigBee, Bluetooth и других систем к Интернету (действует как маршрутизатор). У устройств этого уровня обычно мало памяти и вычислительной мощности. Атаки на этот слой в основном исходят из внешних источников.

Канальный уровень. Канальный уровень обеспечивает создание, передачу и прием кадров данных [51, с. 212-223]. Этот уровень обслуживает запросы сетевого уровня и использует службу физического уровня для приема и отправки пакетов. Канальный уровень делится на подуровень управления логическим каналом (logical channel management, LLC) и подуровень управления доступом к среде (media access control, MAC). LLC предоставляет услуги сетевого уровня, а подуровень MAC регулирует доступ к общей физической среде. Атака на этот уровень может привести к нарушению MAC-адресов, что может привести к сбою идентификации устройства.

Сетевой уровень. На этом уровне пакеты маршрутизируются на основе

преобразования MAC-адресов в сетевые адреса. Он использует протоколы IPv4/IPv6 и RPL (маршрутизация Ripple). Атаки, приводящие к выходу из строя сенсоров и исполнительных механизмов, в свою очередь, приводят к смене информации и источника, из которого она была получена. Впоследствии это может привести к механической поломке.

Транспортный уровень. На этом уровне пакеты разбиваются на мелкие фрагменты. Наиболее распространенные протоколы транспортного уровня включают TCP, UDP и ICMP. Атаки на этом уровне приводят к снижению скорости работы сетевого оборудования и отказу сервисов.

Сеансовый уровень управляет диалогом (сеансом связи). Отслеживает порядок передачи сообщений по сети; в случае сбоя, чтобы не начинать заново, вставляет метки в длинные сообщения. Протоколы сеансового уровня обычно являются неотъемлемой частью функций трех верхних уровней модели.

Уровень представления координирует представление данных (синтаксис) при взаимодействии двух прикладных процессов: преобразование данных из внешнего формата во внутренний; шифрование и дешифрование данных. Примером такого протокола является протокол Secure Socket Layer (SSL), который обеспечивает обмен секретными сообщениями для протоколов прикладного уровня стека TCP/IP [51, с. 212-223].

Прикладной уровень охватывает различные домены. Этот уровень хранит, анализирует и обновляет информацию, полученную от предыдущих уровней. Он принимает управляющие решения, которые можно визуализировать с помощью интерфейса виртуального прототипа. Защита конфиденциальности данных является наиболее важным вопросом этого уровня.

Архитектура помогает определить и объяснить общую структуру ИТ КФС, описать взаимодействие ее компонентов. Безопасность должна выполняться на всех уровнях ее архитектуры, от физического уровня до уровня приложений.

Более высокий уровень безопасности снижает риск раскрытия конфиденциальной информации, обеспечивает анонимность данных и скрывает

важные информационные детали. Безопасность КФС защищает систему от вторжений и снижает вероятность рисков безопасности.

Дерево атак на КФС [51, с. 215]. Согласно стандарту ISO/IEC 27001:2013 угрозы могут быть преднамеренными, случайными или связанными с окружающей средой. Примеры типичных угроз включают в себя: физический ущерб, стихийные бедствия, потерю основных услуг, радиационные сбои, компрометацию информации (например, прослушивание, вмешательство в программное обеспечение и т. д.), технические сбои, несанкционированные действия (например, повреждение данных), компрометацию функций (например, подделка и злоупотребление правами).

По результатам анализа существующих исследований в области безопасности на рис. 1.3.1 предлагается «дерево атак» и угроз на основе функциональной модели КФС [204, с. 39-50; 14, с. 517-528].



Рис. 1.3.1. Древовидная диаграмма атак и угроз кибербезопасности КФС [51, с. 215]

Ветви «дерева» включают следующие виды атак: а) атаки на сенсорные устройства (Sensing); б) атаки на исполнительные механизмы (Actuation); в) атаки на вычислительные компоненты (Computing); г) атаки на коммуникации (Communication); д) атаки на обратную связь (Feedback).

а) Исследователи выявили угрозы и уязвимости, влияющие на сенсоры КФС (такие как «Ввод ложных сигналов радара», «Ослепление камер светом», «Спуфинг GPS» и т. д.) [114]. Поскольку КФС тесно связаны с физическим процессом, в который они встроены, необходимо обеспечить надежность и точность процесса сбора данных. Сенсорная безопасность нуждается в методах поощрения физической аутентификации, чтобы можно было доверять любым данным, полученным от физического процесса [253, с. 133-144].

б) Djouadi и соавторы [169, с. 3659-3664] проанализировали влияние кибератак на актуаторы и рассмотрели два класса, которые охватывают широкий спектр потенциальных атак: атака конечной энергии, включающая, например, потерю и модификацию личных пакетов, атака конечного времени и атака конечного времени.

Импульсные атаки и ограниченная атака, приводящая к подавлению управляющего сигнала. Безопасность управления срабатыванием относится к тому факту, что во время пассивно-активного или активного режима работы никакие действия не могут выполняться без соответствующего разрешения. Спецификация разрешений должна быть динамической, так как требования КФС со временем меняются.

с) Атаки на вычислительные ресурсы были подробно рассмотрены и включают в себя трояны, вирусы, черви и DoS-атаки [346, с. 51-60]. В работе [286, с. 199-210] представлена информация о методах интеллектуального анализа данных (data mining, DM), которые можно использовать для повышения кибербезопасности.

Вредоносная атака может тайно повредить КФС. Поскольку в системах управления имеют место нарушения и ошибки измерений, механизмы

обнаружения должны обеспечивать, чтобы эти регулярные ошибки не вызывали ложной тревоги. Это дает злоумышленнику возможность спрятаться.

Если злоумышленник изменяет реальные данные путем получения ключа для защищенной связи (коммуникационного ключа) или захвата некоторых сетевых устройств, это называется атакой на целостность [73, с. 94-110].
Безопасность хранения включает в себя разработку решений для обеспечения безопасности хранимых данных на платформах КФС от физического или кибервзлома. С точки зрения атакующего построение стратегии ложных атак, как правило, связано с рядом факторов, ресурсов и ограничений безопасности.

г) Атаки на коммуникации атаки включают выборочную пересылку (Selective Forwarding), подмену «пакетов» (Packet Spoofing), повторное воспроизведение «пакетов» (Packet Replaying), атаки Sybil и т. д. (могут использоваться для нарушения распределения ресурсов между узлами в пользу вредоносных программ), которые нарушают маршрутизацию системных пакетов [127, с. 2041-2051; 253, с. 133-144; 94, с. 54-59]. Любое вмешательство в данные может привести к ошибкам в будущих требованиях к их обработке. Если злоумышленник может захватывать и пересылать только настоящие пакеты данных, то эффективным методом атаки является запись некоторых «обычных» данных и их воспроизведение, чтобы избежать обнаружения. Безопасность связи требует разработки протоколов для обеспечения связи между источниками помех (активными) и подслушивающими устройствами (пассивными) между КФС и внутри нее.

д) В работе [204, с. 39-50] предложены трехуровневая логическая модель КФС и метамодель кибератак, где система подвергается атаке типа Feedback Integrity Attack (при этом лишь часть управляющих сигналов сохраняет свою целостность).

Безопасность обратной связи относится к тому факту, что системы управления в КФС, которые обеспечивают необходимую обратную связь для срабатывания, защищены. Современные решения безопасности ориентированы

только на безопасность данных, но их влияние на алгоритмы оценки и управления необходимо изучить, чтобы обеспечить всестороннюю защиту КФС [119, с. 1-6].

По мере развития технологий многие организации интегрировали традиционные сети ОТ с сетями ИТ. Этот подход направлен на снижение затрат на техническое обслуживание и интеграцию сбора и анализа данных [254, с. 38-42]. Однако эта интеграция сопряжена с высокими рисками безопасности по следующим причинам: (1) ОТ-сети имеют другие операционные приоритеты, чем ИТ-сети (например, доступность или конфиденциальность), и одна модель может не подходить обеим; (2) большинство устройств и протоколов КФС не предназначены для поддержки функций безопасности, таких как шифрование данных или контроль доступа, и часто поддерживают удаленный доступ через радиомодемы; (3) дорогостоящие устаревшие устройства КФС предоставляют ограниченные возможности визуализации для внедрения и оценки модификаций безопасности, и (4) важные бизнес-операции в режиме реального времени в ОТ, наряду с правилами безопасности, предотвращают немедленный ремонт, который может потребовать прерывания работы системы.

В свете вышеизложенного эксперты по безопасности предложили некоторые обходные пути для ограничения доступа пользователей к сети ОТ. Другие усилия по обеспечению безопасности КФС включают постоянное совершенствование стандартов, протоколов и устройств для поддержки функций безопасности.

Наконец, характеристики КФС отличаются от традиционных ИТ-сред [228, с. 1-17]. Известно, что операторы этих систем отдают предпочтение доступности, а не безопасности.

Как правило, КФС имеют более длительный жизненный цикл и меньшее количество обновлений, которые подвергают компоненты SCADA известным уязвимостям в течение более длительных периодов времени.

Аномальное состояние КФС может быть вызвано неисправными

компонентами, временными отказами, неправильной настройкой, кибератаками или их сочетанием [257, с. 49-51; 36, с. 14-22]. Злоумышленник вмешивается в КФС, чтобы манипулировать показаниями датчиков или исполнительных механизмов, что приводит к нештатной работе системы.

Обнаружение аномалий в промышленном сценарии имеет важное значение, поскольку необнаруженные сбои могут привести к критическим повреждениям. Раннее обнаружение аномалий может повысить надежность подверженного сбоям промышленного оборудования и снизить затраты на эксплуатацию и техническое обслуживание [69, с. 389-405].

Необходимо определить нормальное состояние КФС и понять логику управления процессом для выявления аномалий. Но сложность и изощренность кибератак, призванных нанести значительный ущерб КФС, растет с каждым годом.

Кибератака может заключаться во внедрении вируса в КФС с использованием хакерских технологий для проникновения в систему [350, с. 166-178]. Злоумышленник затем может обмануть человеко-машинный интерфейс системы, отображая нормальное состояние [191, с. 729-736; 339, с. 1040-1049]. Он изменяет показания сенсоров или исполнительных механизмов, что может нарушить работу КФС. Все это может привести к плачевным последствиям.

К основным кибератакам на КФС, наблюдаемым в реальных инцидентах, относятся атаки на SCADA и протоколы, несанкционированный доступ к проприетарному ПО, удаленная смена пароля, отказ в обслуживании (DoS) и другие [184, с. 1454-1467; 315, с. 2715-2729; 348, с. 90-95; 356, с. 5979-5984; 362, с. 55-64].

DoS-атаки включают различные риски, такие как прерывание бизнес-операций, финансовые потери, и даже ставят под угрозу общественную безопасность. В результате таких атак сервер или сеть передачи данных ограничивает услуги, а это, в свою очередь, может привести к отказу КФС.

Развитие Индустрии 4.0 привело к появлению новых технологий эффективного и надежного мониторинга таких систем.

Таким образом, в состав современных КФС входят устройства, формирующие мультисенсорную конфигурацию. Эти системы упрощают процесс сбора данных, что приводит к доступности больших наборов данных. Следовательно, наблюдается рост развития методов интеллектуального анализа данных для обнаружения аномалий [37, с. 101-113; 17, с. 38-40].

Классический подход к таким проблемам обычно включает методы обработки сигналов, которые извлекают полезную информацию из данных сенсоров. В последние годы для обнаружения механических неисправностей стали применяться глубокие нейронные сети. Они нашли применение в выделении признаков сигналов, решении проблем несбалансированных данных и в качестве методов классификации [117, с. 246-260; 211]. Таким образом, был предложен метод обучения признаков на основе конволюционной нейронной сети (convolutional neural network, CNN) для обнаружения ошибок [230, с. 331-345]. Метод нашел применение в системах диагностики неисправностей на основе изображений [130, с. 1-13; 402, с. 213-237].

Глубокая нейронная сеть (deep neural network, DNN) также нашла применение в решении проблемы дисбаланса данных при диагностике механических отказов.

Чтобы увеличить количество моделей отказа оборудования в несбалансированном наборе данных, был предложен подход, основанный на генеративно-состязательной сети Вассерштейна в сочетании с разреженным автоэнкодером [373, с. 111168-111180].

В [130, с. 1-13] глубокая CNN (deep CNN, DCNN) была принята для извлечения признаков из звуковых сигналов сенсоров. Предлагаемый метод сравнивался с машинами опорных векторов (support vector machines, SVM), наивным байесовским классификатором (naïve Bayes, NB) и нейронной сетью с обратным распространением (back-propagation). Частотно-временные данные в

сочетании с DCNN показали наилучшие результаты.

Был предложен метод диагностики неисправностей с использованием DCNN и обучения радиочастотного ансамбля на основе двумерных изображений в градациях серого, полученных с помощью непрерывного вейвлет преобразования (continuous wavelet transform, CWT) [386, с. 1-21]. Предложенный метод сравнивали с BP, SVM, глубокой сетью доверия (deep belief network, DBN), автоэнкодером с шумоподавлением (Denoising Autoencoder) и CNN. Параллельный ансамбль на основе DNN и CNN, где признаки во временной области (глобальные признаки) извлекаются с помощью DNN, объединяются с локальными признаками, извлеченными CNN из вибрационных сигналов был предложен в [264, с. 1-18].

Учитывая вышеизложенное, возникает необходимость разработки нового подхода к выявлению аномалий в КФС, вызванных кибератаками. Это позволило бы оценить ее устойчивость к кибератакам на разных уровнях реализации.

Вопросы исследований в области обеспечения кибербезопасности ИТ и защищенности ОТ КФС. В настоящее время доступно множество передовых технологий кибербезопасности для КФС систем. Однако исследования и разработки новых методов продолжаются, и некоторые из них описаны ниже.

Анализ рисков необходим для устранения новых угроз для КФС. В настоящее время существует несколько работ по анализу существующих методов оценки рисков.

Среди перспективных направлений исследований выделяются безопасность компонентов, общесистемные оценки, моделирование взаимозависимости, оценки рисков в реальном времени и эффективность средств контроля безопасности.

Текущие стратегии, такие как моделирование для определения эффективности системы против кибератак, были проанализированы. Тестирование на проникновение также является перспективным методом,

который может помочь выявить скрытые уязвимости КФС. Однако исследования показывают, что применение методов на основе ИИ поможет с требуемым анализом безопасности.

Для оценки воздействия или потери информационных активов можно использовать следующие методы. Описаны структура безопасности и передовые инструменты для надлежащего управления уязвимостями и своевременного реагирования на угрозы КФС [33, с. 72-82]. Архитектура ATENA включает современные инструменты для оценки и снижения критичности компонентов КФС и зависимости от критических инфраструктур [331, с. 50-67]. Она ориентирована на управление эффективностью физического потока при повышении устойчивости взаимосвязанных КФС к киберфизическим атакам.

Для выявления рисков в КФС необходимо разработать систему раннего оповещения. Наиболее важными индикаторами риска являются количество попыток взлома, сбоев системы, временные интервалы между сбоями и т. д. Создание базы знаний с их использованием поможет решить проблемы кибербезопасности путем устранения рисков. В работе [44, с. 230-236] был предложен метод оценки рисков в КФС, который определяет возможные пути атак на критические компоненты системы. Этот подход учитывает взаимодействие между операторами ИТ и ОТ и помогает выявлять сложные и запутанные пути атак на систему.

Обнаружение вторжений. Любая мелкомасштабная атака может привести к серьезным потерям, поскольку в КФС задействован широкий спектр различных подсистем и устройств. Таким образом, механизм обнаружения вторжений особенно важен [102, с. 233-239; 352, с. 157-173].

Необходимо разработать систему обнаружения вторжений, которая обнаруживает атаки в киберпространстве, но также учитывает их отражение в физическом пространстве.

Также разработка методов стандартизации данных о сетевых потоках

поможет в своевременном обнаружении аномалий. Тестирование на проникновение или этичный взлом используется для поиска уязвимостей безопасности во многих исследованиях. В связи с этим разработка новых методов обнаружения кибератак нулевого дня необходима, поскольку большинство современных подходов сосредоточены на выявлении уже известных уязвимостей. Это реализуется с помощью стратегической политики реагирования на оповещения APT (advanced persistent threat). Моделирование сценариев атак нулевого дня позволило бы оценить защиту от атак уклонения с использованием нескольких детекторов противника и улучшить скорость обнаружения атак.

Предиктивная аналитика на основе искусственного интеллекта. В настоящее время разработаны различные методы предиктивной аналитики, но они либо не автоматизированы, либо имеют высокий уровень ложных срабатываний [331, с. 50-67].

Предиктивная аналитика позволяет автоматизировать принятие решений по таким задачам кибербезопасности КФС, как прогнозирование вредоносного ПО, классификация уязвимостей, прогнозирование путей атак, разработка системы защиты от спама и т. д.

Вредоносное ПО — это тип кибероружия, обладающий следующими характеристиками [297, с. 1-12]:

• Атаки нулевого дня;

• Маскировка и самоуничтожение;

• Сложность кода;

• Обновление, поддержка и управление инфраструктурой;

и т. д.

Так в работе [237, с. 1-15] был предложен подход к анализу и классификации вредоносного ПО КФС с использованием текстурных признаков. Двоичный файл преобразуется в изображение в оттенках серого. Для извлеченного изображения вычисляется матрица совпадений в оттенках

серого. Метод основан на текстурных признаках изображения Харалика и классификаторах машинного обучения.

Классификация уязвимостей и прогнозирование пути атаки. В работе [324, с. 1-12] была исследована оценка уязвимостей КФС при кибератаках. В качестве индексов уязвимости используются электрическая промежуточность и напряжение. Оценка уязвимости узла информационной системы электропитания основана на количественной формуле и комплексном индексе.

Оптимальный алгоритм прогнозирования пути атаки на основе двойного обучения с подкреплением был разработан в [263, с. 1-15]. Ранее предложенные методы основаны на предположении монотонности, в отличие от предложенного.

Предполагается, что компоненты системы имеют несколько уязвимостей и могут обнаруживать возможные источники и пути атаки в КФС. Хотя все вышеперечисленные подходы основаны на искусственном интеллекте, у каждого есть свои ограничения. Наиболее перспективными являются нейронные графовые сети на основе времени, глубокое байесовское прогнозирование, глубокое генеративное моделирование с временными ограничениями, обнаружение всплесков и т. д. Разработка стратегий на основе отраслевых политик, задач и потребностей повысит точность прогнозирования и кибербезопасности.

Автоматизированное исправление и смягчение. Основываясь на аналитических данных, предприятия могут внедрять и регулировать безопасное управление, которое включает моделирование угроз, автоматизированное исправление и смягчение, автоматизированное исправление и сегментацию сети. Это поддерживают различные современные методы машинного обучения, такие как продвинутые агенты, отобранные защитные механизмы, методы состязательного обучения, обучение с подкреплением, байесовские сети и метод «актора-критика».

В работе [388, с. 1748-1771] была смоделирована приоритизация

исправлений уязвимостей в сложных и взаимозависимых КФС системах. Для решения последовательности исправлений в КФС был разработан трехэтапный метод SmartPatch.

Метод включает в себя теорию игр, системное моделирование и управление рисками. Его эффективность подтверждена на основе IEEE 5-Bus. Разработка таких моделей может помочь экспертам предсказать действия кибератаки, выбрать подходящую киберзащиту и максимально использовать защиту.

КФС используются для автоматизации действий на многих объектах, таких как нефтехимические и водоочистные заводы, нефтеперерабатывающие заводы, транспорт, здравоохранение и т. д.

Производственные активы и КФС уязвимы для потенциальных злоумышленников. Киберпреступники склонны нацеливаться на системы, в которых хранятся данные о цепочке поставок и личные данные клиентов, платежная информация и другая интеллектуальная собственность. Эти проблемы связаны с устаревшим программным обеспечением и открытыми ИТ-системами.

1.4. Анализ и исследование методов и алгоритмов оценки киберустойчивости киберфизических систем

Применение информационных технологий к КФС и их доступность в Интернете делает их привлекательной мишенью для деструктивных кибератак злоумышленников [184, с. 1454-1467; 315, с. 2715-2729].

Наиболее важной задачей является сохранение способности КФС корректно функционировать даже в условиях деструктивных информационных воздействий, поскольку успешное осуществление кибератак на такие системы может привести к негативным финансовым последствиям и экологическим катастрофам, а также к гибели людей.

В нескольких исследованиях за последние два десятилетия были предложены различные определения киберустойчивости.

Устойчивость – это многомерное свойство системы, требующее

управления нарушениями, возникающими в результате отказов физических компонентов, сбоев в работе киберкомпонентов и кибератак [79, с. 1060-1069].

Устойчивость КФС – это защита критически важной функциональности системы при воздействии возмущений и восстановление после сбоев [329, с. 38-47].

Котт и Линков [247, с. 200-202] определяют киберустойчивость как «способность готовиться и планировать, поглощать, восстанавливаться и более успешно адаптироваться к неблагоприятным событиям».

Wei и Ji [375, с. 15-22] выделили следующие характеристики киберустойчивости КФС: минимизация нежелательных инцидентов, возможность восстановления нормальной работы в течение короткого времени.

Согласно [406, с. 1-6; 407, с. 1053-1082; 408, с. 394-403], кривые устойчивости дают возможность одновременно сравнивать устойчивость нескольких сценариев бедствий в условиях неопределенности, чтобы они могли надлежащим образом отображать динамические изменения, связанные с аварийным восстановлением. Этот подход основан на идее измерения площади под такой кривой.

На рис. 1.4.1 представлен пример кривой устойчивости КФС системы, которая показывает, как меняется производительность системы с течением времени в случае ухудшения ее состояния [313, с. 58-66]. Различные состояния устойчивости характеризуются следующим образом: устойчивое состояние; ухудшение состояния в случае незапланированного события; состояние восстановления и состояние после восстановления.

Здесь R_0 представляет начальное значение устойчивости, R_{pe} - значение устойчивости после успешно завершенной кибератаки, R_{pr} - значение устойчивости после подавления кибератаки, t_e - время начала кибератаки, t_{pe} - время окончания кибератаки, t_r - время начала защиты от кибератаки, t_{pr} - время окончания защиты от кибератаки и t_{ir} - время начала восстановления инфраструктуры.

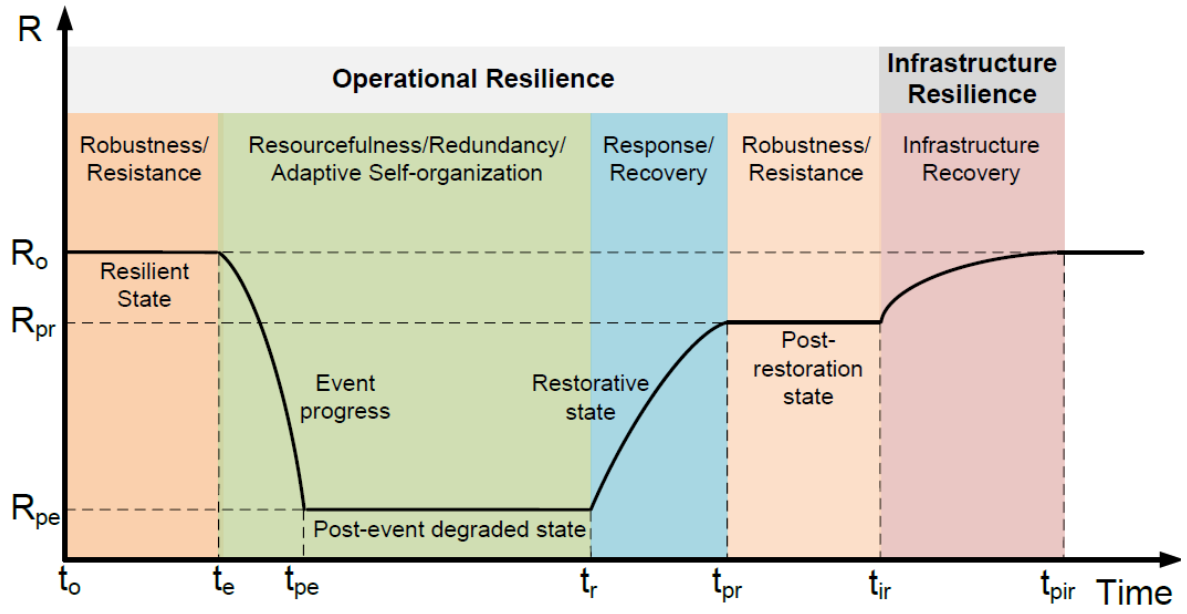


Рис. 1.4.1. Кривая киберустойчивости [313, с. 60]

Поскольку злоумышленники часто ищут и нацеливаются на самое слабое звено – наиболее уязвимый функциональный компонент в КФС. Самые уязвимые устройства могут стать хорошей отправной точкой для эффективного исследования уязвимостей системы безопасности.

Не существует универсального решения, которое адекватно отвечало бы процессам управления исправлениями в сетях ИТ и ОТ. Существуют некоторые различия в реализации исправлений в системах информационных технологий и промышленных системах управления.

Существующие методы не могут всесторонне анализировать распространенные многоэтапные кибератаки в среде КФС.

Граф уязвимостей – многообещающий метод, который перечисляет все возможные пути кибератак с помощью серии эксплойтов. Следовательно, исследователи и практики проявляют интерес к анализу киберустойчивости КФС на основе графов.

В настоящее время предложено множество подходов к анализу защищенности КФС, однако мало внимания уделяется корреляции между устройствами и уязвимостями, особенно при выявлении уязвимостей нулевого дня [81, с. 1-16; 181; 140, с. 41-53; 404, с. 1-5].

Мощным инструментом для оценки киберустойчивости КФС является байесовский граф атак (Bayesian attack graph, BAG). Он предоставляет информацию о связях между уязвимостями, что является существенным фактором.

Графы уязвимостей зависят от операционных систем и сетевых протоколов, правил управления доступом к КФС, идентификации программного обеспечения и информации об уязвимостях [287, с. 273–294]. Уязвимость определяется количественно с использованием Общей системы оценки уязвимостей (Common Vulnerability Scoring System, CVSS) и оценок Национальной базы данных уязвимостей (NVD). Среду CVSS можно комбинировать с BAG для обеспечения показателей безопасности в отношении конфиденциальности, целостности и доступности [172, с. 1230-1237].

Большинство современных работ сосредоточено в основном на теоретических исследованиях и содержит лишь краткую оценку киберустойчивости КФС на основе методов машинного обучения [20, с. 128-131].

Киберугрозы против КФС и их смягчение являются важными задачами, которые необходимо учитывать. Широко изучаемая область исследований – это попытка сделать КФС более устойчивым к кибератакам.

Существует множество исследований, в которых предлагаются защитные методы для предотвращения атаки или обнаружения злонамеренного поведения.

Целью киберустойчивости является всеобъемлющая защита всей КФС с охватом всех доступных киберресурсов.

1.5. Постановка научной задачи исследования

Необходимо разработать методы и алгоритмы обеспечения киберустойчивости и информационной безопасности киберфизических систем (КФС) на основе интеллектуального анализа больших данных о кибератаках и отказах систем. Для этого необходимо разработать методы и алгоритмы

интеллектуальной обработки больших данных КФС, метод обнаружения кибератак на ИТ КФС, алгоритм обнаружения вредоносного ПО в ИТ КФС, метод обнаружения аномалий и отказов в ОТ КФС, метод определения критичности функциональных уязвимостей компонентов КФС и метод оценки рисков КФС.

Входными данными для предлагаемых методов и алгоритмов являются большие ОТ- и ИТ-данные от сенсоров и исполнительных механизмов, которые объединяются и передаются для осуществления их аналитики и обработки. Интеллектуальный анализ больших данных предоставляет возможность потоковой и пакетной обработки. Это приводит к минимуму действия человека в производственном процессе, а экспертная оценка данных (выбор информативных признаков, параллельная обработка, обнаружение выбросов, кластерный анализ и оптимизация процесса кластеризации) может проводиться в режиме реального времени небольшой группой экспертов или с применением технологий машинного обучения.

Компоненты оценки киберустойчивости КФС включает диагностику функционирования устройств и оценку киберрисков. Данные компоненты отслеживают состояние топологии ИТ и обнаруживают инциденты при подключении любых неавторизованных сетевых устройств, анализируют сетевой трафик, который передается с использованием протоколов, определяющих аномальную активность, обнаруживают сетевые атаки, запускаемые из внутренних источников (например, вредоносное ПО на компьютере удаленного пользователя).

Предположим, что существует множество устройств КФС (сенсоров и исполнительных механизмов):

$$\{\boldsymbol{D_1}, \boldsymbol{D_2}, \dots, \boldsymbol{D_m}\} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_m}\} \subset \boldsymbol{X}. \quad (1.5.1)$$

Также имеется множество меток классов состояний кибербезопасности $\{C_0, C_1\} \subset C$. Требуется построить метод M, способный соотнести элементы множества X с одним из известных классов C_0 (множество меток классов

«нормального» состояния КФС) или C_1 (множество меток классов вызванных кибератаками):

$$M: X \to C. \quad (1.5.2)$$

Оценивание киберустойчивости и информационной безопасности КФС при различных кибератаках с применением методов машинного обучения и искусственного интеллекта можно свести к задаче выявления значений, относящихся к классу $\boldsymbol{C_1}$.

Особый акцент в данном научном исследовании состоит в разработке методов и алгоритмов повышения точности, полноты и F-меры при оценивании киберустойчивости и выявлении рисков КФС с целью защиты от потенциальных киберугроз.

1.6. Концептуальная модель обеспечения киберустойчивости киберфизических систем

Предлагаемая концептуальная модель – это иерархическая архитектура, которая состоит из трех основных уровней: уровень краевых вычислений, уровень туманных вычислений и уровень облачных вычислений (Рис. 1.6.1).

Уровень краевых вычислений включает в себя такие интеллектуальные устройства как сенсоры и исполнительные механизмы, которые используются для сбора данных, их предварительной обработки. Этот уровень временно хранит данные и может генерировать первоначальные решения об обнаружении аномального состояния.

На уровне краевых вычислений также выполняется аутентификация кибербезопасности на основе протоколов. Несмотря на то, что данный уровень выполняет предварительную обработку данных сенсоров и исполнительных механизмов в режиме реального времени, необходима их дополнительная обработка для уточнения полученных результатов и последующего принятия решений. С этой целью полученные данные передаются на уровень туманных вычислений.

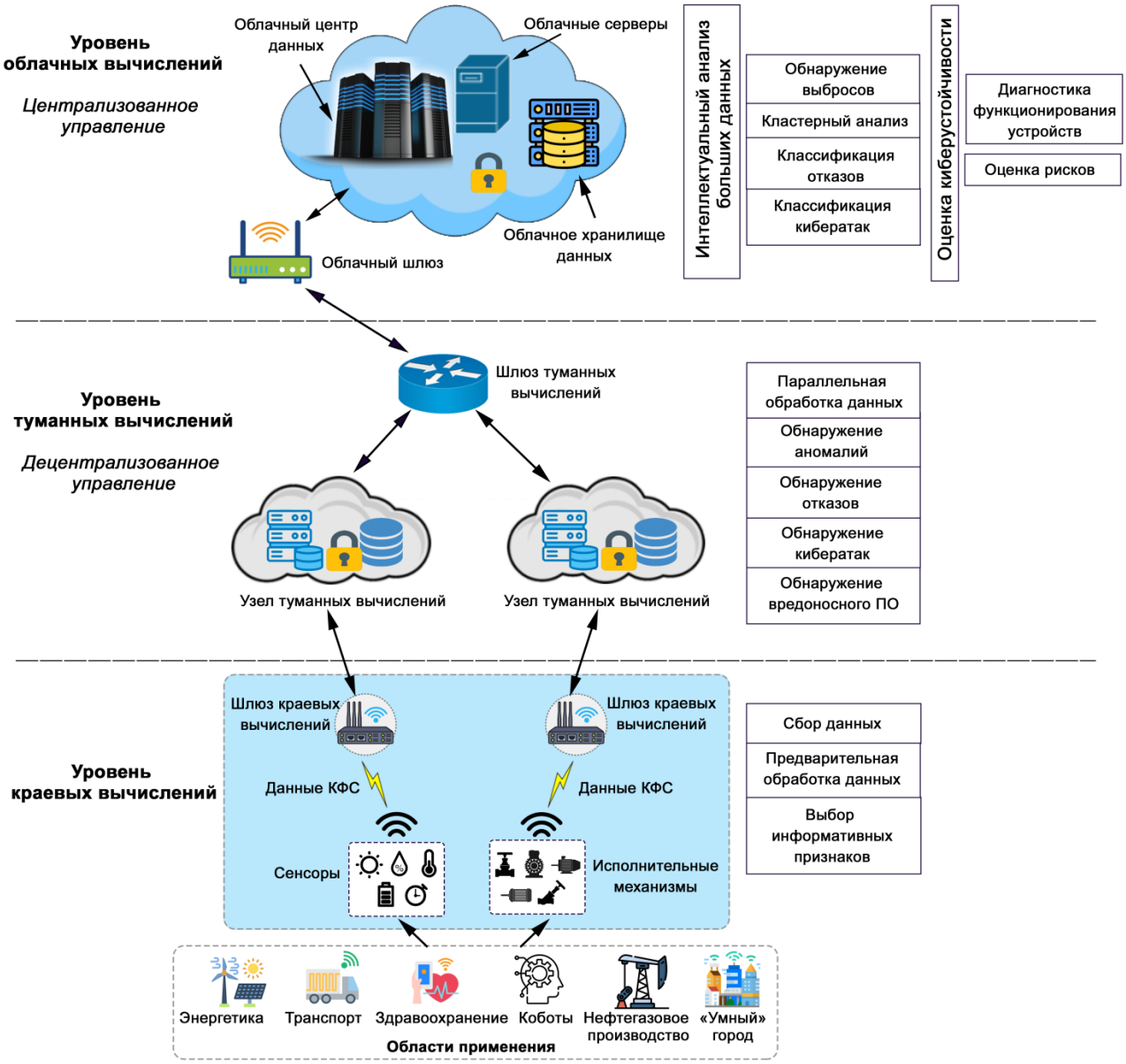


Рис. 1.6.1. Концептуальная модель обеспечения киберустойчивости КФС

Уровень туманных вычислений состоит из большого количества распределенных узлов, содержащих шлюзы, сервисы туманных вычислений, маршрутизаторы, точки доступа, базовые станции и коммутаторы. Туманные вычисления предоставляют децентрализованную платформу, направленную на обнаружение аномальной работы КФС, отказов системы, вызванных кибератаками и обнаружение вредоносного ПО, и расширяют облачные сервисы до обработки и анализа границ сети.

Уровень туманных вычислений расположен на границе сети, где узлы туманных вычислений размещены на конечных устройствах и облачных слоях. Это позволяет снизить вычислительную сложность и потребление энергии, обеспечить использование меньшего количества ресурсов и высокую эффективную мощность. При этом вычисления могут быть статичными, а также динамичными (мобильными).

Уровень облачных вычислений состоит из нескольких серверов, хранилища данных, облачного центра данных и шлюзов. Данный уровень производит высокопроизводительные вычисления и предоставляет услуги для таких областей как энергетика, транспорт, здравоохранение, коботы, нефтегазовое производство, «умный» город и т.д.

Уровень облачных вычислений осуществляет централизованное управление, обеспечивает постоянное хранение больших объемов данных и осуществляет их интеллектуальный анализ, в том числе оценку киберустойчивости КФС.

Интеллектуальный анализ больших данных предоставляет возможность потоковой и пакетной обработки. Это приводит к минимуму действия человека в производственном процессе, а экспертная оценка данных (параллельная обработка, обнаружение выбросов, кластерный анализ, классификация отказов и кибератак) может проводиться в режиме реального времени небольшой группой экспертов или с применением технологий машинного обучения. Устройства облачных вычислений подключены к узлам туманных вычислений с помощью проводного соединения и беспроводных сред (Bluetooth, 5G, ZigBee, WiFi и беспроводной локальной сети).

Уровень облачных вычислений производит обучение больших и сложных моделей на основе глубоких нейронных сетей, которые могут передаваться на узлы краевых и туманных вычислений по запросу. Это позволяет снизить потребление вычислительных ресурсов на уровнях туманных и краевых вычислений.

Предлагаемая архитектура может предоставить необходимое руководство исследователям, промышленникам и специалистам при анализе каждого слоя архитектуры КФС с точки зрения моделирования системы, управления, мониторинга, сбора данных и их анализа для достижения требуемой надежности, кибербезопасности, киберустойчивости и эффективности.

Выводы к главе I

- КФС системы являются перспективной парадигмой для разработки современных и будущих инженерных систем и, как ожидается, окажут важное влияние на реальный мир. Идея КФС фокусируется на проектировании сложных систем, а не кибер- или физической системы по отдельности;
- Исследованы особенности становления и развития КФС;
- Проведен анализ методов и алгоритмов интеллектуальной обработки больших данных КФС;
- Проведен анализ методов и алгоритмов обеспечения кибербезопасности информационных технологий и защищенности операционных технологий КФС. Влияние киберугроз на киберустойчивость КФС было отражено в виде дерева атак и угроз на сенсорные устройства, исполнительные механизмы, вычислительные компоненты, коммуникации и обратную связь;
- Проведен анализ методов и алгоритмов оценки киберустойчивости КФС;
- Представлена постановка научной задачи исследования;
- Предложена трехуровневая концептуальная модель обеспечения киберустойчивости КФС.

ГЛАВА II. РАЗРАБОТКА МЕТОДОВ И АЛГОРИТМОВ ИНТЕЛЛЕКТУАЛЬНОЙ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ ДЛЯ ОБЕСПЕЧЕНИЯ КИБЕРУСТОЙЧИВОСТИ КИБЕРФИЗИЧЕСКИХ СИСТЕМ

При анализе данных качество информации имеет первостепенное значение. Эта задача усложняется ростом больших объемов собираемой информации. Работа с бо74льшими данными требует больших вычислительных ресурсов. В связи с этим исследователи уделяют особое внимание разработке эффективных методов и алгоритмов интеллектуального анализа больших данных.

Высокая степень важности решаемых задач привела к тому, что в этой области появилась целая плеяда различных методов [20, с. 128-131]. Методы отличаются друг от друга простотой реализации, пригодностью для обработки данных и лежащими в их основе базовыми принципами.

В этой главе будут описаны различные методы интеллектуального анализа больших данных КФС и предложенные новые методы и алгоритмы.

2.1. Алгоритм обнаружения выбросов в больших данных на основе метода к-средних для снижения риска блокировки киберфизических систем

В настоящее время обнаружение аномалий является одной из важных проблем во многих областях. Быстрый рост объемов данных требует наличия инструмента для обработки и анализа данных КФС самых разных типов. Методы обнаружения аномалий предназначены для обнаружения отклонений объекта от нормального поведения. Однако сложно выбрать один инструмент для всех типов аномалий из-за возрастающей вычислительной сложности и характера данных КФС.

Рост объемов собираемой информации, разработка средств и методов ее надежного хранения приводят к повышению актуальности разработки новых методов и алгоритмов анализа больших и сверхбольших наборов данных. В

современной науке и технике появилось отдельное направление, связанное с анализом больших и сверхбольших наборов данных, известное как «Большие данные» (Big Data).

Термин «Большие данные» применяется к информации, которую невозможно обработать или проанализировать с помощью традиционных процессов или инструментов.

Сегодня организации все чаще сталкиваются со все большими проблемами, связанными с большими объемами данных. Они имеют доступ к огромному объему информации, но не знают, как извлечь из этих данных пользу, поскольку они находятся в сыром виде или в неструктурированном формате.

В результате организации даже не знают, есть ли в данных какая-либо ценная информация и стоит ли ее хранить. Вопросы данных включают в себя сбор, хранение, извлечение, передачу, анализ и визуализацию. Все эти факторы в совокупности превращаются в настоящую информационную проблему поколения.

Анализ таких объемов данных требует привлечения технологий и инструментов для реализации высокопроизводительных вычислений. В последние годы мировое сообщество рассматривает большой объем данных как набор данных, характеризующийся следующими основными признаками: объем, скорость, разнообразие, достоверность и ценность. В настоящее время успешно развиваются методы представления и интеллектуального анализа наборов данных большого объема. Существует множество подходов к классификации основных научных и инженерных направлений в решении различных актуальных задач в этой области.

Обнаружение аномалий (выбросов) является распространенной проблемой при анализе больших объемов данных. Аномалия может представлять, как отдельные наблюдения, так и объединяться в определенные группы. Наличие аномалий в наборе данных может быть связано с появлением так называемых

«смещенных» значений, связанных с систематической ошибкой, ошибками ввода, ошибками сбора данных и т. д.

Для анализа данных было предложено множество статистических подходов, таких как дисперсионный анализ, линейная/мультилинейная регрессия, дискриминантный анализ, корреляционный анализ, факторный анализ, анализ главных компонент и кластерный анализ. Кластерный анализ можно использовать для получения представления о данных, выдвижения гипотез, обнаружения аномалий и для классификации.

Технология кластеризации используется во многих областях: медицине, археологии, информационных системах и др. Кластеризация данных также часто используется в качестве начального шага для анализа данных [40, с. 577-582]. Алгоритм для обнаружения выбросов в больших данных на основе плотности был предложен для адаптивного получения параметра, названного Natural Value. Для измерения выбросов рассматривался Natural Outlier Factor. Кроме того, Natural Value может использоваться в других алгоритмах обнаружения выбросов, таких как локальный фактор выбросов и алгоритм обнаружения выбросов с использованием фактора нестабильности [113, с. 93-104] для повышения точности обнаружения выбросов в больших данных. A Liu и соавторы в 2014 году представили новый подход к обнаружению выбросов для обработки данных с несовершенными метками и включения ограниченных аномальных образцов в обучение. Для работы с такими данными авторы ввели значения правдоподобия для каждого входного набора данных, которые обозначают степень принадлежности образца к нормальному и аномальному классу соответственно. Данный подход объединяет локальное и глобальное обнаружения выбросов и повышает производительность их обнаружения. В работе [118, с. 56-69] было предложено эффективное приближение к задаче k-средних, предназначенное для больших наборов данных.

Подход рекурсивно разбивает весь набор данных на небольшое количество подмножеств, каждое из которых характеризуется своим центром масс и весом,

что может сократить количество вычисляемых расстояний. Он превосходит методы k-средних++ и мини-пакетные методы на основе k-средних. Jiang и соавторы в 2016 году представили два различных алгоритма инициализации для кластеризации k-мод: 1) с применением традиционной техники обнаружения выбросов на основе расстояния; 2) с применением техники обнаружения выбросов на основе энтропии разбиения. Выбор начальных центров кластера в кластеризации k-мод с обнаружением выбросов был объединен. Для расчета расстояния между двумя объектами, описанными категориальными признаками, была принята метрика взвешенного расстояния на основе соответствия.

Также был предложен алгоритм k-средних с удалением выбросов (KMOR), мотивированный работой [152, с. 270-293]. Идея алгоритма заключается во введении дополнительного «кластера», содержащего все выбросы. Алгоритм KMOR естественным образом относит все выбросы к определенной группе в процессе кластеризации. Результаты экспериментов показали, что алгоритм KMOR способен одновременно кластеризовать данные и обнаруживать выбросы и превосходит другие алгоритмы по точности и времени выполнения.

Целью данного исследования является разработка кластерного подхода для обнаружения аномалий в реальных больших данных. Были разработаны алгоритмы, минимизирующие компактность кластеров и максимально разделяющие кластеры друг от друга по расстояниям между их центрами, и удаляющие центры кластеров от выбранного общего центра точек в наборе данных [55, с. 325-340].

Количество кластеров заранее неизвестно и устанавливается в соответствии с некоторыми субъективными критериями. Поэтому в данной работе количество кластеров определяется согласно [62, с. 79-84].

Для иллюстрации жизнеспособности разработанных алгоритмов обнаружения аномалий представлены результаты на примерах малых, средних и больших реальных наборах данных.

Предлагаемые алгоритмы. Пусть $x_i \in R^n$ $(i = \overline{1,n})$ – точка из набора данных, где n – общее количество точек во входном наборе данных, $x_i \in c_p \in R^k$ $(p = \overline{1,k})$ – номер кластера, где k – количество кластеров, S_W – компактность кластеров, S_{BW} – разделимость кластеров друг от друга и S_B является мерой удаленности центра каждого кластера (O_p) от центра всех точек (O) во входном наборе данных

$$S_W = \sum_{p=1}^{k} \sum_{i=1}^{n} (x_i - O_p)(x_i - O_p)^T, \quad (2.1.1)$$

$$S_{BW} = \sum_{p=1}^{k-1} \sum_{q=p+1}^{k} (O_p - O_q)(O_p - O_q)^T, \quad (2.1.2)$$

$$S_B = \sum_{p=1}^{k} (O_p - O)(O - O_p)^T, \quad (2.1.3)$$

$$O = \frac{1}{n}\sum_{i=1}^{n} x_i, \;\; O_p = \frac{1}{n_p}\sum_{x_i \in C_p} x_i \;, \;\; n_p = |C_p|, \;\; p = 1,2,\dots,k. \quad (2.1.4)$$

Алгоритм первого предложенного подхода для обнаружения выбросов следующий:

Вход: $X = \{x_1, x_2, \dots, x_n\}$

k: количество кластеров.

Выход: $IDX = \{idx_1, idx_2, \dots, idx_n\}$ вектор индексов кластера.

Шаг 1. Нахождение центра всех точек набора данных (O).

Шаг 2. $s = 0$

Шаг 3. Вычисление компактности (S_W) согласно (2.1.1)

Шаг 4. Вычисление разделимости кластеров (S_{BW}) согласно (2.1.2)

Шаг 5. Вычисление удаленности (S_B) согласно (2.1.3)

Шаг 6. Вычисление значения функции (2.1.5) с учетом (2.1.1)-(2.1.4)

$$F_1^{(s)}(x) = \frac{S_B^{(s)} + S_{BW}^{(s)}}{S_W^{(s)}} \longrightarrow \max \quad (2.1.5)$$

Шаг 7. $s = s + 1$

Шаг 8. Повторение шагов 3-7 до тех пор, пока не будет выполнено условие сходимости:

$$\left| \frac{F_1^{(s+1)} - F_1^{(s)}}{F_1^{(s)}} \right| \leq \varepsilon, \quad (2.1.6)$$

где s - количество итераций

Шаг 9. Вывод значений IDX

Конец

Во втором алгоритме задача заключается в максимизации целевой функции:

Вход: $X = \{x_1, x_2, \dots, x_n\}$

k: количество кластеров.

Выход: $IDX = \{idx_1, idx_2, \dots, idx_n\}$ вектор индексов кластера.

Шаг 1. Нахождение центра всех точек набора данных (O).

Шаг 2. $s = 0$

Шаг 3. Вычисление компактности (S_W) согласно (2.1.1)

Шаг 4. Вычисление разделимости кластеров (S_{BW}) согласно (2.1.2)

Шаг 5. Вычисление удаленности (S_B) согласно (2.1.3)

Шаг 6. Вычисление значения функции (2.1.5) с учетом (2.1.1)-(2.1.4)

$$F_2^{(s)}(x) = \frac{S_B^{(s)} * S_{BW}^{(s)}}{S_W^{(s)}} \longrightarrow \max \qquad (2.1.7)$$

Шаг 7. $s = s + 1$

Шаг 8. Повторение шагов 3-7 до тех пор, пока не будет выполнено условие сходимости:

$$\left| \frac{F_2^{(s+1)} - F_2^{(s)}}{F_2^{(s)}} \right| \leq \varepsilon, \qquad (2.1.8)$$

где s - количество итераций

Шаг 9. Вывод значений IDX

Конец

В третьем алгоритме задача состоит в максимизации целевой функции по параметру регуляризации (α), который будет определен экспериментально.

Шаги алгоритма следующие:

Вход: $X = \{x_1, x_2, \dots, x_n\}$

α: параметр регуляризации

k: количество кластеров.

Выход: $IDX = \{idx_1, idx_2, \dots, idx_n\}$ вектор индексов кластера.

Шаг 1. Нахождение центра всех точек набора данных (O).

Шаг 2. $s = 0$

Шаг 3. Вычисление компактности (S_W) согласно (2.1.1)

Шаг 4. Вычисление разделимости кластеров (S_{BW}) согласно (2.1.2)

Шаг 5. Вычисление удаленности (S_B) согласно (2.1.3)

Шаг 6. Вычисление значения функции (2.1.5) с учетом (2.1.1)-(2.1.4)

$$F_3^{(s)}(x) = \frac{1}{S_W^{(s)}}\left(\alpha S_B^{(s)} + (1-\alpha) S_{BW}^{(s)}\right) \longrightarrow \max \qquad (2.1.9)$$

Шаг 7. $s = s + 1$

Шаг 8. Повторение шагов 3-7 до тех пор, пока не будет выполнено условие сходимости:

$$\left|\frac{F_3^{(s+1)} - F_3^{(s)}}{F_3^{(s)}}\right| \leq \varepsilon, \qquad (2.1.10)$$

где s - количество итераций

Шаг 9. Вывод значений IDX

Конец

Экспериментальные наборы данных. Эксперименты проводились на трех наборах данных из репозитория UCI [174, с. 1001-1005; 267], включая набор данных Phishing, набор данных NSL-KDD [35, с. 842-851] и набор данных Spam.

Набор данных сигнатур атак NSL-KDD построен на основе базы данных KDD-99 [283, с. 262-294]. База данных содержит обучающую (125973 образцов) и тестовую (22544 образцов) выборки. Метки назначаются каждому экземпляру либо как тип «аномалия», либо как «нормальное» поведение. Все образцы (148517) рассматривались в данной работе.

Набор данных Spam содержит электронные письма со спамом и без него [267]. Он включает в себя признаки, которые указывают, часто ли встречалось конкретное слово или символ в электронном письме, и измеряют длину

последовательностей заглавных букв. Набор данных содержит два класса: спам («аномалия») или «нормальное» состояние.

Phishing содержит 11055 фишинговых веб-сайтов [288]. Он включает в себя 30 атрибутов (IP-адрес, длины URL-адреса и аномального URL-адреса, перенаправления веб-сайта и т. д.).

Метрики оценки методов кластеризации. Предположим, что набор данных N разделен на классы $C^+ = \left(C_1^+, \dots, C_{k^+}^+\right)$ («истинная» кластеризация). С помощью процедуры кластеризации в наборе данных можно найти кластеры $C = (C_1, \dots, C_k)$ [48, с. 371-400].

Сравнение решений кластеризации основано на подсчете пар точек. По результатам будет принято решение: «нормальное»/аномальное поведение. Наиболее известными метриками кластеризации на основе расстояния, основанными на парах точек данных, являются чистота (purity) [105, с. 376-381; 335, с. 849-865], метрика Миркина [285, с. 35-47], коэффициент разделения (partition coefficient) [99, с. 301-315], вариация информации (variation of information) [316, с. 902-916], F-мера [334, с. 410-420] и V-мера [334, с. 410-420].

Метрика чистоты (Purity). Чистота кластера C_p дает отношение размера доминирующего класса в кластере к размеру самого кластера [70, с. 3583-3602; 105, с. 376-381; 335, с. 849-865]. Значение чистоты всегда находится в интервале $\left[\frac{1}{k^+}, 1\right]$. Чистоту всего набора кластеров можно оценить как взвешенную сумму значений «чистоты» отдельных кластеров:

$$purity(C) = \frac{1}{n}\sum_{p=1}^{k} \max_{p^+=1,\dots,k^+} \left|C_p \cap C_{p^+}^+\right|, \quad (2.1.11)$$

где k^+ - начальное количество классов, k – искомое количество кластеров. Более высокое значение «чистоты» указывает на лучшее решение для кластеризации.

Метрика Миркина (Mirkin). Метрика Миркина определяется следующим образом [285, с. 35-47]:

$$M(C, C^+) = \frac{1}{n^2}\left(\sum_{p=1}^{k}\left|C_p\right|^2 + \sum_{p^+=1}^{k^+}\left|C_{p^+}^+\right|^2 - 2\sum_{p=1}^{k}\sum_{p^+=1}^{k^+}\left|C_p \cap C_{p^+}^+\right|^2\right). \quad (2.1.12)$$

Метрика Миркина масштабируется с коэффициентом $\frac{1}{n^2}$, чтобы ограничить ее диапазон интервалом [0, 1]. Эта метрика, очевидно, равна 0 для идентичных кластеризаций и положительна в противном случае. Следовательно, чем меньше значение метрики, тем лучше кластеризация.

F-мера. Другая оценочная мера, также известная как «точность кластеризации», основана на F-значении кластера C_p и класса $C_{p^+}^+$, то есть на среднем гармоническом значении точности (precision) и полноты (recall). Точность и полнота вычисляются следующим образом [48, с. 371-400]:

$$P\left(C_p, C_{p^+}^+\right) = \frac{\left|C_p \cap C_{p^+}^+\right|}{|C_p|}, \tag{2.1.13}$$

$$R\left(C_p, C_{p^+}^+\right) = \frac{\left|C_p \cap C_{p^+}^+\right|}{\left|C_{p^+}^+\right|}. \tag{2.1.14}$$

Т.о., F-мера имеет следующий вид:

$$F\left(C_p, C_{p^+}^+\right) = \frac{2P\left(C_p, C_{p^+}^+\right)R\left(C_p, C_{p^+}^+\right)}{P\left(C_p, C_{p^+}^+\right)+R\left(C_p, C_{p^+}^+\right)}. \tag{2.1.15}$$

F-мера кластера C_p – это максимально достигаемое F-значение в любом классе из всего набора классов $C^+ = \left(C_1^+, \dots, C_{k^+}^+\right)$. F-мера всего набора данных считается взвешенной суммой F-мер отдельных кластеров. То есть,

$$F(C) = \sum_{p=1}^{k} \frac{|C_p|}{n} \max_{C_{p^+}^+} \cdot F\left(C_p, C_{p^+}^+\right). \tag{2.1.16}$$

Чем выше F-мера, тем лучше решение для кластеризации. Эта мера имеет значительное преимущество перед purity, поскольку она измеряет как однородность, так и полноту кластерного решения.

Коэффициент разделения (Partition coefficient, PC). Данный коэффициент используется для сравнения $C = (C_1, \dots, C_k)$ и $C^+ = \left(C_1^+, \dots, C_{k^+}^+\right)$ распределений [99, с. 301-315. Согласно [180, с. 207-222], он рассчитывается как:

$$PC(C, C^+) = \frac{1}{kk^+} \sum_{p=1}^{k} \sum_{p^+=1}^{k^+} \left(\frac{\left|C_p \cap C_{p^+}^+\right|}{|C_p|}\right)^2. \tag{2.1.17}$$

Более высокое значение $PC(C, C^+)$ указывает на лучшее решение для кластеризации.

Вариация информации (Variation of information, VI). Эта метрика измеряет количество информации, которая получается и теряются при переходе от одной кластеризации C к другой C^+ [180, с. 207-222; 316, с. 902-916]:

$$VI(C, C^+) = \frac{1}{n \log n} \sum_{p=1}^{k} \sum_{p^+=1}^{k^+} \left|C_p \cap C_{p^+}^+\right| \log\left(\frac{\left|C_p\right|\left|C_{p^+}^+\right|}{\left|C_p \cap C_{p^+}^+\right|^2}\right). \qquad (2.1.18)$$

Максимальное значение вариации информации равно $\log n$, что достигается, когда разбиения находятся как можно дальше друг от друга. В этом случае это означает, что одно из них помещает все точки в один кластер, а другой помещает каждую точку в отдельный кластер. Максимальное значение увеличивается с ростом n, поскольку большие наборы данных содержат больше информации, но если это свойство нежелательно, то можно просто нормализовать по $\log n$.

Как правило, чем меньше VI, тем лучше решение для кластеризации.

V-мера. V-мера — это основанная на энтропии мера, которая явно измеряет, насколько успешно выполняются критерии однородности (homogeneity) и полноты (completeness) [335, с. 849-865]. Однородность можно определить как

$$\hom(C) = \begin{cases} 1, & \text{если } H(C^+|C) = 0 \\ 1 - \frac{H(C^+|C)}{H(C^+)}, & \text{в противном случае} \end{cases}, \qquad (2.1.19)$$

где

$$H(C^+|C) = -\sum_{p=1}^{k} \sum_{p^+=1}^{k^+} \frac{\left|C_p \cap C_{p^+}^+\right|}{n} \log\left(\frac{\left|C_p \cap C_{p^+}^+\right|}{\sum_{p^+=1}^{k^+}\left|C_p \cap C_{p^+}^+\right|}\right), \qquad (2.1.20)$$

$$H(C^+) = -\sum_{p^+=1}^{k^+} \frac{\sum_{p=1}^{k}\left|C_p \cap C_{p^+}^+\right|}{k^+} \log\left(\frac{\sum_{p=1}^{k}\left|C_p \cap C_{p^+}^+\right|}{k^+}\right). \qquad (2.1.21)$$

$H(C^+|C)$ равен 0, когда каждый кластер содержит только члены одного класса (идеальная однородная кластеризация).

В вырожденном случае, когда $H(C^+)$ равен 0, и когда имеется только один класс, однородность принимается равной 1.

Полнота симметрична однородности. Полнота может быть определена как

$$\text{comp}(C) = \begin{cases} 1, & \text{если } H(C|C^+) = 0 \\ 1 - \frac{H(C|C^+)}{H(C)}, & \text{в противном случае} \end{cases}, \quad (2.1.22)$$

где

$$H(C|C^+) = -\sum_{p^+=1}^{k^+}\sum_{p=1}^{k}\frac{\left|C_p \cap C_{p^+}^+\right|}{n}\log\left(\frac{\left|C_p \cap C_{p^+}^+\right|}{\sum_{p=1}^{k}\left|C_p \cap C_{p^+}^+\right|}\right), \quad (2.1.23)$$

$$H(C) = -\sum_{p=1}^{k}\frac{\sum_{p^+=1}^{k^+}\left|C_p \cap C_{p^+}^+\right|}{k^+}\log\left(\frac{\sum_{p^+=1}^{k^+}\left|C_p \cap C_{p^+}^+\right|}{k^+}\right). \quad (2.1.24)$$

V-мера решения кластеризации вычисляется путем нахождения среднего гармонического однородности и полноты следующим образом:

$$V(C) = \frac{2\,\text{hom}(C)\text{comp}(C)}{\text{hom}(C)+\text{comp}(C)}. \quad (2.1.25)$$

Вычисления однородности, полноты и V-меры полностью не зависят от количества классов и кластеров, размера набора данных и алгоритма кластеризации.

Экспериментальные результаты. Эксперименты на рассмотренных наборах данных Phishing, NSL-KDD и Spam для оценки производительности предложенных алгоритмов были реализованы в Matlab 2016a на 64-битной системе на базе Windows с процессором Intel Core (i7) с тактовой частотой 2.5 ГГц и 8 Гбайт оперативной памяти.

Характеристики экспериментальных наборов данных представлены в таблице 2.1.1.

Таблица 2.1.1

Описание наборов данных

Набор данных	Количество образцов	C_1^+	C_2^+	Количество атрибутов
Phishing	11055	4898	6157	30
NSL-KDD	148517	71463	77054	41
Spam	4601	2788	1813	57

Для анализа были выбраны шесть метрических показателей различной природы: Purity, Mirkin, F-мера, PC, VI и V-мера.

Наборы данных были разделены на два класса: C_1^+ и C_2^+. В ходе предварительной обработки значения в наборах данных были стандартизированы. Образцы в классе C_1^+ были приняты за аномалии.

Результаты предложенных алгоритмов на основе шести метрик представлены ниже. Лучшие результаты были отмечены жирным шрифтом.

Предлагаемые подходы сравниваются с алгоритмом k-средних. В таблице 2.1.2 показаны экспериментальные результаты для всех наборов данных при применении алгоритма k-средних. Наилучший результат по метрикам по метрикам F-мера и VI был получен для набора данных NSL-KDD и составил 68.01% и 6.37%, соответственно. Purity и Mirkin показали лучший результат на наборе данных Spam. По показателям Purity (55.69%) и PC (25.23%) набор данных Phishing превзошел набор данных NSL-KDD.

Таблица 2.1.2

Оценка производительности алгоритма k-средних для всех наборов данных

Набор данных	C_1	C_2	Purity	Mirkin	F-мера	VI	PC	V-мера
Phishing	9263	1792	0.5569	0.4953	0.5888	0.1213	0.2523	1.0000
NSL-KDD	1834	146683	0.5188	0.4993	**0.6801**	**0.0637**	0.2502	1.0000
Spam	244	4357	**0.6359**	**0.4630**	0.5385	0.0999	**0.2981**	1.0000

Наилучшие результаты для первого предложенного алгоритма были получены для набора данных Spam согласно метрикам Purity (62.42%), PC (27.13%) и Mirkin (46.91%), для набора данных Phishing соласно метрике F-мера (70.81%), а для набора даных NSL-KDD согласно метрике VI (6.25%) (Таблица 2.1.3).

Анализ показывает, что второй предложенный алгоритм обеспечивает высокое качество кластеризации для набора данных NSL-KDD по всем шести рассмотренным метрикам (Таблица 2.1.4). Наименьшие результаты были получены для набора данных Spam согласно Purity, Mirkin и PC.

Таблица 2.1.3

Оценка производительности первого предложенного алгоритма на экспериментальных наборах данных

Набор данных	C_1	C_2	Purity	Mirkin	F-мера	VI	PC	V-мера
Phishing	601	10454	0.5717	0.4897	**0.7081**	0.0955	0.2615	1.0000
NSL-KDD	1415	147102	0.5226	0.4990	0.6833	**0.0625**	0.2700	1.0000
Spam	252	4349	**0.6242**	**0.4691**	0.5462	0.1026	**0.2713**	1.0000

Было рассмотрено влияние параметра регуляризации α на производительность третьего предложенного алгоритма на различных наборах данных (Таблицы 2.1.5-2.1.15). Для α использовались значения 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 и 1.

Таблица 2.1.4

Оценка производительности второго предложенного алгоритма на экспериментальных наборах данных

Набор данных	C_1	C_2	Purity	Mirkin	F-мера	VI	PC	V-мера
Phishing	3023	8032	0.6715	0.4412	0.6357	0.1122	0.3403	1.0000
NSL-KDD	85159	63358	**0.8215**	**0.2932**	**0.8176**	**0.0774**	**0.3570**	1.0000
Spam	4330	271	0.6405	0.4605	0.7531	0.1011	0.3016	1.0000

Из таблиц 2.1.5-2.1.14 можно сделать вывод, что α=0, α=0.1, α=0.2, α=0.3, α=0.4, α=0.5, α=0.6, α=0.7, α=0.8 и α=0.9 дают наилучшие результаты для наборов данных NSL-KDD и Spam.

Таблица 2.1.5

Оценка производительности третьего предложенного алгоритма для $\alpha = 0$

Набор данных	C_1	C_2	Purity	Mirkin	F-мера	VI	PC	V-мера
Phishing	705	10350	0.5952	0.4819	**0.7187**	0.0953	0.2983	1.0000
NSL-KDD	8186	140331	0.5305	0.4981	0.6394	**0.0716**	**0.3503**	1.0000
Spam	1820	2781	**0.7003**	**0.4198**	0.6726	0.1425	0.2892	1.0001

Кроме того, при α=0.6, α=0.7, α=0.8, α=0.9 и α=1 по метрикам Mirkin и Purity были достигнуты наилучшие показатели для набора данных Spam, а по метрикам VI и PC – для набора данных NSL-KDD.

Таблица 2.1.6

Оценка производительности третьего предложенного алгоритма для α=0.1

Набор данных	C_1	C_2	Purity	Mirkin	F-мера	VI	PC	V-мера
Phishing	705	10350	0.5952	0.4819	**0.7187**	0.0953	0.2983	1.0000
NSL-KDD	10527	137990	0.5457	0.4958	0.6270	**0.0736**	**0.3538**	1.0000
Spam	1733	2868	**0.7022**	**0.4182**	0.6679	0.1416	0.2887	1.0001

Таблица 2.1.7

Оценка производительности третьего предложенного алгоритма для α=0.2

Набор данных	C_1	C_2	Purity	Mirkin	F-мера	VI	PC	V-мера
Phishing	786	10269	0.6025	0.4790	**0.7203**	0.0963	0.3051	1.0000
NSL-KDD	11197	137320	0.5495	0.4951	0.6238	**0.0742**	**0.3529**	1.0000
Spam	1683	2918	**0.7005**	**0.4196**	0.6617	0.1415	0.2875	1.0001

Таблица 2.1.8

Оценка производительности третьего предложенного алгоритма для α=0.3

Набор данных	C_1	C_2	Purity	Mirkin	F-мера	VI	PC	V-мера
Phishing	753	10302	0.5995	0.4802	**0.7196**	0.0959	0.3024	1.0000
NSL-KDD	11726	136791	0.5525	0.4945	0.6214	**0.0747**	**0.3521**	1.0000
Spam	1584	3017	**0.7033**	**0.4173**	0.6556	0.1399	0.2878	1.0001

Таблица 2.1.9

Оценка производительности третьего предложенного алгоритма для α=0.4

Набор данных	C_1	C_2	Purity	Mirkin	F-мера	VI	PC	V-мера
Phishing	830	10225	0.6065	0.4773	**0.7211**	0.0968	0.3084	1.0000
NSL-KDD	12292	136225	0.5556	0.4938	0.6188	**0.0753**	**0.3512**	1.0000
Spam	1535	3066	**0.7018**	**0.4186**	0.6487	0.1395	0.2870	1.0001

Таблица 2.1.10

Оценка производительности третьего предложенного алгоритма для α=0.5

Набор данных	C_1	C_2	Purity	Mirkin	F-мера	VI	PC	V-мера
Phishing	853	10202	0.6086	0.4764	**0.7215**	0.0971	0.3100	1.0000
NSL-KDD	13089	135428	0.5601	0.4928	0.6152	**0.0759**	**0.3505**	1.0000
Spam	1495	3106	**0.7009**	**0.4193**	0.6433	0.1391	0.2865	1.0001

Таблица 2.1.11

Оценка производительности третьего предложенного алгоритма для α=0.6

Набор данных	C_1	C_2	Purity	Mirkin	F-мера	VI	PC	V-мера
Phishing	1011	10044	0.6071	0.4770	**0.7154**	0.1017	0.2917	1.0000
NSL-KDD	14228	134289	0.5667	0.4911	0.6100	**0.0768**	**0.3499**	1.0000
Spam	1466	3135	**0.7025**	**0.4180**	0.6419	0.1384	0.2871	1.0001

Таблица 2.1.12

Оценка производительности третьего предложенного алгоритма для α=0.7

Набор данных	C_1	C_2	Purity	Mirkin	F-мера	VI	PC	V-мера
Phishing	1214	9841	0.6034	0.4786	**0.7064**	0.1070	0.2765	1.0000
NSL-KDD	15980	132537	0.5764	0.4883	0.6023	**0.0781**	**0.3485**	1.0000
Spam	1441	3160	**0.7018**	**0.4186**	0.6382	0.1381	0.2868	1.0001

Таблица 2.1.13

Оценка производительности третьего предложенного алгоритма для α=0.8

Набор данных	C_1	C_2	Purity	Mirkin	F-мера	VI	PC	V-мера
Phishing	1428	9627	0.6013	0.4795	**0.6977**	0.1117	0.2688	1.0000
NSL-KDD	18888	129629	0.5927	0.4828	0.5900	**0.0799**	**0.3469**	1.0000
Spam	1426	3175	**0.7016**	**0.4187**	0.6361	0.1379	0.2867	1.0001

Таблица 2.1.14

Оценка производительности третьего предложенного алгоритма для α=0.9

Набор данных	C_1	C_2	Purity	Mirkin	F-мера	VI	PC	V-мера
Phishing	1925	9130	0.5959	0.4816	**0.6769**	0.1208	0.2605	1.0000
NSL-KDD	24685	123832	0.6249	0.4688	0.5674	**0.0825**	**0.3455**	1.0000
Spam	1397	3204	**0.7027**	**0.4178**	0.6339	0.1371	0.2872	1.0001

Таблица 2.1.15

Оценка производительности третьего предложенного алгоритма для α=1

Набор данных	C_1	C_2	Purity	Mirkin	F-мера	VI	PC	V-мера
Phishing	8319	2736	0.5597	0.4929	0.0787	0.1333	0.2530	1.0000
NSL-KDD	130195	18322	0.5979	0.4808	**0.5784**	**0.0837**	**0.3036**	1.0000
Spam	1380	3221	**0.7020**	**0.4184**	0.6310	0.1369	0.2870	1.0001

V-мера не обладает дискриминирующей способностью, т.е. ее значение на разных наборах данных практически одинаково для методов. Отсюда можно

сделать вывод, что использование V-меры бесполезно для оценки результатов кластеризации. Поэтому в следующих сравнениях она не учитывалась.

Чтобы проиллюстрировать жизнеспособность разработанных алгоритмов обнаружения аномалий, результаты представлены на рис. 2.1.1-2.1.5. На рис. 2.1.1 показаны результаты третьего реализованного алгоритма кластеризации по метрике Purity.

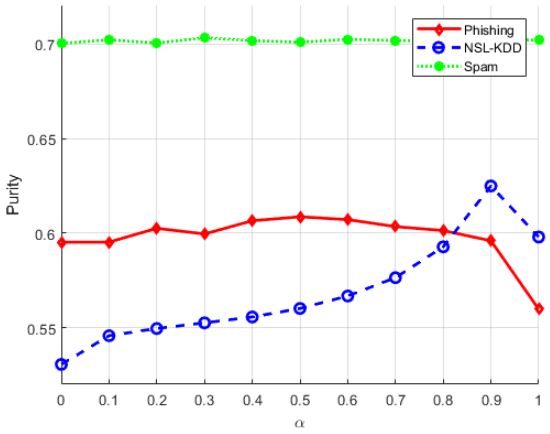


Рис. 2.1.1. Влияние значений α на показатель чистоты для разных наборов данных [51]

На основании экспериментальных результатов можно сделать вывод, что предложенный подход показывает наилучшие результаты для наборов данных NSL-KDD и Spam. Самые низкие значения наблюдались для Phishing, но при α=0.9 можно увидеть улучшение.

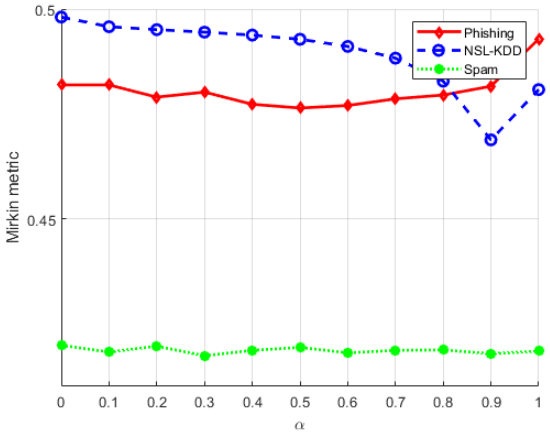


Рис. 2.1.2. Влияние значений α по метрике Миркина для разных наборов данных [51]

Самые низкие результаты по метрике Mirkin для всех значений α были получены для набора данных Spam (Рис. 2.1.2). Для набора данных Phishing значение метрики практически не менялось при значении α от 0 до 0.9.

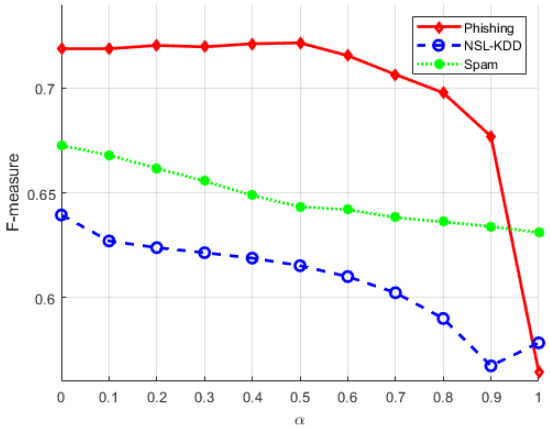


Рис. 2.1.3. Влияние значений α по метрике F-мера для разных наборов данных [51]

На рис. 2.1.3 показано, что самые высокие результаты были получены для набора данных Phishing при α от 0 до 0.9, тогда как при α=1 значение резко падало и составило 7.87%. Наиболее низкие результаты по метрике F-мера были получены для набора данных NSL-KDD.

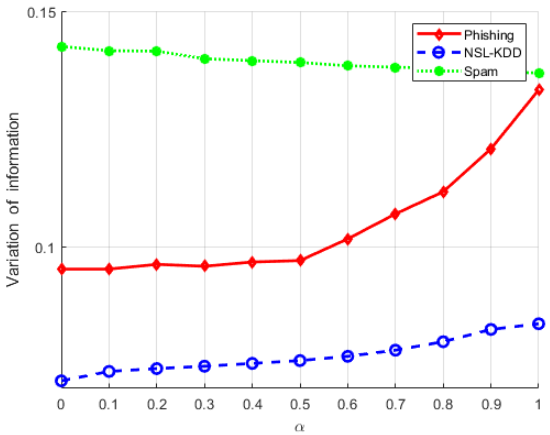


Рис. 2.1.4. Влияние значений α на вариацию информации для разных наборов данных [51]

Наилучшие результаты, то есть минимальные значения метрики VI, были получены для набора данных NSL-KDD (Рис. 2.1.4). Значения VI резко выросли с увеличением значения α для набора данных Phishing.

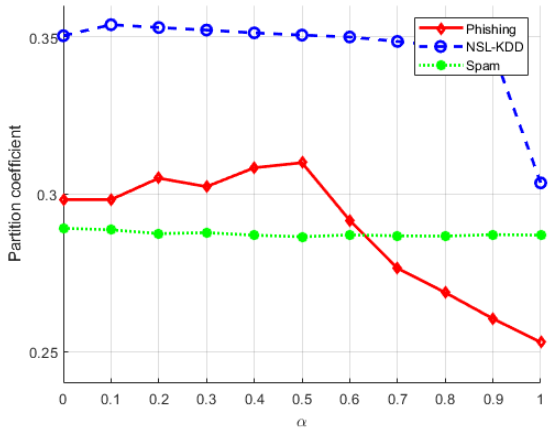


Рис. 2.1.5. Влияние значений α на коэффициент разделения для разных наборов данных [51]

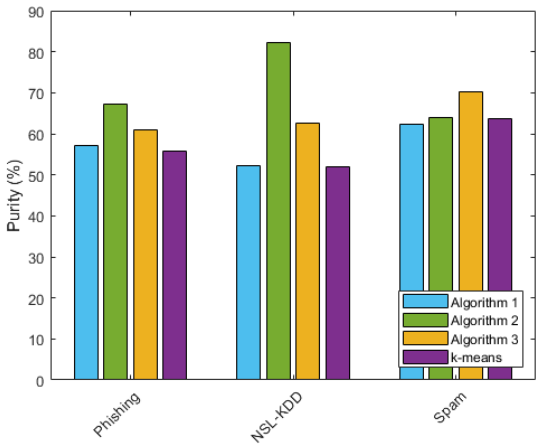
Согласно рис. 2.1.5 результаты экспериментов показали, что набор данных NSL-KDD достигает наилучших результатов практически для всех значений α. Для набора данных Spam значение метрики было практически постоянным. Значения метрики PC резко падают при α от 0.5 до 1 для набора данных Phishing и при α=1 для набора данных NSL-KDD.

Сравнение значений метрик оценки для первого (голубые столбцы), второго (зеленые столбцы), третьего (желтые столбцы) предложенных алгоритмов и алгоритма k-средних (фиолетовые столбцы) на трех наборах данных более наглядно показано на рис. 2.1.6. Для третьего алгоритма были выбраны лучшие результаты по каждой метрике для каждого значения α.

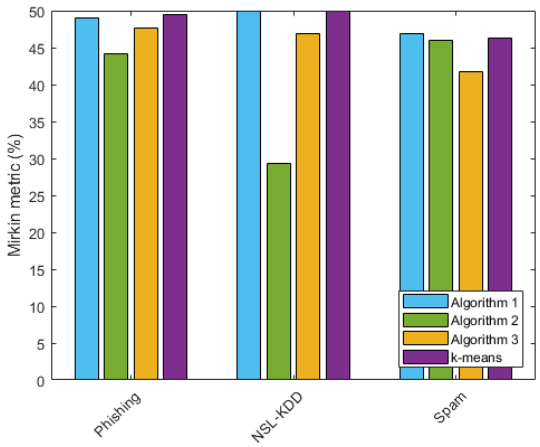
На основании экспериментальных результатов можно сделать вывод, что три предложенных алгоритма превосходят алгоритм k-средних по метрикам Purity, F-мера и PC для набора данных Phishing.

Для набора данных NSL-KDD наилучшие результаты были получены для первого и второго предложенных алгоритмов согласно приведенным выше трем метрикам.

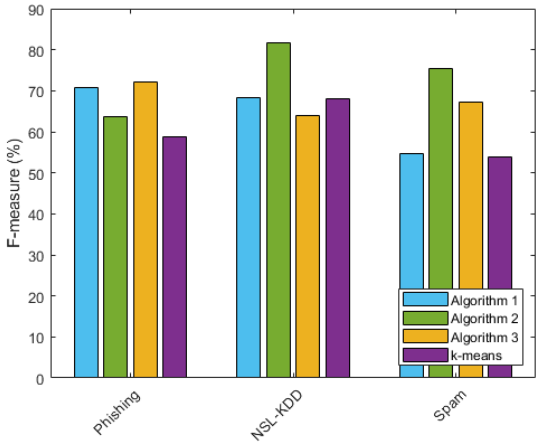
По всем метрикам набор данных Phishing показал лучший результат для второго алгоритма. Purity, Mirkin, F-мера и PC показали хорошие результаты для набора данных NSL-KDD при применении второго алгоритма.



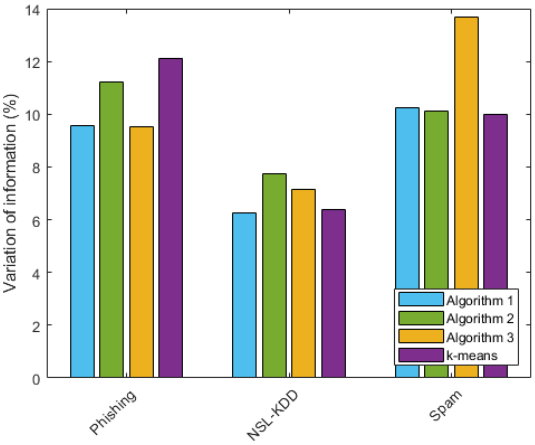
а)



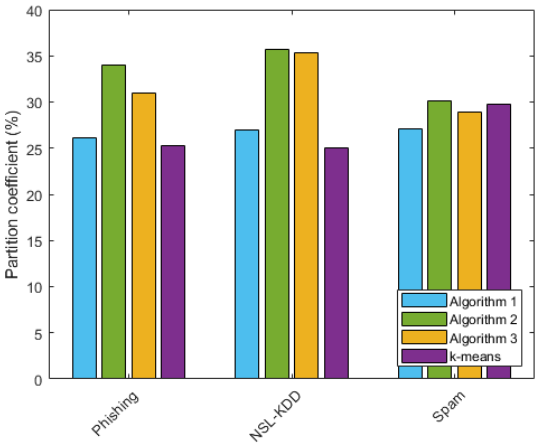
б)



в)



г)



д)

Рис. 2.1.6. Сравнение предложенных алгоритмов с алгоритмом k-средних на основе метрик [51]

Значения метрик Purity, Mirkin и F-мера показали наилучшие результаты для второго и третьего алгоритмов, в то время как метрика PC была лучшей только для второго алгоритма на наборе данных Spam. Можно сделать вывод, что третий алгоритм хорошо работает на наборах данных малого и большого размера, тогда как второй алгоритм показал наилучшие результаты на наборах данных среднего размера [51].

В данном исследовании были предложены новые алгоритмы кластеризации для обнаружения аномалий в больших данных КФС. Целью предложенных алгоритмов является улучшение обнаружения аномалий в системах.

Представленные алгоритмы минимизируют компактность кластеров и максимизируют разделение кластеров друг от друга в соответствии с расстояниями между их центрами и удаленностью центров кластеров от выбранного общего центра точек в наборе данных. Сравнение проводилось с использованием трех наборов данных, содержащих аномальные значения. Качество результата кластеризации оценивалось с использованием шести метрик оценок. Важной особенностью предложенных подходов является то, что они повышают точность обнаружения аномальных значений в КФС на основе кластеризации.

Сравнивалась производительность предложенных алгоритмов с алгоритмом k-средних. Подводя итоги можно сделать вывод, что предложенные алгоритмы эффективно работают на реальных наборах данных различной размерности.

Дальнейшие исследования будут сосредоточены на разработке и применении ансамблей алгоритмов кластеризации для повышения точности обнаружения выбросов в больших данных КФС.

2.2. Алгоритм анализа больших данных киберфизических систем на основе взвешенной кластеризации

С появлением компьютерных технологий объем данных, с которыми может работать человек, значительно увеличился. Ученые долгое время разрабатывали алгоритмы, направленные на упрощение работы с данными, выявление новых, ранее неизвестных знаний, хранящихся в данных. Вычислительные системы ограничены в хранении, анализе и обработке

Больших данных из-за их объема, скорости или разнообразия. Большие объемы данных можно описать, добавив достоверность и ценность к объему, разнообразию и скорости. Таким образом, с помощью пяти V можно найти новое понимание существующих Больших данных. Однако сегодня объем хранимых данных становится слишком большим для обработки традиционными алгоритмами.

Исследователям приходится прибегать к различным уловкам, таким как работа с частями имеющихся данных, использование априорных знаний об имеющихся данных. Таким образом, работа с большими данными КФС вызывает необходимость формализации новых методов, используемых исследователями, создания новых алгоритмов и программных средств, использующих мощь ранее созданных средств для работы с данными больших объемов и размерностей.

Проблема обнаружения аномалий является одной из проблем анализа больших данных. Она широко применяется в следующих областях: обнаружение вторжений в компьютерные сети, обнаружение мошенничества в банковских транзакциях, медицина, мониторинг движения поездов, обнаружение отказов систем космических аппаратов и т. д. [124, с. 1-58].

Одним из самых «удобных» и понятных подходов к обработке больших данных является кластеризация. Проблема кластеризации становится все более актуальной во многих сферах.

В настоящее время разработано множество различных алгоритмов кластеризации [125, с. 224-228; 90, с. 702-719; 106, с. 153-161]. Их сложность зависит от размерности данных, объема набора кластеров, масштаба приложений и т. д. [91, с. 23-27].

Кластерный анализ можно использовать для получения представления о данных, выдвижения гипотез и обнаружения аномалий и классификаций. В 2003 году были сформулированы три требования к алгоритмам кластеризации потоков данных [91, с. 23-27]: 1) сжатие данных и представление сжатых данных; 2) обработка новых точек данных быстрым и инкрементальным способом; 3) быстрое и четкое выделение выбросов. Приложения часто определяются в терминах выбросов (например, в случае обнаружения мошенничества, обнаружения аномалий в сети и т. д.), и в этом случае прямой подход, вероятно, будет более эффективным [248, с. 749-754; 258, с. 25-36; 113, с. 93-104].

Решение проблемы обнаружения аномалий не является тривиальным. Трудность обнаружения аномалий (выбросов) заключается также в том, чтобы присвоить метки шаблонам нормального и аномального поведения, которые нелегко получить [298, с. 1-8; 382, с. 1302-1325]. Однако выбранный подход к обнаружению аномалий может подойти только для определенного круга задач, но не для всех [155, с. 353-375]. Таким образом, когда типы аномалий априори неизвестны, вполне реалистично предположить, что выбор метода для их обнаружения непрост.

Эффективность и точность методов анализа данных напрямую зависит от их качества, которое, в свою очередь, зависит от количества выбросов в данных.

Также был предложен новый алгоритм, использующий модели стохастических дифференциальных уравнений для обнаружения аномалий в стохастических процессах. Rajabzadeh и соавторы указали на наличие аномалий в тестовых данных. Для оценки полученных коэффициентов была применена

локальная полиномиальная регрессия на основе ядра. Была построена логарифмическая функция правдоподобия параметров модели по скользящему окну с использованием формулы Гирсанова. Значение логарифмической функции правдоподобия использовалось как индикатор аномалии для принятия решения о наличии аномалии в тестовых данных. Метод был оценен на реальных данных и использован для различения речевых/неречевых образцов.

Как в смоделированных, так и в реальных данных предложенный алгоритм превзошел другие методы [49, с. 87-96]. Алгоритм k-средних является наиболее широко используемым алгоритмом для приложений интеллектуального анализа данных [382, с. 1302-1325]. Он прост, масштабируем, прост в понимании и может быть адаптирован для работы с данными высокой размерности [118, с. 56-69].

Однако алгоритм k-средних работает с наборами данных только с числовыми атрибутами, что ограничивает его использование при решении задач кластеризации категориальных данных. Для решения этой проблемы был предложен алгоритм k-мод. Благодаря рандомизированному преобразованию предварительной обработки можно достичь точных и надежных оценок в процессе разрежения данных.

В 2017 году был предложен разреженный алгоритм k-средних, который возвращает как назначения, так и центры кластеров за один проход по данным, в то время как современные алгоритмы на основе признаков требуют как минимум двух проходов. Метод предварительной обработки и выборки может использоваться либо для ускорения вычислений и решения проблем памяти в ядре, либо для создания вариантов с одним проходом к проблемам вне ядра или потоковой передачи. Tsay в 2016 году разработал новый подход, объединяющий алгоритм k-средних и классификацию на основе деревьев, для анализа и визуализации временных рядов высокой размерности. Расстояние Хеллингера между функциями плотности широко использовалось в данном анализе. Разумные результаты были получены в результате экспериментов на

реальных наборах данных. Chandola и соавторы [124, с. 1-58] попытались предоставить структурированный и всесторонний обзор исследований в области обнаружения аномалий для дискретных последовательностей. Кроме того, данное исследование может быть применено к непрерывным последовательностям или временным рядам. Например, методы, основанные на евклидовом расстоянии, кросс-корреляции, методы на основе окна и скрытые марковские модели. Экспериментальные результаты показали устойчивость плохих инициализаций и ее эффективность по сравнению с алгоритмами k-средних, k-средних++ [80, с. 1027-1035] и K-гармоническим средним [90, с. 702-719].

Модифицированный алгоритм k-средних KMOR был предложен для кластеризации данных и обнаружения выбросов. Эксперименты проводились на синтетических и реальных наборах данных. Предложенный алгоритм сравнивался с алгоритмом обнаружения и кластеризации выбросов [40, с. 577-582]. Результаты показали превосходство алгоритма KMOR по точности и времени выполнения. Soheily-Khah и соавторы подчеркнули, что разработанные решения позволяют оценивать не только временной центроид данных, но и его весовой вектор, что указывает на репрезентативность элементов центроида. Также обсуждалось влияние изотропии и изоляции кластеров на эффективность методов кластеризации. Предложенные решения могут быть напрямую применимы к любым другим вариациям k-средних.

Данное исследование направлено на разработку кластерного подхода для обнаружения аномалий в больших данных [49, с. 87-96; 53, с. 178-188]. Работа с большими данными требует значительных вычислительных ресурсов. Согласно [62, с. 79-84] предлагается алгоритм, основанный на взвешенной кластеризации (ВК) для заранее известного числа кластеров.

Предлагаемый подход. Пусть $X = (x_1, x_2, \dots, x_n)$ – точки в наборе данных, где n - общее количество точек в наборе данных, $x_i = (x_{i1}, x_{i2}, \dots, x_{im}) \in R^m$ – точка в наборе данных, где m - размерность точек данных, $C = (C_1, C_2, \dots, C_k)$ –

кластеры, где C_p $(p = \overline{1,k})$ – p-ый кластер и k – количество кластеров. Задача состоит в том, чтобы минимизировать функцию для обнаружения аномалий в наборе данных следующим образом:

$$f(x) = \sum_{p=1}^{k} \sum_{x_i \in C_p} \left|C_p\right|_W * \left\|x_i - O_p\right\|^2 \longrightarrow \min, \qquad (2.2.1)$$

где $\left|C_p\right|_W$ – вес p-го кластера.

В этом случае вес кластера определяется как сумма весов всех точек кластера:

$$\left|C_p\right|_W = \sum_{x_i \in C_p} w(x_i),\ p = 1,2,\dots,k, \qquad (2.2.2)$$

где веса точек рассчитываются на основе их расстояния от центра всех точек в наборе данных

$$w(x_i) = \|x_i - O\|,\ O = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad (2.2.3)$$

а центр p-го кластера (O_p) определяется как

$$O_p = \frac{1}{n_p}\sum_{x_i \in C_p} x_i\,,\ n_p = \left|C_p\right|,\ p = 1,2,\dots,k. \qquad (2.2.4)$$

Алгоритм предлагаемого метода обнаружения аномалий следующий:

Вход: $X = (x_1, x_2, \dots, x_n)$

$w = (w_1, w_2, \dots, w_n)$

k: количество кластеров.

Выход: $IDX = \{idx_1, idx_2, \dots, idx_n\}$ вектор индексов кластера.

Шаг 1. Нахождение центра всех точек набора данных (O).

Шаг 2. Вычисление весов всех точек x_i согласно (2.2.3)

Шаг 3. $s = 0$

Шаг 4. Вычисление значения функции согласно (2.2.1) с учетом (2.2.2)

$$f^{(s)} = \sum_{p=1}^{k} \sum_{x_i \in C_p} \left|C_p\right|_W * \left\|x_i - O_p^{(s)}\right\|^2$$

Шаг 5. $s = s + 1$

Шаг 6. Повторение шагов 3-4 до тех пор, пока не будет выполнено условие сходимости:

$$\left|\frac{f^{(s+1)}-f^{(s)}}{f^{(s)}}\right| \leq \varepsilon,$$

где s – количество итераций.

Шаг 7. Вывод значений IDX

Конец

В алгоритме каждый кластер представлен своим центром, и цель состоит в том, чтобы найти решение, которое минимизирует расстояние между каждой точкой и центром кластера, которому она назначена [35, с. 842-851].

Экспериментальные результаты. Для оценки производительности предложенного подхода был проведен ряд экспериментов в Matlab 2016a на 64-битной системе на базе Windows с процессором Intel Core (i7) с тактовой частотой 2.5 ГГц и 8 Гбайт оперативной памяти.

В качестве исходных данных использовались экспериментальные наборы данных NSL-KDD и Phishing. Характеристики наборов данных представлены в таблице 2.2.1.

Таблица 2.2.1

Описание наборов данных

Набор данных	Количество образцов	C_1^+	C_2^+	Количество атрибутов
NSL-KDD	148517	71463	77054	41
Phishing	11055	7262	3793	30

Во время предварительной обработки значения в наборах данных были стандартизированы, чтобы иметь среднее значение равное 0 и стандартное отклонение равное 1. Все наборы данных содержат два класса: C_1^+ и C_2^+. Образцы, включенные в класс C_1^+, принимаются за аномалии.

Результаты предложенного подхода на основе шести метрик представлены в таблице 2.2.2. В качестве оценочных метрик рассматривались Purity, метрика Mirkin, PC, VI, F-мера и V-мера. Также предложенный подход сравнивался с алгоритмом k-средних. В предлагаемом алгоритме каждый кластер представлен своим центром, и задача состоит в том, чтобы найти решение, минимизирующее расстояние между каждой точкой и центром кластера, к

которому она относится, с учетом весов кластеров [180, с. 207-222].

Наилучшие результаты для предлагаемого подхода по показателям Purity и Mirkin были получены для набора данных Phishing, а для набора данных NSL-KDD по метрикам F-мера и PC (Таблица 2.2.2).

В таблице 2.2.2 показаны экспериментальные результаты для всех наборов данных при применении алгоритма k-средних (КС) и предложенного алгоритма (ПА).

Таблица 2.2.2

Оценка эффективности предлагаемого подхода и алгоритма k-средних для разных наборов данных

Метрики / Набор данных	Purity		Mirkin		F-мера		VI		PC	
	ПА	КС	ПА	КС	ПА	КС	ПА	КС	ПА	КС
NSL-KDD	0.5384	0.5188	0.4971	0.4993	**0.6846**	0.6801	0.0720	**0.0637**	**0.2766**	0.2502
Phishing	**0.5880**	0.5569	**0.4845**	0.4953	0.5817	0.5888	0.0998	0.1213	0.2747	0.2523

Наилучший результат по метрике Purity с применением предложенного подхода был получен для набора данных Phishing и составил 58.80%. Согласно метрике Mirkin (48.45%), наименьшее значение было достигнуто для набора данных Phishing.

Наилучший результат для набора данных NSL-KDD с применением k-средних по F-мере составил 68.01%. Набор данных NSL-KDD показал наилучшие результаты для кластеризации k-средних на основе показателя VI (27.66%).

V-мера не обладает дискриминирующей способностью, т.е. ее значение на разных наборах данных практически одинаково для методов. Отсюда можно сделать вывод, что использование V-меры для оценки результатов кластеризации не является необходимым. Поэтому в следующих сравнениях она не учитывалась.

Сравнение предложенного подхода с алгоритмом k-средних показано в таблице 2.2.3.

Таблица 2.2.3

Оценка производительности предложенного алгоритма в сравнении с алгоритмом k-средних

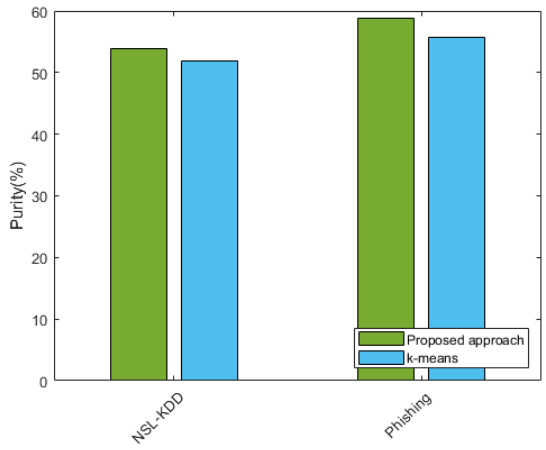
Метрики / Набор данных	Purity (%)	Mirkin (%)	F-мера (%)	VI (%)	PC (%)
NSL-KDD	3.78 (+)	0.44 (+)	0.66 (+)	13.03 (-)	10.55 (+)
Phishing	5.58 (+)	2.18 (+)	1.21 (-)	17.72 (+)	8.88 (+)

Далее для сравнения используется относительное улучшение:

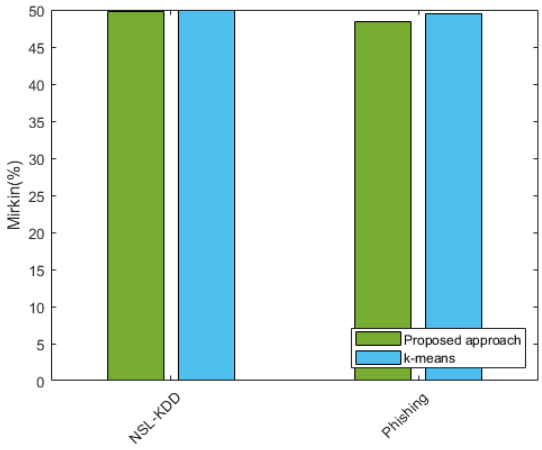
$$\frac{\text{предложенный_метод} - \text{метод_k-средних}}{\text{метод_k-средних}} \times 100\%. \qquad (2.2.5)$$

В таблице 2.2.3 знак «+» означает превосходство результата, а знак «-» означает обратное. Сравнение значений показателей оценки для предложенного подхода (зеленые столбцы) и алгоритма k-средних (голубые столбцы) на двух наборах данных более наглядно показано на рис. 2.2.1.

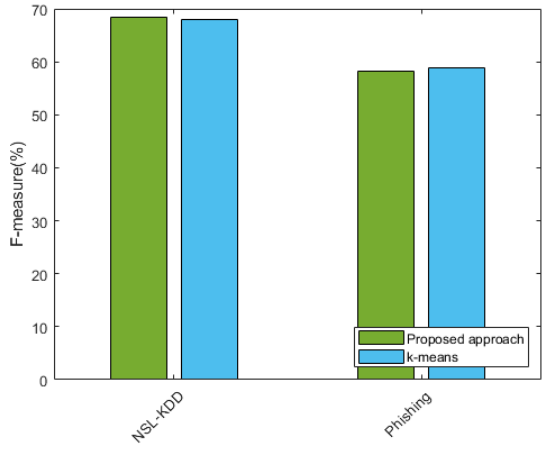
На основании экспериментальных результатов можно сделать вывод, что предложенный подход превосходит алгоритм k-средних по четырем метрикам (Purity, Mirkin, F-мера и PC) для набора данных NSL-KDD. Purity, Mirkin, PC и VI показали хорошие результаты для набора данных Phishing. Значения показателей F-мера были достаточно близкими для обоих подходов к наборам данных NSL-KDD и Phishing.



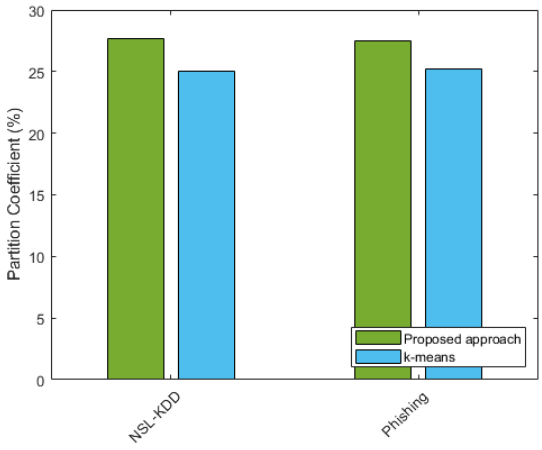
а)



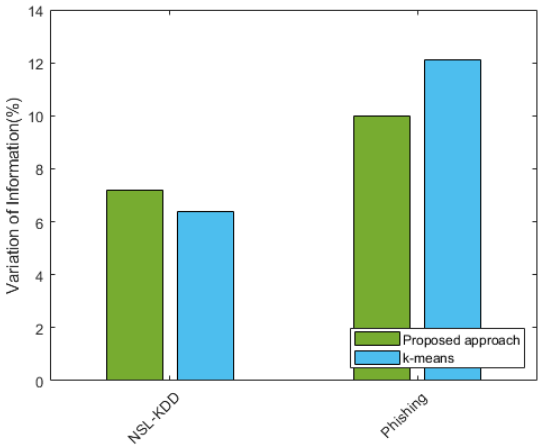
б)



в)



г)



д)

Рис. 2.2.1. Экспериментальные результаты сравнения [49]

Высокие результаты оценки для набора данных Phishing были получены на основе метрик Purity, Mirkin, PC и VI. Для набора данных NSL-KDD улучшение было получено на основе показателей Purity (3.78%), Mirkin (0.44%), PC (10.55%) и F-мера (0.66%), соответственно.

Был предложен алгоритм обнаружения аномальных значений в КФС на основе взвешенной кластеризации. Целью алгоритма, представленного в данном исследовании, было улучшение процесса обнаружения аномалий в больших наборах данных КФС.

Веса, полученные путем суммирования весов каждой точки из набора данных, были назначены кластерам. Вес каждой точки определялся ее положением относительно центра во всем наборе данных. Видно, что взвешивание улучшает решение кластеризации.

Сравнение проводилось на экспериментальных наборах данных с алгоритмом k-средних. Качество результата кластеризации оценивалось с использованием шести метрик оценок. Важной особенностью подхода является то, что он повышает точность обнаружения аномалий на основе кластеризации. Экспериментальные результаты показали, что предложенный алгоритм более точно обнаруживает аномалии в КФС по сравнению с алгоритмом k-средних и имеет практическое значение.

Применяя предложенный подход к кластеризации данных КФС, можно повысить надежность разбиения кластеров на группы. Исследовалось влияние весов кластеров на производительность и точность обнаружения аномалий в больших данных.

Можно сделать вывод, что предлагаемый подход становится более эффективным с увеличением размера анализируемого набора данных. Важно, что предложенный подход может быть применен для повышения эффективностии обнаружения аномалий в реальных больших наборах данных КФС, а также в различных областях исследований кибербезопасности таких систем.

2.3. Метод параллельной обработки больших данных для снижения риска выхода из строя киберфизических систем

Алгоритмы кластеризации широко применяются в анализе больших данных для демонстрации внутренней взаимосвязи между данными. Наиболее популярным среди них является алгоритм k-средних. Он достаточно быстр и не требует больших вычислительных ресурсов для небольших наборов данных. Однако метод показывает неудовлетворительные результаты на больших наборах данных.

Анализ больших наборов данных требует наличия больших вычислительных мощностей, что не всегда осуществимо. А также результат кластеризации k-средних зависит от положения начальных центроидов кластеров, в связи с этим он может легко попасть в локальный оптимум.

Однако алгоритм k-средних довольно чувствителен к проблемам инициализации. Поэтому в [88, с. 622-633] был предложен подход к инициализации алгоритма k-средних. Он довольно прост в реализации, нетривиален и достаточно быстро сходится за небольшое количество итераций. Однако этот подход требует больших вычислительных затрат.

Для выполнения кластеризации многомерных данных необходимо использовать параллельную технологию. В связи с этим растет потребность в распараллеливании при кластеризации больших данных с сохранением основной структуры кластеризации и снижением вычислительных затрат. Поэтому параллельные алгоритмы на основе алгоритма k-средних используются для кластеризации данных в различных приложениях [148, с. 262-274]. Параллельная реализация на основе алгоритма кластеризации k-средних на кластере персональных компьютеров была описана Othman и соавторами. Предложенный алгоритм был распараллелен на основе присущего ему параллелизма данных, особенно в операциях вычисления расстояний и обновления центроида для большого набора данных. Временная сложность этого метода сильно зависит от количества итераций. Рост больших данных, которые занимают слишком много ресурсов памяти, привел к разработке различных методов и алгоритмов их кластеризации для решения этой проблемы [34]. Исследователи предлагают различные методы, основанные на k-средних, для решения этой проблемы в различных приложениях.

В 2018 году Ismkhan описал метод I-k-средних−+. Метод можно применять к образцам с большим количеством точек данных или большим количеством кластеров. Использование некоторых эвристик эффективно ускоряет I-k-средние−+. Однако, метод в 1.22 раза работает медленнее, чем алгорит k-средних для большого набора данных. В работе [53, с. 178-188] была предложена взвешенная кластеризация, которая была применена для улучшения качества кластеризации с применением алгоритма k-средних.

Результат суммирования весов каждой точки путем присвоения ее кластерам повышает точность кластеризации данных [49, с. 87-96].

Исходя из приведенного выше анализа исследований, постоянный рост объемов данных подтверждает необходимость разработки новых подходов для параллельной кластеризации. Однако анализ литературы в этой области выявил ограничения существующих подходов: а) хотя рассмотренные работы используют параллельную обработку данных, не все из них оценивают скорость кластеризации данных; б) большинство экспериментов проводились на небольших и средних наборах данных; в) многие методы имеют высокую вычислительную сложность по сравнению с алгоритмом k-средних.

В настоящее время метод параллельной обработки данных нашел широкое применение [148, с. 262-274; 229, с. 674-679; 337, с. 200-209]. Разбиение данных на небольшие блоки (батчи) позволяет решить проблему кластеризации за разумное время [62, с. 79-84].

При этом результат каждого батча должен быть достоверным и демонстрировать хороший результат для всего набора данных [56, с. 9-14]. Количество кластеров в отдельном батче может отличаться от их количества в исходном наборе данных, что является проблемой. Идеальную кластеризацию трудно реализовать, потому что трудно определить естественное количество кластеров [360, с. 1395-1412]. Также сложно выбрать оптимальный размер батча.

Данное исследование направлено на решение проблемы ускорения процесса кластеризации. Это было достигнуто за счет связывания параллелизации и разделения набора данных на батчи с использованием кластеризации на основе k-средних. В этом случае набор данных разбивается на несколько равных батчей. Результирующие кластеры создаются параллельно без полной загрузки памяти, что значительно ускоряет кластеризацию.

Актуальность исследования заключается в том, что использование батчей малой размерности снижает вычислительные затраты и увеличивает скорость сходимости алгоритма кластеризации.

Для оценки производительности предлагаемого подхода используются реальные большие наборы данных. Предлагаемый подход сравнивается с алгоритмами k-средних.

Мотивация данного исследования заключается в преодолении вышеуказанных ограничений. Основной вклад состоит в следующем:

•Предлагается новый метод кластеризации, основанный на параллельной пакетной кластеризации с использованием алгоритма k-средних. Основная идея заключается в том, что набор данных делится на несколько частей одинаковой размерности, которые затем параллельно кластеризуются для обнаружения центроидов кластеров и преодоления «проклятия размерности», присущего алгоритму k-средних при кластеризации больших данных. Затем алгоритм k-средних снова применяется к полученному массиву центроидов из всех батчей.

•Метод, предложенный Thompson в 1987 году, рассматривается для определения оптимального размера батчей.

• Эксперименты на наборах данных различной размерности показывают, что предлагаемый подход значительно сокращает время кластеризации по сравнению с классическим алгоритмом k-средних.

•Тест суммы рангов Уилкоксона применяется для проверки статистической значимости предлагаемого метода параллельной кластеризации больших данных КФС.

Предложенный подход. Предложен параллельный пакетный алгоритм на основе k-средних для кластеризации больших данных [59, с. 1-11]. После разделения набора данных на пакеты (батчи), каждый батч кластеризуется отдельно. Параллельное выполнение процесса продолжается до тех пор, пока не будет достигнуто условие сходимости. Сумма квадратов ошибок минимизируется для всех кластеров. Для каждого батча центроиды

объединяются в набор данных, который затем также группируется. Разбиение данных на батчи перед кластеризацией и применение параллельной обработки позволяет сократить время вычислений.

Пусть $X = \{x_1, x_2, \dots, x_n\}$ – конечное число точек, заданных в m-мерном пространстве, q – размер батча, максимальное значение $q (q < n)$ которого определяется параметрами ПК, а также обрабатывается за разумное время, $C = \{C_1, C_2, \dots, C_k\}$ – множество кластеров, где $C_p^q (p = 1,2, \dots, k)$ — p-ый кластер q-го батча, а k – количество кластеров, O_p^q — центроид p-го кластера в q-ом батче.

Целевая функция имеет следующий вид:

$$f(x) = \sum_{p=1}^{k} \sum_{x_i \in C_p} \left\| x_i - O_p^q \right\|^2 \longrightarrow \min, \quad (2.3.1)$$

$$O_p^q = \frac{\sum_{x_i \in C_p^q} x_i}{\left| C_p^q \right|}, \; p = 1,2, \dots, k, \quad (2.3.2)$$

где $\|.\|$ - Евклидова норма в $\Re^m$, $\left| C_p^q \right|$ – число точек данных в кластере C_p^q.

Результирующий центроид, полученный после применения k-средних к набору центроидов всех батчей, обозначается как $O_p^* \; (p = 1,2, \dots, k)$.

Предлагаемый алгоритм включает следующие шаги:

Вход: $X = \{x_1, x_2, \dots, x_n\}$: набор конечных точек

k: число искомых кластеров

Выход: Набор k кластеров $C = \{C_1, C_2, \dots, C_k\}$

Шаг 1: Определение оптимального размера батча q.

Шаг 2: Разделение набора данных из n элементов на набор батчей равного размера q.

Шаг 3: Применение кластеризации к каждому батчу:

Шаг 4: Инициализация центроидов кластеров.

Шаг 5: Повторить

Шаг 6: Вычислить значение функции (2.3.1) для каждой точки данных x_i.

Шаг 7: Пересчитать новые центроиды кластера согласно (2.3.2)

Шаг 8: пока не будет выполнено условие сходимости

Шаг 9: Создать набор центроидов, полученных на шаге 2.

Шаг 10: Повторить

Шаг 11: Вычислить значение функции (2.3.1) для каждого центроида из шага 3.

Шаг 12: Пересчитать новые центроиды кластера

Шаг 13: пока не будет выполнено условие сходимости

Шаг 14: Сопоставить точки данных с кластерами.

Шаг 15: Набор кластеров C.

На рис. 2.3.1 показана общая схема предлагаемого подхода. Набор данных разбит на батчи. В качестве метода кластеризации рассматривается алгоритм k-средних. Далее измеряется расстояние между точками данных и полученными центроидами кластеров каждого батча.

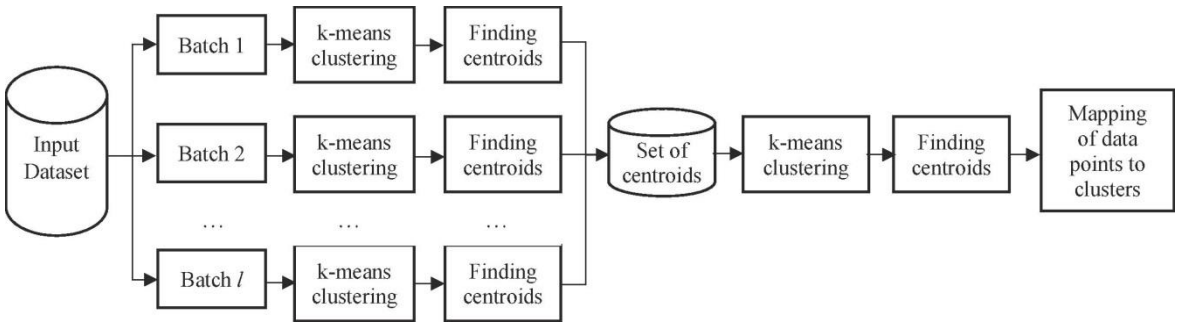


Рис. 2.3.1. Блок-схема предлагаемого алгоритма [54]

Алгоритм работает параллельно и итеративно. Обнаруженные центроиды каждого батча объединяются и повторно кластеризуются для получения результирующих центроидов.

Расчет размера батча. Набор данных делится на равные батчи. Для определения оптимального размера батча был рассмотрен метод, предложенный в [314, с. 1229-1244]:

$$q = \frac{\nu(\alpha)\cdot k^2}{r^2}, \qquad (2.3.3)$$

где α – желаемый уровень значимости, $\nu(\alpha)$ – значение, полученное из таблицы в [363, с. 42-46], k – количество кластеров, r – «относительная разница».

Предполагается, что α=0.05 (с вероятностью 95%), $\nu(\alpha)$=1.27359 и r=0.08. Согласно (2.3.3) объем батча при k=5 будет равен $q = \frac{1.27359 \cdot 5^2}{0.08^2} = 4975$, а при k=10 – $q = \frac{1.27359 \cdot 10^2}{0.08^2} = 19900$. Таким образом, можно увидеть применимость этого подхода для всех кластеров набора данных.

В случае, когда рассматриваемый набор данных невозможно разделить на равные батчи, оставшаяся часть данных переходит на Шаг 5 алгоритма, где точки сопоставляются с каждым кластером, полученным на Шаге 4.

Результаты экспериментов. Рассматриваются два реальных набора данных для оценки эффективности предложенного подхода (Таблица 2.3.1). Они были взяты из репозитория машинного обучения UCI [267].

Таблица 2.3.1

Описание экспериментальных наборов данных

Набор данных	Количество образцов	Количество атрибутов	Размер файла	Ссылки
Individual Household Electric Power Consumption	2075259	9	142.50 Gb	[223, с. 164-173]
Phone Accelerometer	1048575	4	32.00 Mb	[357]

Экспериментальный набор данных Individual Household Electric Power Consumption содержит измерения потребления активной энергии, используемой каждую минуту (в ватт-часах) в течение почти четырех лет в одном домашннем хозяйстве. Общее количество образцов составило 2075259. Доступны девять различных электрических величин и некоторые значения измерений, такие как напряжение, global_active_power, global_intensity, global_reactive_power и другие. Все квлендарные временные метки присутствуют в наборе данных.

Набор данных Phone Accelerometer содержит 1048575 образцов и 4 атрибута.

В экспериментах эти наборы данных разбиваются на равные фрагменты (батчи). Каждый батч имеет одинаковый размер (5000, 10000, 15000 и 20000

образцов). k образцов случайным образом выбираются в батче для инициализации центроидов.

В работе предложенный подход сравнивается с алгоритмом k-средних при количестве кластеров, равным 2, 3, 5, 10 и 15.

Алгоритмы выполнялись по десять раз, чтобы сравнить эффективность предложенного алгоритма и k-средних. Учитывались средние значения целевой функции и времени. Все эксперименты проводились в одной и той же вычислительной среде для получения объективных результатов. Все алгоритмы были запрограммированы в Matlab 2018a. Алгоритмы выполнялись на виртуальной машине Intel Xeon (R), процессор X5670 @ 2,93 ГГц * 24 с 24 ГБ RAM.

Сравнение с алгоритмом k-средних. В целом наилучшая производительность наблюдается для алгоритма k-средних по значению целевой функции. Однако по сравнению с предложенным подходом вычислительная эффективность k-средних ниже.

Предложенный подход требует небольших вычислительных ресурсов и является достаточно быстрым.

Однако алгоритм k-средних не может сгруппировать большие наборы данных с большим количеством кластеров (k=10 и k=15) [54]. Превосходство предложенного подхода может показать только время выполнения CPU.

Предложенный подход значительно сокращает время вычислений. Эффективность кластеризации и время выполнения (время CPU для алгоритма за один запуск) сравнивались между предоженным подходом и алгоритмом k-средних на реальных наборах данных (Таблица 2.3.2). $F_{k-means}^{mean}$ и $T_{k-means}^{mean}$ обозначают средние значения целевой функции и времени, полученные с помощью k-средних, а F^{mean} и T^{mean} – средние значения, полученные с помощью предложенного подхода.

Результаты экспериментов на наборах данных Phone Accelerometer и Individual Household Electric Power Consumption с использованием

предложенного подхода представлены в таблицах 2.3.2-2.3.3. В таблицах показано относительное улучшение (%) предложенного подхода по сравнению с алгоритмом k-средних по средним значениям целевой функции и времени. Знак «+» указывает на улучшение кластеризации при применении предложенного подхода, а знак «-» указывает на ее ухудшение по сравнению с алгоритмом k-средних.

Таблица 2.3.2

Результаты экспериментов для набора данных Phone Accelerometer

Количество кластеров	$F^{mean}_{k-means}$	$T^{mean}_{k-means}$ (сек)	Размер батча	F^{mean}	%	T^{mean} (сек)	%
***k*=2**	4.1942E+10	5.22	5000	4.1934E+10	+0.02	1.15	+353.91
			10000	4.1930E+10	+0.03	1.08	+383.33
			15000	4.1911E+10	+0.07	0.99	+427.27
			20000	4.1947E+10	-0.01	0.97	+438.14
***k*=3**	2.5774E+10	6.99	5000	2.5749E+10	+0.10	2.72	+156.99
			10000	2.5760E+10	+0.05	2.57	+171.98
			15000	2.5774E+10	0.00	2.47	+183.00
			20000	2.5734E+10	+0.16	2.55	+174.12
***k*=5**	1.5673E+10	22.41	5000	1.5671E+10	+0.01	7.81	+186.94
			10000	1.5676E+10	-0.02	7.35	+204.90
			15000	1.5690E+10	-0.11	10.10	+121.88
			20000	1.5702E+10	-0.18	9.75	+129.85
***k*=10**	7.8899E+9	160.91	5000	7.8868E+9	+0.04	30.19	+432.99
			10000	7.9019E+9	-0.15	35.64	+351.49
			15000	8.0193E+9	-1.61	123.81	+29.97
			20000	8.0872E+9	-2.44	121.84	+32.07
***k*=15**	5.2791E+9	339.77	5000	5.2758E+9	+0.06	84.68	+301.24
			10000	5.3325E+9	-1.00	272.66	+24.61
			15000	5.4164E+9	-2.53	301.75	+12.60
			20000	5.3990E+9	-2.22	324.70	+4.64

В таблице 2.3.2 оцениваются средние значения целевой функции и среднее время выполнения алгоритма для набора данных Phone Accelerometer. Показано увеличение производительности предложенного подхода для различных значений k и размера батча для данного набора данных. При k=3 и размере батча равном 20000, среднее значение целевой функции меньше, чем у алгоритма k-средних.

Метод достигает хорошей производительности после шести итераций в соответствии со значениями целевой функции.

Относительное улучшение среднего времени выполнения по сравнению с алгоритмом к-средних составляет 174.12%. При размере батча 10000 и k=5 относительное улучшение среднего времени выполнения составляет 204.90%.

Таблица 2.3.3 показывает, что среднее время вычислений для предложенного параллельного алгоритма значительно меньше, чем для k-средних. Из таблицы видно, что изменение количества кластеров не оказывает существенного влияния на значение целевой функции, а при k=3 и размере батча 5000 наблюдается улучшение (0.02%).

Таблица 2.3.3

Результаты экспериментов для набора данных Individual Household Electric Power Consumption

Количество кластеров	$F_{k-means}^{mean}$	$T_{k-means}^{mean}$ (сек)	Размер батча	F^{mean}	%	T^{mean} (сек)	%
***k*=2**	1.0145E+9	4.28	5000	1.0146E+9	-0.01	2.53	+69.17
			10000	1.0151E+9	-0.06	2.25	+90.22
			15000	1.0149E+9	-0.04	2.34	+82.91
			20000	1.0154E+9	-0.09	2.41	+77.59
***k*=3**	9.2519E+8	6.51	5000	9.2501E+8	+0.02	3.87	+68.22
			10000	9.2559E+8	-0.04	3.82	+70.42
			15000	9.2608E+8	-0.10	3.72	+75.00
			20000	9.2796E+8	-0.30	4.06	+60.34
***k*=5**	6.5133E+8	103.75	5000	6.5435E+8	-0.46	7.49	+1285.18
			10000	6.5482E+8	-0.53	7.46	+1290.75
			15000	6.5760E+8	-0.95	7.68	+1250.91
			20000	6.5463E+8	-0.50	11.29	+818.95
***k*=10**	4.6718E+8	394.33	5000	4.6853E+8	-0.29	13.02	+2928.65
			10000	4.6940E+8	-0.47	21.26	+1754.80
			15000	4.7084E+8	-0.78	59.95	+557.76
			20000	4.7103E+8	-0.82	72.21	+446.09
***k*=15**	3.7862E+8	489.97	5000	3.7995E+8	-0.35	17.82	+2649.55
			10000	3.8031E+8	-0.44	69.33	+606.72
			15000	3.8207E+8	-0.90	97.37	+403.20
			20000	3.8250E+8	-1.01	99.29	+393.47

Среднее время работы алгоритма значительно улучшается с увеличением количества кластеров для набора данных Individual Household Electric Power Consumption.

Согласно таблицам 2.3.2-2.3.3, средние значения целевой функции практически совпадают с алгоритмом k-средних для всех рассмотренных наборов данных.

Сравнение времени выполнения предложенного подхода и k-средних для двух реальных наборов данных показано на рис. 2.3.2. Каждая диаграмма суммирует результаты десяти последовательных запусков предлагаемого подхода с размером батча равном 20000. Экспериментальные результаты показывают, что время выполнения предложенного алгоритма уменьшилось по сравнению с алгоритмом k-средних. Из рис. 2.3.2 (а, б) видно, что разница между результатами предложенного подхода и алгоритмом k-средних не сильно отличается для набора данных Phone Accelerometer при k=2, k=3 и k=5. Меньше времени на кластеризацию было потрачено для данного набора данных по сравнению с наборов данных Individual Household Electric Power Consumption при небольшом количестве кластеров.

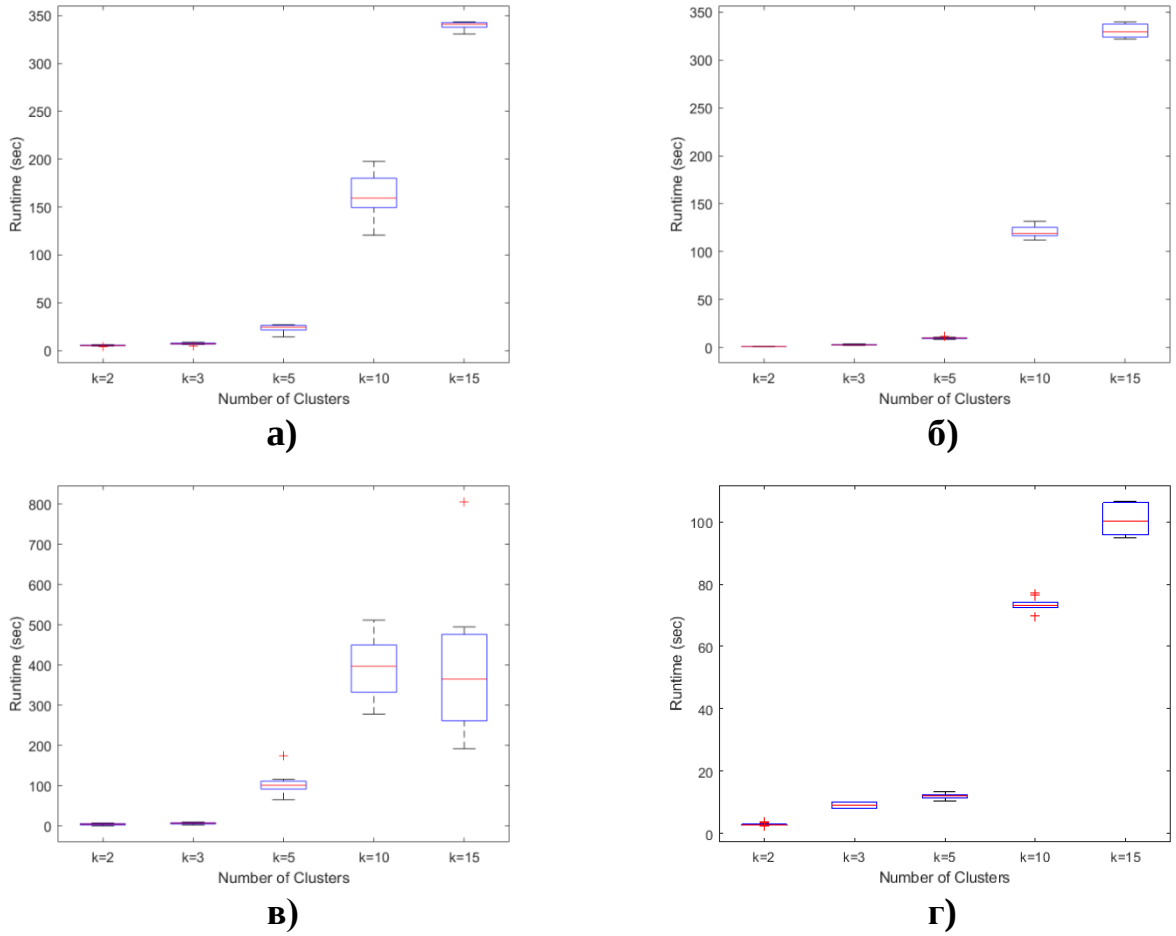


Рис. 2.3.2. Сравнение производительности предложенного подхода и алгоритма k-средних на наборах данных Phone Accelerometer (а, б) и Individual Household Electric Power Consumption (в, г) [54]

Таким образом, с увеличением количества кластеров, требующих огромных ресурсов памяти, производительность предложенного подхода улучшается по сравнению с алгоритмом k-средних.

Оценка производительности параллельной пакетной кластеризации. Был рассмотрен анализ предложенной параллельной пакетной кластеризации на основе k-средних для различного количества компьютерных узлов на двух наборах данных при k=15 и размере батча равном 20000.

Изменение количества компьютерных узлов (1, 2, 4, 6 и 8) показано на рис. 2.3.3.

Таким образом, время выполнения уменьшается с увеличением количества узлов для предложенного метода. Время почти постоянно после четырех узлов, а при восьми узлах наблюдается значительное улучшение.

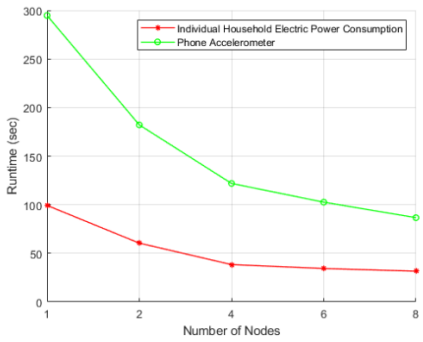


Рис. 2.3.3. Производительность предложенного метода при различном количестве узлов для двух наборов данных [54]

Это доказывает, что использование параллельной пакетной кластеризации повышает скорость предложенного метода примерно в два раза по сравнению с его последовательной версией. Для дополнительной оценки предложенного метода был проведен анализ при различном количестве батчей на примере набора данных Individual Household Electric Power Consumption (Рис. 2.3.4).

При k=2 и k=3 предложенный метод мало зависит от изменения количества батчей. Предлагаемый подход обеспечивает более заметное улучшение с увеличением количества кластеров и батчей. Таким образом, можно увеличить количество батчей (>200), что и было сделано ранее.

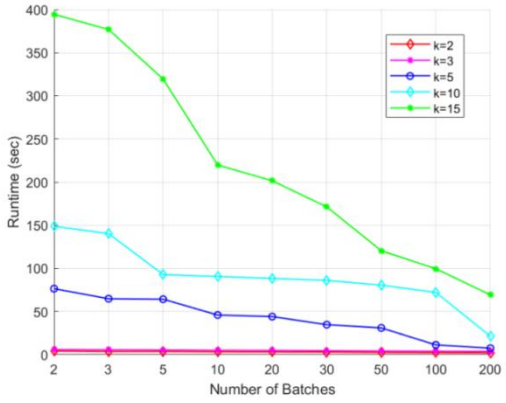


Рис. 2.3.4. Производительность предложенного метода при различном количестве батчей для набора данных Individual Household Electric Power Consumption [54]

Таким образом, на рис. 2.3.5 показаны временные характеристики предложенного метода для наборов данных различной размерности.

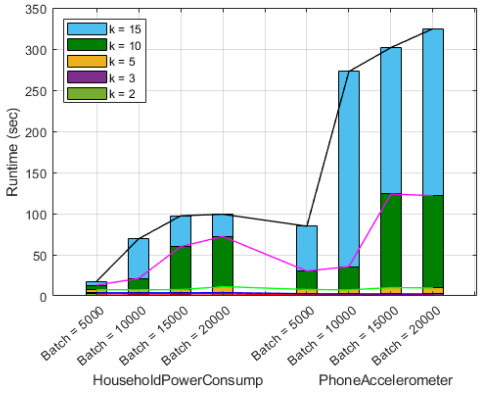


Рис. 2.3.5. Среднее время работы предложенного метода для различных наборов данных [54]

Размеры наборов данных варьируются от 1048575 выборок до 2075259. Приводится время выполнения предлагаемого метода в зависимости от изменения размера батча.

Временные шкалы при k=2 и k=3 почти перекрывают друг друга из-за их схожих временных характеристик. При количестве кластеров равном 10 и 15 временные шкалы начинают резко отклоняться друг от друга с увеличением размера батча.

На рис. 2.3.5 показано, что предложенный метод демонстрирует значительное преимущество в эффективности при использовании батчей

небольшой размерности. Например, при k=5 и размере батча, равном 5000, предлагаемый подход требует на выполнение в среднем 7.81 сек. для набора данных Phone Accelerometer и 7.49 сек. для набора данных Individual Household Electric Power Consumption, соответственно.

Статистический анализ результатов. Также был проведен статистический анализ результатов, полученных с помощью критерия суммы рангов Уилкоксона для 5% уровня значимости [212, с. 573-601]. Полученные p-значения для предложенного метода и алгоритма k-средних с использованием разного количества кластеров при размере батча равном 20000 для всех наборов данных показаны в таблице 2.3.4.

Таблица 2.3.4

***p*-значения теста Уилкоксона при оценке производительности предложенного метода (размер батча 20000) для алгоритма k-средних**

Набор данных	Алгоритм k-средних				
	k=2	k=3	k=5	k=10	k=15
Phone Accelerometer	1.8267E-4	1.8267E-4	1.8267E-4	5.8284E-4	0.0058
Individual Household Electric Power Consumption	0.0452	0.0091	1.8267E-4	1.8267E-4	1.8267E-4

В качестве нулевой гипотезы предполагается, что разницы между медианными значениями для двух методов практически нет. Полученные p-значения, показанные в таблице, малы в соответствии с уровнем значимости, что является доказательством против нулевой гипотезы. Следовательно, результаты предложенного метода для всех наборов данных являются статистически значимыми и возникли не случайно.

Анализ наборов данных большого объема требует наличия больших вычислительных мощностей, что не всегда осуществимо. Поэтому необходимость создания новых методов анализа больших наборов данных КФС для сокращения временных затрат привела к разработке подхода кластеризации данных на основе параллельных разбиений на батчи. Разбиение данных на батчи помогает сократить время отклика алгоритма кластеризации и

вычислительные затраты, несмотря на увеличение количества кластеров. Было отмечено увеличение примерно в два раза скорости параллельной реализации предлагаемого метода на многих узлах.

В этом исследовании для сравнения с целью изучения производительности предложенного подхода использовались наборы данных различной размерности. На основании результатов экспериментов была доказана вычислительная эффективность метода по сравнению с алгоритмом k-средних. Эффективный выбор начальных центроидов, который остается важной проблемой кластеризации, повышает точность алгоритма параллельной пакетной кластеризации. Анализируя влияние размера батча на среднее значение функции и среднее время выполнения метода, можно сделать вывод, что эффективность достигается даже при небольшом размере батча. Несмотря на увеличение количества кластеров, скорость предложенного подхода значительно выше по сравнению с алгоритмом k-средних. Метод способен кластеризовать большие данные КФС за короткое время и может применяться для анализа больших данных с целью снижения риска выхода системы из строя.

2.4. Метод кластерного анализа больших данных киберфизических систем на основе консенсусного ансамбля

Для извлечения знаний из существующей растущей информации актуальна необходимость разработки подходов на основе машинного обучения. Практика показывает, что кластеризация применяется для исследования и оценки кластеризованных знаний. Таким образом, кластеризованные экспертные системы разрабатываются для объяснения данных при диагностике и принятии окончательных решений.

Объем информации, которую необходимо обработать экспертам, стремительно растет с развитием технологий. Необходимо разрабатывать новые интеллектуальные подходы на основе машинного обучения, чтобы помочь экспертам принимать консенсусные решения из возможного набора альтернатив. Сейчас наблюдается тенденция использования ансамблевых

методов в кластерном анализе, когда сложно применить определенный алгоритм для кластеризации данных.

Одной из проблем кластеризации является высокая вычислительная стоимость, что затрудняет получение хороших результатов на сложных больших данных КФС.

Кластеризация данных требует разработки общего алгоритма и выбора наилучших критериев принятия решений. Однако из-за большого разнообразия существующих методов сложно выбрать наиболее эффективный. Исследователи разработали множество подходов к кластеризации, которые учитывают преимущества и недостатки различных методов для построения одной окончательной кластеризации. В 2007 году Nguyen и Caruana представили три алгоритма, подобных методу максимизации ожиданий (консенсус итерационного голосования, консенсус итерационного вероятностного голосования и консенсус итерационного попарного голосования) для решения проблемы консенсусной кластеризации [302, с. 607-612].

Другим примером применения консенсусной кластеризации являются медицинские экспертные системы. Интегративная статистическая модель для одновременной оценки как консенсусной кластеризации, так и кластеризации, специфичной для источника, была предложена Lock и Dunson в 2013 году. Подход показал гибкие и масштабируемые в вычислительном отношении результаты при кластеризации больших обьемов данных. Ограничение метода - высокие вычислительные затраты [60, с. 1-15].

Предложенные алгоритмы являются вариациями алгоритма k-средних, использующими различные меры расстояния, применяемые к вектору кластеризации базового уровня. Однако ограничением этого подхода являются высокие вычислительные затраты. Гомогенный кластерный ансамбль на основе алгоритма кластеризации роя частиц был предложен De Oliveira и соавторами в 2017 году [160, с. 141-153]. Его можно рассматривать как двухэтапную

процедуру: сначала из данных получают несколько базовых разбиений, затем окончательное разбиение вычисляется с помощью консенсусной функции, которая принимает базовые разбиения в качестве входных данных. Предложенный кластерный ансамбль использует алгоритма роя частиц как для генерации базовых разбиений, так и в консенсусной функции и поддерживает как непересекающиеся, так и перекрывающиеся разбиения. Ограничения метода включают чувствительность к вопросам инициализации, а также высокую вычислительную стоимость.

Консенсусный подход широко используется для повышения точности и устойчивости результатов кластеризации [97, с. 427-436; 302, с. 607-612]. Подход заключается в поиске согласованного решения за счет возможности совместного использования методов кластерного анализа. Построить наиболее подходящую схему кластеризации для конкретной области можно, применив консенсусный подход к разному набору алгоритмов в соответствии с их преимуществами и отличительными особенностями. При выработке окончательного решения рассматриваются разные точки зрения, которые не только не противоречат, но, наоборот, компенсируют недостатки каждого метода.

Консенсусная кластеризация помогает создавать «надежные» разбиения, обрабатывать шумы и выбросы [302, с. 607-612; 160, с. 141-153]. Она действует как многообещающее решение для кластеризации данных. Много работ было посвящено консенсусной кластеризации на основе k-средних, но исследования все еще носят предварительный и фрагментарный характер. В то же время консенсусная кластеризация ищет общее разбиение, которое согласуется с существующими методами одиночной кластеризации. В данном случае проблемой является разработка функции полезности и ее эффективная оптимизация. С этой целью предлагается функция полезности, основанная на чистоте [63, с. 1-4; 64, с. 17-20].

Подводя итоги анализа состояния исследований по применению

консенсусной кластеризации к большим наборам данных, можно сделать следующие выводы. Работы, направленные на повышение качества кластеризации, зачастую требуют больших вычислительных ресурсов. Также консенсусная кластеризация является достаточно популярным направлением исследований в связи с непрерывным ростом объемов данных. Это подтверждает актуальность данного исследования.

Данное исследование включает:

•Взвешенную консенсусную кластеризацию для эффективного объединения отдельных разбиений для кластерного анализа больших данных КФС. Результаты исследования показывают, что взвешенная консенсусная кластеризация на основе функции полезности чистоты является эффективной, понятной и надежной.

•Предложенный подход может быть применен к любой экспертной системе. Он разработан для любого количества альтернатив и может быть реализован в любой архитектуре КФС. Кластеризация с взвешенным консенсусом позволит экспертам использовать ее при принятии групповых решений.

•Экспериментальные результаты на реальных наборах данных разного размера с использованием метрик Евклида, Минковского, квадрата Евклида, косинуса и расстояния Чебышева показывают, что кластеризация с применением взвешенного консенсуса является высокоэффективной, сопоставимой и превосходящей современные подходы по качеству кластеризации данных КФС.

Консенсусная кластеризация – это многообещающее решение для поиска кластеров в многомерных данных из-за его надежности и универсальности [97, с. 427-436].

Предлагаемый подход. Пусть $X = \{x_1, x_2, \dots, x_n\}$ – точки в наборе данных, где n – общее количество точек в наборе данных [60, с. 1-15]. $x_i = \{x_{i1}, x_{i2}, \dots, x_{im}\} \in R^m$ — точка в наборе данных, где m — размерность точек

данных. Разбиение X на K четких кластеров представляется как набор K подмножеств объектов в $C = \{C_k | k = 1, \dots, K\}$ с $C_k \cap C_{k'} = \emptyset \; \forall k \neq k'$ и $\bigcup_{k=1}^{K} C_k = X$ или как вектор меток $\pi = (L_\pi(x_1), L_\pi(x_2), \dots, L_\pi(x_n))^T$ для любого $i \;\; x_i \xrightarrow{L_\pi} \{1,2, \dots, K\}$.

Существующие методы консенсусной кластеризации можно разделить, как правило, на две категории: методы с глобальными целевыми функциями и без них [368, с. 337-372]. Учитывая базовые разбиения r от X до $\Pi = \{\pi_1, \pi_2, \dots, \pi_r\}$ (базовое разбиение является результатом одного алгоритма кластеризации (например, DBSCAN, OPTICS и т. д.). И есть K_i-кластеров в π_i для $1 \leq i \leq r$. Цель состоит в том, чтобы найти консенсусное разбиение π путем решения следующей оптимизационной задачи:

$$\pi^* = \underset{\pi}{\operatorname{argmax}}(w_i U(\pi, \pi_i)), \tag{2.4.1}$$

где π^* является функцией консенсуса, U является функцией полезности, которая измеряет сходство между π и любым π_i, а разбиение π_i является базовым разбиением и $w_i \in [0,1]$, $\sum_{i=1}^{r} w_i = 1$.

Другими словами, необходимо найти оптимальное разбиение, наиболее соответствующее базовому разбиению. Различные функции полезности измеряют сходство двух разбиений в разных аспектах, предоставляя разные целевые функции для консенсусной кластеризации. Предлагаемый метод использует функцию полезности, основанную на чистоте, для объединения всех основных разбиений в согласованное, которое принимает решения по общему согласию:

$$f = (1 - \lambda) \cdot \sum_{i=1}^{r} w_i U(\pi, \pi_i) + \lambda \cdot \|w\|^2 \longrightarrow \max \tag{2.4.2}$$

при условии

$$\sum_{i=1}^{r} w_i = 1, w_i \geq 0, \forall i, \tag{2.4.3}$$

где $0 \leq \lambda \leq 1$ – параметр регуляризации, который указывает на компромисс между максимизацией взвешенной функции полезности и гладкостью, обеспечиваемой w. В экспериментах λ будет определяться экспериментально.

Используется евклидово расстояние, расстояние Минковского для p=3 и p=4, квадрат евклидова расстояния, косинусное расстояние и расстояние Чебышева.

Первый член в (2.4.2) используется для минимизации взвешенного расстояния между отдельными разбиениями и π. Второй член в уравнении (2.4.3) является членом регуляризации для обеспечения гладкости весов.

Цель метода состоит в том, чтобы объединить эти основные разбиения в метод PWCC (Partitions Weighted Consensus Clustering), который гарантирует высокую эффективность и надежность алгоритма консенсусной кластеризации. Как сформулировано в уравнении (2.4.1), функция полезности определяется для двух разбиений π и π_i и измеряет их сходство на уровне разбиения. Для расчета функции полезности чистоты можно использовать следующую ассоциативную таблицу (Таблица 2.4.1).

Таблица 2.4.1

Матрица коассоциаций

		π_i			
		$C_1^{(i)}$	$C_2^{(i)}$	$\cdots$	$C_{K_i}^{(i)}$	$\sum$
π	C_1	$n_{11}^{(i)}$	$n_{12}^{(i)}$	$\cdots$	$n_{1K_i}^{(i)}$	n_{1+}
	C_2	$n_{21}^{(i)}$	$n_{22}^{(i)}$	$\cdots$	$n_{2K_i}^{(i)}$	n_{2+}
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	C_K	$n_{K1}^{(i)}$	$n_{K2}^{(i)}$	$\cdots$	$n_{KK_i}^{(i)}$	n_{K+}
	$\sum$	$n_{+1}^{(i)}$	$n_{+2}^{(i)}$	$\cdots$	$n_{+K_i}^{(i)}$	n

Здесь π содержит K кластеров и π_i - K_i кластеров, $n_{kj}^{(i)}$ обозначает количество точек, содержащихся как в кластере $C_j^{(i)}$ в π_i, так и в кластере C_k в π, $n_{k+} = \sum_{j=1}^{K_i} n_{kj}^{(i)}$ – это количество точек в C_k, $n_{+j}^{(i)} = \sum_{k=1}^{K} n_{kj}^{(i)}$ – это количество точек в $C_j^{(i)}$, n – это общее количество точек.

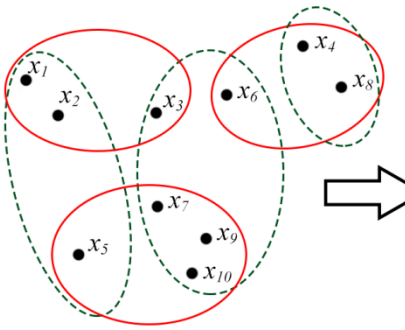
На рис. 2.4.1 показан пример матрицы коассоциации для набора данных из 10 точек. На нем показано одно базовое разбиение π_1, выделенное зеленым, и истинная кластеризация, отмеченная красным.

Пусть $m_{jk}^{(i)} = \left(\max n_{j1}^{(i)}, \max n_{j2}^{(i)}, \ldots, \max n_{jK_{(i)}}^{(i)}\right)$ и $m_k^{(i)} = n_{+k}^{(i)}$, тогда можно определить функцию полезности для измерения сходства между двумя разбиениями, представленными в таблице 2.4.1.

Функция полезности чистоты имеет следующий вид:

$$U(\pi, \pi_i) = \sum_{k=1}^{K_i} \frac{n_{+k}^{(i)}}{n} \max_j n_{jk}^{(i)} = \sum_{k=1}^{K_i} \frac{m_k}{n} m_{jk}^{(i)}. \quad (2.4.4)$$

Пусть имеются базовые разбиения r. Задача состоит в том, чтобы найти взвешенный консенсус π с набором весов $\{w_1, w_2, \ldots, w_r\}$, присвоенных каждому методу и $w_i = \frac{1}{r}, \ \forall i$.



	$C_1^{\pi_1}$	$C_2^{\pi_1}$	$C_3^{\pi_1}$	Σ
C_1^{π}	2	1	0	3
C_2^{π}	1	3	0	4
C_3^{π}	0	1	2	3
Σ	3	5	2	10

Рис. 2.4.1. Иллюстрация матрицы коассоциаций для трех кластеров [55]

Метод DBSCAN (Density-Based Spatial Clustering of Applications with Noise) основан на концепции внутренних и граничных точек, плотности достижимости, D-связности, порога ε и минимального количества точек в кластере (MinPts) [178, с. 226-231]. Под ε-соседями точки $p \in D$ понимается множество точек, расстояние до которых не превосходит ε.

$$N_\varepsilon(p) = \{q \in D | dist(p, q) \leq \varepsilon\}. \quad (2.4.5)$$

MinPts подбирается экспериментально так, чтобы $|N_\varepsilon(q)| \geq MinPts$. *MinPts* регулирует порог «шума». Этот алгоритм кластеризации прост в вычислительном отношении и устойчив к помехам.

Чувствительность к выбору параметров алгоритма DBSCAN породила ряд его модификаций. Одним из них является OPTICS (Ordering Points to Identification the Clustering Structure) [78, с. 49-60], который позволяет

упорядочить исходный набор и упростить процесс кластеризации.

В этом методе строится диаграмма достижимости (reachability diagram), благодаря которой появляется возможность при фиксированном значении *MinPts* обрабатывать не только заданное значение ε, но и все $\varepsilon^* < \varepsilon$.

Для упорядочения набора *D* для каждого его элемента вычисляются два параметра – расстояние ядра и расстояние достижимости. OPTICS позволяет решать задачи кластеризации в условиях, когда кластеры имеют не только разную форму, но и разное распределение плотности данных в каждом классе.

CLARANS (Clustering Large Applications with Randomized Search) был предложен для кластеризации больших приложений с рандомизированным поиском и сочетает в себе преимущества алгоритмов PAM и CLARA [239, с. 123-150]. Основная идея PAM состоит в том, чтобы выбрать одну точку из каждого кластера в качестве медоида. В отличие от k-средних, метод включает медоид вместо среднего, что делает алгоритм более эффективным.

CLARA — лучшее решение для обработки больших данных по сравнению с PAM. Но его недостаток в том, что он рассматривает образцы данных, а не полные наборы данных. В связи с этим для повышения эффективности кластеризации был предложен алгоритм CLARANS. Он использует метод выборки для сокращения пространства поиска [28, с. 243-248]. Этот подход выполняется динамически на каждой итерации. При этом процесс кластеризации представляет собой поиск по графу, вершиной которого является множество медоидов. Каждому узлу назначается функция стоимости [98, с. 78-95]:

$$Cost(X, M) = \frac{\sum_{i=1}^{n} dist(x_i, rep(M, x_i))}{n}, \quad (2.4.6)$$

где *M* - набор медоидов, $rep(M, x_i)$ возвращает ближайший к x_i медоид *M*.

CLARANS, как и PAM, перемещается от одного узла к одному из своих соседей, пока не найдет решение с минимальной стоимостью. Алгоритм CLARANS динамически сокращает пространство поиска.

Алгоритм k-средних строит k кластеров, расположенные таким образом,

чтобы минимизировать стандартное отклонение выборок объекта от центров кластеров [236, с. 881-892; 80, с. 1027-1035; 88, с. 622-633]. В то же время исходное расположение кластеров сильно влияет на алгоритм [118, с. 56-69]. Целевой функцией алгоритма является среднеквадратичное расстояние (евклидова метрика) между выборками объектов и центрами их кластеров:

$$f(X, C) = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i - O_j \right\|^2, \quad (2.4.7)$$

где O_j - центр кластера C_j, вычисляемый согласно (2.4.8)

$$O_j = \frac{1}{|C_j|} \sum_{i=1}^{n} x_i. \quad (2.4.8)$$

Алгоритм SNNC (Shared Nearest Neighbor Clustering) был предложен Jarvis и Patrick, в котором связь создается между парой точек p и q тогда и только тогда, когда p и q расположены близко друг к другу в их k-ближайшем соседе [232, с. 1025-1034; 176, с. 1-15]. Этот алгоритм является расширением DBSCAN.

Основная идея SNNC опирается на определение основных точек, вокруг которых строятся кластеры различных размеров и форм, не определяя их количества [278, с. 76-85; 144, с. 21-27]. Подсчет количества точек, общих между двумя точками p и q в их списке k-ближайших соседей на основе метрики расстояния, позволяет определить сходство между ними. Чем больше количество общих точек, тем выше сходство между p и q.

Метрики оценки. Рассматриваются внешние и внутренние индексы для оценки результатов кластеризации. Внешний индекс относится к сравнению решения кластеризации с реальной кластеризацией. Это важно при оценке производительности алгоритма кластеризации на больших наборах данных. Внутренний индекс оценивает кластеры по их структурным свойствам. Внутренняя проверка кластера направлена на измерение качества кластеризации в реальных условиях, когда нет знаний о реальной кластеризации.

Внешние метрики. Метрики оценки методов кластеризации с учетом

факторов компактности и разделимости указывают на правильность разделения на кластеры.

Пусть набор данных D разделен на классы $C^+ = (C_1^+, \dots, C_{k^+}^+)$ (истинная кластеризация). И с помощью процедуры кластеризации к этому набору данных были получены кластеры $C = (C_1, \dots, C_k)$.

Сравнение решений методов кластеризации основано на подсчете количества совпадений C^+ и C. По результатам принимается решение: аномальное/нормальное поведение.

Наиболее известными метриками для оценки по расстоянию методов кластеризации на основе пар точек данных являются Purity [105, с. 376-381; 335, с. 849-865], метрика Mirkin [285, с. 35-47], PC [99, с. 301-315], VI [316, с. 902-916], F-мера [334, с. 410-420].

Внутренние метрики. Индекс Davies-Bouldin (DB) и индекс Calinski-Harabasz (CH) рассматривались в качестве внутренних индексов для оценки немаркированных наборов данных.

Индекс DB основан на соотношении расстояний внутри кластера и между кластерами [153, с. 224-227]. DB рассчитывается следующим образом:

$$DB = \frac{1}{k}\sum_{i=1}^{k} R_i \tag{2.4.9}$$

$$R_i = \max_{i \neq j}\left(\frac{\delta(C_i)+\delta(C_j)}{dist(C_i,C_j)}\right), \tag{2.4.10}$$

где k – количество кластеров, δ – дисперсия внутри кластера, $dist$ – расстояние между i и j кластерами. Целевое значение — это минимум индекса.

Индекс CH [116, с. 1-27] характеризуется следующей функцией:

$$CH = \frac{B(k)/(k-1)}{W(k)/(n-k)}, \tag{2.4.11}$$

где

$$B(k) = \sum_{i=1}^{k} n_i \, dist^2(O_i, O), \tag{2.4.12}$$

$$W(k) = \sum_{i=1}^{k} \sum_{x \in C_i} dist^2(x, O_i), \tag{2.4.13}$$

где k – количество кластеров, n – количество объектов в рассматриваемом

наборе данных D, C_i – i-ый кластер, n_i – количество объектов в C_i, O – центр набора данных D, O_i – центр C_i в наборе данных, $W(k)$ – сумма внутрикластерных дисперсий для всех кластеров и $B(k)$ представляет собой взвешенную сумму квадратов расстояний между C_i и набором данных D. Наиболее вероятным числом кластеров является значение k, при котором индекс *CH* достигает своего максимального значения [116, с. 1-27].

Результаты экспериментов. Был проведен ряд экспериментов для оценки продуктивности предложенного подхода. Эксперименты проводились на платформе операционной системы Windows® 10–64 бит с процессором Core i7 2,5 ГГц, ОЗУ 8,0 ГБ. Предлагаемый подход был оценен на R 3.4.1.

Три набора данных из репозитория UCI [174, с. 1001-1005; 267], включая наборы данных Phishing, Phone accelerometer и NSL-KDD [35, с. 842-851] использовались в качестве входных данных. Для экспериментов обучающие и тестовые наборы в последнем наборе данных (148517 образцов) были объединены. Эти наборы данных имеют малый, средний и большой размер. Характеристики наборов данных представлены в таблице 2.4.2.

Таблица 2.4.2

Описание наборов данных

Набор данных	Количество образцов	C_1^+	C_2^+	Количество атрибутов
Phishing	11,055	4,898	6,157	30
NSL-KDD	148,517	71,463	77,054	41
Phone Accelerometer	13,062,475	6,240,983	6,821,492	6

Все наборы данных содержат два класса: C_1^+ и C_2^+. Образцы в классе C_1^+ считаются аномалиями. Значения в наборах данных были стандартизированы во время предварительной обработки.

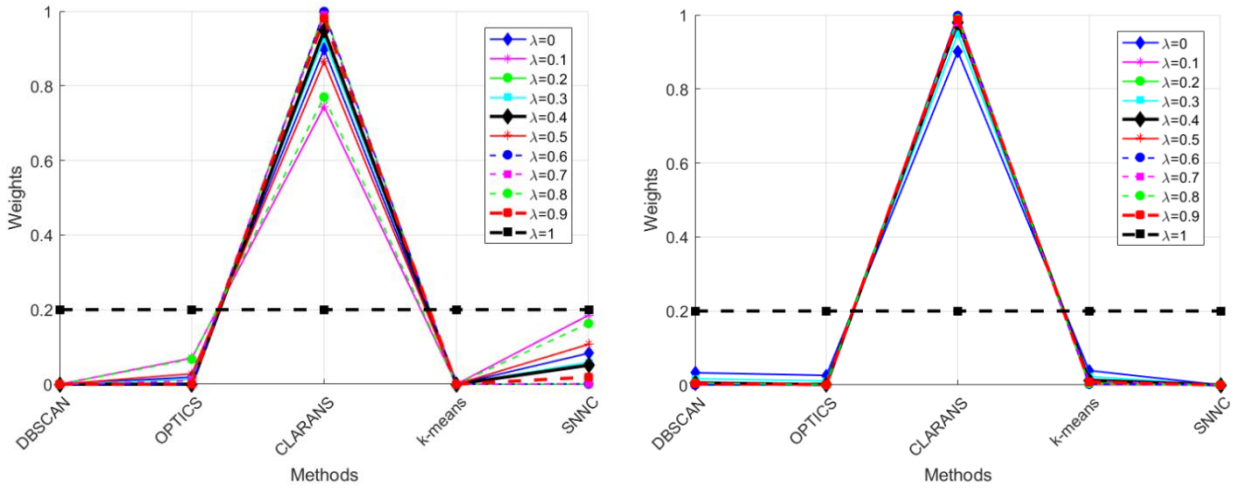
Эксперименты были сосредоточены на сравнении результатов кластеризации предложенного подхода с различными метриками расстояния (евклидово расстояние (Eucl), косинусное расстояние (Cos), квадрат евклидова расстояния (SqEucl), расстояние Минковского при p=3 (Mink3), расстояние

Минковского при p=4 (Mink4) и расстояние Чебышева (Cheb)) с отдельными методами кластеризации (DBSCAN [178, с. 226-231], OPTICS [78, с. 49-60], CLARANS, k-means и SNNC [318, с. 247-276]).

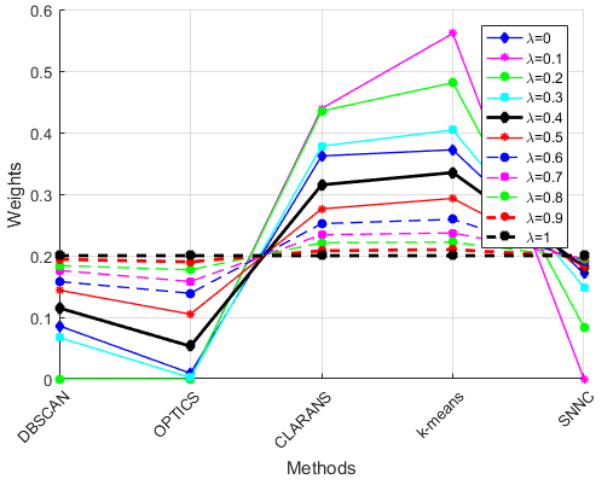
Рассмотрено влияние параметра регуляризации на производительность алгоритмов кластеризации на разных наборах данных. Значения λ приняты раными от 0.1 до 1.0.

Веса получены в Matlab 2016a с использованием Optimization Toolbox для решения задачи оптимизации (2.5.2) – (2.5.3). Начальные значения весов w_i для всех основных разбиений принимаются равными $w_i = 0.2, i = 1, \dots, 5$.

На рис. 2.4.2 показана зависимость весов отдельных методов кластеризации от параметра для всех наборов данных. На рис. 2.4.2 (а) можно увидеть влияние параметра λ на веса пяти методов кластеризации для набора данных NSL-KDD. На рисунке показано, что веса методов DBSCAN, OPTICS и k-средних меньше, чем веса методов CLARANS и SNNC для набора данных NSL-KDD. При этом наилучший результат для $\lambda = 0.2$, $\lambda = 0.6$, $\lambda = 0.7$ и $\lambda = 0.9$ показал метод CLARANS. А методы DBSCAN и k-средних оказались неэффективными.



а) **б)**



в)

Рис. 2.4.2. Зависимость весов методов кластеризации от параметра λ [55]

Для набора данных Phishing (Рис. 2.4.2 (б)) веса пяти методов кластеризации для разных значений λ довольно близки, за исключением $\lambda = 0$ и $\lambda = 0.2$. В этом случае наибольший вес имеет метод CLARANS.

Согласно рис. 2.4.2 (в), наибольшие веса достигаются методами CLARANS и k-средних при значениях $\lambda > 1$. В этом случае наибольший вес для набора данных Phone accelerometer получается при применении алгоритма k-средних при $\lambda = 0.1$.

Таким образом, резюмируя вышеизложенное, метод CLARANS показал наилучшие результаты для наборов данных NSL-KDD и Phishing. Значения $w_i = 0.2$ $(i = 1, \dots, 5)$ является оптимальным решением для $\lambda = 1$, что подтверждается экспериментами.

Наилучшие результаты для наборов данных Phishing и NSL-KDD были получены при $\lambda = 0.6$, а для набора данных Phone Accelerometer при $\lambda = 0.1$. Набор весов, полученных с помощью предлагаемого подхода для всех наборов данных, показан в таблице 2.4.3. Видно, что значения весов сопоставимы с результатами предложенного подхода.

Максимальные веса были присвоены методам DBSCAN и OPTICS при применении метрики евклидова расстояния к предложенному подходу для набора данных NSL-KDD, а для набора данных Phishing наибольший вес

получил метод SNNC (Таблица 2.4.3 (а)).

Таблица 2.4.3

Сравнение весов отдельных методов кластеризации для $\lambda = 0.6$

Набор данных	DBSCAN	OPTICS	CLARANS	k-means	SNNC
Phishing	0.178	0.182	0.258	0.174	0.208
NSL-KDD	0.202	0.197	0.225	0.201	0.175
Phone Accelerometer	0.161	0.125	0.263	0.262	0.189

а)

Набор данных	DBSCAN	OPTICS	CLARANS	k-means	SNNC
Phishing	0	0	0	0	1
NSL-KDD	0	0	0	1	0
Phone Accelerometer	0.200	0.203	0.198	0.197	0.202

б)

Набор данных	DBSCAN	OPTICS	CLARANS	k-means	SNNC
Phishing	0	0	0.998	0.002	0
NSL-KDD	0	0	0.999	0	0.001
Phone Accelerometer	0	0	0.195	0.761	0.045

в)

Набор данных	DBSCAN	OPTICS	CLARANS	k-means	SNNC
Phishing	0.180	0.187	0.250	0.179	0.204
NSL-KDD	0.201	0.199	0.222	0.200	0.178
Phone Accelerometer	0.160	0.134	0.253	0.263	0.190

г)

Набор данных	DBSCAN	OPTICS	CLARANS	k-means	SNNC
Phishing	0.180	0.190	0.247	0.182	0.201
NSL-KDD	0.200	0.200	0.220	0.200	0.180
Phone Accelerometer	0.167	0.137	0.248	0.255	0.193

д)

Набор данных	DBSCAN	OPTICS	CLARANS	k-means	SNNC
Phishing	0.181	0.204	0.236	0.193	0.186
NSL-KDD	0.185	0.224	0.214	0.184	0.193
Phone Accelerometer	0.201	0.164	0.213	0.214	0.208

е)

Применение метрики косинусного расстояния позволило точно определить наилучший подход для NSL-KDD и Phishing в соответствии с таблицей 2.4.3(б). Таблица 2.4.3(в) показала почти 100% определение лучших методов на всех наборах данных для предлагаемого подхода с метрикой квадрата евклидова расстояния. Согласно таблице 2.4.3(г) и таблице 2.4.3(д), метод CLARANS получил наибольший вес для наборов данных Phishing, Phone Accelerometer и NSL-KDD для метрик расстояния Mink3 и Mink4. По метрике расстояния Чебышева (Таблица 2.4.3(е)) лучший результат для набора данных Phishing

показал метод CLARANS, а для набора данных Phone Accelerometer – k-средних и CLARANS.

Результаты предлагаемого подхода на основе метрик Purity, Mirkin, PC, VI и F-мера представлены в таблицах 2.4.4-2.4.6. Лучшие результаты в таблицах выделены жирным шрифтом.

Из таблицы 2.4.4 можно сделать вывод, что предложенный подход с квадратом евклидова расстояния показал наилучший результат по всем пяти метрикам и совпал с результатом метода CLARANS для NSL-KDD.

Таблица 2.4.4

Сравнение производительности предлагаемого подхода с методами одиночной кластеризации для набора данных NSL-KDD

Метод	Purity	Mirkin	F-мера	VI	PC
DBSCAN	0.5134	0.4987	0.6716	0.0587	0.3753
OPTICS	0.6400	0.4608	0.5489	0.0868	0.3148
CLARANS	**0.8489**	**0.2565**	**0.8533**	**0.0563**	**0.4111**
k-means	0.5188	0.4993	0.6832	0.0582	0.3752
SNNC	0.5470	0.4956	0.6018	0.0924	0.2553
PWCC(Eucl)	0.7552	0.3697	0.7734	0.0690	0.3897
PWCC(Cosine)	0.5188	0.4993	0.6832	0.0582	0.3752
PWCC(SqEucl)	**0.8489**	**0.2565**	**0.8533**	**0.0563**	**0.4111**
PWCC(Mink3)	0.7552	0.3697	0.7734	0.0690	0.3897
PWCC(Mink4)	0.7552	0.3697	0.7734	0.0690	0.3897
PWCC(Cheb)	0.7552	0.3697	0.7734	0.0690	0.3897

Таблица 2.4.5

Сравнение производительности предлагаемого подхода с отдельными методами кластеризации для набора данных Phishing

Метод	Purity	Mirkin	F-мера	VI	PC
DBSCAN	0.5569	0.4987	0.5312	0.1468	0.2549
OPTICS	0.5569	0.4987	0.5309	0.1468	0.2549
CLARANS	**0.7166**	**0.4062**	**0.7283**	0.1189	**0.3053**
k-means	0.5569	0.4957	0.6550	0.1229	0.2524
SNNC	0.5569	0.4988	0.6739	0.1016	0.2745
PWCC(Eucl)	0.5905	0.4836	0.7174	**0.0948**	0.2927
PWCC(Cosine)	0.5569	0.4988	0.6739	0.1016	0.2745
PWCC(SqEucl)	**0.7166**	**0.4062**	**0.7283**	0.1189	**0.3053**
PWCC(Mink3)	0.5905	0.4836	0.7174	**0.0948**	0.2927
PWCC(Mink4)	0.5905	0.4836	0.7174	**0.0948**	0.2927
PWCC(Cheb)	0.5905	0.4836	0.7174	**0.0948**	0.2927

Анализируя таблицу 2.4.5 для набора данных Phishing, можно сделать

вывод, что предложенный консенсусный подход с квадратом евклидова расстояния дал наилучший результат для Purity (0.7166), Mirkin (0.4062), F-меры (0.7283) и PC (0.3053). Результаты основных разбиений превзошли VI и составили 0.0948.

Исследования показали, что предложенный метод выявляет больше аномалий в кластерах, чем каждый отдельный метод кластеризации.

Таблица 2.4.6

Сравнение производительности предлагаемого подхода с отдельными методами кластеризации для набора данных Phone Accelerometer

Метод	Purity	Mirkin	F-мера	VI	PC
DBSCAN	0.5136	0.4999	0.6611	**0.0600**	0.2583
OPTICS	0.5089	0.4998	0.5348	0.0904	0.2501
CLARANS	0.7393	0.3858	**0.7499**	0.0660	0.3435
k-means	**0.7403**	**0.3845**	0.7257	0.0619	**0.3613**
SNNC	0.6325	0.4649	0.6257	0.0856	0.2687
PWCC(Eucl)	0.6932	0.4191	0.6829	0.0732	0.3168
PWCC(Cosine)	0.5872	0.4693	0.5939	0.0856	0.2671
PWCC(SqEucl)	**0.7403**	**0.3845**	0.7257	0.0620	**0.3613**
PWCC(Mink3)	0.6891	0.4206	0.6826	0.0741	0.3121
PWCC(Mink4)	0.6932	0.4206	0.6826	0.0741	0.3121
PWCC(Cheb)	0.7109	0.4043	0.7050	0.0704	0.3263

Для набора данных Phone Accelerometer (Таблица 2.4.6) предложенный подход превзошел такие методы кластеризации, как DBSCAN и OPTICS по Purity, Mirkin, F-мере и PC, но немного уступил по VI.

По рангам, полученным для показателей Purity, Mirkin, F-мера, VI и PC, был рассчитан результирующий ранг [153, с. 224-227] для всех методов кластеризации (методы одиночной кластеризации и предлагаемый консенсусный подход) (Таблица 2.4.7).

Из таблицы 2.4.7 предложенный подход с использованием квадратичной евклидовой метрики показал наилучший результат для всех наборов данных по трем метрикам (Purity, Mirkin и PC), а самый низкий результат показал метод SNNC по метрике Purity.

Предложенный подход с косинусной метрикой имеет второй ранг по Purity, Mirkin и PC при рассмотрении всех наборов данных в целом, несмотря

на то, что он получил неудовлетворительные результаты для наборов данных NSL-KDD и Phishing.

Из таблиц 2.4.4-2.4.7 можно сделать вывод:

1) Выбор метрики расстояния существенно влияет на качество консенсусной кластеризации.

2) Наилучшие результаты среди методов кластеризации показал предложенный подход консенсусной кластеризации на основе чистоты.

3) Метод с использованием метрик расстояний евклида, минковского (p=3 и p=4) и Чебышева имеет одинаковые ранги равные 23.1429 (Таблица 2.4.7).

4) Предлагаемый подход, основанный на метриках SqEucl и Cosine, превосходит отдельные методы кластеризации.

Таблица 2.4.7

Результирующий ранг методов кластеризации

Метод	Сколько раз метод имеет ранг *s*							Результирующий ранг
	s=1	*s*=2	*s*=3	*s*=4	*s*=5	*s*=6	*s*=7	
PWCC (SqEucl)	24	2	4	0	0	0	0	28.5714
PWCC (Cosine)	15	4	7	1	2	1	0	25.1429
CLARANS	14	4	7	1	4	0	0	24.7143
PWCC (Eucl)	8	8	9	1	1	3	0	23.1429
PWCC (Mink3)	8	8	9	1	1	3	0	23.1429
PWCC (Mink4)	8	8	9	1	1	3	0	23.1429
PWCC (Cheb)	8	8	9	1	1	3	0	23.1429
OPTICS	8	3	7	7	3	2	0	21.4286
DBSCAN	2	9	7	7	3	1	1	20.4286
k-means	4	6	8	4	4	4	0	20.0000
SNNC	6	5	4	6	5	4	0	19.8571

Статистический анализ предлагаемого метода. Для оценки статистической значимости обобщения результатов был проведен тест статистической значимости, известный как критерий суммы рангов Уилкоксона для независимых выборок, при уровне значимости 5%. Для каждого набора данных было создано пять групп, соответствующих пяти методам (PWCC(Eucl), PWCC(Cos), PWCC(Mink3), PWCC(Mink4) и PWCC(Cheb)). Одновременно сравниваются две группы, одна из которых соответствует методу PWCC(SqEucl). Каждая группа состоит из значений Purity для наборов

данных, полученных в результате десяти последовательных запусков соответствующего метода.

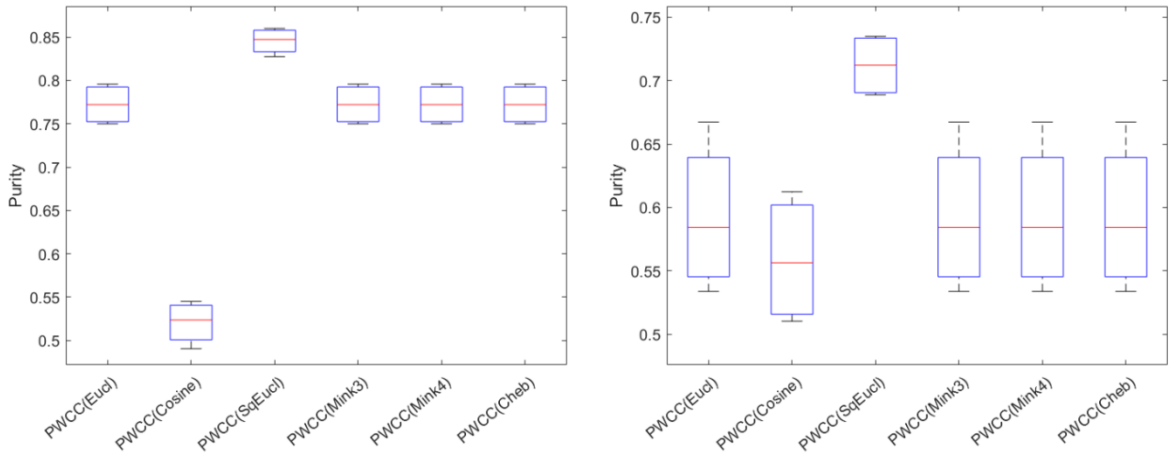
Чтобы установить, что это качество является статистически значимым, в таблице 2.4.8 представлены *p*-значения, полученные с помощью критерия суммы рангов Уилкоксона для сравнения двух групп одновременно. В качестве нулевой гипотезы предполагается, что между средними значениями двух групп нет существенных различий, тогда как альтернативная гипотеза состоит в том, что существует значительная разница в средних значениях двух групп [47, с. 1-13]. Из таблицы видно, что *p*-значения намного меньше 0.05 (уровень значимости 5%) для большого набора данных Phone Accelerometer.

Таблица 2.4.8

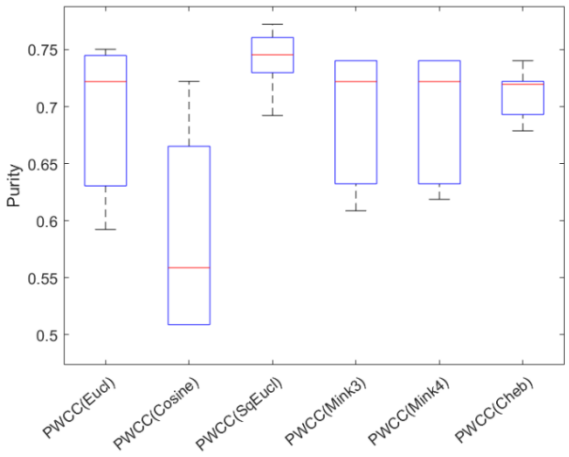
***p*-значения полученные по критерию суммы рангов Уилкоксона путем сравнения PWCC(SqEucl) с другими методами**

Набор данных	PWCC (Eucl)	PWCC (Cos)	PWCC (Mink3)	PWCC (Mink4)	PWCC (Cheb)
Phishing	0.0286	1	0.0286	0.0286	0.0286
NSL-KDD	0.0286	0.0286	0.0286	0.0286	0.0286
Phone Accelerometer	0.0261	1.5114e-04	0.0076	0.0076	0.0024

Это указывает на то, что наилучшие медианные значения, полученные с помощью PWCC(SqEucl), являются статистически значимыми и не являются случайными. Ниже приведены диаграммы для наглядного сравнения характеристик рассмотренных методов (Рис. 2.4.3).



а) **б)**



в)

Рис. 2.4.3. Диаграмма «ящики с усами» оценки производительности для всех рассмотренных наборов данных [55]

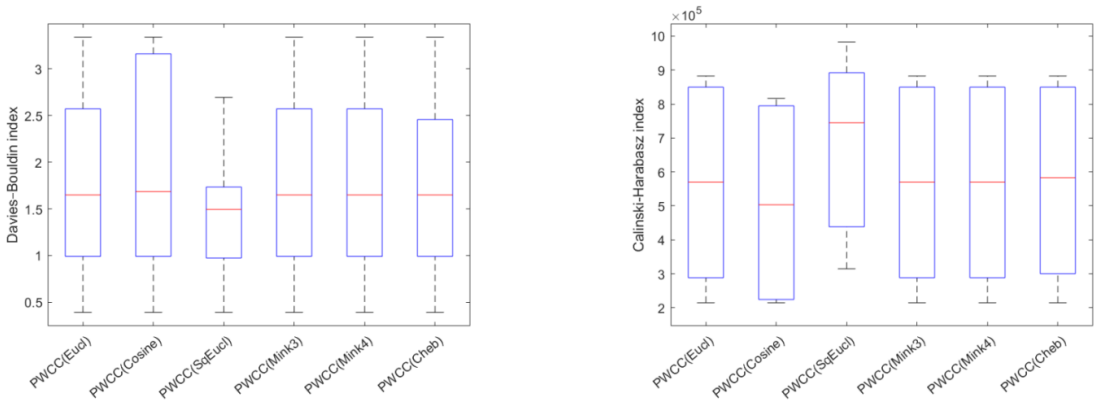
Видно, что максимальная точность и наибольшая минимальная точность наблюдаются для PWCC(SqEucl). А для наборов данных Phishing, NSL-KDD и Phone Accelerometer метод PWCC(SqEucl) также показал высокие результаты.

Статистический анализ кластеризации больших данных без меток демонстрирует возможность применения взвешенной консенсусной кластеризации к анализу больших данных КФС без меток. С этой целью функция полезности, основанная на Purity, была заменена функцией, основанной на индексе DB.

В этом случае задача состоит в том, чтобы минимизировать целевую функцию (2.4.2) для нахождения оптимального консенсусного разбиения. Превосходство предложенного метода с использованием квадрата евклидовой метрики видно из рис. 2.4.4.

На рисунке представлены результаты оценки предлагаемого подхода на основе индекса DB и индекса CH. Предложенная модель также превзошла известные методы при использовании других показателей расстояния для более чем 50% результатов на наборе данных Phone Accelerometer (Рис. 2.4.4). Самые низкие показатели по индексам DB и CH (Рис. 2.4.4) наблюдались у предложенного метода с косинусной метрикой. Таким образом, предложенный подход показал близкие результаты для всех рассмотренных метрик расстояния

[55]. Вышеприведенный анализ позволяет сделать вывод, что предложенный метод обеспечивает высокую точность кластеризации больших данных КФС.



а) **б)**

Рис. 2.4.4. Диаграмма "ящики с усами" оценки производительности для набора данных Phone Accelerometer [55]

Таким образом, в данном исследовании был предложен метод на основе взвешенной консенсусной кластеризации для эффективной интеграции методов кластеризации. Цель состояла в том, чтобы показать как взвешивание улучшает решение для кластеризации больших наборов данных.

Сравнение проводилось с использованием трех наборов данных, содержащих аномальные значения. Результаты, полученные с помощью алгоритмов DBSCAN, OPTICS, CLARANS, k-средних и SNNC, используются при формировании окончательного решения. Качество результата кластеризации оценивалось с использованием пяти метрик (Purity, Mirkin, PC, VI и F-мера).

На основании результатов эксперимента можно сделать следующие выводы:

• Результаты экспериментов показали, что предложенный метод точнее обнаруживает аномалии по сравнению с отдельными методами кластеризации. PWCC с использованием метрик Евклида, Минковского (p = 3 и p = 4), квадрата Евклида, косинуса и расстояния Чебышева сравнивался с методами кластеризации. Наилучшее разбиение было определено при применении

предложенного алгоритма с метрикой квадрата евклидова расстояния к рассматриваемым наборам данных.

На основании экспериментальных результатов можно сделать вывод, что PWCC компенсирует недостатки каждого из рассмотренных методов и повышает эффективность кластеризации. Предложенный подход рассчитан на любое количество альтернатив и может быть реализован в любой архитектуре КФС. Метод может применяться экспертами при принятии группового решения.

• Несмотря на это, предложенный взвешенный консенсус зависит от априорного знания количества кластеров и инициализации центров кластеров.

Однако в этой области остается много нерешенных проблем, что должно стимулировать исследователей к разработке новых и совершенствованию существующих подходов к консенсусной кластеризации. Возможные будущие направления исследований включают:

• Существующие методы, основанные на консенсусной кластеризации, рассматривают фиксированное количество базовых разбиений, которые играют важную роль в принятии окончательного решения. Небольшое количество базовых разбиений может затруднить принятие экспертом правильного окончательного решения. Интерактивный выбор оптимального количества базовых разбиений позволит повысить точность кластеризации и сократить вычислительные затраты. Актуальность использования методов глубокого обучения позволит решить эту задачу.

• Распараллеливание предлагаемого метода позволит ускорить процесс кластеризации, выполняя его в несколько потоков. Желаемую кластеризацию можно получить для разных пар кластеров.

• В данном исследовании количество кластеров для наборов данных было известно заранее. Однако необходимо разработать новый подход для определения оптимального количества кластеров в больших наборах данных.

• Существующие методы инициализации центров кластеров имеют

высокую вычислительную сложность. Разработка нового метода, определяющего начальные центры и повышающего точность кластеризации, является интересным направлением исследований.

• Интерпретируемость результатов кластеризации важна для извлечения знаний из полученной информации в помощь эксперту. Разработка подходов в этом направлении является перспективным направлением исследований и может быть полезна во многих приложениях экспертных систем.

Выводы к главе II

- Разработан алгоритм обнаружения выбросов в больших данных на основе метода k-средних для снижения риска блокировки КФС;
- Предложен алгоритм анализа больших данных КФС на основе взвешенной кластеризации;
- Разработан метод параллельной обработки больших данных для снижения риска выхода КФС из строя;
- Предложен метод кластерного анализа больших данных КФС на основе консенсусного ансамбля.

ГЛАВА III. РАЗРАБОТКА МЕТОДОВ И АЛГОРИТМОВ ОБЕСПЕЧЕНИЯ КИБЕРБЕЗОПАСНОСТИ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ КИБЕРФИЗИЧЕСКИХ СИСТЕМ

Кибератаки являются одной из причин аномальных явлений, наблюдаемых в работе ИТ КФС, а также при передаче трафика по сети. Аномалии сетевого трафика могут привести к некорректной работе отдельного канала или целых сегментов сети, что приведет к отказу в обслуживании на данном сетевом оборудовании. Сетевые атаки постоянно меняются, поскольку злоумышленники используют индивидуальные подходы. На это также влияют изменения в программном и аппаратном обеспечении.

В данной главе предлагаются метод классификации кибератак на КФС с применением экстремального машинного обучения, метод обнаружения DoS атак на КФС с применением ансамбля классификаторов и алгоритм обнаружения злонамеренного ПО в ИТ КФС на основе визуальных представлений.

3.1. Метод классификации кибератак на киберфизические системы с применением экстремального машинного обучения

Обеспечение безопасности ИТ КФС позволяет выявлять различные виды вторжений и предотвращать кибератаки. Для этого проводится непрерывный анализ событий и обнаружение отклонений от «нормального» состояния в сети. Системы обнаружения вторжений являются одним из наиболее важных средств защиты от сетевых атак на КФС.

В последнее время с развитием сетевых технологий угрозы безопасности КФС значительно возросли [171, с. 78-91]. Таким образом, повышение уровня сетевой безопасности является одним из актуальных вопросов для исследователей [196, с. 21-29; 122, с. 1062-1072].

Методы машинного обучения используются для анализа параметров сетевого трафика на наличие признаков кибератаки [15, с. 40-45]. В работе [221, с. 418-421] предлагается метод экстремального машинного обучения для

обнаружения вторжений в сетевом трафике. Экспериментальные результаты позволяют сделать вывод о практической значимости предложенного подхода для обнаружения кибератак в сетевом трафике КФС.

Решение проблемы обнаружения аномалий нетривиально, так как сама природа аномалий изменчива. Дать исчерпывающее определение аномального или «нормального» поведения в контексте КФС довольно сложно [298, с. 1-8; 382, с. 1302-1325]. Другая причина заключается в том, что некоторые методы обнаружения аномалий требуют меток нормального состояния и аномального поведения, которые трудно получить [155, с. 353-375; 188, с. 13-23]. Кроме того, выбрать правильный инструмент для обнаружения аномалий непросто. Предлагаемые инструменты могут подойти только для одного типа аномалий, но не для всех [124, с. 1-58]. Таким образом, когда типы аномалий априори неизвестны, очень реалистично предположить, что выбор метода обнаружения аномалий не является тривиальным. Кроме того, проблемой является масштаб сети.

При обнаружении аномалий необходимо учитывать балансировку нагрузки (распределение процесса выполнения задач между несколькими серверами сети с целью увеличения общей производительности) и устойчивость к сбоям (способность системы работать при выходе из строя отдельных ее элементов), особенно с ростом размера существующей сети [166, с. 39-47; 74, с. 68-79; 38, с. 19-31].

Более подробную информацию об обнаружении аномалий можно получить с помощью методов машинного обучения. Теоретически алгоритмы машинного обучения могут достигать высокой производительности, т. е. могут минимизировать количество ложных срабатываний и максимизировать точность обнаружения. Однако для этого обычно требуется бесконечное количество обучающих выборок (теоретически). На практике это условие невыполнимо из-за ограниченных вычислительных возможностей и необходимости реагирования в реальном времени.

Предлагаемый подход. В данном исследовании предлагается метод для обнаружения кибератак на ИТ КФС с целью обеспечения их кибербезопасности с применением экстремального машинного обучения (Extreme Learning Machine, ELM).

Общая схема предлагаемого метода обнаружения кибератак с применением экстремального машинного обучения показана на рис. 3.1.1. Этот подход состоит из следующих этапов: предварительная обработка, обучение модели, обнаружение аномалий и классификация кибератак [101, с. 38-200].

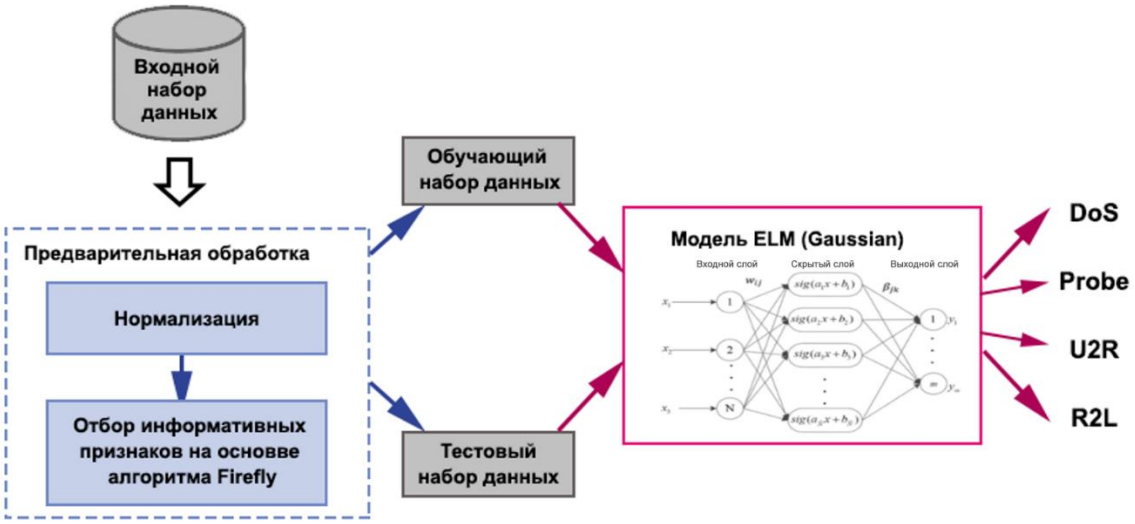


Рис. 3.1.1. Блок-схема предлагаемого подхода

На этапе предварительной обработки производится отбор наиболее информативных признаков. Для этого был применен генетический алгоритм поведения «светлячков» с целью повышения точности и увеличения скорости работы алгоритма классификации.

При анализе данных сетевого трафика проблема размерности стоит очень остро [104, с. 1-6]. Размерность имеющихся данных, характеризующаяся различным числом признаков, достигает большого числа показателей. В силу этого необходимо снизить размерность признакового пространства и выделить из них наиболее важные.

Отбор признаков помогает улучшить точность классификации и ее способность к обобщению [104, с. 1-6]. Другим мотивом для отбора признаков

является то, что меньшее количество признаков приводит к более интерпретируемым классификаторам, что важно во многих областях. Кроме того, измерение некоторых переменных признаков может быть довольно дорогостоящим с точки зрения требований к хранению и передаче данных или времени на обучение.

Данные с меньшей размерностью также могут быть достаточно просто визуализированы [197, с. 676-677]. Таким образом, отбор признаков является важной задачей во многих КФС системах.

Во многих исследованиях был сделан вывод о том, что различные алгоритмы отбора признаков имеют различное поведение на различных наборах данных, и, следовательно, опасно использовать только один алгоритм [289, с. 306-313].

Методы машинного обучения широко применяются для отбора признаков с целью анализа сетевого трафика на предмет наличия кибератак. Теоретически алгоритмы машинного обучения могут получить высокую производительность, т.е. могут минимизировать уровень ложных тревог и максимизировать точность обнаружения. Однако обычно требуется бесконечное число обучающих образцов. На практике это условие невозможно в силу ограничения вычислительной мощности КФС и требования ответа в режиме реального времени.

При классификации набор данных обычно включает большое количество признаков, которые могут быть релевантными, нерелевантными или избыточными. Избыточные и нерелевантные признаки не пригодны для классификации, и они могут даже снизить эффективность классификатора в отношении большого пространства поиска, которое также известно, как «проклятие размерности» [104, с. 1-6].

Преимущества отбора признаков включают в себя сокращение вычислительных затрат, экономию дискового пространства, упрощение процедур выбора модели для точного прогнозирования и интерпретации

комплексных зависимостей между переменными. Отобранные признаки не только оптимизируют точность классификации, но также уменьшают количество необходимых данных для достижения оптимального уровня производительности процесса обучения [104, с. 1-6].

Методы отбора признаков обычно включают в себя стратегию поиска, меру оценки, критерий остановки и валидацию результатов. Среди двух подходов, используемых для отбора признаков, а именно метода фильтров (filter approach) и метода «обертки» (wrapper approach), первый работает лучше при анализе данных высокой размерности [104, с. 1-6].

Генетический алгоритм является одним из недавних разработок для отбора признаков. В настоящее время он является очень эффективным в научно-технической оптимизации.

Классификация на основе протоколов была предложена с использованием генетического алгоритма с логистической регрессией и применена к набору данных KDD'99 в работах [43, с. 10-20].

Существует множество алгоритмов работающих на основе имитации поведения природных агентов, таких как рыбы, птицы, насекомые и т.д. Среди них алгоритм поведения «светлячков» является одним из тех, который может приводить к эффективным решениям большого числа задач [89, с. 473-480; 18, с. 113-120; 392, с. 169-178]. В работе [89, с. 473-480] представлен современный подход для отбора признаков с применением алгоритма Firefly доказавший свою применимость.

Алгоритм «светлячков» был предложен Xin She Yang и основан на поведении светлячков [392, с. 169-178]. Данный алгоритм предполагает, что существуют P светлячков y_k $(k = 1,2, \dots, p)$, первоначально произвольно размещенных в пространстве. Интенсивность света I каждого светлячка определяется целевой функцией $f(x)$. В простейшей форме, интенсивность света $I(r)$ изменяется в зависимости от расстояния r монотонно и экспоненциально, как это показано в (3.1.1):

$$I = I_0 e^{-\gamma r}, \quad (3.1.1)$$

где I_0 – исходная интенсивность света и γ – коэффициент поглощения света. Если $I_i > I_j$, $j \neq i$, то менее яркий светлячок j будет двигаться в направлении более яркого светлячка i.

Привлекательность изменяется в зависимости от расстояния $r_{ij} = d(y_i, y_j)$:

$$r_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}. \quad (3.1.2)$$

Движение одного светлячка к другому, более привлекательному светлячку определяется как (3.1.2). Если яркость j больше, чем i, то передвигаем i к j. Таким образом, сходимость светлячка определяется его движением.

В (3.1.2) второе слагаемое обусловлено привлекательностью. В третьем члене α – параметр рандомизации, а Q_t представляет собой распределение Гаусса случайных чисел. Если $\beta_0 = 0$, то движение становится произвольным. Если $\gamma = 0$, то задача сводится к оптимизации роя частиц.

Обозначим через Θ область допустимых значений. Строки матрицы $X \in R^{m \times n}$ при этом представляют элементы обучающей выборки, $\xi(x_{ij})$ - номер класса, которому соответствует значение x_{ij} j-го признака на i-ом элементе выборки, а c – количество распознаваемых классов. Далее производится оценка информативности $I(x_{ij})$ $(i = 1, \dots, m)$ j-го признака с областью определения Θ. Признаки сортируются в порядке убывания оценки их важности, наименее информативные не рассматриваются.

Далее на основе алгоритма «светлячков» необходимо сгенерировать популяцию светлячков P, где каждый светлячок соответствует отобранным признакам [104, с. 1-6].

При этом необходимо определить изменчивость интенсивности света и формулировку привлекательности. Привлекательность светлячка пропорциональна интенсивности света, которая меняется с расстоянием r и задается в виде:

$$\beta = \beta_0 e^{-\gamma r^2}, \quad (3.1.3)$$

где β_0 превращается в привлекательность при $r = 0$. Движение светлячка k привлекает другого более яркого светлячка l и определяется как

$$y_k^{t+1} = y_k^t + \beta_0 e^{-\gamma r_{kl}^2}(y_l^t - y_k^t) + \alpha t Q_k^t. \quad (3.1.4)$$

Требуется оценить информативность признаков сетевого трафика для повышения скорости работы систем обнаружения вторжений, сохраняя при этом достаточно хорошие результаты.

На этапе обучения методы машинного обучения используются для создания «нормального» профиля трафика при отсутствии кибератак на КФС. На этапе обнаружения входные события помечаются как аномалия, если записи о событиях отличаются от обычного профиля. Впоследствии на этапе классификации можно определить тип кибератаки, а также обнаружить ранее неизвестные кибератаки.

Однако обнаружение кибератак на ИТ КФС осложняется высоким уровнем ложных срабатываний. В этих случаях злоумышленник может использовать эти шаблоны, чтобы система распознавала их как «нормальное» состояние.

Алгоритм экстремального машинного обучения был предложен Huang и соавторы в 2006 г. [218, с. 489-501]. Этот алгоритм основан на однослойной нейронной сети с прямой связью (single-layer feedforward neural network, SLFN). ELM также решает задачи классификации и регрессии [32, с. 1303-1310; 164, с. 512-525; 215, с. 32-48; 336, с. 199-207]. Следовательно, ELM можно использовать для обнаружения кибератак на ИТ КФС. Он широко используется во многих исследованиях [32, с. 1303-1310; 164, с. 512-525; 95, с. 1-13]. ELM имеет высокую скорость вычислений и не требует итеративной настройки параметров для обучения.

В модели ELM обучение по входным данным выполняется в течение секунд или минут даже для больших наборов данных и сложных приложений, чего невозможно достичь с помощью обычных методов.

SLFN с нейронами в скрытом слое и функцией активации можно описать как [218, с. 489-501]:

$$f_L(x) = \sum_{i=1}^{L} \varphi_i h_i(x), \quad (3.1.5)$$

где $\varphi = [\varphi_1, \varphi_2, \dots, \varphi_L]^T$ – матрица выходных весов из скрытого слоя к выходному слою и $h(x) = [h_1(x), h_2(x), \dots, h_L(x)]$ – нелинейное пространство отображения, которое содержит выходной вектор относительно входного x.

Функция активации $G(a, b, x)$ представляет собой нелинейную кусочно-непрерывную функцию [218, с. 489-501; 217, с. 3056-3062]. В модели ELM параметры скрытых узлов (a, b) генерируются случайным образом, и поэтому она более эффективна, чем машины опорных векторов или искусственные нейронные сети [218, с. 489-501; 215, с. 32-48].

Выходной вектор Y_j в матричной форме вычисляется следующим образом:

$$Y = H\varphi, \quad (3.1.6)$$

где H – матрица значений выходных нейронов скрытого слоя, которая определяется выражением

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} G(a_1, b_1, x_1) & \cdots & G(a_L, b_L, x_L) \\ \vdots & \ddots & \vdots \\ G(a_1, b_1, x_N) & \cdots & G(a_L, b_L, x_N) \end{bmatrix}, \quad (3.1.7)$$

где N – число входных образцов. А целевая выходная матрица Y имеет вид:

$$Y = \begin{bmatrix} y_1^T \\ \vdots \\ y_N^T \end{bmatrix} = \begin{bmatrix} y_{11} & \cdots & y_{1m} \\ \vdots & \ddots & \vdots \\ y_{1N} & \cdots & y_{Nm} \end{bmatrix}. \quad (3.1.8)$$

Обучение сети заключается в вычислении матрицы выходных весов скрытого слоя согласно

$$\varphi = H^+ Y, \quad (3.1.9)$$

где $H^+ = (H^T H)^{-1} H^T$ – псевдообратная матрица (матрица Мура-Пенроуза) для случая невырожденной матрицы H.

Из теории матриц [398, с. 128-155] известно, что (3.1.9) является наилучшим приближенным решением (3.1.6) методом наименьших квадратов, позволяющим избежать локальных минимумов при обучении сети.

При обучении нейронной сети, т. е. по (3.1.9) вычисляется матрица φ выходных весов скрытого слоя, выходной вектор вычисляется на основе (3.1.6). И к сети

применяется вектор входных признаков (которого нет в обучающей выборке). Затем вычисляется количество классов, к которым принадлежит входной вектор.

Описание экспериментального набора данных. Для проведения экспериментов был рассмотрен набор данных сигнатур NSL-KDD [35, с. 842-851], собранный на основе базы KDD-99 по инициативе американской Ассоциации перспективных оборонных научных исследований DARPA [283, с. 262-294]. Он охватывает широкий спектр различных вторжений, смоделированных в среде, имитирующей сеть Военно-воздушных сил США. NSL-KDD включает 41 признак и имеет метки для каждого конкретного типа атаки. Каждый вектор обозначается либо как нормальный, либо как кибератака. Отклонение от «нормального поведения» рассматривается как кибератака [155, с. 353-375].

Рассмотренный набор данных NSL-KDD имеет следующие преимущества [361, с. 53-58]: нет избыточных записей в обучающем наборе, так что классификатор не покажет какой-либо предвзятый результат; нет дубликата записей в тестовом наборе (имеются некоторые атаки, которые не присутствуют в обучающем наборе); количество выбранных записей из каждой группы уровней сложности обратно пропорционально доле записей в исходном наборе данных NSL-KDD.

Обучающий набор данных состоит из 21 различных кибератак, а тестовый – 37. Известные виды кибератак содержатся в обучающем наборе, в то время как новые атаки – это дополнительные атаки в тестовом наборе данных (они отсутствуют в обучающем наборе). Кроме того, количество записей в обучающем наборе составляет 125973 образца, а в тестовом – 22544. Это преимущество делает его доступным для проведения экспериментов на полных данных без необходимости случайным образом выбирать небольшую часть.

Все атаки в NSL-KDD поделены на четыре группы [361, с. 53-58]: DoS (Denial of Service Attack), атаки пользователей на корневой компьютер (Users to

Root Attack, U2R), удаленные атаки на локальный компьютер (Remote to Local Attack, R2L) и атаки зондирования (Probe). Каждая запись имеет 41 признак (Таблица 3.1.1). Протоколы, которые рассматриваются в NSL-KDD, включают TCP, UDP (User Datagram Protocol) и ICMP (Internet Control Message Protocol).

Таблица 3.1.1

Список признаков для каждой записи базы данных NSL-KDD

№	Название признака	№	Название признака	№	Название признака
1	duration	15	su_attempted	29	same_srv_rate
2	protocol_type	16	num_root	30	diff_srv_rate
3	service	17	num_file_creations	31	srv_diff_hast_rate
4	flag	18	num_shells	32	dst_host_count
5	scr_bytes	19	num_access_files	33	dst_host_srv_count
6	dst_bytes	20	num_outbound_cmds	34	dst_host_same_srv_rate
7	land	21	is_host_login	35	dst_host_diff_srv_rate
8	wrong_fragments	22	is_quest_login	36	dst_host_same_src_port_rate
9	urgent	23	count	37	dst_host_srv_diff_host_rate
10	hot	24	srv_count	38	dst_host_serror_rate
11	num_failed_logins	25	serror_rate	39	dst_host_srv_serror_rate
12	logged_in	26	srv_serror_rate	40	dst_host_rerror_rate
13	num_compromised	27	rerror_rate	41	dst_host_srv_rerror_rate
14	root_shell	28	srv_rerror_rate		

Метки присваиваются каждой записи либо в качестве типа «кибератаки», либо как «нормальное» состояние [155, с. 353-375].

Метрики оценки производительности методов. Все четыре возможных показателя обнаружения вторжений в ИТ КФС (истинно-положительный, ложно-положительный, ложно-отрицательный и истинно-отрицательный) показаны на рис. 3.1.2 в виде матрицы ошибок (confusion matrix) [179, с. 157-165].

Задача состоит в том, чтобы свести к минимуму ложно-отрицательные и ложно-положительные результаты.

Наиболее часто используемой метрикой оценки эффективности методов обнаружения кибератак является частота истинно положительных результатов (true positive, TP) (также известная как вероятность обнаружения), которая представляет собой количество правильно классифицированных вторжений.

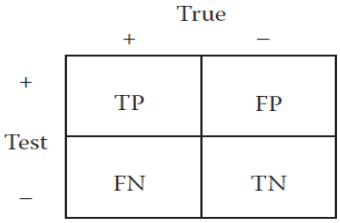


Рис. 3.1.2. Матрица ошибок для оценки эффективности предложенного метода

Еще одним важным показателем является частота ложноположительных результатов (false positive, FP) (или ложная тревога), которая обозначает количество аномальных образцов, классифицированных как нормальные образцы. Частота ложноотрицательных результатов (false negative, FN) — это количество нормальных образцов, классифицированных как вторжение. Частота истинно отрицательных результатов (true negative, TN) – это количество правильно классифицированных вторжений.

Кроме того, другие параметры, такие как ROC-кривая (receiver operating characteristic curve), используются для оценки системы обнаружения вторжений (intrusion detection system, IDS). ROC-кривая — это показатель обнаружения кибератак на КФС, который создается путем построения TP и FP при различных пороговых значениях. ROC-кривая сравнивает частоту TP с FP [179, с. 157-165]. Одно важное ограничение этой метрики состоит в том, что она вычисляет общую производительность системы обнаружения вторжений на всех исходных данных. Чем выше значение площади под ROC-кривой (area under ROC curve, AUC), тем лучше производительность метода.

Выражение (3.1.10) представляет собой меру полноты (recall), которая определяется как доля «нормального поведения»:

$$recall = \frac{number\ of\ true\ positives}{number\ of\ true\ positives + number\ of\ false\ negatives}. \qquad (3.1.10)$$

Тем не менее, мера полноты недостаточно содержательна, так как она может быть получена тривиальным образом путем классификации всех типов поведения как вредоносных.

Существует еще одна метрика, называемая метрикой «правильности» (precision), которая решает эту проблему:

$$precision = \frac{number\ of\ true\ negatives}{number\ of\ true\ negatives + number\ of\ false\ positives}. \quad (3.1.11)$$

При классификации всего трафика как нормального, данная метрика достигается полностью.

F-мера является показателем, который сочетает в себе обе приведенные выше метрики:

$$F\text{-}measure = \frac{2 \times recall \times precision}{recall + precision}. \quad (3.1.12)$$

Результаты экспериментов. Эксперименты были проведены в ОС Windows® 10–64 с процессором Core i7 (2.5 ГГц), 8.0 Гб RAM. Оценка информативности признаков проводилась в среде Matlab 2016a на наборе данных NSL-KDD. В ходе тестирования были использованы реализации данных алгоритмов в программном решении Weka 3.8.0.

Параметры алгоритма поведения светлячков, использованные в эксперименте, приведены в таблице 3.1.2.

Таблица 3.1.2

Параметры алгоритма оценки информативности признаков

Параметр	Значение
Число светлячков	7
Параметр рандомизации (α)	0.2
Привлекательность	2
Коэффициент поглощения света (γ)	1

Были отобраны 20 наиболее информативных признаков (Таблица 3.1.3) для набора данных NSL-KDD.

Разработанный подход был проанализирован с использованием следующих классификаторов: наивного Байесовского классификатора (NaiveBayes), деревьев решений (J48), аддитивной логистической регрессии

(Additive Logistic Regression – LogitBoost), Байесовского классификатора (BayesNet) и метода к-ближайших соседей (IBk).

Таблица 3.1.3

Результаты оценки информативности признаков с применением алгоритма «светлячков»

Метод	Отобранные информативные признаки
Алгоритм «светлячков»	22,26,30,5,37,14,7,13,27,21,10,18,24,23,11,12,31,1,20,36

В таблицах 3.1.4 и 3.1.5 приводятся результаты работы алгоритмов классификации сетевых атак на основе 41 признака и отобранных информативных признаков из имеющегося набора данных, соответственно.

Таблица 3.1.4

Сравнение производительности алгоритмов классификации для 41 признака

Метод	TPR (%)	FPR (%)	Precision (%)	Recall (%)	F-мера (%)	AUC (%)
J48	75.8	13.2	76.7	75.8	**74.0**	81.8
NaiveBayes	70.2	**11.7**	75.8	70.2	70.7	86.5
BayesNet	71.5	19.2	78.6	71.5	67.0	**92.5**
LogitBoost	74.5	15.8	78.0	74.5	73.3	90.6
IBk	**76.8**	16.3	**81.2**	**76.8**	72.6	82.3

Таблица 3.1.5

Сравнение производительности алгоритмов классификации для отобранных признаков

Метод	TPR (%)	FPR (%)	Precision (%)	Recall (%)	F-мера (%)	AUC (%)
J48	76.5	14.7	80.9	76.5	74.3	82.1
NaiveBayes	59.5	7.9	73.4	59.5	64.9	84.7
BayesNet	73.9	16.9	80.2	73.9	69.8	93.6
LogitBoost	78.8	11.5	81.8	78.8	78.7	93.5
IBk	**99.6**	**0.2**	**99.6**	**99.6**	**99.6**	**100**

Наилучшие результаты были отмечены жирным шрифтом. В качестве метрик были рассмотрены TPR, FPR, Precision, Recall, F-мера и AUC.

Из таблиц можно сделать заключение, что наилучшие результаты были получены методом к-ближайших соседей. Согласно метрике FPR наименьший процент ошибки классификации для 41 признака был достигнут методом

NaiveBayes, а для отобранных 20 признаков – методом к-ближайших соседей [104, с. 1-6].

Сравнение производительности алгоритмов (Таблица 3.1.4) показало, что метод BayesNet превосходит остальные по метрике AUC (92.5%). Анализируя полученные данные по метрике F-мера, уменьшение размерности вектора признаков согласно информативности привело к улучшению работы методов J48, NaiveBayes, LogitBoost и IBk.

Сравнение значений AUC для рассмотренных классификаторов более наглядно демонстрируется на рис. 3.1.3 (красным цветом обозначены результаты для 41 признака, а синим – 20 признаков на основе предложенного метода). Наилучшие результаты были получены для метода к-ближайших соседей, однако он требует больших временных затрат. В силу этого предпочтение можно было бы отдать методам BayesNet или LogitBoost [104, с. 1-6].

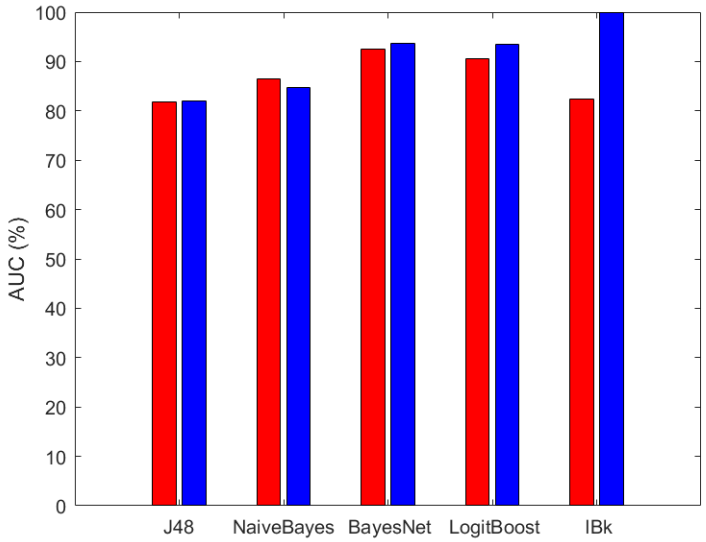


Рис. 3.1.3. Сравнение производительности методов классификации для набора данных NSL-KDD

Предложенный метод на основе ELM сравнивался при различных функциях активации (функция активации радиального базиса (radial basis activation function, RBF), функция активации треугольного базиса (triangular

basis activation function, TriBas) и функция активации Гаусса) и тестировался с использованием базы данных NSL-KDD (Таблица 3.1.6). Скрытый слой имеет 30 нейтронов с разными функциями активации. Такое количество нейронов обеспечивает высокую точность классификации на обучающей выборке и приемлемую точность на тестовой выборке для структуры сети и метода обучения.

В исследовании применялась функция активации Гаусса для различных значений параметров дисперсии (σ) и математического ожидания (μ) [109, с. 211-215]. Наилучшая точность обнаружения DoS-, U2R- и R2L-атак была достигнута при использовании функции активации Гаусса с $\sigma = 0.1$ и $\mu = 4$ (DoS – 86.89 %, U2R – 99.99 % и R2L – 99.94 %). В случае Probe-атак высокая точность (99.06%) была получена для функции активации Гаусса при $\sigma = 0.2$ и $\mu = 0$.

Следует отметить, что метод ELM при использовании функции активации Гаусса продемонстрировал высокие результаты, как по скорости, так и по точности классификации (Таблица 3.1.6).

Таблица 3.1.6

Сравнение точности обнаружения вторжения при различных функциях активации

Функция активации	DoS	Probe	U2R	R2L
RBF	80.01%	91.79%	99.88%	**99.94%**
TriBas	81.60%	95.62%	99.65%	99.57%
Gaussian (σ=0.2, $\boldsymbol{\mu}$=0)	84.26%	**99.06%**	99.95%	99.70%
Gaussian (σ=0.1, $\boldsymbol{\mu}$=4)	**86.89%**	90.01%	**99.99%**	**99.94%**

Рассмотренный классификатор на основе метода ELM обеспечивает приемлемое качество классификации кибератак и высокую производительность, что делает его привлекательным решением для систем обнаружения вторжений в ИТ КФС в режиме реального времени. В таблице 3.1.7 показаны результаты сравнения предложенного метода с различными

известными методами, такими как BayesNet, LogitBoost, IBk, SVM, алгоритмом «случайного леса» (random forest, RF) [112, с. 5-32; 347, с. 147-163], DNN и MLP.

Таблица 3.1.7

Оценка производительности предлагаемого подхода для NSL-KDD

Метод	BayesNet	LogitBoost	IBk	SVM	RF	DNN	MLP	ELM (Gaussian)
Accuracy (%)	69.9	78.8	99.6	93.80	97.93	91.50	81.43	**99.78**

Таким образом, можно сделать вывод о практической значимости предложенного подхода к обнаружению кибератак в КФС.

На основании результатов проведенных экспериментов можно сделать следующие выводы: кибератаки на сети ИТ в КФС являются одной из причин аномальных явлений, наблюдаемых в работе сетевого оборудования, а также при передаче трафика по сети. Обнаружение аномалий возможно с помощью методов машинного обучения для повышения надежности существующих КФС.

В данном исследовании был предложен метод обнаружения аномалий в ИТ КФС с их последующей классификацией на основе ELM. Для выбора наиболее информативных признаков, используемых для обнаружения атак в сетевом трафике был рассмотрен алгоритм «светлячков».

В связи с вышеизложенным можно выделить следующие направления исследований в области обнаружения и классификации кибератак на КФС:

- увеличение доли корректирующих процессов, не требующих вмешательства человека-эксперта, что снижает уровень эвристического принятия решений и позволяет качественно улучшить время реагирования на атаку (например, за счет автоматической генерации сигнатур для нового вредоносного ПО через несколько минут после подтверждения его аномально быстрого распространения в сети);

- устойчивость к новым технологиям на основе технологий искусственного интеллекта, используемым злоумышленником с целью сокрытия факта вторжения, активного воздействия на саму систему обнаружения вторжений КФС путем создания условий отказа в обслуживании или генерации избыточного потока ложных срабатываний, что делает невозможным ее использование.

3.2. Метод обнаружения DoS атак на киберфизические системы с применением ансамбля классификаторов

Анализ больших данных при обнаружении вторжений и при решении задач сетевой безопасности привлекает все большее внимание, так как облегчает изучение больших объемов сложных и разрозненных данных, выявляет сетевые вторжения и способствует борьбе с кибератаками на КФС [71, с. 10-11].

Уязвимости в стеке протоколов связи (TCP/IP) приводят к преднамеренным или непреднамеренным распределенным атакам типа «отказ в обслуживании» (DDoS) [126, с. 337-348]. DDoS-атаки можно обнаружить с помощью существующих методов машинного обучения.

Последние исследовательские работы были основаны на методах бинарной классификации, которые позволяют различать два состояния («нормальное» или «аномальное»). В случае конфликта между бинарными классификаторами окончательное решение достигается путем сравнения их точности. Альтернативное решение возможно путем разработки ансамбля классификаторов. Требуется новый подход к объединению таких классификаторов в ансамбль.

Много исследований и обзорных статей было посвящено технологии обнаружения вторжений [122, с. 1062-1072] или интеллектуальному анализу данных для конкретных приложений. С момента введения принципов обнаружения вторжений Деннингом в 1987 году было разработано большое количество систем реактивной защиты [124, с. 1-58]. Методы обнаружения

вторжений можно разделить на три категории: одиночные, гибридные и ансамблевые. SVM и искусственные нейронные сети являются наиболее популярными подходами среди одиночных классификаторов. Несколько классификаторов объединяются для более высокой цели значительного повышения эффективности классификатора, известного как ансамбль классификаторов. Большинство голосов, бэггинг и бустинг являются одними из распространенных стратегий объединения классификаторов. Хотя известно, что недостатки компонентов классификаторов накапливаются в ансамбле классификаторов, но он способен работать очень эффективно в различных комбинациях.

Таким образом, исследователи все больше и больше интересуются применением ансамбля классификаторов с каждым днем. Точность классификации была улучшена путем объединения мнений нескольких экспертов в одно с использованием ансамблевого подхода в [29, с. 360-372]. Метод построения ансамбля использует сгенерированные веса метода оптимизации роя частиц для создания ансамбля классификаторов с более высокой точностью для обнаружения вторжений. Эта работа была основана на методах бинарной классификации, которые могут различать два состояния. В случае конфликта между бинарными классификаторами окончательное решение принималось путем сравнения их точности. Архитектура интеллектуального фильтра ложных тревог с использованием метода ансамблевого выбора на основе голосования, направленного на поддержание точности снижения ложных результатов также была предложена в 2013 году. Эксперимент проводился с использованием алгоритмов машинного обучения SVM, дерева решений и KNN. Работа [30, с. 225-246] направлена на выявление многоклассовых моделей SVM, наиболее подходящих для задачи обнаружения вторжений. Был разработан новый подход, основанный на наборе оптимальных или почти оптимальных весовых коэффициентов, которые определяют взаимосвязь между правилами принятия решений бинарных классификаторов

SVM. Общая архитектура для автоматизированной системы обнаружения и реагирования на DDoS-атаки с использованием машинного обучения также была исследована Kumar и соавторами. Основной целью было минимизировать стоимость ошибок классификации обнаружения вторжений.

В связи с этим в данном исследовании был предложен новый метод обнаружения DoS-атак на основе совокупности классификаторов [54, с. 73-82].

Предлагаемый подход. Теоретически алгоритмы классификации могут достигать высокой производительности, то есть они могут минимизировать количество ложных срабатываний и максимизировать точность обнаружения. Одной из наиболее привлекательных особенностей алгоритмов является способность отличать нормальное поведение от аномального [171, с. 153-160]. В контексте обнаружения вторжений алгоритм классификации обычно представляет собой карту, которая адаптируется к невидимым аномалиям сети [303, с. 231-245].

Формально каждый экземпляр данных представляет собой вектор признаков X, измеренный во времени t и обозначаемый как $X(t)$.

Алгоритм классификации предназначен для обучения функции, которая отображает все образцы $X(t)$ в их собственные состояния. Для достижения своей цели они используют набор экземпляров данных в сети. Этот набор известен как обучающий набор данных. Некоторые алгоритмы изучают функцию отображения с помощью помеченных обучающих наборов, где каждый образец в обучающем наборе помечен как одно из состояний. Эти алгоритмы называются алгоритмами обучения с учителем. Целью использования этих алгоритмов является достижение высокой точности классификации.

К наиболее популярным методам машинного обучения с учителем относятся SVM, деревья решений, байесовские сети, алгоритм KNN и т. д.

Машины опорных векторов (SVM). Целью SVM является классификация точек данных X n-мерного пространства с использованием $(n-1)$-мерной

гиперплоскости [143, с. 21-27]. Гиперплоскость удовлетворяет $w^T x + b = 0$, где X – множество точек, w - нормальный вектор, перпендикулярный гиперплоскости, и b является смещением гиперплоскости $w^T x + b = 0$ от начальной точки вдоль направления w [30, с. 225-246].

Расстояние от точки данных до разделяющей гиперплоскости $w^T x + b = 0$ можно рассчитать как $r = \frac{w^T x + b}{\|w\|}$. Ближайшие к гиперплоскости точки данных называются опорными векторами. Расстояние между опорными векторами известно как зазор (margin). Линейный SVM может быть достигнут квадратичной оптимизацией:

$$\underset{\substack{w,b \\ y(w^T x + b) \geq 1.}}{\operatorname{argmin}\left(\frac{1}{2}\|w\|^2\right)}, \tag{3.2.1}$$

SVM может точно находить линейные, нелинейные и сложные границы классификации даже при небольшом объеме обучающей выборки. SVM широко используются для передачи различных типов данных путем переключения функций ядра. Наиболее часто используемые функции ядра включают линейную, полиномиальную, радиальную базисную функцию и сигмовидную.

Однако выбор функции ядра и подгонка соответствующих параметров в SVM все еще находятся в процессе проб и ошибок. SVM работает быстро, но его продолжительность увеличивается в четыре раза при удвоении размера выборки. К сожалению, корень алгоритмов SVM лежит в двоичной классификации. Для решения задач многоклассовой классификации несколько SVM для бинарных классов могут быть объединены классификацией каждого класса или классификацией каждой пары классов.

Деревья решений. Дерево решений (decision tree, DT) — это модель с древовидной структурой, в которой есть листья, представляющие классы или решения, и ветви, представляющие сочетания признаков, которые ведут к этим классификациям [12, с. 12-18].

Древовидная классификация входного вектора выполняется путем обхода

дерева от корневого узла до вершины с листом. Каждый узел дерева вычисляет неравенство на основе одной из входных переменных. Каждый лист относится к определенному классу. Каждое неравенство, используемое для разделения входного пространства, основано только на одной из входных переменных. Линейные DT аналогичны бинарным DT, за исключением того, что неравенство, вычисляемое в каждом узле, имеет случайный линейный вид, который может зависеть от нескольких переменных. DT зависит от правил «если-то», но не требует никаких параметров и метрик. Эта простая и интерпретируемая структура позволяет DT решать проблему для различных типов атрибутов.

DT также могут управлять отсутствующими значениями или зашумленными данными [12, с. 12-18]. Однако они не могут гарантировать оптимальную точность, в отличие от других методов машинного обучения. Хотя DT просты в изучении и реализации, они не часто используются для обнаружения вторжений. Это связано с тем, что найти наименьшее дерево решений NP-трудно.

Байесовские сети. Байесовский сетевой классификатор основан на правиле Байеса, которое дает гипотезу H о классах и данных x.

$$P(H|x) = \frac{P(x|H)P(H)}{P(x)}, \quad (3.2.2)$$

где $P(H)$ представляет собой априорную вероятность каждого класса без информации о переменной x, $P(H|x)$ является апостериорной вероятностью переменной x по возможным классам, $P(x|H)$ является условной вероятностью x при этой вероятности H [103, с. 993-1022].

Узлы байесовской сети представлены случайными величинами и дугами, показывающими вероятностные отношения между переменными и условными вероятностями. Узел всегда вычисляет апостериорные вероятности, представляя доказательство наследования для выбранных узлов [142, с. 309-347; 136, с. 462-467].

Naïve Bayes (NB) — это простая байесовская сетевая модель, в которой

предполагается, что все переменные независимы. Необходимо найти гипотезу максимального правдоподобия, определяющую метку класса для тестовых данных, для классификации по NB [194, с. 1-7].

Классификатор NB может быть разрешен с помощью гипотезы максимальной апостериорной вероятности (maximum a posteriori probability, MAP) для данных следующим образом:

$$\underset{c_j \in C}{\operatorname{argmax}} P(x|c_j)P(c_j), \tag{3.2.3}$$

где x — наблюдаемые данные, а $C = \{c_j\}$ — набор меток классов.

Наивный Байес эффективен для задач с логическим завершением. Однако наивный байесовский метод основан на сильном предположении о независимости переменных.

Алгоритм K-ближайших соседей (KNN). Количество ближайших соседей k и меры расстояний являются ключевыми компонентами алгоритма KNN. Выбор количества должен основываться на перекрестной проверке. При увеличении числа k снижается влияние шума в данных при классификации, и это может стереть разницу между классами. На практике k должно быть меньше квадратного корня из общего числа обучающих выборок.

В случае множественной классификации метод KNN основан на измерении расстояния от одной выборки данных до каждой обученной выборки [399, с. 718-721]. Вычисляются k-наименьшие расстояния, и наиболее распространенный класс, основанный на этих KNN, считается меткой выходного класса. KNN легко реализовать, но он требует много памяти и времени.

Обнаружение DoS атак с применением ансамбля классификаторов. Для решения задачи обнаружения DoS-атак в ИТ КФС предлагается ансамбль классификаторов, превосходящий по точности одиночные классификаторы [12, с. 12-18].

Пусть $x_i \in R^n$ $(i = \overline{1,n})$ – точка из набора данных, где n – общее количество точек во входном наборе данных, $M = \{M_1, M_2, \dots, M_m\}$ – набор

методов классификации, впоследствии объединенных в ансамбль, m – количество методов классификации, a_{ij} – значение классификации для точки x_i, k – количество классов (Таблица 3.2.1).

Таблица 3.2.1

Оценка классификаторов по каждой точке

Точка данных	Классификатор			Ансамблевое значение
X	M_1	…	M_m	
x_1	a_{11}	…	a_{1m}	$\max_j(a_{1j})$
⋮	⋮	⋱	⋮	⋮
x_n	a_{n1}	…	a_{nm}	$\max_j(a_{nj})$

Предлагаемый алгоритм, указывающий вероятность принадлежности к определенным классам, возвращает вектор оценок классификации для каждой точки данных.

Разработанный подход состоит в том, что для каждой точки из набора данных прогнозируемая метка класса соответствует максимальному значению среди всех оценок, полученных методами кластеризации для данной точки. Алгоритм предлагаемого подхода для обнаружения сетевых атак на основе ансамбля классификаторов представлен ниже:

Вход: $x_i \in R^n$: набор точек данных

n: количество точек данных

m: количество классификаторов

$M = \{M_1, M_2, \dots, M_m\}$: методы классификации

k: количество классов

Выход:

$A = \{a_{ij}\}_{n \times m}$: значения классификации для каждой точки набора данных

P: вектор значений ансамбля

for i=**1 to** n **do**

for j**=1 to** k **do**

Вычислить значение a_{ij} для M_i

end

$P_i = \max_j a_{ij}$

end

Требуется повысить точность обнаружения DoS-атак за счет использования ансамбля классификаторов.

Результаты экспериментов. Для экспериментов была рассмотрена база данных сигнатур NSL-KDD, построенная на основе базы данных KDD-99 по инициативе Американской ассоциации агентства перспективных оборонных исследовательских проектов DARPA [361, с. 53-58]. Для проведения исследования в области обнаружения вторжений был собран набор данных связей, который охватывает широкий спектр вторжений, смоделированных в среде, имитирующей сеть ВВС США.

Статистический анализ показал, что в базах данных есть важные проблемы, которые сильно влияют на производительность систем, а также приводят к очень плохой оценке подходов к обнаружению аномалий. Рассматриваемая база данных NSL-KDD имеет следующие преимущества:

1. Отсутствие избыточных записей в обучающей выборке, чтобы классификатор не показывал результатов смещения.

2. В тестовом наборе нет повторяющихся записей. Он содержит некоторые атаки, которых нет в обучающем наборе данных.

3. Количество записей, выбранных из каждой группы уровней сложности, обратно пропорционально количеству записей в исходном наборе данных KDD.

Обучающий набор данных состоит из 21 различных атак из 37, представленных в тестовом наборе. Кроме того, допустимо количество записей в обучающей (125973 выборки) и тестовой выборке (22544 выборки) NSL-KDD. Это преимущество делает его доступным для экспериментов с исчерпывающими данными без необходимости случайного выбора небольшой

части данных. Следовательно, оценка результатов различных исследовательских проектов непротиворечива и сопоставима.

Все атаки в NSL-KDD делятся на четыре группы:

• DoS-атаки включают: «neptun», «back», «smurf», «pod», «land» и «teardrop».

• Атаки U2R включают: «buffer_overflow», «loadmodule», «rootkit» и «perl».

• Атаки R2L включают: «warezclient», «multihop», «ftp_write», «imap», «guess_passwd», «warezmaster», «spy» и «phf».

• Атаки Probe (Probing Attack) включают: «portsweep», «satan», «nmap» и «ipsweep».

К основным задачам, выдвигаемым при обнаружении сетевых вторжений, относятся распознавание атак редких типов, повышение точности обнаружения подозрительной активности, а также повышение эффективности моделей обнаружения вторжений в режиме реального времени. Каждая запись имеет 41 атрибут, описывающий различные признаки.

Для оценки производительности классификаторов используются следующие метрики: точность, precision, recall и F-мера. Для любого алгоритма классификации возможны четыре случая классификации, и это помогает понять разницу между следующими показателями: истинно-положительные (TP), ложно-положительные (FP), истинно-отрицательные (TN) и ложно-отрицательные (FN) результаты.

Точность классификации можно определить как долю правильных результатов, которая достигается классификатором. Precision показывает какая часть объектов, идентифицированных классификатором как положительные, действительно положительна. Recall определяет, какая часть положительных объектов была отобрана классификатором. F-мера – это метрика, которая сочетает в себе precision и recall.

Частота ошибок (или ошибка неправильной классификации) измеряет отношение неправильно классифицированных образцов к общему количеству классифицированных образцов.

$$Error\ rate = \frac{FP+FN}{TP+TN+FP+FN}. \quad (3.2.4)$$

Эксперименты проводились с использованием платформы операционной системы Windows® 10–64 бит с процессором Core i7 2,5 ГГц, 8,0 ГБ ОЗУ. Предлагаемый подход был оценен на R 3.4.1. Набор данных NSL-KDD был разделен на 3 класса (DoS, «нормальное» состояние и другие атаки (в них вошли атаки U2R, R2L и Probe)).

Проведен сравнительный анализ предлагаемого подхода с одиночными классификаторами. Ансамбль классификаторов состоял из комбинаций DT, SVM с различными функциями ядра, NB и KNN [12, с. 12-18].

Точность, precision, recall и F-мера рассматривались в качестве метрик оценки. Результаты классификации представлены в таблицах 3.2.2-3.2.5. Лучшие результаты были отмечены жирным шрифтом. Из этих таблиц были получены ранги (показаны в квадратных скобках) для каждой метрики каждого класса, результаты которых приведены в таблицах 3.2.6-3.2.8.

Таблица 3.2.2

Сравнение точности классификации предлагаемого алгоритма с другими классификаторами

Класс / Метод	DoS	«Нормальное» состояние	Другие атаки
DT	86.32% [7]	77.19% [7]	64.25% [8]
KNN	88.21% [3]	79.67% [3]	65.80% [6]
SVM(Linear)	87.21% [5]	77.96% [6]	66.22% [5]
SVM(Polynom)	86.64% [6]	79.50% [4]	68.31% [3]
SVM(RBF)	87.25% [4]	78.84% [5]	65.38% [7]
NB	80.45% [8]	71.25% [8]	66.50% [4]
DT+KNN+SVM(Polynom)	90.74% [2]	84.77% [2]	74.53% [2]
DT+KNN+SVM(Polynom)+NB+ SVM (Linear)	**92.33% [1]**	**88.58% [1]**	**83.35% [1]**

Из таблицы 3.2.2 можно сделать вывод, что наибольшая точность (92.33%) обнаружения DoS-атак была достигнута для ансамбля из пяти классификаторов (DT+KNN+SVM(Polynom)+NB+SVM(Linear)), что превзошло результат классификатора KNN на 4.12%.

Несмотря на то, что NB показывает самый низкий результат (80.45%), при добавлении его в ансамбль классификаторов точность предложенного подхода увеличилась и составила 92.33% для пяти классификаторов.

Сравнение значений recall и precision при обнаружении DoS-атак показано в таблицах 3.2.3 и 3.2.4, соответственно.

Таблица 3.2.3

Сравнение методов классификации согласно precision

Класс / Метод	DoS	«Нормальное» состояние	Другие атаки
DT	95.94% [5]	63.17% [8]	88.15% [4]
KNN	95.88% [6]	65.87% [4]	86.05% [6]
SVM(Linear)	96.65% [3]	64.37% [7]	84.47% [7]
SVM(Polynom)	96.14% [4]	65.70% [5]	88.12% [5]
SVM(RBF)	85.07% [7]	64.92% [6]	96.17% [2]
NB	73.03% [8]	74.78% [2]	41.80% [8]
DT+KNN+SVM(Polynom)	96.74% [2]	71.95% [3]	94.41% [3]
DT+KNN+SVM(Polynom)+NB+ SVM (Linear)	**99.98% [1]**	**76.81% [1]**	**100% [1]**

Таблица 3.2.4

Сравнение методов классификации согласно recall

Класс / Метод	DoS	«Нормальное» состояние	Другие атаки
DT	74.19% [8]	97.33% [6]	29.75% [8]
KNN	78.08% [3]	97.62% [3]	33.28% [6]
SVM(Linear)	75.72% [5]	96.19% [7]	34.42% [5]
SVM(Polynom)	74.77% [6]	97.50% [5]	38.23% [4]
SVM(RBF)	76.00% [4]	97.57% [4]	32.54% [7]
NB	74.51% [7]	57.07% [8]	58.51% [2]
DT+KNN+SVM(Polynom)	82.86% [2]	98.62% [2]	49.99% [3]
DT+KNN+SVM(Polynom)+NB+ SVM (Linear)	**84.67% [1]**	**100% [1]**	**66.70% [1]**

Таблица 3.2.5

Сравнение методов классификации по метрике F-мера

Класс / Метод	DoS	«Нормальное» состояние	Другие атаки
DT	83.67% [7]	76.62% [7]	44.48% [8]
KNN	86.07% [3]	78.66% [3]	48.00% [6]
SVM(Linear)	84.91% [4]	77.13% [6]	48.91% [4]
SVM(Polynom)	84.12% [6]	78.50% [4]	53.33% [3]
SVM(RBF)	84.90% [5]	77.97% [5]	47.07% [7]
NB	73.76% [8]	64.74% [8]	48.76% [5]
DT+KNN+SVM(Polynom)	89.27% [2]	83.20% [2]	65.37% [2]
DT+KNN+SVM(Polynom)+NB+ SVM (Linear)	**91.69% [1]**	**86.88% [1]**	**80.02% [1]**

Таблица 3.2.6

Результирующий ранг методов класса DoS

Метод	Сколько раз метод находится в ранге *s*, *s* =								Результирующий ранг	
	1	2	3	4	5	6	7	8		
DT+KNN+SVM(Polynom)+NB +SVM (Linear)	4	0	0	0	0	0	0	0	4.0	1
DT+KNN+SVM(Polynom)	0	4	0	0	0	0	0	0	3.5	2
KNN	0	0	3	0	0	1	0	0	2.6	3
SVM(RBF)	0	0	0	2	1	0	1	0	2.3	4
SVM(Linear)	0	0	1	1	2	0	0	0	1.9	5
SVM(Polynom)	0	0	0	1	0	3	0	0	1.8	6
DT	0	0	0	0	1	0	2	1	1.1	7
NB	0	0	0	0	0	0	1	3	0.4	8

Таблица 3.2.7

Результирующий ранг методов класса «нормальное» состояние

Метод	Сколько раз метод находится в ранге *s*, *s* =								Результирующий ранг	
	1	2	3	4	5	6	7	8		
DT+KNN+SVM(Polynom)+NB +SVM (Linear)	4	0	0	0	0	0	0	0	4.0	1
DT+KNN+SVM(Polynom)	0	3	1	0	0	0	0	0	3.4	2
KNN	0	0	3	1	0	0	0	0	2.9	3
SVM(Polynom)	0	0	0	2	2	0	0	0	2.3	4
SVM(RBF)	0	0	0	1	2	1	0	0	2.0	5
SVM(Linear)	0	0	0	0	0	2	2	0	1.3	6
NB	0	1	0	0	0	0	0	3	1.3	7
DT	0	0	0	0	0	1	2	1	1.0	8

В таблице 3.2.5 представлены результаты по метрике F-мера, где оценивается эффективность методов классификации в целом, объединяющая значения precision и recall.

Для класса DoS ансамбль DT+KNN+SVM(Polynom)+NB+SVM(Linear) показал лучший ранг (Aliguliyev, 2009) по всем четырем метрикам accuracy, precision, recall и F-мера. Из таблицы 3.2.6 наихудший результат показал метод NB по трем метрикам (из четырех), а метод DT имел худший ранг по одной метрике (recall).

Из таблицы 3.2.7 ансамбль DT+KNN+SVM(Polynom)+NB+SVM (Linear) показал наилучшие результаты для класса «нормальное» состояние.

Наилучший ранг в классе «Другие атаки» получен для ансамбля DT+KNN+SVM(Polynom)+NB+SVM(Linear) по четырем метрикам accuracy, recall, precision и F-мера (Таблица 3.2.8).

Таблица 3.2.8

Результирующий ранг методов для класса «Другие атаки»

Метод	Сколько раз метод находится в ранге *s*, *s* =								Результирующий ранг	
	1	2	3	4	5	6	7	8		
DT+KNN+SVM(Polynom)+NB +SVM (Linear)	4	0	0	0	0	0	0	0	4.0	1
DT+KNN+SVM(Polynom)	0	2	2	0	0	0	0	0	3.3	2
SVM(Polynom)	0	0	2	1	1	0	0	0	2.6	3
NB	0	1	0	1	1	0	0	1	2.1	4
SVM(Linear)	0	0	0	1	2	0	1	0	1.9	5
SVM(RBF)	0	1	0	0	0	0	3	0	1.6	6
KNN	0	0	0	0	0	4	0	0	1.5	7
DT	0	0	0	1	0	0	0	3	1.0	8

Из таблиц 3.2.6-3.2.8 можно сделать вывод:

1) Наилучшие результаты среди методов обнаружения сетевых атак показал ансамбль DT+KNN+SVM(Polynom)+NB+SVM (Linear).

2) Методы SVM(Linear) и NB имеют равные ранги (Таблица 3.2.7).

3) Предлагаемый подход с различными комбинациями классификаторов превосходит одиночные классификаторы.

Для каждого классификатора были рассчитаны коэффициенты ошибок [50]. Отдельные классификаторы (DT, NB, SVM и KNN) сравнивались с различными ансамблями классификаторов с использованием предложенного подхода.

KNN, SVM(Linear) и SVM(RBF) дали самую низкую частоту ошибок (Error rate) среди отдельных классификаторов для класса DoS (11.79%, 12.79% и 12.75%, соответственно). Наименьший результат по классу DoS показал метод NB и составил 19.55%. Несмотря на то, что классификатор NB имеет самый высокий процент ошибок, в сочетании с другими классификаторами в предлагаемом подходе он показал хороший результат.

Таким образом, в настоящее время обработка и анализ больших данных важны для обеспечения информационной безопасности КФС. Обнаружение вторжений является одной из серьезных проблем в области сетевой безопасности. В данном исследовании для противодействия кибератакам на ИТ КФС был успешно применен ансамбль классификаторов. Ансамбль повышает точность распознавания за счет объединения различных отдельных классификаторов. Ансамбль классификаторов состоял из комбинаций DT, SVM с различными функциями ядра, алгоритмов NB и KNN. В целом рассмотренные методы классификации показали высокую точность обнаружения DoS-атак. Наиболее точный результат показал ансамбль из пяти классификаторов (DT+KNN+SVM(Polynom)+NB+SVM(Linear)). Можно сделать вывод, что разработанный метод к обнаружению сетевых атак может иметь практическое значение.

3.3. Алгоритм обнаружения вредоносного программного обеспечения в киберфизических системах на основе изображений

Повышенная подверженность вредоносному ПО является одним из самых серьезных последствий, с которыми сталкиваются КФС системы в связи с ростом цифровизации, такой как облачные технологии, электронная коммерция и конвергенция ИТ и ОТ [320, с. 176-192; 252, с. 213-232; 237, с. 1-15].

В последние годы вредоносное программное обеспечение для КФС претерпело изменения. КФС состоят из большого количества подключенных устройств (например, сенсоров, интеллектуальных электрических счетчиков и т. д.), на которые нацелены многие семейства вредоносных программ, такие как Tsunami, Bashlite и Mirai [96, с. 1-8; 301, с. 280-286]. Они используют слабую аутентификацию, устаревшие прошивки и сканеры, предназначенные для поиска открытых портов и компрометации системных устройств [219, с. 3171-3189].

Вредоносные программы имеют много общего друг с другом, а именно схожий исходный код [319, с. 1-15; 390, с. 1-14]. Они постоянно развиваются благодаря поддержке DDoS-атак для более критического заражения системных устройств [394, с. 1-12].

Самым слабым звеном в цепочке безопасности КФС является человеческий фактор. Киберпреступники используют этот фактор для получения несанкционированного доступа, кражи персональных данных и заражения систем вредоносным ПО [241, с. 1-35; 129, с. 4677-4694].

Таргетированное вредоносное ПО как вид кибероружия имеет следующие особенности [297, с. 1-12]: эксплуатация уязвимостей, включая атаки нулевого дня; маскировка и самоуничтожение; поддержка, обновление и управление инфраструктурой; широкий функционал в области решения целевых задач; высокое качество кода и т. д.

Среди различных вариантов вредоносных программ можно выделить криптомайнеры, вирусы, программы-вымогатели, черви, шпионские программы и т. д. Их основной целью является незаконный сбор информации, блокировка сервисов и кибершпионаж.

Таким образом, на основе анализа характеристик вредоносного ПО можно сделать вывод о наличии статических и динамических характеристик [151, с. 1-12; 299, с. 21-30]. Их можно использовать в качестве признаков для обнаружения вредоносного кода.

Обнаружение вредоносного ПО полностью зависит от векторов кибератак, которые можно наблюдать на основе анализа полученных образцов [31, с. 3432-3447]. Новые подходы рассматривают исполняемый файл как последовательность инструкций языка ассемблера (файл Asm) или последовательность байтов [333, с. 206303-206324]. Байт-файл вредоносного ПО – это шестнадцатеричное представление переносимого исполняемого вредоносного файла [340, с. 107-119; 403, с. 438-451]. Характеристики, извлеченные из таких файлов, не всегда предоставляют классификатору полезную информацию [333, с. 206303-206324]. Однако в этом легко убедиться, визуально анализируя бинарный код.

В последнее время визуализация стала использоваться как альтернативный и практичный подход к анализу вредоносного ПО, не требующий глубокого исследования [333, с. 206303-206324; 388, с. 1748-1771]. Было проведено несколько исследований по обнаружению и классификации вредоносных программ в КФС с использованием машинного обучения. При анализе вредоносных образцов они преобразуются в текстовые данные или данные по полученным изображениям для последующего применения методов машинного обучения. В текстовых данных байт кодирует информацию в последовательности букв или цифр. Более того, в изображениях байт представляет интенсивность пикселя. В этом случае структура двоичных образцов преобразуется в двумерные изображения, а полученные из них признаки затем используются для классификации вредоносных программ.

Так, Nataraj и соавторы предложили стратегию полутоновых изображений для классификации вредоносных программ [299, с. 21-30]. Двоичные образцы вредоносных программ считываются как векторы беззнаковых 8-битных целых чисел и преобразуются в матрицу. Она сохраняется как изображение, из которого извлекаются признаки GIST. В этом подходе классификация основана на алгоритме KNN с использованием метрики евклидового расстояния. Этот подход можно применять для анализа больших наборов данных и

классификации упакованных и неупакованных двоичных образцов вредоносных программ. Ali и соавторы разработали метод обнаружения вредоносных программ на основе N-грамм и логистической регрессии (LR) [68, с. 1-21]. Для извлечения индикатора компрометации из вредоносных файлов была применена методика динамического анализа. Недавняя работа была сосредоточена на применении подхода трансферного обучения к классификации вредоносных программ на основе изображений. Yan и соавторы предложили алгоритм парного вращения, инвариантного к совместному возникновению локального бинарного шаблона для классификации вредоносного кода с лучшей дискриминирующей способностью и обнаружили, что он также обладает линейной разделимостью между различными семействами вредоносного кода [389, с. 852-859]. Метод обнаружения вредоносных программ основан на использовании двух глубоких нейронных сетей: CNN и LSTM, за которыми следует ансамбль стекирования для их последующего слияния. Модель CNN на основе глубокого обучения, которая не требует проектирования признаков, была предложена в работе [255, с. 334-351]. Xiao и соавторы предложили метод визуализации вредоносных программ, который объединяет цветные поля меток (CoLab), VGG16 и SVM. CoLab отмечает разделы PE-файла, чтобы дополнительно подчеркнуть информацию о распределении разделов в преобразованном изображении вредоносного ПО [380, с. 1-14]. Carletti и соавторы обсудили анализ устойчивости к обфускации с использованием ResNet50, InceptionV3, VGG16 и MobileNet для систем классификации вредоносных программ на основе изображений. Модель Mobilenet показала наилучший результат [121, с. 414-423]. В работе [388, с. 1748-1771] была разработана структура на основе EfficientNetB0 в сочетании с SVM и RF. Однако, в результате экспериментов наблюдалась неправильная классификация некоторых образцов вредоносных программ из-за их схожести. Lachtar и соавторы предложили решение, которое использует методы визуализации для преобразования инструкций приложения в изображения и

применили три предварительно обученные архитектуры CNN: LeNet, AlexNet и InceptionV3 [256, с. 134-137].

Предложенный подход преобразует вредоносный код в изображения в оттенках серого [52, с. 1-14]. Рассматриваются две предварительно обученные модели глубокого обучения: AlexNet и MobileNet. Используются как изображения вредоносных программ в оттенках серого, так и на основе преобразования Радона [233, с. 1-12; 405, с. 2097-2111]. Байтплоты в оттенках серого отправляются в AlexNet. В то время как изображения, полученные после преобразования Радона, используются в качестве входных изображений для MobileNet. Затем эмпирически выполняется выбор признаков. Далее полученные признаки подаются на два полностью связанных слоя для классификации в соответствии с рассматриваемыми семействами вредоносных программ. Эксперименты проводятся на трех больших наборах данных вредоносного ПО [46, с. 888-898; 193, с. 3948-3952; 330]. Предложенный подход классифицирует вредоносное ПО на семейства по общим визуальным характеристикам и может быть использован для последующего принятия решений.

В целом, основные вклады текущего исследования:

• Разработка ансамблевой модели на основе трансферного обучения, которая объединяет визуализацию вредоносных программ и применение преобразования Радона.

• Модель не требует разработки новых признаков и основана на визуальном анализе изображений вредоносных программ.

• Экспериментальные наборы данных различных размеров (Microsoft malware, IoT_Malware и MalNet-Image) рассматривались для классификации вредоносного ПО в КФС.

• Сравнение производительности с другими известными методами машинного обучения доказало превосходство предлагаемого подхода с точки зрения эффективности.

• Экспериментальные результаты показывают, что объединение признаков из двух глубоких нейронных сетей, AlexNet и MobileNet, может эффективно классифицировать вредоносное ПО в КФС даже при небольших изменениях в изображениях.

Предлагаемый подход включает три основных этапа: предварительную обработку, извлечение признаков и классификацию вредоносного ПО. Общая схема предлагаемого подхода представлена на рис. 3.3.1.

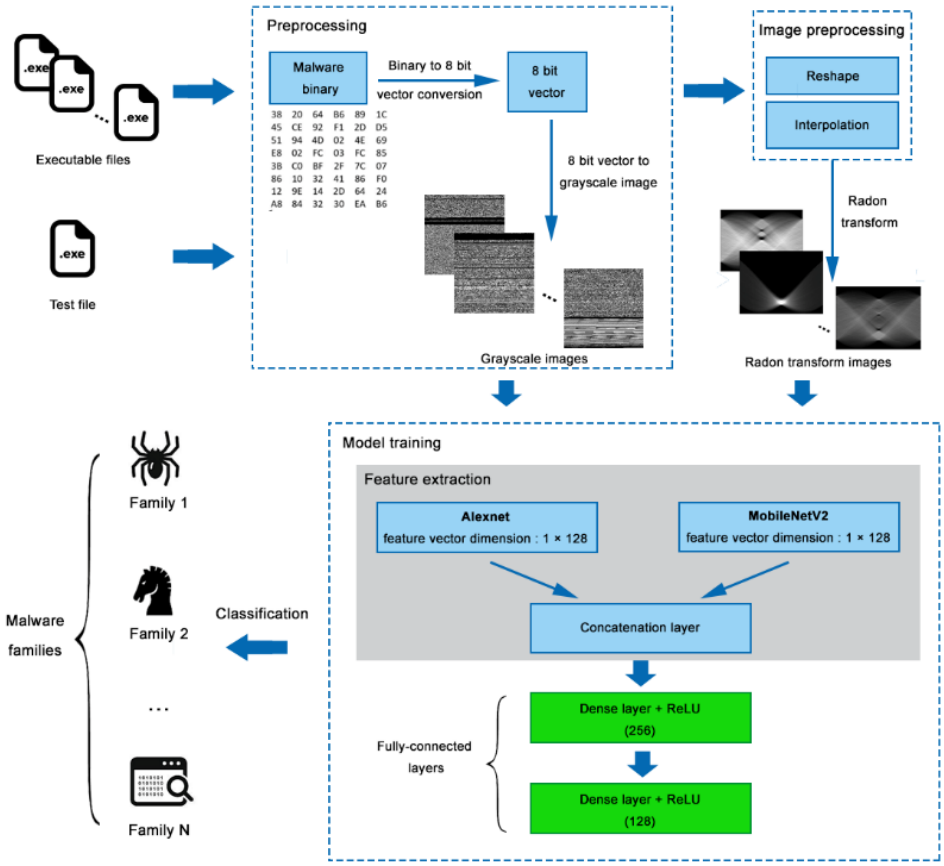


Рис. 3.3.1. Блок-схема предлагаемого подхода [48]

На первом этапе образцы бинарных файлов преобразуются в изображения. Для этого последовательность байтов преобразуется в двоичную форму PNG изображения в оттенках серого с использованием подхода, предложенного Nataraj и соавторы [299, с. 21-30]. Он преобразует двоичный файл вредоносного ПО в последовательность 8-битных строк, причем каждая строка преобразуется в десятичное число, представляющее один пиксель канала от 0 до 255.

Синусоидальные текстурные признаки извлекаются из изображений в оттенках серого с помощью преобразования Радона.

Поскольку изображения вредоносных программ не имеют фиксированной формы, их размер изменяется до 224×224 пикселей с использованием алгоритма билинейной интерполяции с сохранением текстуры изображения. Далее образцы предварительно обрабатываются и нормализуются.

На следующем этапе преобразование Радона применяется к полученным изображениям вредоносных программ в оттенках серого. В отличие от большинства методов, преобразование Радона является обратимым преобразованием изображения и поэтому может рассматриваться как метод представления текстуры изображения [400, с. 8-16]. Алгоритм также параллельно применяет преобразование Радона к изображениям вредоносных программ.

Преобразование Радона широко используется в компьютерной томографии, медицине, геодезии и др. [405, с. 2097-2111]. Двумерное преобразование Радона имеет следующий вид:

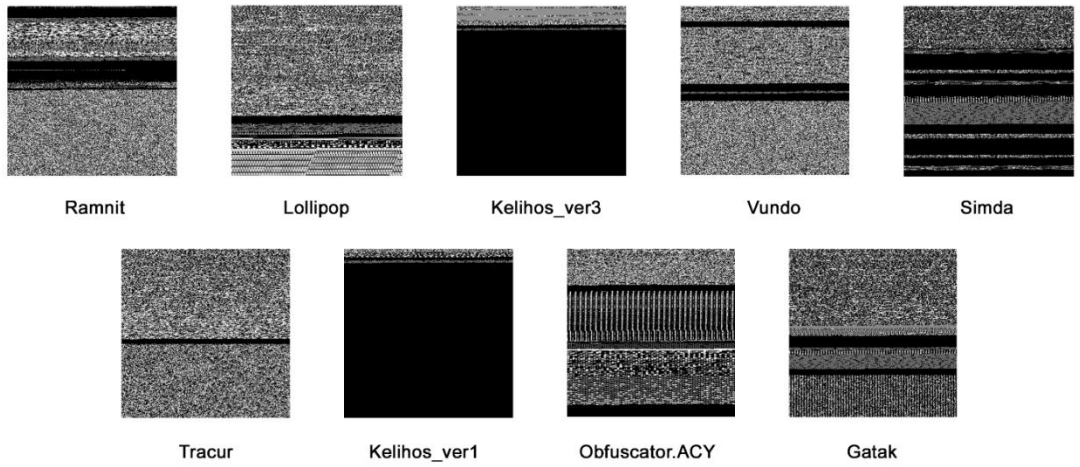
$$RT(\eta, \mu) = \iint_{R^2} f(x, y)\partial(\eta - x\cos\mu - y\sin\mu)dxdy. \quad (3.3.1)$$

Здесь η, μ – полярные координаты. В этом случае обратное преобразование Радона можно рассчитать как

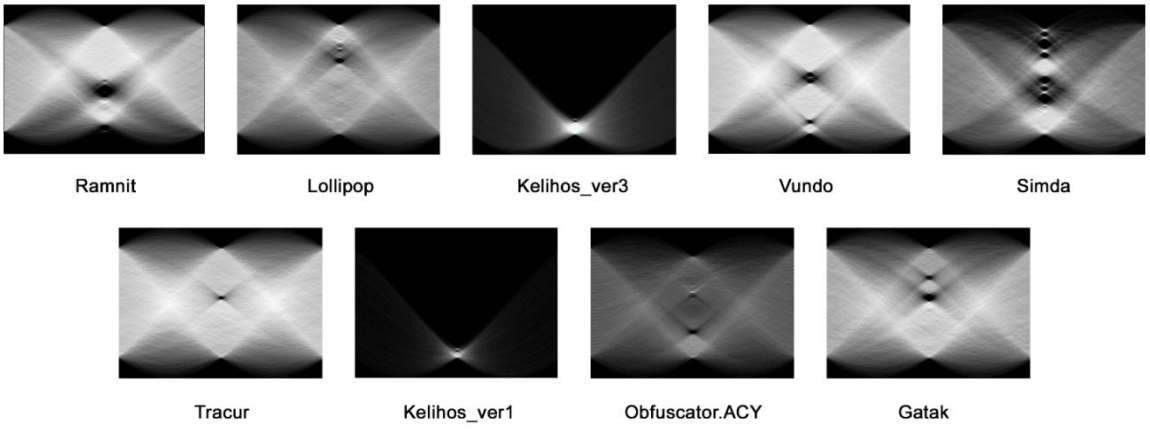
$$f(x, y) = \frac{1}{2p^2}\int_0^{2p}\int_{-\infty}^{\infty}\frac{\frac{\partial}{\partial\eta}RT(\eta,\mu)d\eta d\mu}{x\cos\mu + y\sin\mu - \eta}. \quad (3.3.2)$$

Результат включает образцы изображений вредоносного ПО с характерными текстурированными характеристиками в оттенках серого. Изображения, полученные после применения преобразования Радона, помогают интерпретировать обнаруженные вредоносные программы.

На рис. 3.3.2, 3.3.3 и 3.3.4 показаны полутоновые изображения двоичных файлов вредоносного ПО и соответствующие изображения, полученные после применения преобразования Радона для рассматриваемых наборов данных.



а)



б)

Рис. 3.3.2. Примеры изображений вредоносного ПО (а) и изображений, полученных после применения преобразования Радона (б) для набора данных Microsoft Malware BIG [48]

Видно, что неупакованные бинарные образцы, принадлежащие к разным семействам вредоносных программ, сильно различаются [147, с. 3187-3196].

Однако найти различия в структуре упакованных бинарных образцов разных семейств вредоносного ПО можно только с помощью методов машинного обучения.

Следующий шаг включает извлечение признаков из изображений в оттенках серого. Для этого к наборам данных изображений применяется трансферное обучение. В отличие от других методов машинного обучения, глубокое обучение позволяет быстро и автоматически извлекать признаки на основе визуального анализа [400, с. 8-16]. В этом исследовании были

рассмотрены две модели на основе CNN (AlexNet [251, с. 84-90] и MobileNetV2 [213]), для извлечения признаков из изображений вредоносных программ.

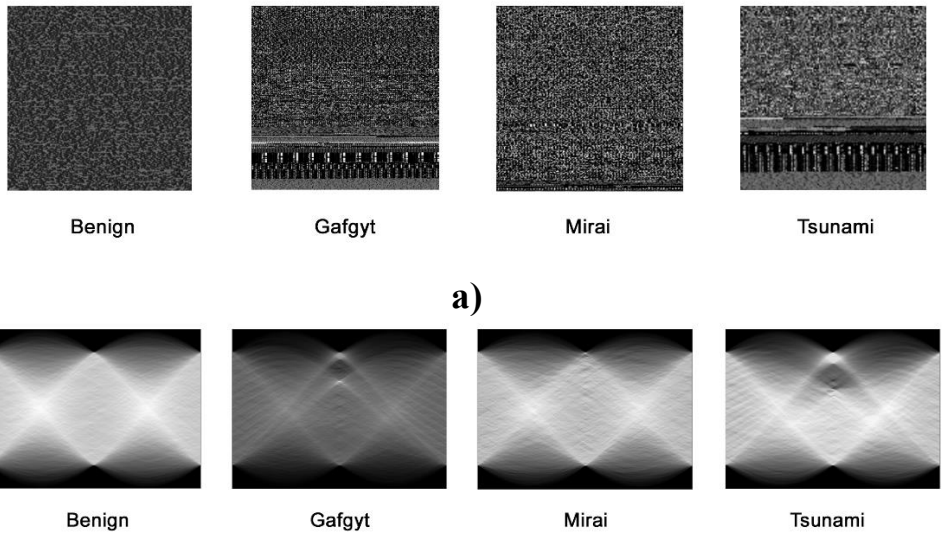
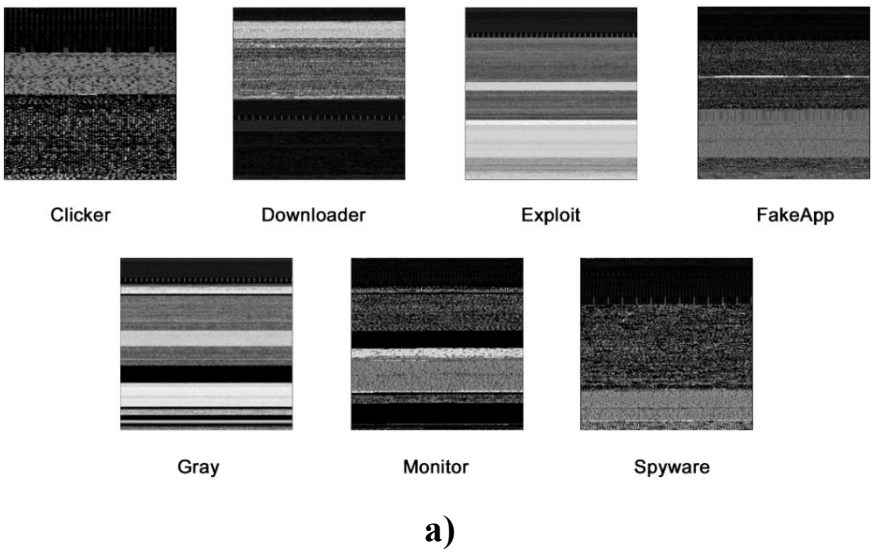
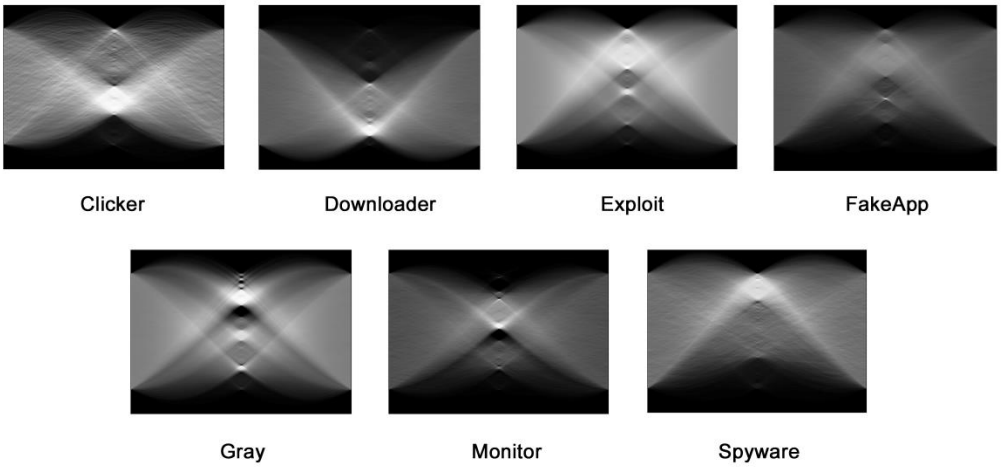


Рис. 3.3.3. Примеры изображений вредоносного ПО (а) и изображений, полученных после применения преобразования Радона (б) для набора данных IoT_Malware [48]

Увеличение числа данных выполняется, поскольку исходный набор данных несбалансирован. Процесс включает в себя несколько преобразований изображения: изменение размера, обрезку, поворот, переворот по горизонтали и др. [251, с. 84-90].





б)

Рис. 3.3.4. Примеры изображений вредоносного ПО (а) и изображений, полученных после применения преобразования Радона (б) для набора данных MaLNet-Images [48]

Модель CNN используется потому, что она имеет отличную производительность при обработке изображений. CNN основана на операции свертки, которая уменьшает размерность пространства признаков и извлекает из изображения более сложные абстрактные признаки. Его можно описать следующим образом:

$$h(i) = f(\sum_{x=1}^{l} \sum_{y=1}^{r} \xi_{xy} \times w_{xy}^{i} + b^{i}), \quad (3.3.3)$$

где $f(\cdot)$ – функция активации (а именно, ReLU), ξ_{xy} — значение узла (x, y) входных данных в фильтре; и w_{xy}^{i} показывает вес фильтра; b^{i} — параметр смещения; l — длина и ширина фильтра.

AlexNet — это легкая глубокая модель с широкими возможностями определения характеристик, которую можно легко обучить. Она быстро обучается и имеет высокую скорость прогнозирования. AlexNet доказал свою эффективность и показал многообещающие результаты в следующих задачах: диагностика неисправностей КФС [263, с. 1-15], обнаружение сетевых вторжений [170, с. 1-4] и т. д. Именно поэтому ее можно применять для решения задач классификации вредоносных программ. Модель AlexNet содержит пять сверточных слоев в модуле извлечения признаков и три полносвязных слоя в модуле классификации [251, с. 84-90]. AlexNet трудно

переобучить, поэтому она в основном используется для приложений идентификации и классификации [261, с. 144-153].

MobileNet – это небольшая модель с несколькими параметрами, разработанная для различных приложений кибербезопасности (например, аутентификация пользователей [255, с. 334-351; 380, с. 1-14], обнаружение мошенничества [242, с. 29-37], обнаружение вредоносного ПО [121, с. 414-423] и др.) при ограниченных ресурсах. AlexNet и MobileNet обучаются на миллионах изображений из базы данных ImageNet [115, с. 273-280]. Модель учитывает предварительно обученные веса изображений.

Сочетание преимуществ AlexNet и MobileNet требует больше вычислительных ресурсов и увеличивает время выполнения модели, но в то же время повышает эффективность с точки зрения точности.

В предложенной модели AlexNet рассматривает визуальное представление изображения вредоносного ПО в оттенках серого. Кроме того, MobileNet учитывает изображения, полученные после применения преобразования Радона.

Предложенная модель заменила последние слои AlexNet и MobileNet полносвязными слоями по 128 нейронов в каждом.

Далее полученные характеристики двух рассматриваемых моделей объединяются и подаются на два полносвязных слоя, состоящих из 256 и 128 нейронов, для классификации по семействам вредоносных программ. Каждый из нейронов полносвязного слоя связан со всеми нейронами предыдущего слоя и содержит большое количество параметров [41, с. 4776-4794].

Гиперпараметры предложенной модели, основанной на трансферном обучении, можно увидеть в таблице 3.3.1. Оптимизатор Адама показал многообещающие результаты при классификации изображений и применяется на этапе обучения, а функция активации softmax используется для расчета прогнозирующих вероятностей для всех семейств вредоносных программ в выходном слое.

Прогнозируемые вероятности рассчитываются следующим образом:

$$p(\hat{x} = j|x) = \frac{\exp(x_{b_i}(j))}{\sum_{s=1}^{S} \exp(x_{b_i}(s))}, \quad (3.3.4)$$

где s - число семейств вредоносного ПО.

Таблица 3.3.1

Гиперпараметры предложенной модели

Параметр	Значение
Тип оптимизатора	Adam, RMSProp, SGD
Количество эпох	30
Размер батча	128, 256
Скорость обучения	0.001, 0.0001
Функция активации	ReLU
Функция ядра SVM	RBF
Гиперпараметр gamma для SVM	0.1
Параметр регуляризации SVM	10
Число соседей в KNN	12
Число деревьев в RF	200

Производительность подхода оценивается в конце каждой эпохи и выбирается модель с наименьшей ошибкой валидации. В таблице 3.3.2 показаны результаты настройки гиперпараметров предложенной модели.

Алгоритм предлагаемого подхода приведен ниже:

Алгоритм

Вход: Вектор признаков $FA = \{fa_1, fa_2, \dots, fa_z\}$

Вектор признаков $FM = \{fm_1, fm_2, \dots, fm_w\}$

Выход: Классификация семейств вредоносного ПО $MF = \{mf_1, \dots, mf_k\}$

Шаг 1. Извлечение вектора признаков с помощью AlexNet

Шаг 2. Извлечение вектора признаков с помощью Mobilene

Шаг 3. Объединение признаков

Шаг 4. repeat

Шаг 5. for each epoch do

Шаг 6. Рассчитать значение функции потерь

Шаг 7. until the convergence condition is met

Шаг 8. Выбор наилучшей модели

Шаг 9. Обучение модели

Шаг 10. Оценка обученной модели на тестовых данных

Шаг 11. return MF

Категориальная кросс-энтропия определяет значение потерь для каждого класса [244, с. 1-15]. Она доказывает правильную классификацию входных данных и рассчитывается как

$$LogLoss = -\frac{1}{N}\sum_{k=1}^{K}\sum_{j=1}^{N} f(j,k)\log(f(j,k)), \qquad (3.3.5)$$

где N – количество образцов, K – количество классов и $f(j,k)$ - вероятность того, что образец j принадлежит классу k.

Таблица 3.3.2

Настройка гиперпараметров

Размер батча	Скорость обучения	Тип оптимизатора	Точность (%)		Значение потерь	
			Обучение	Валидация	Обучение	Валидация
128	0.001	Adam	86.26	85.16	0.4477	0.4857
		SGD	91.80	89.96	0.2950	0.2969
		RMSProp	83.98	82.63	0.5744	0.4884
	0.0001	Adam	87.30	87.08	0.4002	0.4071
		SGD	89.96	89.94	0.3033	0.3168
		RMSProp	90.68	87.99	0.2892	0.3787
256	0.001	Adam	92.06	91.25	0.5335	0.5573
		SGD	90.62	88.31	0.3870	0.2846
		RMSProp	86.51	82.46	0.8341	0.9391
	0.0001	**Adam**	**99.89**	**99.51**	**0.1932**	**0.1887**
		SGD	98.04	97.30	0.3524	0.3690
		RMSProp	92.03	89.53	0.6135	0.5139

Описание экспериментальных наборов данных. Рассматриваются три набора данных для проведения экспериментов: Microsoft Malware BIG, IoT_Malware и MaLNet-Image.

Набор данных Microsoft Malware BIG размещен на Kaggle и доступен для исследований [330]. Его размер составляет 500 Гб, и он содержит 9 различных семейств вредоносных программ, включая Rammit, Lollipop, Kelihos_ver3,

Vundo, Simda, Tracur, Kelihos_ver1, Obfuscator и Gatak. Каждое семейство вредоносных программ принадлежит к одному из шести типов: черви, рекламное ПО, бэкдоры, трояны, трояны-загрузчики и обфусцированное вредоносное ПО (Таблица 3.3.3).

Обучающий набор содержит 10868 образцов, а тестовый набор — 10873. Поскольку тестовый набор не содержит меток семейств вредоносных программ, в исследовании используется только обучающий набор данных. Каждый образец данных имеет шестнадцатеричное представление двоичного содержимого вредоносного ПО и файлов исходного кода на языке ассемблера, созданных с помощью дизассемблера IDA. Эта информация включает в себя последовательности сборки, строки, вызовы функций и т. д.

Таблица 3.3.3

Описание набора данных Microsoft malware BIG

	Тип вредоносного ПО	Число образцов	Категория	Описание
1	Ramnit	1541	Червь	Способен быстро распространяться и самовоспроизводиться без вмешательства человека.
2	Lollipop	2478	Рекламное ПО	Предназначен для рекламы и является источником дохода.
3	Kelihos_ver3	2942	бэкдор	Скрытно пускает злоумышленника в систему, предоставляя права администратора.
4	Simda	42		
5	Kelihos_ver1	398		
6	Gatak	1013		
7	Vundo	475	троян	Внедряется в устройство незаметно для пользователя и передает персональные данные третьему лицу.
8	Tracur	751	троян-загрузчик	Троян, который перенаправляет на другую страницу с помощью результатов поиска в Интернете.
9	Obfuscator.ACY	1228	скрытое вредоносное ПО	Используется для обхода антивирусных сканеров.

Набор данных IoT_Malware для обнаружения вредоносного ПО в КФС был собран Alasmary и соавторами [46, с. 888-898]. Образцы были случайным образом отобраны из CyberIOC с января 2018 года по конец февраля 2019 года. Набор данных содержит 3016 «нормальных» образцов и 13798 вредоносных файлов. Образцы в наборе данных IoT_Malware были оценены с помощью VirusTotal [370]. Вредоносное ПО для КФС включает три класса: Gafgyt (11128 образцов), Mirai (2408 образцов) и Tsunami (262 образца) (Таблица 3.3.4).

Большой общедоступный набор данных MaLNet-Image содержит более 1.2 миллиона двоичных образов, включающих 47 типов и 696 семейств вредоносного ПО [193, с. 3948-3952]. Для проведения экспериментов были выбраны 60437 образцов вредоносного ПО и 19736 «нормальных» образцов из набора данных (Таблица 3.3.5).

Таблица 3.3.4

Описание набора данных IoT_Malware

	Тип вредоносного ПО	Число образцов	Категория	Описание
1	Benign	3,016		
2	Gafgyt	11,128	бэкдор	Это семейство вредоносных программ используется для DDoS-атак.
3	Mirai	2,408	бэкдор	Это вредоносное ПО в Linux, которое получает команды от серверов управления для проведения кибератак.
4	Tsunami	262	бэкдор	Backdoor.Linux.Tsunami предоставляет злоумышленнику полный доступ к зараженному компьютеру, который становится частью ботнета.

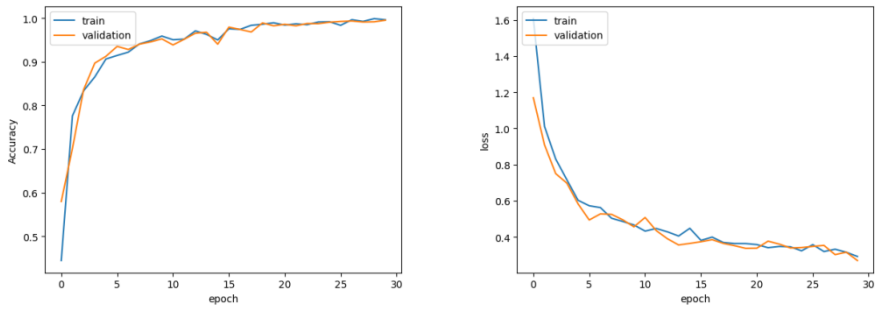
Результаты экспериментов. Предложенный подход был реализован в среде Python 3.7.12 с использованием библиотеки Keras с Tensorflow. Эксперименты проводились на виртуальной машине с процессором Intel Xeon (R) X5670 @ 2,93 ГГц * 24 и 24 ГБ оперативной памяти. Чтобы эффективно оценить производительность предлагаемого алгоритма, наборы данных случайным образом разделяются на обучающий (80%) и валидационный набор (20%). Подход выполняется 20 раз, чтобы уменьшить ошибку классификации.

Таблица 3.3.5

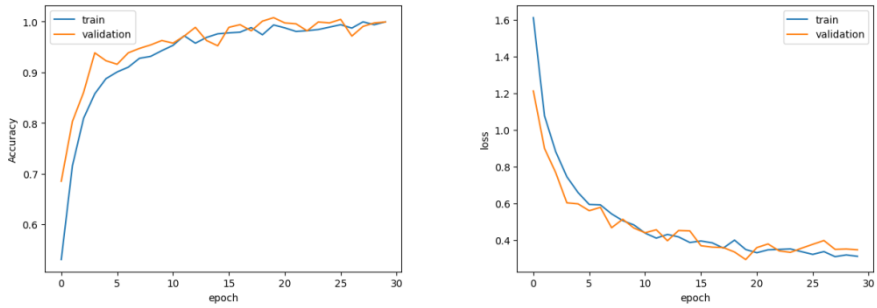
Описание набора данных MaLNet-Image

	Тип вредоносного ПО	Число образцов	Категория	Описание
1	Benign	19736		
2	Addisplay	17458	Adware	Это вредоносное ПО, распространяемое через несколько приложений.
3	Adload	533	Adware	Использует атаку «Человек посередине» для внедрения рекламных объявлений на веб-страницы.
4	Clicker	591	Adware	Получает информацию об URL-адресе сканирования через сообщения Firebase.
5	Exploit	5571	Exploit	Может использовать известную уязвимость для получения несанкционированного доступа или контроля над устройством.
6	Riskware++SMSsend	547	Riskware	Это приложение с модулем для отправки SMS-сообщений, которое может предоставить больше контроля или доступа к устройству, чем разрешено.
7	SPR	13832	Riskware	Вредоносное ПО, требующее потенциально опасные разрешения, угрожающие конфиденциальности пользователей.
8	Spyware	6590	Spyware	Может повлиять на конфиденциальность, производительность пользователя или контроль над компьютером и устройством.
9	SMSsend++Trojan	4524	Trojan	Используется для получения прибыли, отправляя SMS на платные номера.
10	Monitor	1357	Spyware	Отслеживает действия на контролируемом устройстве и сохраняет информацию удаленно.
11	ROG	1975	Ransomware	Шифрует личные документы на компьютере жертвы, а затем предлагает расшифровать данные после оплаты.
12	Gray	930	Grayware	Находится в серой зоне между обычным программным обеспечением и вирусом.
13	Hacktool	542	HackTool	Используется для отправки потока сетевых пакетов на целевой компьютер на основе известных уязвимостей.
14	FakeApp	425	Riskware	Копирует функции других приложений.
15	Virus	465	Virus	Перенаправляет пользователя на веб-страницу, похожую на Office 365, после производится загрузка вредоносного файла.
16	Downloader	4997	Trojan-Downloader	Тайно загружает вредоносное ПО с удаленного сервера, затем устанавливает и запускает файлы.

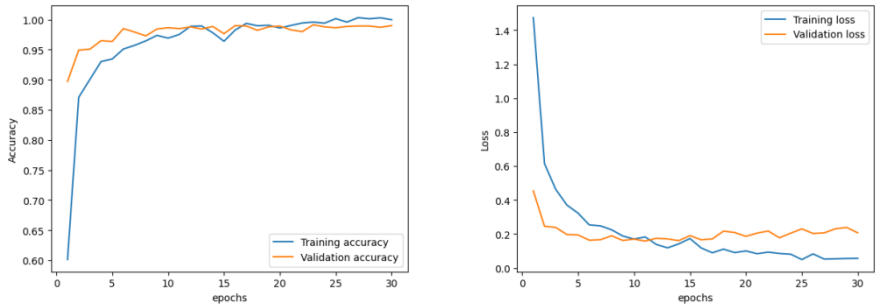
Эксперименты проводятся на наборах данных Microsoft Malware BIG, IoT_Malware и MaLNet-Image. Полученные изображения в оттенках серого были нормализованы. Изображения были изменены до размера 224x224 и отправлены на вход моделей AlexNet и MobileNet. Матрица связности $N \times N$ используется для расчета четырех типов метрик для проверки предлагаемого подхода для классификации вредоносного ПО. Этими метриками являются точность, precision, recall и F-мера [147, с. 3187-3196]. Кривые потерь при обнаружении и кривые точности классификации изображений из наборов данных Malware BIG, IoT_Malware и MaLNet-Image показаны на рис. 3.3.5.



а) Microsoft malware BIG



б) IoT_Malware



в) MaLNet-Image

Рис. 3.3.5. Кривые точности обнаружения вредоносных программ и потерь для наборов данных вредоносных программ Microsoft malware BIG (а), IoT_Malware dataset (б) и MaLNet-Image (в) [48]

Для минимизации проблемы переобучения использовалась k-кратная перекрестная проверка.

Эксперименты показали оптимальное значение *k*, равное 5. По мере увеличения количества эпох наблюдается устойчивое снижение потерь при валидации.

Предложенная модель показала значение потерь равное 0.1932 на этапе обучения и 0.1887 на этапе валидации для набора данных Microsoft malware BIG. Сложность классификации наблюдалась для семейства Simda, что связано с небольшим количеством образцов для этого класса. При этом точность для класса Kelijos_ver3 составила 100%. В таблицах 3.3.6, 3.3.7 и 3.3.8 показана эффективность предложенной модели для различных семейств вредоносных программ на основе показателей precision, recall и F-мера [185, с. 27-38].

Таблица 3.3.6

Результаты классификации вредоносных программ предложенной модели для набора данных Microsoft malware BIG

Вредоносное ПО	Precision	Recall	F-мера
Ramnit	0.76	0.83	0.80
Lollipop	0.95	0.92	0.94
Kelihos_ver3	**1.00**	**1.00**	**1.00**
Vundo	0.93	0.93	0.93
Simda	0.40	0.29	0.33
Tracur	0.82	0.83	0.83
Kelihos_ver1	0.96	0.95	0.95
Obfuscator.ACY	0.93	0.90	0.92
Gatak	0.90	0.89	0.90

По данным метрики F-мера классы Lollipop, Vundo, Kelihos_ver3, Kelihos_ver1 и Obfuscator.ACY показали более 90% для набора данных Microsoft malware BIG. Однако семейства вредоносных программ Ramnit, Tracur и Gatak были менее точно распознаны по показателю precision и составили 76%, 82% и 90%, соответственно.

Таблица 3.3.7

Результаты классификации вредоносных программ предложенной модели для набора данных IoT_Malware

Вредоносное ПО	Precision (%)	Recall (%)	F-мера (%)
Benign	**99.50**	99.20	**99.35**
Mirai	97.88	**99.57**	98.72
Tsunami	96.09	92.81	94.42
Gafgyt	97.99	98.23	98.11

Таблица 3.3.8

Результаты классификации вредоносных программ предложенной модели для набора данных MalNet-Image

Вредоносное ПО	Precision (%)	Recall (%)	F-мера (%)
Benign	**100**	**100**	**100**
Addisplay	**100**	**100**	**100**
Adload	98.12	92.86	95.42
Clicker	97.78	**100**	98.88
Exploit	**100**	98.18	99.08
Riskware++SMSsend	92.00	**100**	95.83
SPR	**100**	**100**	**100**
Spyware	97.13	96.98	97.05
SMSsend++Trojan	**100**	98.88	99.44
Monitor	97.72	98.17	97.94
ROG	99.01	**100**	99.50
Gray	99.20	**100**	99.60
Hacktool	99.80	89.74	94.50
FakeApp	**100**	96.55	98.25
Virus	95.73	96.88	96.30
Downloader	99.32	**100**	99.66

Для четырех классов набора данных IoT_Malware по метрике precision были получены следующие результаты: Benign (99.50%), Gafgyt (97.99%),

Tsunami (96.09%) и Mirai (97.88%). Наиболее точно по метрике recall был идентифицирован класс семейства вредоносных программ Mirai, нацеленный на заражение КФС. В то же время метрики precision и F-measure позволяют определить класс программного обеспечения «Benign» почти со 100% точностью. Для набора данных MaLNeT+Images, согласно метрикам precision, recall и F-мера, Benign, Addisplay и SPR были классифицированы с 100% точностью. При этом другие рассмотренные классы вредоносного ПО также были распознаны с достаточно высокой точностью, как показано в Таблице 3.3.8.

В таблице 3.3.9 сравнивается предложенный подход к классификации вредоносного ПО с различными методами машинного обучения на наборах данных Microsoft malware BIG и IoT_Malware, включая MALGRA [68, с. 1-21], SVM, случайный лес, KNN и PRICoLBP [389, с. 852-859].

Таблица 3.3.9

Сравнение производительности предложенного подхода с различными методами машинного обучения

Методы	Признаки	Набор данных	Точность (%)
LR [64]	N-gram	Microsoft malware dataset	98.40
LBP, KNN, SVM, RF, Gradient boost classifier [389]	PRICoLBP		98.60
RF [94]	bytecode		99.44
Предложенный подход (AlexNet)	Изображения		90.23
Предложенный подход (Mobilenet)	Изображения		89.52
Предложенный подход (AlexNet+Mobilenet)	Изображения		**99.89**
RF	N-gram	IoT_Malware dataset	98.46
SVM	N-gram		97.53
KNN	N-gram		95.02
Предложенный подход (AlexNet)	Изображения		96.71
Предложенный подход (Mobilenet)	Изображения		90.28
Предложенный подход (AlexNet+Mobilenet)	Изображения		**99.95**

По сравнению с экспериментальными результатами рассматриваемых методов предложенная структура хорошо работает при извлечении признаков из изображений вредоносного ПО с использованием глубоких нейронных моделей AlexNet и MobileNet, как показано в Таблице 3.3.10.

Таким образом, приведенные выше экспериментальные результаты доказывают эффективность и применимость предложенного подхода к обнаружению вредоносного ПО в КФС.

Таблица 3.3.10

Оценка эффективности предлагаемого подхода с использованием современных методов, основанных на глубоком обучении

Модель	Набор данных	Точность (%)
VGG16 [380]	Microsoft malware dataset	98.94
MobileNet [122]		99.25
Xception [273]		99.17
LeNet, AlexNet, InceptionV3 [256]		99.70
CNN [255]		98.64
InceptionV3 [35]		99.60
Предложенный подход		**99.89**
Adversarial learning [27]	IoT_Malware dataset	97.67
CNN [94]		95.00
Предложенный подход		**99.95**
Vision Transformer [338]	MalNet-Image dataset	97.00
EfficientNetB0+SVM+RF [388]		92.90
Предложенный подход		**99.20**

В целом, в данном исследовании был разработан алгоритм для классификации вредоносного ПО в КФС на основе визуального сходства изображений и извлечения признаков с использованием моделей AlexNet и MobileNet. Была продемонстрирована возможность применения в предложенном преобразования Радона к изображениям в оттенках серого подходе для повышения точности классификации новых двоичных файлов вредоносных программ. Однако были проанализированы только изображения образцов вредоносных программ в оттенках серого без рассмотрения цветных изображений. Эксперименты проводились на трех наборах данных вредоносного ПО с девятью, четырьмя и шестнадцатью классами.

Предложенный подход сравнивается с различными методами машинного обучения, включая RF, KNN, LSTM, SVM и CNN. Оценка разработанного алгоритма показала, что он способен эффективно распознавать такие семейства вредоносных программ, как Gafgyt, Tsunami, Mirai и Kelihos для набора данных IoT_Malware. Классы Addisplay, Adload, Exploit, Riskware++SMSsend, Spyware, SMSsend++Trojan, ROG и Downloader также были распознаны с высокой точностью для большого набора данных MalNet-Image. В исследовании рассматривались неупакованные бинарные образцы вредоносных файлов. Исследование продемонстрировало стабильную производительность с точностью 99.89%, 99.95% и 99.20% для наборов данных вредоносных программ Microsoft, IoT_Malware и MalNeT-Images соответственно. Были также проведены различные исследования по применению различных типов признаков, включая N-граммы, байт-коды и изображения. Также было проведено сравнение с различными моделями на основе глубокого обучения, такими как VGG16, Xception, InceptionV3 и EfficientNetB0. Так было проведено сравнение с методом на основе CNN для классификации вредоносных программ с использованием изображений с применением трансферного обучения предложенным Lachtar и соавторами [256, с. 134-137]. Ограничением этого метода является сложность извлечения признаков. Он показал более низкую точность по сравнению с предложенным алгоритмом. Abusnaina и соавторы исследовали задачу классификации вредоносных программ в КФС и разработали модель, которая извлекает потенциальные вредоносные поведенческие шаблоны [31, с. 3432-3447]. Сравнение доказало, что предложенный ансамблевый подход на основе моделей AlexNet и Mobilenet показывает относительное улучшение на 2.24% по сравнению с данной работой. Таким образом, точность предложенного подхода превзошла рассмотренные методы классификации вредоносных программ на наборах данных Microsoft Malware classification и IoT_Malware. А эксперименты на большом наборе данных MalNeT-Images доказали превосходство

разработанного алгоритма по сравнению с другими методами, основанными на глубоком обучении. Seneviratne и соавторы [338, с. 103121-103135] предложили самоконтролируемую модель на основе архитектуры Vision Transformer. Сравнение предложенного подхода с этой моделью [338, с. 103121-103135] показало относительное улучшение на 2.27%.

Таким образом, учитывая масштаб угроз кибербезопасности КФС и большое количество незащищенных устройств, это исследование является важным шагом на пути к разработке новых методов защиты систем и практических инструментов. Эта работа может быть полезна исследователям и практикам для своевременного обнаружения и устранения киберугроз КФС.

Выводы к главе III

- Разработан метод классификации кибератак на КФС с применением экстремального машинного обучения;
- Разработан метод обнаружения DoS атак на КФС с применением ансамбля классификаторов;
- Предложен алгоритм обнаружения вредоносного ПО в КФС на основе изображений.
- Учитывая масштабы угроз на КФС и большое количество незащищенных устройств, разработанные методы являются важным шагом на пути к разработке новых практических инструментов защиты таких систем.

ГЛАВА IV. РАЗРАБОТКА МЕТОДОВ И АЛГОРИТМОВ ОБЕСПЕЧЕНИЯ ЗАЩИЩЕННОСТИ ОПЕРАЦИОННЫХ ТЕХНОЛОГИЙ КИБЕРФИЗИЧЕСКИХ СИСТЕМ

Аномальное состояние КФС может быть вызвано неисправными компонентами, временными отказами, неправильной настройкой, кибератаками или их сочетанием [257, с. 49-51; 36, с. 14-22]. Злоумышленник вмешивается в КФС, чтобы манипулировать показаниями сенсоров или исполнительных механизмов, что приводит к нештатной работе системы.

Обнаружение аномалий в промышленном сценарии имеет важное значение, поскольку необнаруженные сбои могут привести к критическим повреждениям [86, с. 2687-2698]. Раннее обнаружение аномалий может повысить надежность подверженного сбоям промышленного оборудования и снизить затраты на эксплуатацию и техническое обслуживание.

Данная глава посвящена разработке метода обнаружения аномалий в ОТ КФС по данным физических устройств на основе иерархических скрытых Марковских моделей, алгоритму обнаружения аномалий по акустическим сигналам КФС с применением трансферного обучения, методу обнаружения атак на устройства КФС с применением глубокой гибридной модели и метода классификации отказов КФС на основе визуальных представлений.

4.1. Метод обнаружения аномалий в операционных технологиях киберфизических систем на основе иерархических скрытых Марковских моделей

Современные КФС представляют собой сложные компьютерные сети со множеством взаимосвязанных компонентов, которые выполняют различные функции, обеспечивая безопасную и эффективную работу промышленных объектов, транспортных систем и коммунальных служб. При этом наиболее важными являются обеспечение доступности систем управления, безопасности человеческой жизни и целостности обрабатываемых данных. Проблема безопасности КФС весьма актуальной во всех сферах их применения.

Необходимо определить нормальное состояние КФС и понять логику управления технологическим процессом для выявления аномалий в системе [108, с. 1-12]. Но сложность и изощренность кибератак, призванных нанести значительный ущерб КФС, растет с каждым годом. Кибератака может заключаться во внедрении вируса в КФС с использованием хакерских технологий для проникновения в систему [350, с. 166-178]. Злоумышленник затем может обмануть человеко-машинный интерфейс системы, отображая нормальное состояние [191, с. 729-736; 339, с. 1040-1049]. Он изменяет показания сенсоров или исполнительных механизмов, что может нарушить работу КФС [146, с. 2236-2246]. Все это может привести к плачевным последствиям.

К основным кибератакам на КФС, наблюдаемым в реальных инцидентах, относятся атаки на SCADA и протоколы, несанкционированный доступ к проприетарному ПО, удаленная смена пароля, отказ в обслуживании (DoS) и другие [184, с. 1454-1467; 315, с. 2715-2729; 321, с.125-130; 348, с. 90-95; 356, с. 5979-5984; 362, с. 55-64].

Учитывая вышеизложенное, возникает необходимость разработки нового подхода к выявлению аномалий КФС, вызванных кибератаками. Это позволило бы оценить ее устойчивость к атакам на разных уровнях реализации.

Предлагается подход, основанный на иерархической скрытой марковской модели (hierarchical hidden Markov model, HHMM), состоящей из простых HMM [321, с. 48-55]. В этом случае наблюдаемые события моделируются с помощью HMM. HHMM подходит для задач со сложной иерархической структурой и зависящих от времени [65, с. 39-41]. Модель требует значительно меньше обучающих данных для достижения производительности, сравнимой с HMM. Кроме того, слои в нижней части HHMM, которые более чувствительны к изменениям окружающей среды, таким как тип сенсора, частота дискретизации и другие, могут быть переобучены отдельно без изменения более высоких уровней HHMM.

Иерархическая скрытая Марковская модель. ННММ представляет собой многоуровневый стохастический процесс, который можно описать следующим образом [187, с. 41-62; 150].

Пусть имеется последовательность наблюдений $\boldsymbol{O} = (\boldsymbol{o_1}, \boldsymbol{o_2}, \dots, \boldsymbol{o_T})$ размерности $\boldsymbol{T}$. $\boldsymbol{o_t}$ – $\boldsymbol{t}$-ая мера, полученная блоком КФС. В этом случае параметры модели равны $\boldsymbol{\lambda} = \{\{\boldsymbol{A^r}\}_{\boldsymbol{r}\in\{\boldsymbol{1},\dots,\boldsymbol{R-1}\}}, \{\boldsymbol{\pi^r}\}_{\boldsymbol{r}\in\{\boldsymbol{1},\dots,\boldsymbol{R-1}\}}, \{\boldsymbol{B}\}\}$.

Здесь $\boldsymbol{s_t^r} = \boldsymbol{i}$ – состояние ННММ на шаге $\boldsymbol{t}$, где $\boldsymbol{i}$ – индекс состояния, а $\boldsymbol{r} \in \{\boldsymbol{1}, \dots, \boldsymbol{R}\}$ – индекс иерархии, принимающий значение $\boldsymbol{r} = \boldsymbol{1}$ для корневого состояния, $\boldsymbol{r} = \{\boldsymbol{2}, \dots, \boldsymbol{R-1}\}$ – для остальных внутренних состояний и $\boldsymbol{r} = \boldsymbol{R}$ – для производственных состояний [150].

Модель характеризуется вероятностью перехода между внутренними состояниями и выходным распределением производственных состояний. Имеется матрица вероятности перехода между состояниями $\boldsymbol{A^r}$ для каждого внутреннего состояния $\boldsymbol{s_t^r}$ при $\boldsymbol{r} \in \{\boldsymbol{1}, \dots, \boldsymbol{R-1}\}$, где $\boldsymbol{A_{ij}^r} = \boldsymbol{p}(\boldsymbol{s_t^{r+1}} = \boldsymbol{j} | \boldsymbol{s_j^{r+1}} = \boldsymbol{i})$ – вероятность горизонтального перехода из состояния i в состояние j на уровне $\boldsymbol{r}$. Вектор начальных распределений подсостояний $\boldsymbol{\pi^r}$ включает элементы $\boldsymbol{\pi_j^r} = \boldsymbol{p}(\boldsymbol{s_t^{r+1}} = \boldsymbol{j} | \boldsymbol{s_t^r})$ для $\boldsymbol{r} \in \{\boldsymbol{1}, \dots, \boldsymbol{R-1}\}$ и есть вероятность того, что состояние s_t^r активирует состояние $\boldsymbol{s_t^{r+1}}$. Если $\boldsymbol{s_t^{r+1}}$ – внутреннее состояние, то $\boldsymbol{\pi_j^r}$ – вероятность вертикального перехода. Каждое производственное состояние $\boldsymbol{s_t^R}$ параметризуется вектором выходного параметра $\boldsymbol{B^i}$, который зависит от особенностей модели наблюдения $\boldsymbol{p}(\boldsymbol{o_t} | \boldsymbol{s_t^R} = \boldsymbol{j}, \boldsymbol{B^j})$, соответствующей $\boldsymbol{j}$-му производственному состоянию. Согласно $\boldsymbol{\pi^r}$, одно из подсостояний для каждого внутреннего состояния выбирается случайным образом. Наблюдение генерируется, когда внутреннее состояние переходит в рабочее состояние $\boldsymbol{s_t^R} = \boldsymbol{j}$. Управление возвращается во внутреннее состояние, которое приводит к текущему производственному состоянию $\boldsymbol{s_t^{R-1}}$. Он выбирает следующее состояние на том же уровне в соответствии с $\boldsymbol{A^{R-1}}$. Процесс завершается, когда управление передается корневому состоянию.

Предлагаемый подход. Далее описывается предлагаемый подход к обнаружению аномалий в ОТ с использованием ННММ. В методе рассматривается НММ, состоящая из $\boldsymbol{N}$ уровней моделей НММ, где НММ на уровне $\boldsymbol{N+1}$ соответствуют символам наблюдения или генераторам вероятностей на уровне $\boldsymbol{N}$. Каждый уровень $\boldsymbol{i}$ в ННММ состоит из $\boldsymbol{K_i}$ моделей НММ (Рис. 4.1.1). На любом заданном уровне $\boldsymbol{N+1}$ ННММ модели последовательность $\boldsymbol{N+1}$ символов наблюдения $\boldsymbol{o_L = \{o_1, o_2, \dots, o_{T_L}\}}$ может использоваться для создания нового наблюдения для уровня $\boldsymbol{L-1}$ моделей ННММ. На самом низком уровне (т. е. уровне $\boldsymbol{N}$) символы наблюдения $\boldsymbol{o_p = \{o_1, o_2, \dots, o_{T_p}\}}$ будут генерироваться непосредственно из наблюдений за моделируемым процессом. В этом случае $\boldsymbol{o_p}$ будут формироваться из квантованных значений сенсоров. Таким образом, на каждом слое ННММ наблюдения генерируются на основе результатов предыдущего слоя. А на самом нижнем слое символы формируются из измерений наблюдаемого процесса.

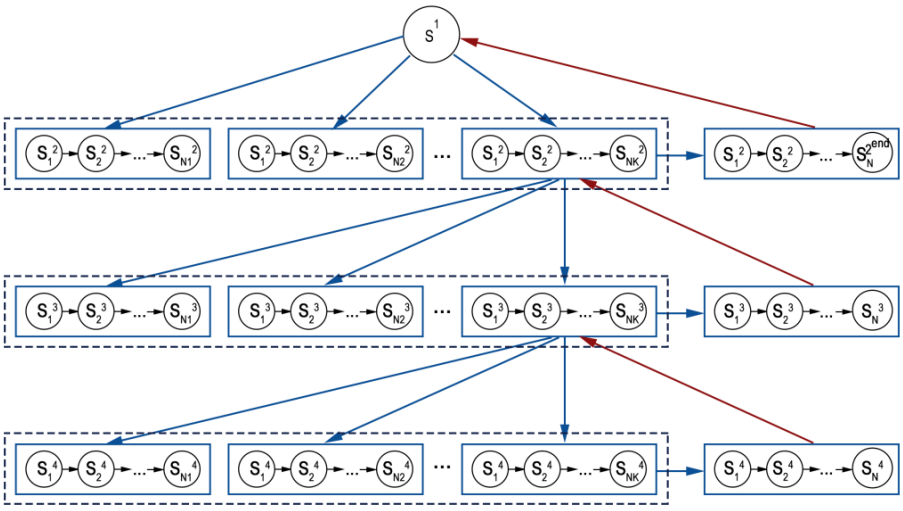


Рис. 4.1.1. Иерархическая скрытая Марковская модель

Нет необходимости проходить все уровни с одинаковым уровнем детализации. Например, можно использовать окна на любом уровне структуры. Наиболее вероятностный НММ на уровне $\boldsymbol{L+1}$ может использоваться как генератор вероятностей, передавая полное распределение вероятностей вверх

по уровням HHMM. В этом случае при наблюдении $\boldsymbol{i}$-го HMM правдоподобие $\boldsymbol{L(i)}$ может быть использовано в модели HMM уровня $\boldsymbol{L}$ для учета неопределенности результатов моделей HMM на уровне $\boldsymbol{L+1}$.

Первым шагом в обнаружении аномалий КФС с использованием модели HHMM является сбор данных сенсоров (Рис. 4.1.2). Затем среди них выбираются наиболее информативные признаки. Использование методов векторного квантования (например, алгоритма k-средних) создает значимые наблюдения для HMM. Алгоритм Баума-Уэлча выполняет коррекцию параметров модели на этапе обучения [93, с. 1554-1563]. Вероятная последовательность состояний HMM на первом уровне приводит к генерации вектора признаков [396, с. 265-286]. Применение векторного квантования к этим признакам создает последовательность наблюдений из HMM второго уровня и т. д. Затем на этапе тестирования определяется последовательность состояний с помощью алгоритма Витерби, который использует параметры этапа обучения.

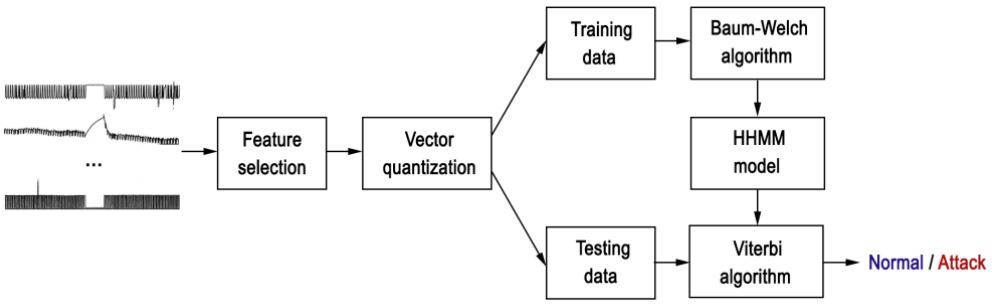


Рис. 4.1.2. Блок-схема предлагаемого подхода

Преимущество, которое можно ожидать от использования HHMM по сравнению с большой одноуровневой HMM моделью, заключается в том, что HHMM с меньшей вероятностью пострадает от переобучения, поскольку отдельные подкомпоненты обучаются независимо на меньших объемах данных. Следствием этого является то, что модель HHMM требует значительно меньше обучающих данных для достижения производительности, сравнимой с HMM.

Еще одним преимуществом является то, что слои в нижней части HHMM,

которые более чувствительны к изменениям окружающей среды, таким как тип сенсора, частота дискретизации и т. д., могут быть переобучены отдельно без изменения более высоких уровней HHMM.

В предлагаемом подходе рассматривается HHMM с тремя уровнями (Рис. 4.1.3). Один HMM используется для имитации задачи на уровнях 1 и 2, где каждое состояние соответствует подзадаче. Существует HMM для каждого возможного состояния сенсора на уровне 3, полученном во время выполнения процесса. Последовательность наблюдений для моделей HMM на уровне 3 генерируется из квантованных значений сенсоров. Индекс HMM с наибольшей вероятностью среди моделей HMM на уровне 3 затем принимается в качестве символа наблюдения для модели HMM на уровне 2. Затем HMM на уровне 2 используется для вычисления вероятности конкретного состояния как функции времени, учитывая последовательность наблюдений, сгенерированных моделями HMM на уровне 3.

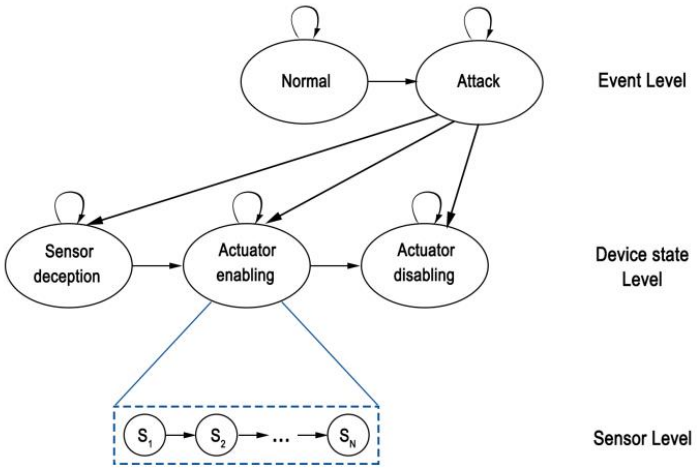


Рис. 4.1.3. Трехуровневая HHMM модель

Наиболее вероятностная модель среди HMM уровня 2 затем берется в качестве символа наблюдения для модели HMM первого уровня, которая вычисляет вероятность конкретного состояния. Каждое состояние модели HMM на уровне 1 может помочь эксперту в принятии окончательного решения. Здесь выбирается лучший HMM на уровне 3 с наибольшей вероятностью, и для

модели HMM на уровне 1 генерируется символ наблюдения, соответствующий этому состоянию сенсора.

Описание набора данных. Рассматривается набор данных безопасной обработки воды (secure water treatment, SWaT) для оценки предлагаемой модели обнаружения аномалий КФС [145, с. 88-99; 222, с. 1058-1065]. Этот набор данных охватывает показания 25 сенсоров и 26 исполнительных механизмов, которые необходимы для раннего обнаружения отказов системы. Он содержит информацию о 36 сценариях атак. Набор данных был собран Сингапурским университетом технологий и дизайна и находится в открытом доступе на веб-сайте [145, с. 88-99]. SWaT содержит записи, включающие семь дней нормальной работы и четыре дня кибератак.

Результаты экспериментов. Предложенный подход на основе HHMM реализован в R4.1.0 с использованием различных библиотек, включая depmixS4. Все эксперименты проводились на Intel Xeon (R), CPU X5670 @ 2,93 ГГц * 4 с 24 ГБ оперативной памяти.

Набор данных содержит 946722 образца с 51 атрибутом [145, с. 88-99]. Он разделен на обучающий набор данных (80%) и тестовый набор данных (20%) для оценки эффективности предложенной модели для обнаружения аномалий.

Оценка производительности HHMM основана на следующих показателях: precision, recall и F-мера. Результаты оценки последовательности скрытых состояний предлагаемого подхода с использованием алгоритма Витерби представлены на рис. 4.1.4.

На рисунке нормальное состояние показано синей кривой. Атаки разбросаны по времени и представлены красной кривой. HMM может определить «истинное» состояние на уровне события, несмотря на сложность обнаружения состояния атаки.

Согласно рассмотренным показателям precision, recall и F-меры, предложенная модель на основе HHMM показала 0.9998, 0.9164 и 0.9563 соответственно. Для всесторонней оценки модели на основе HHMM по

сравнению с предыдущими работами проводится ее сравнение с такими моделями, как глубокая нейронная сеть (DNN) [250, с. 72-83], машины опорных векторов (SVM) [250, с. 72-83], TABOR [268, с. 525-536], 1D CNN (convolutional neural network) [250, с. 72-83], LSTM-CUSUM [262, с. 703-716] и MAD-GAN [262, с. 703-716].

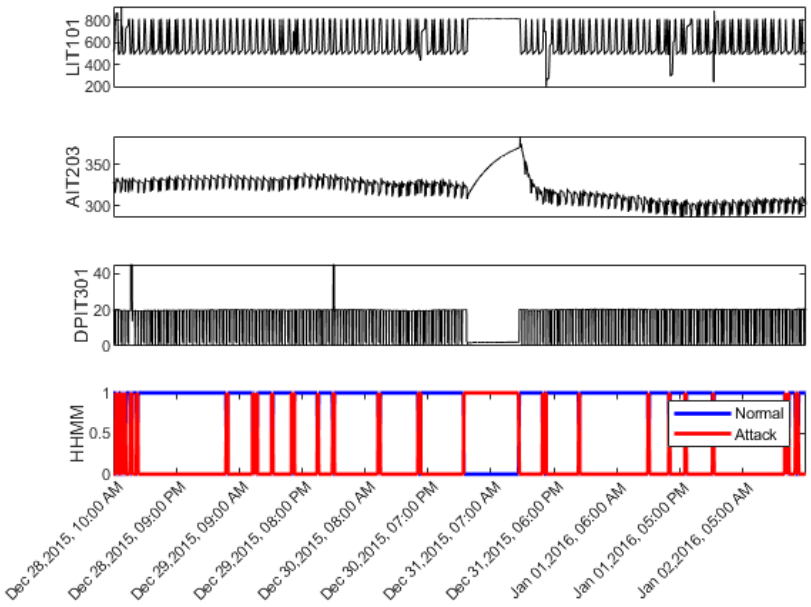


Рис. 4.1.4. Апостериорные вероятности состояний системы

В таблице 4.1.1 представлены полученные результаты. По показателю precision HHMM немного уступает 1D CNN. Таким образом, модель приводит к улучшению recall на 0.0274. Оценка предлагаемого подхода с использованием метрики F-мера также дает более высокие показатели. Это доказывает успешность обнаружения атак на КФС с помощью модели HHMM.

Таблица 4.1.1

Оценка предлагаемого подхода

Метод	Метрики		
	Precision	Recall	F-мера
DNN [250]	0.9830	0.6785	0.8028
SVM [250]	0.9250	0.6990	0.7963
TABOR [268]	0.8617	0.7880	0.8232
1D CNN [250]	1.0000	0.8530	0.9200
LSTM-CUSUM [262]	0.9070	0.6770	0.7750
MAD-GAN [262]	0.9610	0.9420	0.9510
Предложенный подход	0.9998	0.9164	0.9563

В данном исследовании HHMM была применена для обнаружения аномальных элементов в показаниях устройств КФС. Идея предлагаемого подхода заключается в том, что несколько моделей обнаруживают подозрительную активность в соответствии с данными сенсоров на самом низком уровне HHMM. Подход оценивается на наборе данных SWaT и сравнивается с результатами предыдущих работ. Эксперименты продемонстрировали, что предложенный метод показал лучшие результаты с точки зрения метрик recall (0.9164) и F-мера (0.9563). Нормальные и аномальные состояния, полученные из модели HHMM, можно использовать в качестве входных данных для идентификации известных атак и атак нулевого дня.

4.2. Метод обнаружения аномалий в акустических сигналах киберфизических систем с применением трансферного обучения

Аномальное состояние КФС может быть вызвано неисправными компонентами, временными отказами, неправильной настройкой, кибератаками или их сочетанием [257, с. 49-51; 36, с. 14-22]. Злоумышленник вмешивается в КФС, чтобы манипулировать показаниями сенсоров или исполнительных механизмов, что приводит к нештатной работе системы.

Обнаружение аномалий в промышленном сценарии имеет важное значение, поскольку необнаруженные сбои могут привести к критическим повреждениям. Раннее обнаружение аномалий может повысить надежность подверженного сбоям промышленного оборудования и снизить затраты на эксплуатацию и техническое обслуживание.

Развитие Индустрии 4.0 привело к появлению новых технологий эффективного и надежного мониторинга таких систем. Таким образом, в состав современных КФС входят устройства, формирующие мультисенсорную конфигурацию. Эти системы упрощают процесс сбора данных, что приводит к доступности больших наборов данных. Следовательно, наблюдается рост развития методов интеллектуального анализа данных для обнаружения

аномалий [37, с. 101-113; 25, с. 1-12]. Классический подход к таким проблемам обычно включает методы обработки сигналов, которые извлекают полезную информацию из данных сенсоров.

Обнаружение аномалий в ОТ КФС становится важной областью исследований. Сложность здесь заключается в получении информации от нескольких сенсоров, которые отличаются своими определенными акустическими свойствами. Исследователи предлагают новые методы и расширяют существующие алгоритмы для обнаружения неисправностей промышленного оборудования [292, с. 2-4; 201, с. 1-3; 137, с. 337-348; 25, с. 1-12; 293, с. 49-56].

Morita и соавторы [292, с. 2-4] предложили анализ главных компонент с локальным фактором выброса и смешанной гауссовой моделью для обнаружения аномальных звуков в ОТ КФС при наличии ограниченных вычислительных ресурсов и небольшого набора данных. В статье [201, с. 1-3] описан подход, который объединяет предварительно обученные вложения OpenL3 с ошибкой реконструкции интерполяционного автокодировщика, использующего смешанную гауссову модель в качестве окончательного предиктора. Параметры были установлены индивидуально для каждого устройства с использованием результатов из набора данных. А в 2021 году Michau и Fink разработали архитектуру для обучения осмысленному и разреженному представлению высокочастотных сигналов устройств КФС. Они объединили как теорию вейвлетов, так и глубокое обучение, для задач классификации и обнаружения аномалий в КФС. Применение архитектур глубокого обучения автокодировщика для обнаружения аномалий в акустических сигналах без учителя на основе сверточных и полносвязных нейронных сетей было рассмотрено в работе [137, с. 337-348]. Энергетические характеристики мел-спектрограммы были извлечены из необработанных звуковых сигналов. Было проведено несколько предварительных экспериментов для настройки гиперпараметров автокодировщика. В 2020

Tiwari и соавторы предложили ансамбль состоящий из двух систем, способных регистрировать аномальное поведение КФС. В первой системе был предложен метод обнаружения выбросов, основанный на поиске ближайшего соседа. Во второй системе для обнаружения аномалий применялись i-векторы и смешанная гауссова модель. В качестве оценок аномалий использовалось отрицательное логарифмическое правдоподобие. Были предложены сети OutlierNets, семейство очень компактных глубоких сверточных автокодировщиков, адаптированных для обнаружения аномалий в акустических сигналах в режиме реального времени [25, с. 1-12]. Они имеют чрезвычайно низкую сложность и соответствуют или даже превосходят большую сверточную архитектуру автокодировщика по площади под кривой рабочей характеристики приемника, демонстрируя задержку в микросекундном масштабе на встроенном оборудовании. Была изучена эффективность обнаружения аномалий в акустических сигналах на основе трансферного обучения по изображениям [293, с. 49-56]. Авторы рассмотрели различные глубокие нейронные модели. Результаты показали, что признаки, извлеченные с помощью моделей ResNet18 и ResNet34 с смешанной гауссовой моделью и одноклассовой SVM, достигли наилучшего среднего значения по площади под кривой рабочей характеристики приемника. Это подтвердило, что признаки на основе изображений с моделями на основе трансферного обучения могут достигать конкурентоспособных результатов в обнаружении аномалий в акустических сигналах КФС устройств.

Можно сделать следующие выводы, обобщая анализ текущего состояния исследований в области обнаружения аномалий в акустических сигналах ОТ КФС:

1) Небольшой объем работ был сосредоточен на трансферном обучении для извлечения признаков и обнаружения неисправностей в промышленных устройствах КФС.

2) Не все функциональные возможности глубоких нейронных сетей

учитываются.

Все это подтверждает актуальность данного исследования, где объектом исследования является сложное промышленное оборудование КФС с неоднородными сенсорными системами различной природы. Для этого требуется предварительная обработка данных для извлечения наиболее информативных признаков [117, с. 246-260]. Обычно это очень трудоемкая задача, требующая знаний экспертов.

Предметом исследования являются методы обнаружения аномалий в сигналах промышленного оборудования КФС на основе трансферного обучения. В предлагаемом методе рассматриваются изображения спектрограммы и скалограммы сигналов для более точной классификации отказов оборудования КФС [22, с. 115-118; 355, с. 124-132]. Рассматривается функция активации SPOCU (scaled polynomial constant unit) [245, с. 3385-3401] для повышения точности предлагаемого подхода. Применяется алгоритм XGBoost, поскольку он обладает высокой производительностью и требует небольших вычислительных ресурсов на этапе обучения.

Цель работы — использование трансферного обучения в сочетании с алгоритмом XGBoost для повышения точности обнаружения аномального состояния по акустическим сигналам КФС.

Предлагаемый подход. Трансферное обучение использует веса из глубокой нейронной модели, предварительно обученной на большом наборе данных. Его можно применять к небольшому набору данных, обеспечивая сходимость без переобучения.

Предлагаемый подход к обнаружению аномалий машинного сигнала по изображениям на основе трансферного обучения состоит из следующих этапов: предварительная обработка, извлечение признаков с помощью глубокой нейронной сети, объединение признаков и классификация на основе алгоритма XGBoost (Рис. 4.2.1).

Рассматриваемые сигналы предварительно разбиваются на ячейки по 128

отсчетов с перекрытием в 64 отсчета. Из сигналов извлекаются скалограмма, основанная на вейвлет-преобразовании, и спектрограмма, основанная на кратковременном преобразовании Фурье (short-term Fourier transform, STFT). STFT разбивает сигнал на несколько перекрывающихся блоков, умножая их на оконную функцию Ханнинга:

$$\boldsymbol{\Theta(\omega,\tau) = \int_{-\infty}^{+\infty} \theta(t)\sigma(t-\tau)e^{-i2\pi\omega t}dt,} \qquad (4.2.1)$$

где $\boldsymbol{\sigma}$ - оконная функция.

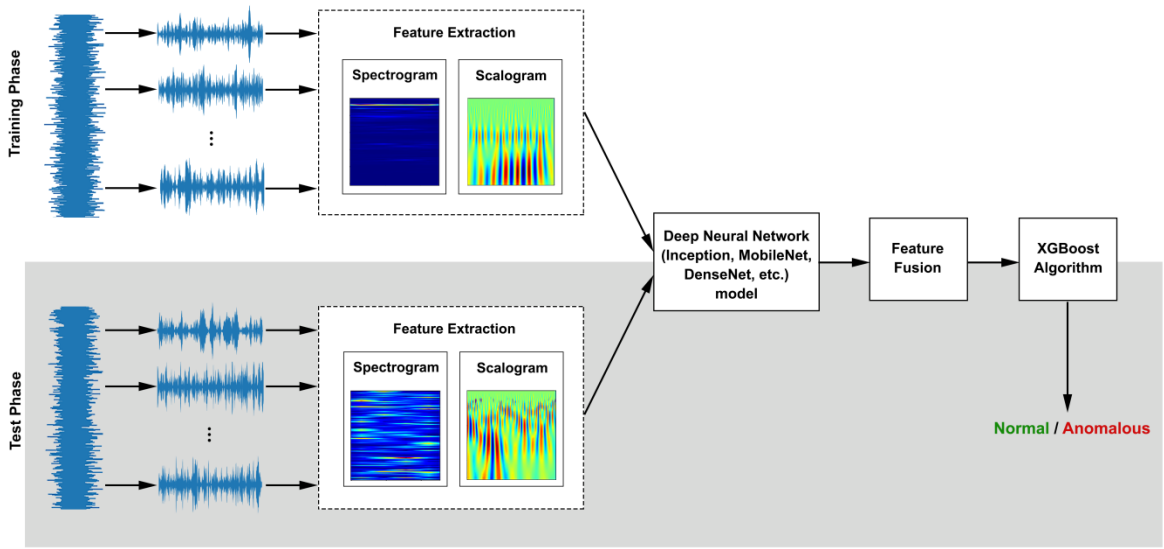


Рис. 4.2.1. Блок-схема предлагаемого подхода [355]

А вейвлет Морле учитывается при вейвлет-преобразовании для получения более информативных изображений:

$$\boldsymbol{\psi(t) = e^{-\beta^2 t^2/2}\cos(\pi t),} \qquad (4.2.2)$$

где $\boldsymbol{\beta}$ - параметр, который контролирует форму материнского вейвлета.

Получение изображений скалограммы и спектрограммы выполняется параллельно. Затем визуальные представления (RGB, Red, Green, Blue) размером 128×128×3 отправляются в глубокую нейронную сеть.

Поскольку модели глубокого обучения обучаются на больших наборах данных различных изображений, их можно применять для обнаружения аномалий в сигналах промышленных объектов. Каждый из слоев модели отвечает за разные функции изображения.

Для извлечения признаков из изображений спектрограмм и скалограмм рассматриваются следующие предварительно обученные глубокие нейронные сети: Xception [135, с. 1800-1807], MobileNet, DenseNet-121 [216, с. 2261-2269] и InceptionV3 [359, с. 2818-2826; 39, с. 11-18].

Модель MobileNet представляет собой небольшую сеть, содержащую свертки с разделением по глубине и повышающую производительность распознавания. Сеть InceptionV3 включает параллельные сверточные слои, которые затем объединяются для получения результата [358, с. 1-9; 359, с. 2818-2826]. Сеть Xception представляет собой линейный набор остаточных сверточных слоев [135, с. 1800-1807; 238, с. 1-6]. Операция суммирования ускоряет переход от одного слоя модели к другому [209, с. 770-778]. Рассматриваемая модель DenseNet используется для сбора информации со всех уровней сети и передачи ее на последующие уровни при недостатке данных для обучения [216, с. 2261-2269].

Полностью связанный слой рассматривался как вектор признаков с использованием предварительно обученной модели. Результаты объединяются для извлечения информации о различных характеристиках и уменьшения ошибок распознавания. Таким образом, общий размер вектора признаков составляет 1x1024.

SPOCU [245, с. 3385-3401] рассматривается в предлагаемой модели как функция активации в скрытых слоях для повышения точности обнаружения аномалий в сигналах на основе изображений и рассчитывается следующим образом:

$$\boldsymbol{G(x) = \rho H(\frac{x}{\zeta} + \xi) - \rho H(\xi)}, \tag{4.2.3}$$

где $\boldsymbol{\xi \in (0, 1)}$, $\boldsymbol{\rho, \zeta > 0}$ и

$$\boldsymbol{H(x) = \begin{cases} r(c), & x \geq c \\ r(x), & x \in [0, c) \\ 0, & x < 0 \end{cases}} \tag{4.2.4}$$

При $\boldsymbol{r(x) = x^3(x^5 - 2x^4 + 2)}$ и $\boldsymbol{1 \leq c < \infty}$. Согласно (4.2.3), при $\boldsymbol{c = \infty}$, $\boldsymbol{\rho =}$

$\mathbf{3.0937}$, $\boldsymbol{\xi = 0.6653}$ и $\boldsymbol{\zeta = 4.437}$.

Затем полученный вектор признаков передается в классификатор XGBoost, предложенный в [128, с. 785-794]. XGBoost — это дерево регрессии, поддерживающее задачу классификации. В основе алгоритма лежит оптимизация значения целевой функции. В этом случае целевая функция оптимизации определяется следующим образом:

$$\widetilde{\boldsymbol{y_i}} = \sum_{\boldsymbol{k=1}}^{\boldsymbol{K}} \boldsymbol{\phi_k}(\boldsymbol{x_i}),\ \boldsymbol{\phi_k} \in \boldsymbol{\Phi}, \tag{4.2.5}$$

где $\boldsymbol{K}$ – число деревьев решений, $\boldsymbol{\phi_k}$ является независимым деревом с оценками листьев, $\boldsymbol{\Phi}$ – пространство дерева регрессии. В этом случае функция потерь определяется следующим уравнением:

$$\boldsymbol{L(\phi_t)} = \sum \boldsymbol{v}(\widetilde{\boldsymbol{y}_i}, \boldsymbol{y_i}) + \sum \boldsymbol{\Omega}(\boldsymbol{\phi_t}), \tag{4.2.6}$$

где $\boldsymbol{v}$ – значение потерь алгоритма XGBoost, $\widetilde{\boldsymbol{y}_i}$ – предсказанный выход, $\boldsymbol{\Omega}$ - член регуляризации, который предотвращает переобучение (4.2.7).

$$\boldsymbol{\Omega(\phi)} = \boldsymbol{\gamma K} + \frac{\boldsymbol{1}}{\boldsymbol{2}}\boldsymbol{\lambda}\|\boldsymbol{w}\|^{\boldsymbol{2}}, \tag{4.2.7}$$

где $\boldsymbol{K}$ - количество узлов, $\boldsymbol{w}$ – оценка на каждом листе, $\boldsymbol{\gamma}$ и $\boldsymbol{\lambda}$ являются константами для контроля степени регуляризации.

Таким образом, получаем следующее:

$$\boldsymbol{L(\phi_t)} \approx \sum_{\boldsymbol{j=1}}^{\boldsymbol{T}} \left[\left(\sum_{\boldsymbol{i}\in \boldsymbol{I_j}} \boldsymbol{g_i}\right)\boldsymbol{w_j} + \frac{\boldsymbol{1}}{\boldsymbol{2}}\left(\sum_{\boldsymbol{i}\in \boldsymbol{I_j}} \boldsymbol{h_i} + \boldsymbol{\lambda}\right)\boldsymbol{w_j^2}\right] + \boldsymbol{\gamma K}, \tag{4.2.8}$$

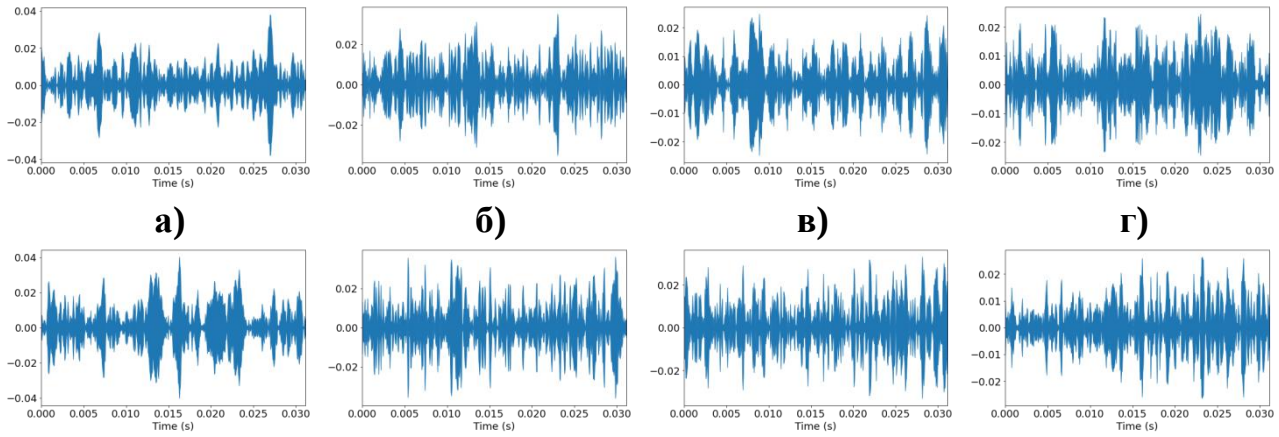
где $\boldsymbol{g_i}$ является первой производной, и $\boldsymbol{h_i}$ - вторая производная функции потерь, соответственно. Для метода XGBoost скорость обучения составляет 0.001, количество деревьев для соответствия 100, максимальная глубина дерева равна 6, $\boldsymbol{\gamma = 0}$ и $\boldsymbol{\lambda = 1}$.

Результаты экспериментов. Для оценки предлагаемого подхода рассматривается набор данных КФС при «нормальных» и аномальных режимах работы (MIMII, Malfunctioning Industrial Machine Investigation and Inspection) [323, с. 209-213]. Набор аудиоданных был собран с использованием круговой микрофонной решетки, состоящей из восьми отдельных микрофонов, в виде 16-битных аудиосигналов с частотой дискретизации 16 кГц [323, с. 209-213]. Он

содержит восемь отдельных каналов для каждого сегмента. Набор данных MIMII содержит звук четырех различных типов устройств: клапанов, насосов, вентиляторов и направляющих.

Для каждого типа оборудования рассматривались различные реальные аномальные сценарии: загрязнение, утечка, дисбаланс вращения, повреждение рельса и т. д. MIMII также содержит данные для четырех идентификаторов устройств (00, 02, 04 и 06). В наборе данных учитывались различные уровни отношения сигнал/шум (6 дБ, 0 дБ и -6 дБ). Он состоит из 26092 «нормальных» и 6065 аномальных звуковых сегментов.

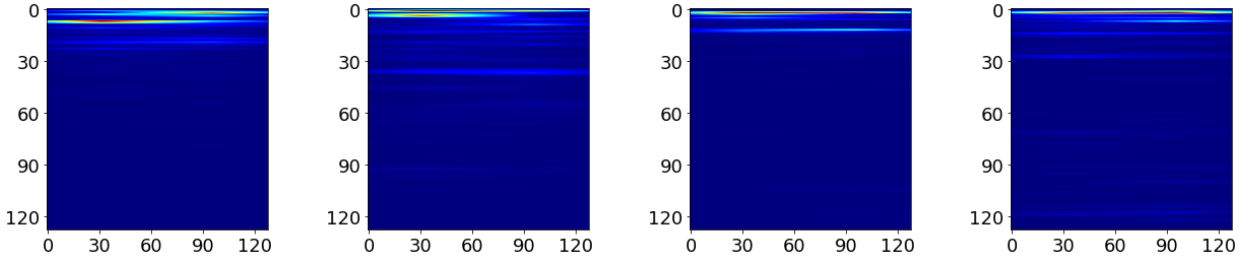
«Нормальные» и аномальные сигнатуры для всех типов устройств во временной области набора данных MIMII показаны на рис. 4.2.2 ((а)-(г)) и рис. 4.2.2 ((д)-(з)), соответственно. Спектрограмма STFT и скалограмма на основе вейвлет-преобразования для вентиляторов, насосов, направляющих и клапанов показаны на рис. 4.2.3 и 4.2.4, соответственно.



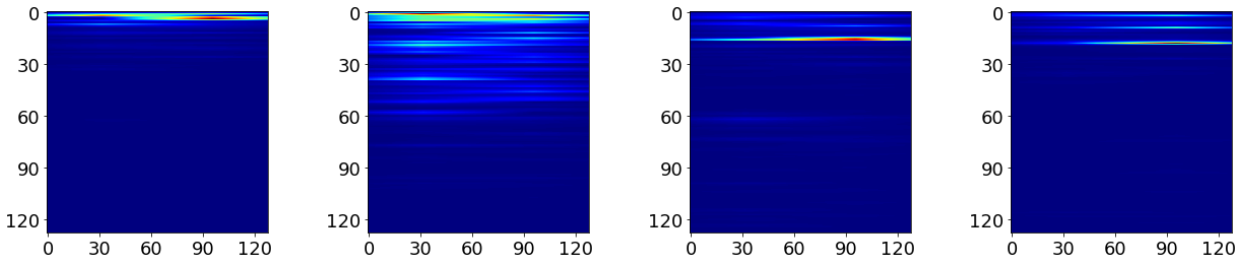
а) **б)** **в)** **г)**

д) **е)** **ж)** **з)**

Рис. 4.2.2. Осциллограммы нормальной ((а)-(г)) и аномальной работы ((д)-(з)) четырех рассматриваемых устройств [355]



а) **б)** **в)** **г)**

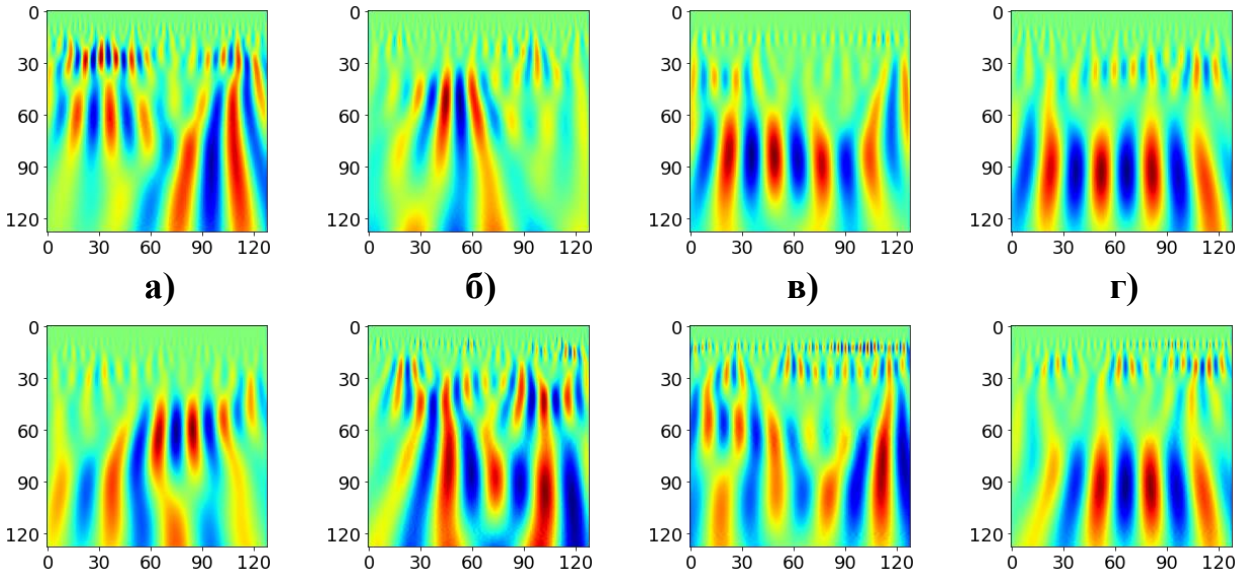


д) е) ж) з)

Рис. 4.2.3. Спектрограмма кратковременного преобразования Фурье четырех рассматриваемых устройств (ID: 00) в нормальных ((а)-(г)) и аномальных ((д)-(з)) условиях при -6 дБ SNR [355]

Эксперименты проводились на Python 2.7.13 с использованием различных библиотек, включая Tensorflow, Librosa и Keras. Использовалась виртуальная машина Intel Xeon (R) CPU X5670 @ 2,93 ГГц * 24 с 24 ГБ ОЗУ.

В этом исследовании анализируются и сравниваются различные модели глубокого обучения (такие как Xception, Inception, DenseNet и MobileNet). Они обучаются на наборе данных MIMII и применяются для извлечения признаков из спектрограммы и скалограммы. В качестве показателей оценки рассматривались precision, recall и F-мера.



а) б) в) г)

д) е) ж) з)

Рис. 4.2.4. Скалограмма непрерывного вейвлет-преобразования четырех рассматриваемых устройств (ID: 00) в нормальных ((а)-(г)) и аномальных ((д)-(з)) условиях при -6 дБ SNR [355]

Оптимизация гиперпараметров проводилась с помощью перекрестной проверки. Комбинация параметров была выбрана на основе наименьших потерь при обучении и наибольшей точности. Наилучшая производительность модели наблюдалась при размере батча, равном 32. Также оценка скорости обучения выполнялась на разных значениях. Точность модели снижалась с увеличением скорости обучения. Важно отметить, что все четыре рассмотренные модели достигли высокой точности за счет снижения скорости обучения до 0.001.

Результаты экспериментов представлены в таблице 4.2.1. Сравнение различных глубоких нейронных моделей показало, что модель Densenet+XGBoost превосходит другие рассмотренные модели в обнаружении аномалий из сигналов оборудования по метрике F-мера. Модели Inception+XGBoost, Xception+XGBoost и Densenet показали хорошие результаты для направляющих с использованием показателей precision и recall. Несмотря на то, что все модели показали лучшие результаты по recall для устройства типа «клапан», они уступали Densenet+XGBoost по метрикам precision и F-мера.

Таблица 4.2.1

Оценка эффективности предлагаемого подхода

Модель	Метрика	Устройство			
		Fan	Pump	Slider	Valve
Inception+XGBoost	Recall	87.0	95.0	98.0	100
	Precision	89.1	87.9	100	94.3
	F-мера	92.3	90.6	98.1	96.3
Xception+XGBoost	Recall	100	88.0	98.0	100
	Precision	84.2	100	100	92.6
	F-мера	96.8	92.8	98.1	95.3
Mobilenet+XGBoost	Recall	96.0	88.0	94.0	100
	Precision	78.7	95.7	100	96.2
	F-мера	92.0	90.9	96.1	97.2
Densenet+XGBoost	Recall	99.9	96.0	98.0	100
	Precision	87.0	100	100	97.1
	F-мера	98.2	97.1	98.1	97.7

Таким образом, модель Densenet+XGBoost оказалась лучшей. Для оценки ее производительности было проведено сравнение с такими моделями, как

Alexnet [293, с. 49-56], ResNet18 [293, с. 49-56], ResNet34 [293, с. 49-56], SqueezeNet [293, с. 49-56], IAEO3_opt [201, с. 1-3] и PCA+LOF+GMM [292, с. 2-4] для всех типов устройств с идентификаторами (00, 02, 04 и 06) в соответствии с метрикой AUC (Таблицы 4.2.2 и 4.2.3).

Таблица 4.2.2

Сравнение среднего значения AUC для различных моделей трансферного обучения в наборе данных MIMII

Устройство / Модель	Fan				Pump			
	00	02	04	06	00	02	04	06
AlexNet+GMM [293]	57.7	61.7	53.9	94.5	84.1	70.8	81.6	66.0
AlexNet+OC-SVM [293]	51.0	73.1	59.7	93.2	77.5	56.4	81.1	60.1
ResNet18+GMM [293]	62.6	64.1	59.3	94.4	84.5	71.3	84.0	68.3
ResNet34+GMM [293]	58.7	65.6	57.0	90.9	78.4	66.8	87.9	63.2
ResNet34+OC-SVM [293]	50.1	67.4	57.5	83.0	64.9	51.5	81.2	60.2
SqueezeNet+OC-SVM [293]	55.6	64.8	46.2	78.8	86.7	49.4	88.4	62.3
GMM [292]	67.4	87.1	79.3	96.2	72.5	70.4	94.2	87.1
IAEO3_opt [201]	65.5	83.3	71.4	98.1	84.4	77.8	98.0	78.9
Densenet+XGBoost	93.6	95.1	87.9	95.9	97.7	96.9	97.8	97.1

Densenet+XGBoost продемонстрировала лучшую производительность среди всех типов устройств, в результате чего средние AUC составили 93.1%, 97.3%, 98.4% и 93% для вентилятора, насоса, ползунка и клапана соответственно.

Таблица 4.2.3

Сравнение различных моделей трансферного обучения в наборе данных MIMII

Устройство / Модель	Slider				Valve			
	00	02	04	06	00	02	04	06
AlexNet+GMM [293]	98.3	80.9	61.4	57.5	60.2	69.2	59.9	53.5
AlexNet+OC-SVM [293]	96.2	81.4	53.6	56.5	61.6	73.6	48.3	48.9
ResNet18+GMM [293]	99.1	85.8	68.8	65.6	58.3	73.3	60.2	56.9
ResNet34+GMM [293]	99.6	90.4	82.5	69.1	73.0	79.1	60.1	61.9
ResNet34+OC-SVM [293]	96.8	85.0	71.4	64.3	75.6	77.8	64.3	53.1
SqueezeNet+OC-SVM [293]	99.2	81.5	59.4	71.6	69.0	71.3	53.1	58.2
GMM [292]	97.7	75.9	96.9	94.2	99.4	91.8	94.2	80.7
IAEO3_opt [201]	95.9	84.0	97.9	85.9	100	99.7	99.8	98.8
Densenet+XGBoost	99.6	99.5	97.2	97.3	93.8	95.7	93.9	88.6

Предложенная модель позволила добиться значительного улучшения обнаружения аномалий по данным сенсоров устройств по AUC 95.45% по сравнению с ранее предложенными моделями [201, с. 1-3; 292, с. 2-4; 137, с. 337-348; 293, с. 49-56].

Эффективность модели Densenet+XGBoost улучшилась примерно на 8% по сравнению с моделью PCA [292, с. 2-4], примененной к логарифмической спектрограмме аудиосигнала в сочетании с LOF и GMM для набора данных MIMII. Grollmisch и соавторы [201, с. 1-3] предложили метод, сочетающий вложения OPENL3 и интерполяционный автоэнкодер (IAEO3_opt) для обнаружения аномалий акустических сигналов.

По сравнению с моделью IAEO3_opt Densenet+XGBoost модель улучшилась примерно на 7% [201, с. 1-3; 292]. Densenet+XGBoost дал сопоставимые результаты (AUC 95.5%) в проверенном наборе данных MIMII. Coelho и соавт. [137, с. 337-348] использовали CNN в сочетаниии с автоэнкодером для задачи обнаружения акустических аномалий, где результаты составили в среднем 72.0%, 73.1%, 91.8% и 78.9% для вентилятора, насоса, ползунка и клапана, соответственно. Точность предлагаемого метода для вентилятора составляет 93.1%, для насоса – 97.3%, для ползунка – 98.4%, для клапана – 93.0%, что значительно превышает приведенный выше результат. Результаты 10-кратной кросс-валидации продемонстрировали надежность и устойчивость предложенного метода.

Таким образом, актуальная проблема обнаружения аномалий в ОТ КФС была решена на основе акустических сигналов полученных от промышленного оборудования. Научная новизна полученных результатов заключается в том, что предложен подход с применением трансферного обучения и классификатора XGBoost.

В этой области проведено мало исследований, поэтому проводятся исследования различных предварительно обученных архитектур глубоких нейронных моделей для обнаружения аномалий в КФС. В качестве входных

данных для предложенного подхода были рассмотрены спектрограмма и скалограмма акустических сигналов. Для повышения точности предложенного метода использовалась функция активации SPOCU.

Практическая значимость полученных результатов заключается в том, что эксперименты на реальном наборе данных MIMII показали эффективность предложенного подхода и могут помочь специалистам в диагностике неисправностей оборудования КФС. Сравнение с другими известными методами доказало превосходство модели Densenet+XGBoost с точки зрения точности обнаружения аномалий на основе трансферного обучения для дополнительного повышения производительности и киберустойчивости КФС на больших реальных наборах данных.

4.3. Метод обнаружения атак на устройства киберфизических систем с применением глубокой гибридной модели

КФС системы, состоящие из распределенных вычислительных элементов, взаимодействующих с физическими процессами, прочно вошли в нашу жизнь [395, с. 111-116; 276, с. 174-216; 57, с. 10211-10226].

Элементы КФС контролируют внутренние и внешние киберфизические показатели. На основе анализа протекающих физических и технологических процессов система выдает команды исполнительным механизмам, вносящим изменения в киберфизическую среду.

Защита КФС, таких как электростанции, электрические сети, транспортные системы, нефте- и газопроводы, от киберугроз является одной из важных задач современного мира. Стремительный рост приложений КФС приводит к необходимости обеспечения информационной безопасности таких систем. Взаимовлияние элементов КФС друг на друга, а также появление новых технологий, их более высокая функциональность, доступность и дешевизна создают условия для появления новых киберугроз [385, с. 525-539]. С точки зрения информационной безопасности глобальная доступность объектов киберпространства породила проблему обеспечения устойчивой работы

современного производства в условиях случайных и целенаправленных компьютерных атак. Так, в 2010 году иранские ядерные объекты подверглись кибератаке червя Stuxnet. Далее последовали атаки на системы водоснабжения SCADA. В Пуэрто-Рико были взломаны интеллектуальные счетчики. А в апреле 2020 года скоординированные кибератаки на ПЛК в секторе водоснабжения Израиля привели к сбоям в работе оборудования. Все это приводит к долгосрочным и труднообнаруживаемым последствиям для управления процессами КФС. Существуют угрозы, направленные не на кражу или искажение информации, а на получение контроля над исполнительными механизмами, прежде всего в промышленной, энергетической и общественно-политической сферах.

Для получения контроля над процессом (системой) злоумышленнику необходимо знать сам физический процесс: принцип работы, условия отказа оборудования, поведение процесса и др. Восстановление после этих атак может занять несколько часов или даже повредить систему и привести к ее отказу.

Обнаружение кибератак на КФС, которые могут привести к ее неисправности и отказу оборудования, является важной областью исследований. Исследователи предлагают новые методы и расширяют существующие алгоритмы для обнаружения кибератак на такие системы [202, с. 1-7]. Обнаружение аномального или аварийного состояния КФС может быть реализовано с использованием многочисленных подходов. Актуальность интеграции технологии машинного обучения с КФС подчеркивается большим количеством научных исследований по этой теме. Традиционные классификаторы машинного обучения, такие как kNN, наивный байесовский классификатор, деревья решений, SVM и другие, широко использовались для обнаружения кибератак на КФС. Эти методы используются, когда тип кибератаки известен заранее. Нормальное состояние системы определено и метод обнаруживает отклонения полученных данных от эталонных. Традиционные методы машинного обучения направлены на выявление

кибератак на КФС, которые могут привести систему к сбоям и повлиять на нормальное функционирование физических процессов. Однако эффективная диагностика кибератак на КФС остается проблемой, когда системы очень сложны, а знаний об исследуемом объекте недостаточно. Глубокие нейронные сети позволяют эффективно решать проблему адаптивности к обнаружению кибератак на КФС [84, с. 1-10; 209, с. 770-778; 183, с. 1915-1929].

Глубокое обучение – это современная техника искусственного интеллекта, состоящая из алгоритмов и решений, значительно расширяющих возможности и эффективность нейронных сетей [42, с. 1-4; 259, с. 436-444; 342, с. 1-21]. Большое количество слоев позволяет нейронной сети строить рассматриваемый объект из простых признаков, постепенно переходя к более сложным [66, с. 18-20; 24, с. 42-46]. Глубокая нейронная сеть, состоящая из LSTM слоя, за которым следуют слои прямой связи с несколькими входами и выходами, сравнивалась с одноклассовой SVM в работе [222, с. 1058-1065]. Глубокое обучение было выбрано, потому что оно может изучать нелинейные отношения без каких-либо предварительных знаний о КФС, а LSTM модель может улавливать ее динамическую природу. Li и соавторы в 2020 году представили модель угроз состязательного машинного обучения в КФС. Этот подход позволяет эффективно генерировать примеры киберугроз, несмотря на ограничения физической системы.

Одномерная CNN модель была изучена и применена для обнаружения аномального поведения КФС в работе [250, с. 72-83]. Также была показана эффективность ансамбля глубоких нейронных сетей. В ходе экспериментов рассматривались конкретные сценарии кибератак, которые отличаются от реальных сценариев. Несмотря на то, что существует множество успешных реализаций диагностики неисправностей оборудования, связанных с кибератаками, методы, основанные на глубоком обучении, показали, что одна и та же модель не всегда стабильна [250, с. 72-83]. Сложность ОТ КФС влияет на ограничения методов.

Подводя итог анализу состояния исследований по обнаружению неисправностей КФС, вызванных кибератаками, можно сделать следующие выводы. Во-первых, необходимо обеспечить информационную безопасность приложений КФС из-за постоянного роста объемов данных. Во-вторых, недостаточно исследований было посвящено использованию ансамблей глубоких нейронных сетей, которые повышают точность обнаружения аномалий в КФС. Кроме того, не учтены все функциональные возможности глубоких нейронных сетей. Это подтверждает актуальность данного исследования.

Цель данного исследования — преодолеть это ограничение и разработать универсальный, независимый от данных подход на основе глубокой гибридной модели DeepGCL. Он объединяет преимущества отдельных глубоких нейронных сетей. Метод должен обеспечить высокую точность обнаружения кибератак и низкую корреляцию ошибок между моделями. Данное исследование объединяет преимущества GRU [134, с. 1724–1734], CNN [259, с. 436–444] и LSTM [357, с. 194–197] для обнаружения кибератак, приводящих к сбоям в ОТ КФС. Рассматривается недавно предложенная функция активации SPOCU [245, с. 3385–3401], которая демонстрирует хорошую производительность. Она сравнивается с функцией активации ReLU (rectified linear unit). Исследование показывает, что технология параллельных вычислений повышает точность и эффективности предложенного подхода на основе глубокого обучения. Оптимизатор AdamW+Amsgrad [274, с. 1-19; 328, с. 1-23] уменьшает ошибку обучения глубокой нейронной модели и увеличивает эффективную скорость обучения [195, с. 1-5; 205, с. 177–185].

Основные вклады данного исследования включают следующее:

(1) Предлагается новый подход к обнаружению кибератак с целью обеспечения защищенности ОТ КФС на основе архитектуры трех параллельных глубоких нейронных сетей, таких как CNN, GRU и LSTM, которые одновременно собирают информацию о состоянии системы через них.

(2) Применение функции активации SPOCU в скрытых слоях предлагаемой модели показывает высокую точность обнаружения кибератак.

(3) Применение предлагаемой модели DeepGCL к установке по очистки сырой воды и процессам контура нагревателя газойля показывает, что подход эффективно обнаруживает кибератаки на КФС.

Предлагаемый подход. В этом исследовании предлагается модель DeepGCL в сочетании с функцией активации SPOCU для повышения точности классификации кибератак на КФС. Предлагаемый подход обнаруживает атаки на физическом уровне, отслеживая поведение системы. Набор обучающих данных и набор тестовых данных подготовлены для оценки модели.

Пусть обучающие данные содержат $\boldsymbol{N}$ устройств, от которых данные об отказах поступают в виде многомерного временного ряда. Каждое устройство содержит $\boldsymbol{m}$ сенсоров. Данные с каждого устройства представлены в виде

$$\boldsymbol{X_n} = [\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_{T_n}}] \in \boldsymbol{R^{m \times T_n}} \ \ (\boldsymbol{n} = \overline{\boldsymbol{1, N}}), \tag{4.3.1}$$

где $\boldsymbol{T_n}$ – время отказа.

Многомерный вектор измерений сенсора во времени $\boldsymbol{t}$ рассчитывается как

$$\boldsymbol{x_t} = [\boldsymbol{\omega_t^1}, \boldsymbol{\omega_t^2}, \dots, \boldsymbol{\omega_t^m}] \in \boldsymbol{R^{m \times 1}} \ \ (\boldsymbol{t} = \overline{\boldsymbol{1, T_n}}. \tag{4.3.2}$$

Входной вектор признаков предварительно обрабатывается. В момент времени $\boldsymbol{t}$ модель DeepGCL получает информацию от сенсора $\boldsymbol{x_t}$ размерности $\boldsymbol{m}$, после чего выходные данные классифицируются как нормальное поведение или атака. Как только модель DeepGCL хорошо обучена, ее можно повторно использовать для классификации новых выборок данных. Обучение используется в автономном режиме, а затем «обученная» модель развертывается в сети для обнаружения различных атак [16, с. 27-28]. Структура предлагаемой модели для обнаружения кибератак на КФС показана на рис. 4.3.1. Она включает модели глубокого GRU, глубокого CNN и глубокого LSTM. GRU и LSTM используются для повышения точности и снижения вычислительной сложности при обнаружении кибератак на КФС. Они рассматриваются из-за наличия блока памяти, который содержит как

новую информацию, так и данные за очень длительный период. 1D-CNN содержит сверточные слои, оптимизированные для извлечения трудноразличимых признаков [250, с. 72-83]. Рассмотренные глубокие нейронные сети используют исторические данные и выявляют информативные признаки без предварительного отбора для решения проблемы нелинейности.

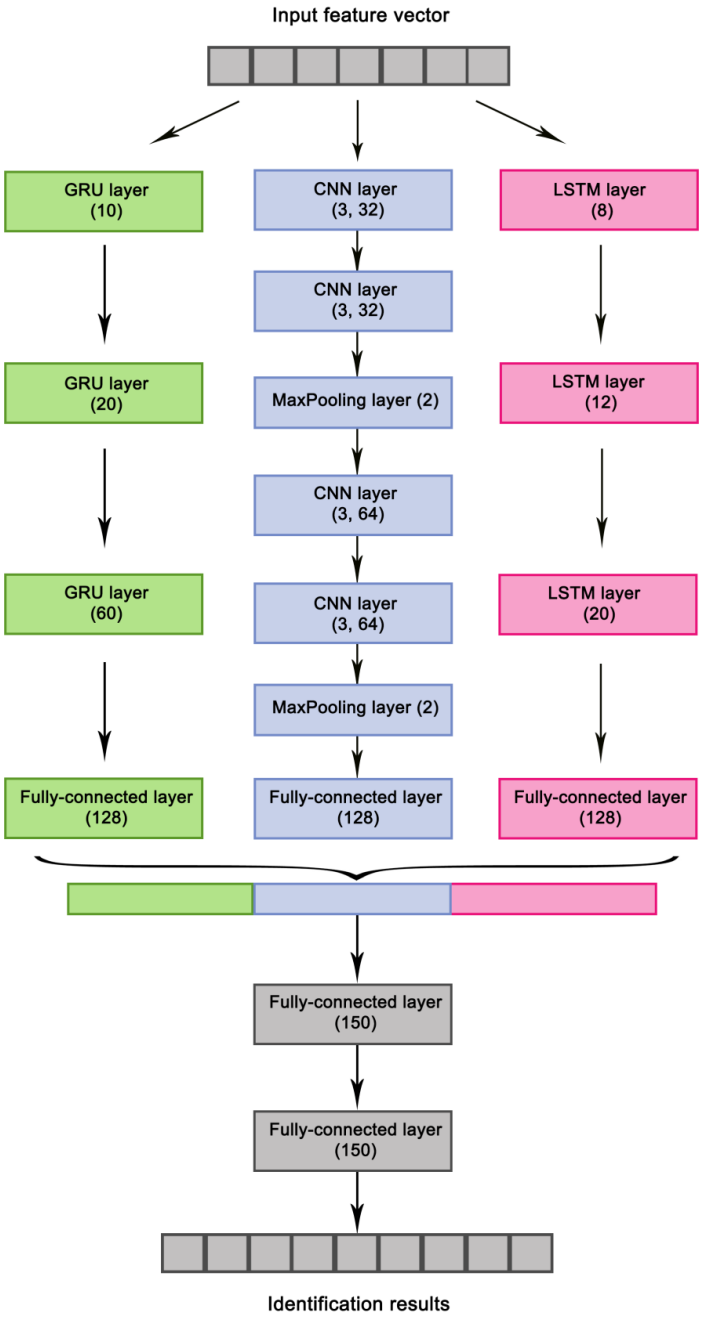


Рис. 4.3.1. Общая архитектура гибридной DeepGCL модели [53]

Глубокая CNN модель состоит из пяти слоев. Первые два слоя – это

сверточные слои (включающие 32 нейрона с 3 ядрами). За ними следует слой maxpooling с фильтром длины 2. Третий и четвертый сверточные слои состоят из 64 нейронов с 3 ядрами. Далее применяется maxpooling. За ним следует один полносвязный слой.

Выходные данные $\boldsymbol{y_t^{CNN}}$ и скрытые слои $\boldsymbol{h_t^{CNN}}$ для вектора признаков $\{\boldsymbol{F_1^{CNN}}, \boldsymbol{F_2^{CNN}}, \dots, \boldsymbol{F_T^{CNN}}\}$ CNN модели определяются следующим образом:

$$\boldsymbol{h_t^{CNN}} = \boldsymbol{\sigma}(\boldsymbol{W_{ih}}\boldsymbol{F_t^{CNN}} + \boldsymbol{W_{hh}}\boldsymbol{h_{t-1}^{CNN}} + \boldsymbol{b_h}), \quad (4.3.3)$$

$$\boldsymbol{y_t^{CNN}} = \boldsymbol{W_{ho}}\boldsymbol{h_t^{CNN}} + \boldsymbol{b_o}, \quad (4.3.4)$$

где $\boldsymbol{W_{ih}}$, $\boldsymbol{W_{hh}}$ и $\boldsymbol{W_{ho}}$ – матрицы весов между входным, выходным и скрытым слоями, $\boldsymbol{b_h}$ и $\boldsymbol{b_o}$ – векторы смещения [214, с. 1-18], $\boldsymbol{t}$ – количество образцов, $\boldsymbol{\sigma}$ – функция активации. В качестве функции активации можно взять ReLU, функцию гиперболического тангенса или SPOCU [245, с. 3385-3401].

Глубокая LSTM модель включает четыре слоя. Она решает проблему исчезающего градиента [357, с. 194-197]. Первый слой состоит из 8, второй из 12 и третий из 20 нейронов. За ними следует один полносвязный слой из 128 нейронов. LSTM модель включает в себя три мультипликативных блока: входной вентиль $\boldsymbol{i_t^{LSTM}}$, вентиль «забывания» $\boldsymbol{f_t^{LSTM}}$ и выходной вентиль $\boldsymbol{o_t^{LSTM}}$, который управляет ячейкой $\boldsymbol{c_t^{LSTM}}$, хранящей историческую и будущую информацию [357, с. 194-197]. Они рассчитываются следующим образом [214, с. 1-18; 381, с. 1-22]:

$$\boldsymbol{i_t^{LSTM}} = \boldsymbol{\sigma}(\boldsymbol{W_i}[\boldsymbol{h_{t-1}^{LSTM}}, \boldsymbol{F_t^{LSTM}}] + \boldsymbol{b_i}), \quad (4.3.5)$$

$$\boldsymbol{f_t^{LSTM}} = \boldsymbol{\sigma}(\boldsymbol{W_f}[\boldsymbol{h_{t-1}^{LSTM}}, \boldsymbol{F_t^{LSTM}}] + \boldsymbol{b_f}), \quad (4.3.6)$$

$$\boldsymbol{o_t^{LSTM}} = \boldsymbol{\sigma}(\boldsymbol{W_o}[\boldsymbol{h_{t-1}^{LSTM}}, \boldsymbol{F_t^{LSTM}}] + \boldsymbol{b_o}), \quad (4.3.7)$$

$$\boldsymbol{c_t^{LSTM}} = \boldsymbol{f_t^{LSTM}} \circ \boldsymbol{c_{t-1}^{LSTM}} + \boldsymbol{i_t^{LSTM}} \circ \tanh(\boldsymbol{W_c}[\boldsymbol{h_{t-1}^{LSTM}}, \boldsymbol{F_t^{LSTM}}] + \boldsymbol{b_c}), \quad (4.3.8)$$

$$\boldsymbol{y_t^{LSTM}} = \boldsymbol{o_t^{LSTM}} \circ \tanh(\boldsymbol{c_t^{LSTM}}), \quad (4.3.9)$$

где $\boldsymbol{i}$ - входной вентиль, $\boldsymbol{f}$ - вентиль «забывания», $\boldsymbol{o}$ - выходной вентиль, $\boldsymbol{W_i}$, $\boldsymbol{W_f}$, $\boldsymbol{W_o}$ и $\boldsymbol{W_c}$ - матрицы весов, $\circ$ - произведение Адамара, $\{\boldsymbol{F_1^{LSTM}}, \boldsymbol{F_2^{LSTM}}, \dots, \boldsymbol{F_T^{LSTM}}\}$ - вектор признаков, $\boldsymbol{y_t^{LSTM}}$ - выход LSTM, $\boldsymbol{h_{t-1}^{LSTM}}$ -

скрытые состояния, а $\boldsymbol{b_i}$, $\boldsymbol{b_f}$, $\boldsymbol{b_c}$ и $\boldsymbol{b_o}$ – векторы смещения.

Глубокая GRU модель состоит из четырех слоев. Первый слой GRU включает 10 нейронов. Второй и третий слои GRU содержат по двадцать нейронов. За ними следует один полносвязный слой, состоящий из 128 нейронов. Слои GRU рассчитываются следующим образом [134, с. 1724-1734]:

$$\boldsymbol{z_t^{GRU}} = \boldsymbol{\sigma}(\boldsymbol{W_z}[\boldsymbol{h_{t-1}^{GRU}}, \boldsymbol{F_t^{GRU}}]), \quad (4.3.10)$$

$$\boldsymbol{r_t^{GRU}} = \boldsymbol{\sigma}(\boldsymbol{W_r}[\boldsymbol{h_{t-1}^{GRU}}, \boldsymbol{F_t^{GRU}}]), \quad (4.3.11)$$

$$\boldsymbol{y_t^{GRU}} = (\mathbf{1} - \boldsymbol{z_t^{GRU}}) \circ \boldsymbol{h_{t-1}^{GRU}} + \boldsymbol{z_t^{GRU}} \circ \tanh(\boldsymbol{W}[\boldsymbol{r_t^{GRU}} \circ \boldsymbol{h_{t-1}^{GRU}}, \boldsymbol{F_t^{GRU}}]), \quad (4.3.12)$$

где $\boldsymbol{z_t^{GRU}}$ - вентиль «обновления», $\boldsymbol{y_t^{GRU}}$ - выход GRU, $\boldsymbol{r_t^{GRU}}$ - вентиль «сброса», $\{\boldsymbol{F_1^{GRU}}, \boldsymbol{F_2^{GRU}}, \dots, \boldsymbol{F_T^{GRU}}\}$ - вектор признаков, $\boldsymbol{h_{t-1}^{GRU}}$ - скрытые состояния, и $\boldsymbol{W_z}$, $\boldsymbol{W_r}$ и $\boldsymbol{W}$ – матрицы весов. В отличие от LSTM модели, у него нет модулей хранения. Выходные размеры глубокой LSTM и глубокой GRU одинаковы, чтобы обеспечить аналогичные вычисления.

Выходные данные моделей глубокой GRU, глубокой CNN и глубокой LSTM объединяются для повышения точности обнаружения кибератак на КФС и передаются на два полносвязных слоя, состоящих из 150 нейронов. За ними следует слой softmax. Выбор функции активации играет важную роль в построении модели, основанной на глубоком обучении. Функция SPOCU используется для повышения надежности предлагаемой гибридной модели. При использовании функции активации SPOCU происходит ускорение конвергенции глубокой нейронной сети, а потери падают значительно быстрее, чем при использовании других функций, таких как ReLU. На этапе обучения используется оптимизатор AdamW+Amsgrad [274, с. 1-19; 328, с. 1-23] с размером батча 128 и числом эпох равном 20 (Таблица 4.3.1). Согласно [195, с. 1-5; 205, с. 177-185; 274, с. 1-19], вес определяется следующим образом:

$$\boldsymbol{w_{t+1}} = (\mathbf{1} - \boldsymbol{\lambda})\boldsymbol{w_t} - \frac{\boldsymbol{\alpha}}{\sqrt{\widehat{\boldsymbol{S}}_t + \boldsymbol{\varepsilon}}}\boldsymbol{V_t}, \quad (4.3.13)$$

где

$$\widehat{\boldsymbol{S}}_t = \mathbf{max}(\widehat{\boldsymbol{S}}_{t-1}, \boldsymbol{S_t}), \quad (4.3.14)$$

$$\boldsymbol{V_t = \beta_1 V_{t-1} + (1 - \beta_1) g_t}, \quad (4.3.15)$$

$$\boldsymbol{S_t = \beta_2 S_{t-1} + (1 - \beta_2) g_t^2}. \quad (4.3.16)$$

Здесь $\boldsymbol{\lambda}$ представляет сокращение весов, $\boldsymbol{\alpha}$ - скорость обучения, $\boldsymbol{V_t}$ - момент 1-го порядка, $\boldsymbol{S_t}$ - момент 2-го порядка, $\boldsymbol{\beta_1}$ - скорость затухания для момента 1-го порядка, $\boldsymbol{\beta_2}$ - скорость затухания для момента 2-го порядка, $\boldsymbol{\widehat{S}_t}$ - компонент адаптивной скорости обучения, $\boldsymbol{g_t}$ - градиент временного шага.

Для расчета потерь используется кросс-энтропия [244, с. 1-15].

Таблица 4.3.1

Параметры обучения предлагаемой модели

Параметр	Значение
Оптимизатор	AdamW+AMSGrad
Размер батча	128
Начальная скорость обучения	0.001
Сокращение весов	$1x10^{-5}$
β_1	0.9
β_2	0.999
Максимальное число эпох	20

Описание наборов данных. Для оценки предлагаемого подхода необходимы реальные наборы данных с нормальным и аномальным поведением КФС. Для экспериментов наборы данных должны содержать достаточное количество аномальных выборок. В связи с этим рассматриваются наборы данных для двух промышленных объектов: набор данных SWaT [145, с. 88-99] и набор данных GHL [186; 202, с. 1-7], которые содержат информацию о нормальном поведении КФС и их отказах, вызванных кибератаками. Характеристики наборов данных представлены в таблице 4.3.2. Эти наборы данных представляют собой работу реальных КФС и используются для оценки алгоритмов машинного обучения. Набор данных о безопасной очистке воды (SWaT) был собран в Сингапурском университете технологий и дизайна. Это установка для очистки сырой воды, которая обычно используется в общественной инфраструктуре и рассматривается исследователями в качестве испытательного стенда для разработки различных мер по обнаружению

кибератак.

Таблица 4.3.2

Описание экспериментальных наборов данных

Набор данных	Количество образцов	Число атрибутов	Количество классов
Secure Water Treatment dataset [146]	449 919	51	36
Gasoil Heating Loop dataset [186; 202]	15 000	16	4

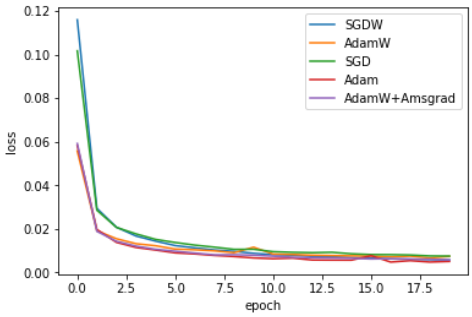
Он был построен на основе сенсоров, исполнительных механизсов, сигналов ввода-вывода ПЛК, человеко-машинных интерфейсов и SCADA для семи дней нормальной работы системы и четырех дней атак [275, с. 154–166]. Система содержит двадцать пять сенсоров и двадцать исполнительных механизмов. Все эти атрибуты имеют решающее значение для производительности предприятия и служат ранними индикаторами отказа. Они помогают предотвратить повреждение дорогостоящего оборудования. SWaT содержит информацию о 36 сценариях атак (например, атака с переполнением резервуара путем отключения насоса, атака на опустошение резервуара, атака с повреждением насоса и другие) и нормальной работой системы [222, с. 1058-1065]. Двадцать шесть атак включают одноэтапные одноточечные (SSSP) атаки, четыре одноэтапные многоточечные (SSMP) атаки, две многоэтапные одноточечные (MSSP) атаки и четыре многоэтапные многоточечные (MSMP) атаки. Общее количество образцов в наборе данных SWaT составляет 449919. Набор данных о контуре нагрева газойля (GHL) был собран в Лаборатории Касперского Future Technologies [186]. Испытательный стенд состоит из трех резервуаров: приемного бака (RT), нагревательного бака (HT) и коллекторного бака (КТ). Часть газойлевой смеси из RT переводится в HT на подогрев, а затем возвращается в RT, и процесс повторяется. Операцию проводят до тех пор, пока температура смеси при КТ не достигнет требуемого значения. Затем подогретую смесь переносят в бак КТ, а бак RT заполняют другой частью газойлевой смеси. Модель газойлевой установки состоит из образцов,

соответствующих различным параметрам процесса. Имеется 19 атрибутов, из которых рассматриваются 16 атрибутов: уровень газойля в RT, температура газойля в RT, управление подачей газойля, управление подогревателем и другие. Аномальное поведение системы наблюдалось из-за кибератак. Они включают несанкционированное изменение максимального уровня RT, максимальной температуры HT, частоты насоса и значения времени релаксации системы [186]. Так атака на переполнение RT характеризуется высоким значением фактического уровня RT, а проблема перевода из-за обрыва клапана RT характеризуется нулевым значением фактического уровня RT.

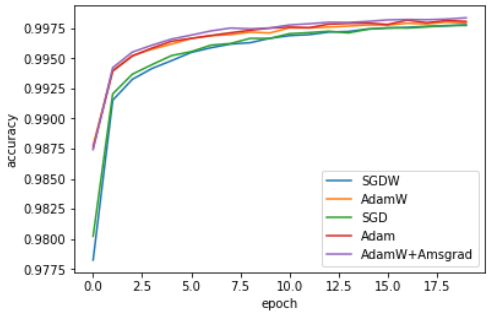
Метрики оценки производительности. Оценка производительности предлагаемой модели основана на следующих показателях: precision, recall, accuracy, F-мера, частота ложных срабатываний и *LogLoss*. Метрика *LogLoss* используется для оценки производительности алгоритмов обучения. Минимизация *LogLoss*, также известная как кросс-энтропия, эквивалентна максимизации логарифмической вероятности наблюдения данных модели. Все рассмотренные метрики являются широко используемыми показателями эффективности в машинном обучении [185, с. 27-38].

Результаты экспериментов. Эксперименты проводились на виртуальной машине Intel Xeon (R), CPU X5670 @ 2,93 ГГц * 24 с 24 ГБ оперативной памяти. Предлагаемый подход был оценен в Python 2.7.13 с использованием различных библиотек, включая Tensorflow и Keras. Эксперименты проводились на двух наборах данных для оценки предложенной модели глубокого гибрида для обнаружения кибератак на КФС, которая сочетает в себе преимущества глубокой CNN, глубокой GRU и глубокой LSTM. Наборы данных разделены на обучающий набор (80%) и тестовый набор (20%). SPOCU и ReLU рассматривались как функции активации. Модель DeepGCL оценивалась с использованием различных оптимизаторов: стохастический градиентный спуск (SGD), SGD с сокращением весов (SGDW), Adam, Adam с сокращением весов (AdamW) и AdamW+Amsgrad. Результаты были получены для 20 эпох и

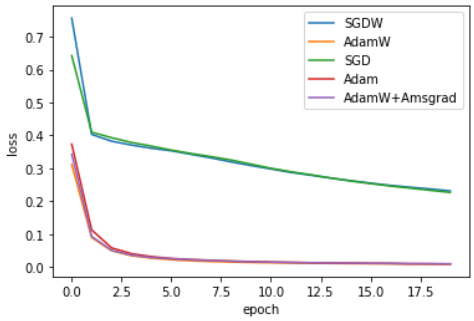
размера батча равного 128. Представленные результаты усреднены по 10 запускам. На рис. 4.3.2-4.3.3 показаны результаты для наборов данных SWaT и GHL соответственно. Погрешность тестирования стабильно уменьшается с увеличением числа эпох. Однако при использовании оптимизатора SGDW предложенная модель показала более высокий уровень ошибок для набора данных SWaT (Рис. 4.3.2(а)), а для набора данных GHL (Рис. 4.3.3(а)) оптимизатор SGD показал более низкий результат. На рис. 4.3.2(б) лучший результат показала предложенная модель с оптимизатором AdamW+Amsgrad, тогда как SGDW оказался менее точным для набора данных SWaT. По точности все оптимизаторы показали высокие результаты при обнаружении кибератак для модели DeepGCL с использованием ReLU, достигнув 99.68%, 99.75%, 99.67%, 99.79% и 99.80% для SGDW, AdamW, SGD, Adam и AdamW+Amsgrad, соответственно. А при применении функции активации SPOCU результаты для оптимизаторов SGDW, AdamW, SGD, Adam и AdamW+Amsgrad составили 95.79%, 99.80%, 95.94%, 99.83% и 99.85%, соответственно (Рис. 4.3.2 (г)).



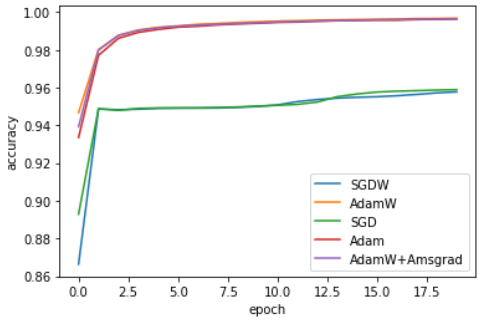
а) ReLU



б) ReLU

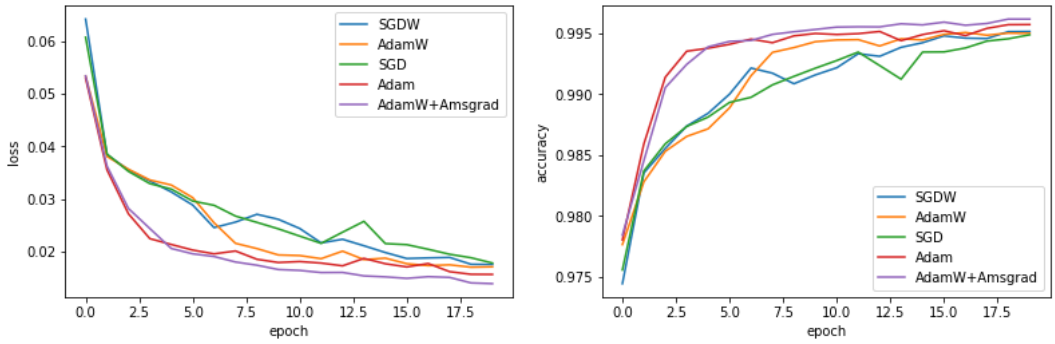


в) SPOCU

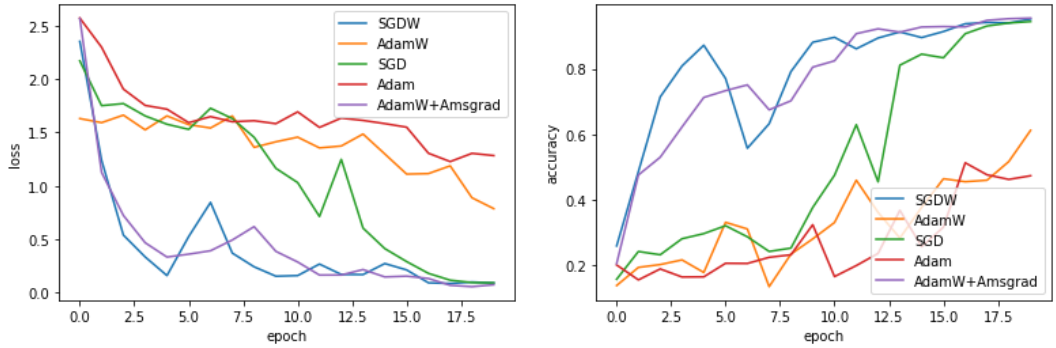


г) SPOCU

Рис. 4.3.2. Кривые точности и потерь при обнаружении кибератак для модели DeepGCL на наборе данных SWAT [53]



а) ReLU **б) ReLU**



в) SPOCU **г) SPOCU**

Рис. 4.3.3. Кривые потерь и точности обнаружения кибератак для модели DeepGCL на наборе данных GHL [53]

Согласно рис. 4.3.3(б), для набора данных GHL оптимизаторы Adam и AdamW+Amsgrad показали близкие результаты. Наихудшие показатели наблюдались для SGD, SGDW и AdamW для набора данных GHL. По точности результаты для модели DeepGCL с использованием ReLU показали следующее: 99.38%, 99.45%, 99.37%, 99.48% и 99.55% для оптимизаторов SGDW, AdamW, SGD, Adam и AdamW+Amsgrad, соответственно. Тогда как при использовании функции активации SPOCU результаты для оптимизаторов SGDW, AdamW, SGD, Adam и AdamW+Amsgrad составили 98.16%, 61.24%, 98.05%, 42.37% и 99.64%, соответственно (Рис. 4.3.3(г)). А самый низкий результат по точности показали оптимизаторы Adam и AdamW. Итак, рис. 4.3.2-4.3.3 доказывают, что при использовании функции активации SPOCU сходимость DeepGCL ускоряется, а потери падают намного быстрее, чем при использовании функции ReLU. В таблице 4.3.3 показано, что наилучшая скорость обучения для модели DeepGCL составляет 0.001 и занимает 312 секунд для набора данных SWaT и 113 секунд для GHL.

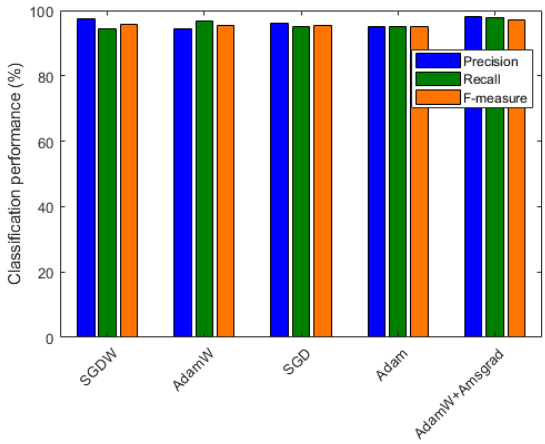
Таблица 4.3.3

Оценка предлагаемой модели

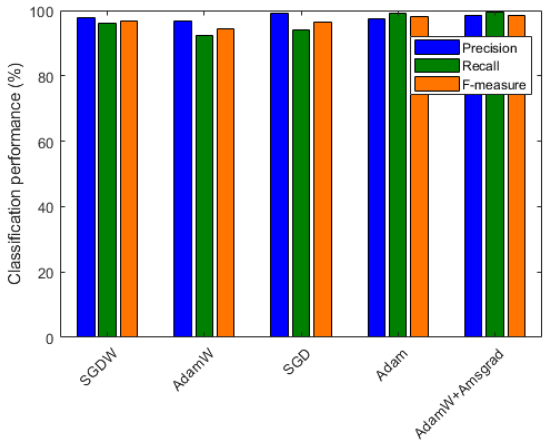
Набор данных	Скорость обучения	Потери при обучении	Потери при тестировании	Точнсть обучения (%)	Точность тестирования (%)	FAR	Время (сек)
SWaT	0.01	0.1338	0.1316	96.40	96.60	0.0032	**309.35**
	0.001	**0.0057**	**0.0068**	**99.76**	**99.75**	**0.0003**	311.95
	0.0001	0.0088	**0.0097**	99.67	99.62	0.0006	312.8
GHL	0.01	1.6103	1.6347	60.18	54.22	0.2000	**111**
	0.001	**0.0181**	**0.0534**	**98.64**	**97.80**	**0.0046**	112.7
	0.0001	0.5624	1.0760	67.59	62.62	0.0781	112.7

DeepGCL при использовании SPOCU является более точным, имеет наименьшие потери при выбранной скорости обучения, чем другие, более низкую частоту ложных срабатываний (столбец FAR (false acceptance rate)) и лучшую скорость обнаружения, о чем свидетельствуют значения точности обучения и точности тестирования. Таким образом, предложенная модель является наиболее последовательной в обнаружении атаки по рассматриваемым наборам данных КФС.

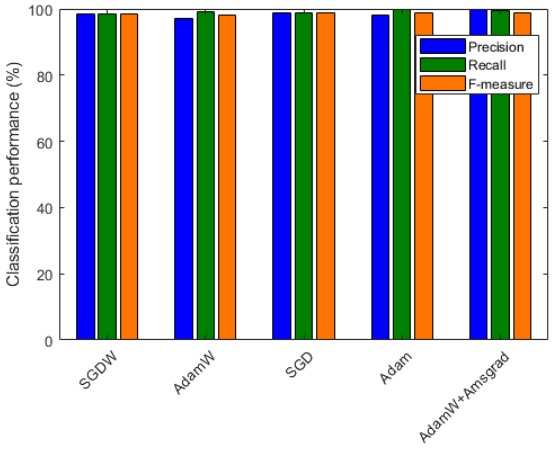
Для набора данных SWaT было проведено исследование для обнаружения четырех типов атак (Рис. 4.3.4). По показателям precision, recall и F-мера атаки были наиболее точно обнаружены с помощью оптимизатора AdamW+Amsgrad.



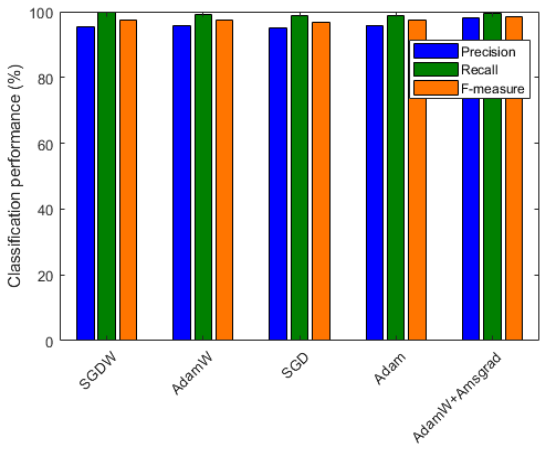
а) SSSP атаки



б) SSMP атаки



в) MSMP атаки



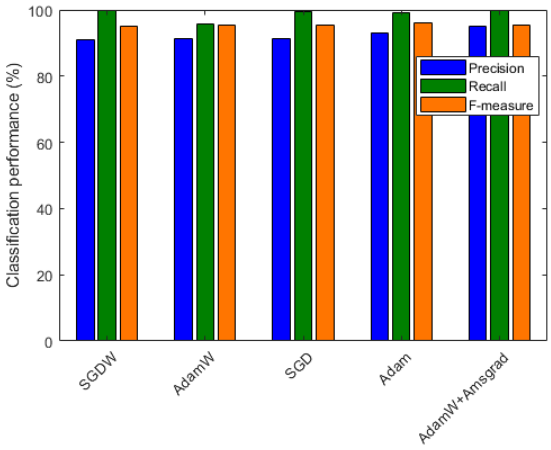
г) MSSP атаки

Рис. 4.3.4. Оценка обнаружения различных атак на SWaT с использованием разных оптимизаторов [53]

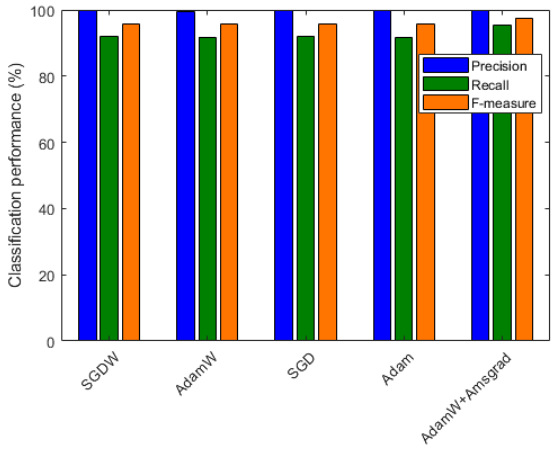
Тогда как этот подход несколько уступал по точности по метрике полноты при распознавании атаки на уровне RT в GHL (Рис. 4.3.5). Однако показал 100% результат для атаки на релаксацию системы.

Результаты обнаружения атак на набор данных SWaT по метрике recall представлены в таблице 4.3.4. Сравнение предложенной гибридной модели с моделями LSTM, 1D-CNN, GRU и известными методами машинного обучения (DNN, SVM и TABOR) также показано.

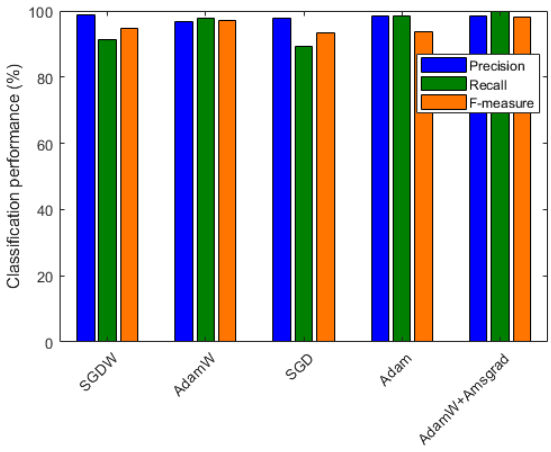
Как показано, была обнаружена атака на ПЛК, направленная на MV-101, что приводит к переполнению бака. Об этом сообщает сенсор LIT-101 (recall равен 96.70%).



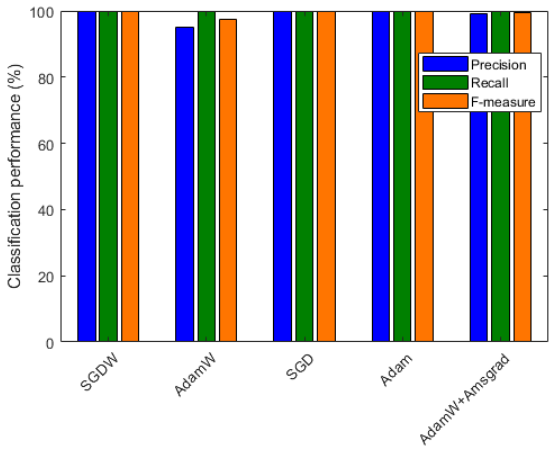
а) Pump frequency



б) RT



в) HT



г) Время релаксации системы

Рис. 4.3.5. Оценка обнаружения различных атак на GHL с использованием разных оптимизаторов [53]

Атака SSSP на закрытие P-302 для остановки притока бака обнаружена со 100% точностью. Атака SSMP, направленная на остановку оттока воды из бака, приведшая к отключению P-101 и P-102, была обнаружена с показателем recall 92%.

По метрике recall (100%) успешно обнаружена аномалия, вызванная атакой типа MSSP на постоянное включение P-101 и переполнение LIT-301. Атака SSMP, направленная на ухудшение качества воды, показала 97.90% (отключения P-203 и P-205). Атака MSMP, направленная на отключение системы, также была точно обнаружена (99.9%).

Также предложенная модель сравнивалась с другими методами машинного обучения. Сравнение с DNN, SVM [222, с. 1058-1065] и TABOR [268, с. 525-536] на наборе данных SWaT показало, что гибридная модель DeepGCL с использованием SPOCU и AdamW+Amsgrad превзошла эти методы и достигла наилучших результатов (выделено жирным шрифтом) в большинстве атак (Таблица 4.3.4).

Экспериментальные результаты на наборе данных GHL (Таблица 4.3.5) показали высокую точность предложенного метода. Были рассмотрены четыре типа атак, приводящих к сбоям в работе КФС.

Таблица 4.3.4

Сравнение модели DeepGCL с другими методами по метрике recall на наборе данных SWaT

	Атака	Метод							
		LSTM	1D CNN	GRU	DeepGCL (ReLU)	DeepGCL (SPOCU)	DNN [222]	SVM [222]	TABOR [268]
SSSP атаки	MV-101	1.000	0.722	0.995	0.971	0.997	0	0	0.049
	P-102	0.946	0.902	0.968	0.989	0.968	0	0	0.930
	LIT-101	0.947	0.179	0.987	0.805	0.967	0	0	0
	MV-504	0.553	0.128	0.882	0.841	0.852	0	0.035	0.328
	DPIT	0.971	0.625	0.985	0.997	0.983	0.927	0.919	0.612
	FIT-401<0.7	0.851	0.945	0.851	0.936	0.936	1.000	0.433	0.994
	FIT-401=0	1.000	0.975	1.000	0.992	1.000	0.978	1.000	0.998
	LIT-301_1	0.786	0.086	0.911	0.893	0.888	0	0	0
	MV-303	0.970	0.999	1.000	0.962	1.000	0	0	0.597
	AIT-504_16	0.755	0.138	0.981	0.949	0.978	0.123	0.130	0.004
	P-302	1.000	0.978	1.000	1.000	1.000	0.936	0.936	1.000
	LIT-401_L	0.971	0.212	1.000	0.998	1.000	0	0	0
	LIT-301>HH	0.945	0.038	0.945	0.920	0.945	0	0.905	0
	LIT-101_H	0.804	0.279	0.891	0.881	0.890	0	0	0.890
	P-101	0.636	0.982	0.500	0.134	0.975	0	0	0.990
	LIT-101_LL	0.827	0.156	0.971	0.912	0.965	0	0.119	0.889
	FIT-401=0	0.994	0.265	0.989	0.921	0.990	0.933	0.927	0.997
	LIT-301_0.5	0.956	0.997	0.962	0.950	0.995	0	0.357	0
SSMP атаки	MV-101, LIT-101	0.904	0.420	0.992	0.949	0.989	0	0.017	0.083
	P-203, P-205	0.985	0.625	0.942	0.810	0.979	0	0	0
	LIT-401, P-401	0.992	0.341	0.983	0.999	0.990	0	0.009	0
	P-101, P-102	0.874	0.557	0.939	0.863	0.920	0	0	0.258
	P-501, FIT-502	0.934	0.265	1.000	0.999	0.999	1.000	1.000	0.998
	P-602, DIT-301, MV-302	1.000	0.341	0.978	0.985	0.999	0.876	0.875	0
	LIT-101, P-101, MV-201	1.000	0.234	0.956	0.995	1.000	0	0.003	0.999
MSSP атаки	P-101, LIT-301	1.000	0.998	0.972	1.000	1.000	0	0	0.999
	P-302, LIT-401	1.000	0.335	0.994	0.965	1.000	0	0	0.196
	AIT-402, AIT-502	0.928	0.265	0.978	0.954	0.977	0.923	0.927	0.996
	FIT-401, AIT-502	0.976	0.800	0.918	0.893	0.972	0.940	0	0.369

Таблица 4.3.5

Сравнение модели DeepGCL с другими методами, использующими метрику precision для набора данных GHL

Атака	Метод						
	LSTM	1D-CNN	GRU	DeepGCL (ReLU)	DeepGCL (SPOCU)	DNN	SVM
Pump frequency	0.813	0.511	**0.998**	**0.998**	**0.998**	0.993	0.922
HT temperature	0.927	**1.000**	0.345	0.998	**1.000**	0.987	0.507
RT level	0.925	0.587	**0.953**	0.924	**0.953**	0.932	0.895
System relaxing time	0.602	**1.000**	0.979	0.997	0.999	0.967	0.757

По метрике recall атака на несанкционированное изменение максимальной температуры HT была точно обнаружена и составила 100%. Атака на клапан в резервуаре RT, приведшая к остановке подачи нагретой жидкости в резервуар RT, была распознана DeepGCL(SPOCU) и превысила LSTM (92.5%), 1D-CNN (58.7%) и DeepGCL(ReLU) (92.4%) и совпала с GRU по метрике recall (95.3%). Атаки на изменение времени релаксации системы выявлялись с высокой точностью и составили 99.9%.

Также была исследована эффективность предложенной модели по сравнению с LSTM, 1D CNN, GRU и DeepGCL(ReLU) для каждой атаки из набора данных GHL. Согласно таблице 4.3.5, точность обнаружения атак по максимальной температуре HT и значению времени релаксации системы для модели DeepGCL(SPOCU) совпала с 1D-CNN; а для атак на уровень RT – с GRU. Несмотря на эффективность методов SVM и DNN, по точности они уступали предложенному подходу.

Подводя итог, рассмотренные процессы были проанализированы по их научной значимости и практической применимости. Также была показана эффективность предложенной глубокой гибридной модели, объединяющей преимущества трех глубоких нейронных сетей для обнаружения кибератак на КФС (Таблица 4.3.6). Таким образом, сравнение DeepGCL с известными

методами машинного обучения показало более точный результат классификации атак.

Таблица 4.3.6

Оценка DeepGCL(SPOCU) на различных наборах данных

Набор данных	Метрики		
	Precision (%)	Recall (%)	F-мера (%)
Secure Water Treatment dataset	99.76	99.75	99.74
Gasoil Heating Loop dataset	97.06	95.66	96.04

Подводя итоги в данном исследовании была предложена гибридная модель DeepGCL, которая объединяет преимущества глубокой GRU, глубокой CNN и глубокой LSTM. Исследование проводилось на двух больших реальных наборах данных КФС, содержащих информацию о кибератаках на основе данных сенсоров для улучшения интерпретации результатов. Модель обнаруживает кибератаки, приводящие к сбоям в ОТ КФС. Для оценки предложенной модели рассматривались наборы данных SWaT и GHL, содержащие информацию о нормальном поведении КФС и ее сбоях, вызванных кибератаками. Для повышения надежности модели DeepGCL использовалась функция активации SPOCU. Она ускорила сходимость глубокой гибридной модели. Модель DeepGCL была протестирована на различных оптимизаторах. Для повышения эффективной скорости обучения использовался оптимизатор AdamW+Amsgrad. Сравнение предложенного метода с известными методами машинного обучения проводилось на основе метрик точности, precision, recall, частоты ложных срабатываний и F-меры. В результате модель DeepGCL в сочетании с функцией активации SPOCU доказала свое превосходство. При этом предложенный подход показал высокую точность обнаружения кибератак, сохранив при этом хорошую обобщающую способность. Согласно результатам эксперимента, можно сделать вывод, что глубокая гибридная модель имеет большой потенциал, чтобы стать эффективным и действенным инструментом для диагностики и прогнозирования кибератак на ОТ КФС и может стать

многообещающей альтернативой для интеллектуальных систем обслуживания в будущем. Предложенный подход может применяться к различным приложениям безопасности ОТ КФС.

4.4. Метод классификации отказов киберфизических систем на основе изображений

Оценка производительности и своевременное обнаружение неисправностей оборудования КФС могут снизить эксплуатационные расходы и расходы на техническое обслуживание. Признаками неисправности оборудования являются изменения сигналов сенсоров. Оценка вибрационных сигналов на основе акселерометров может обнаруживать сбои и позволяет оценивать отказы КФС.

Интеллектуальная система диагностики помогает специалистам выявлять неисправности в оборудовании. Эксперту должна быть предоставлена подтверждающая информация о качестве системы. Поэтому решение интеллектуальной диагностической системы должно учитывать мнение эксперта.

Поиск отклонений от нормальной работы КФС, которые могут привести к неисправностям, является важной областью исследований. Исследователи разрабатывают новые методы и расширяют существующие алгоритмы обнаружения неисправностей в ОТ КФС. В настоящее время предлагаемые подходы включают анализ вибрационных сигналов полученных от сенсоров КФС и диагностику неисправностей для решения данной проблемы [110; 310, с. 415-422].

Сигнал вибрации несет в себе наиболее важную информацию о состоянии механических устройств. Примером может служить различное нефтяное оборудование. Характеризующие неисправности устройств признаки извлекаются для интеллектуального анализа необработанных сигналов и повышения точности диагностики. Так был предложен подход на основе спектральной регрессии для извлечения признаков неисправностей сигналов

сенсоров КФС [378, с. 13694-13719]. Метод k-средних рассматривался для оценки производительности методов спектральной регрессии, анализа главных компонент, факторного анализа, проекций с сохранением локальности, собственных карт Лапласа и линейного дискриминантного анализа. Также Lu и соавторы в 2017 году разработали метод диагностики неисправностей КФС на основе стекированного шумоподавляющего автокодировщика, в котором для получения признаков высокого порядка используются представление разреженности и сжатие данных. Модель сравнивалась с известными методами и показала относительно лучшие результаты из-за возможности сжатия данных для высоконадежного самообучения. Автоматизированный выбор признаков вибрационного сигнала от сенсоров нефтяного оборудования для обнаружения отказов был исследован в работе [158, с. 1807-1812]. Также был предложен алгоритм выбора признаков двоичного ансамбля с KNN, обученным на различных наборах признаков (признаки временной и частотной областей). Применение выбора признаков с подходом классификации «один против одного» повысило точность классификации. Было разработано интеллектуальное решение для диагностики неисправностей перед установкой КФС [310, с. 415-422]. В качестве классификаторов рассматривались KNN, SVM, деревья решений, RF со стандартизацией и без нее. KNN со стандартизацией показал наилучшие результаты при обнаружении различных неисправностей КФС устройств. В работе [159, с. 67-72] была изучена производительность ELM для автоматического обнаружения неисправностей в устройствах связанных с добычей нефти. Для выбора признаков рассматривался алгоритм последовательного прямого выбора. ELM показал довольно хорошие результаты по сравнению с KNN и SVM. Rauber и соавторы в 2017 году сравнили производительность ELM с отображением ядра и без него с другими классификаторами. Также были оценены три типа неисправностей КФС (перекос вала, дисбаланс лопастей насоса и механическое трение) и неисправные сенсоры на основе данных акселерометров [309, с. 381-396]. В

последние годы глубокие нейронные сети применялись для обнаружения механических неисправностей КФС. Они нашли применение в извлечении признаков вибрационных сигналов, в решении проблемы несбалансированных данных и в качестве методов классификации. Таким образом, был предложен метод обучения признакам на основе CNN для обнаружения неисправностей [230, с. 331-345]. Метод нашел применение в системах диагностики неисправностей КФС на основе изображений [130, с. 1-13; 402, с. 213-237]. DNN также нашли применение в решении проблемы дисбаланса данных при диагностике неисправностей в ОТ. Чтобы увеличить количество шаблонов отказов оборудования в несбалансированном наборе данных, Wang и соавторы предложили подход, основанный на генеративно-состязательной сети Вассерштейна [373, с. 111168-111180]. По сравнению с другими методами, данный подход неправильно классифицировал только 2.59% образцов. В работе [130, с. 1-13] глубокая CNN была рассмотрена для извлечения признаков из вибрационных сигналов. Предложенный метод сравнивался с SVM, NB и нейронной сетью обратного распространения. Частотно-временные данные в сочетании с DCNN показали наилучшие результаты. Иерархическая адаптивная к скорости обучения глубокая CNN была предложена Guo и соавторами в 2016 году для обнаружения отказов оборудования. Модель автоматически извлекает признаки неисправностей и показывает лучшие результаты, чем глубокая CNN и метод искусственной опорной векторной регрессии. Был предложен метод диагностики неисправностей КФС с использованием ансамбля из глубокой CNN модели и RF с использованием 2D-изображений в оттенках серого, полученных с помощью преобразования CWT [386, с. 1-21]. Предложенный метод сравнивался с нейронной сетью обратного распространения, SVM, глубокой сетью убеждений и CNN. Li и соавторы в 2017 году описали ансамбль глубокой CNN с улучшенной теорией Демпстера–Шейфера для диагностики неисправностей устройств КФС. Предложенная модель имеет высокую точность диагностики и адаптивность по сравнению с SVM, многослойной

персептронной нейронной сетью, DNN, глубокой CNN с широкими ядрами первого слоя. Параллельный ансамбль DNN и CNN, где признаки временной области извлекаются DNN (глобальные признаки), объединяются с признаками, извлеченными CNN (локальные признаки) из сигналов вибрации [264, с. 1-18].

Следующие выводы можно сделать, обобщая анализ текущего состояния исследований в области создания интеллектуальной системы обнаружения отказов в КФС. Во-первых, очень мало исследований было посвящено применению глубоких нейронных сетей для извлечения признаков и обнаружения отказов КФС, особенно, оборудования связанного с добычей нефти. Во-вторых, модели, основанные на глубоком обучении, используются при диагностике отказов только как замена известных методов классификации. Кроме того, не учитывается вся функциональность глубоких нейронных сетей. Все это подтверждает актуальность данного исследования.

Целью данного исследования является разработка метода классификации отказов ОТ с использованием модели на основе глубокого обучения. Модель сочетает в себе преимущества глубокой нейронной сети (deep neural network, DNN) и CNN. Характеристики сигналов вибрации в частотной и временной областях рассматриваются как входные характеристики для модели DNN.

CNN может обрабатывать двухмерные (2D) изображения по принципу работы человеческого мозга, эффективно извлекает признаки из изображений, а также требует меньше параметров обучения, в отличие от DNN. Спектрограмма кратковременного преобразования Фурье (short-time Fourier transform, STFT) и скалограмма непрерывного вейвлет-преобразования (continuous wavelet transform, CWT) сигнала сенсора считаются входными данными для CNN в этом исследовании. Спектрограмма несет информацию о фиксированном частотно-временном представлении сигнала, что не позволяет получить полное представление о происходящем. В то же время скалограмма дает более детальное представление о вибрационном сигнале. Поэтому спектрограмма и скалограмма отправляются в CNN и объединяются для более информативного

изучения сигнала. Однако CNN дает неудовлетворительные результаты на изображениях высокой размерности. В связи с этим размер спектрограммы и скалограммы уменьшен до 128x128x3 пикселей.

Результаты этого исследования показывают, что предложенная глубокая гибридная модель может автоматически и одновременно извлекать характеристики сигналов сенсоров, которые чувствительны к сбоям во временной, частотной и частотно-временной областях. Таким образом, предложенная гибридная модель с использованием глубоких нейронных сетей может быть применена при диагностике отказов оборудования КФС.

Извлечение признаков из данных сенсоров. Сигналы сенсоров широко используются для обнаружения неисправностей в различных устройствах. Сигналы, полученные сенсорами, обрабатываются для получения важной информации, а затем выделяются доминирующие характеристики сигнала вибрации [110, с. 128-143; 317, с. 164-173; 372, с. 2415-2436; 378, с. 13694-13719]. Они включают в себя особенности временной и частотной областей. Спектрограмма быстрого преобразования Фурье и скалограмма CWT используются для анализа сигналов в частотно-временной области [325, с. 508-517].

Признаки временной области. Признаки временной области часто включают характеристики, чувствительные к ошибкам [308, с. 127-203; 393, с. 2573-2588]. К ним можно отнести такие характеристики, как MN (arithmetic mean), SD (standard deviation), RMS (root mean square), KV (kurtosis value), SV (skewness value), PPV (peak-to-peak value), IF (impulse factor) и SF (shape factor) [226, с. 396-399]. Эти признаки определяются следующим образом:

$$\boldsymbol{X_{mn}} = \frac{\mathbf{1}}{\boldsymbol{N}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{x_i}, \tag{4.4.1}$$

$$\boldsymbol{X_{sd}} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{N-1}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} |\boldsymbol{x_i} - \boldsymbol{X_{mn}}|^{\mathbf{2}}}, \tag{4.4.2}$$

$$\boldsymbol{X_{rms}} = \sqrt{\frac{\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{x_i^2}}{\boldsymbol{N}}}, \tag{4.4.3}$$

$$\boldsymbol{X_{kv}} = \frac{\sum_{i=1}^{N}(x_i - X_{mn})^4}{NX_{sd}^4}, \quad (4.4.4)$$

$$\boldsymbol{X_{sv}} = \frac{\sum_{i=1}^{N}(x_i - X_{mn})^3}{NX_{sd}^3}, \quad (4.4.5)$$

$$\boldsymbol{X_{ppv}} = \mathbf{max}(\boldsymbol{x_i}) - \mathbf{min}(\boldsymbol{x_i}), \quad (4.4.6)$$

$$\boldsymbol{X_{if}} = \frac{\max|\boldsymbol{x_i}|}{\frac{1}{N}\sum_{i=1}^{N}|x_i|}, \quad (4.4.7)$$

$$\boldsymbol{X_{sf}} = \frac{\boldsymbol{X_{rms}}}{\frac{1}{N}\sum_{i=1}^{N}|x_i|}, \quad (4.4.8)$$

где $\boldsymbol{x}$ - входной сигнал, $\boldsymbol{N}$ - число точек данных.

Признаки частотной области. Анализ частотной области позволяет получить необходимую информацию о сигнале, не содержащуюся во временной области [393, с. 2573-2588]. Эти характеристики часто включают SC (spectral centroid), SR (spectral roll-off) и др. и рассчитываются следующим образом:

$$\boldsymbol{X_{sc}} = \frac{\int_0^{+\infty} \boldsymbol{m \cdot s(m) dm}}{\int_0^{+\infty} \boldsymbol{s(m) dm}}, \quad (4.4.9)$$

$$\boldsymbol{X_{sr}} = \boldsymbol{\lambda} \int_0^{+\infty} |\boldsymbol{s(m)}| \boldsymbol{dm}, \quad (4.4.10)$$

где $\boldsymbol{s(m)}$ является магнитудой элемента разрешения (magnitude of bin) $\boldsymbol{m}$ и $\boldsymbol{\lambda}$ является порогом, который принимается равным 0.85 согласно [364, с. 293-302]. Средние значения $\boldsymbol{X_{sc}}$ и стандартного отклонения $\boldsymbol{X_{sr}}$ рассматриваются как признаки.

Для представления вибрационного сигнала также рассматриваются Мел-частотные кепстральные коэффициенты (Mel-frequency cepstral coefficients, MFCC), которые фиксируют некоторые из его важных свойств. Они были предложены Davis и Mermelstein [156, с. 65-74] и нашли широкое применение в области распознавания и анализа речи.

Для их получения рассматриваемый вибрационный сигнал разбивается на фреймы (в данном исследовании их длина составляет 2048 мс со сдвигом 512 мс). Фреймы умножаются на оконную функцию (например, окно Хэмминга). Каждый фрейс подвергается STFT, а затем к полученным данным применяется

банк фильтров MFCC [387, с. 327-340]. Для уменьшения размерности используется дискретное косинусное преобразование. Для каждого фрейма извлекаются двенадцать кепстральных признаков, а затем рассчитываются среднее значение и стандартное отклонение как репрезентативные признаки.

Предлагаемый подход. Предлагается подход к обнаружению неисправностей ОТ по сигналам сенсоров. Рассмотрена модель на основе глубокого обучения, которая показывает высокую точность при проведении экспериментов. Блок-схема предлагаемого подхода, состоящего из следующих этапов: обработка сигналов и выделение признаков, а также применение глубокой гибридной модели для диагностики неисправностей погружного насоса центробежного типа (ЭПН) (electric submersible pump, ESP) в качестве примера, показана на рис. 4.4.1 [61, с. 1-16].

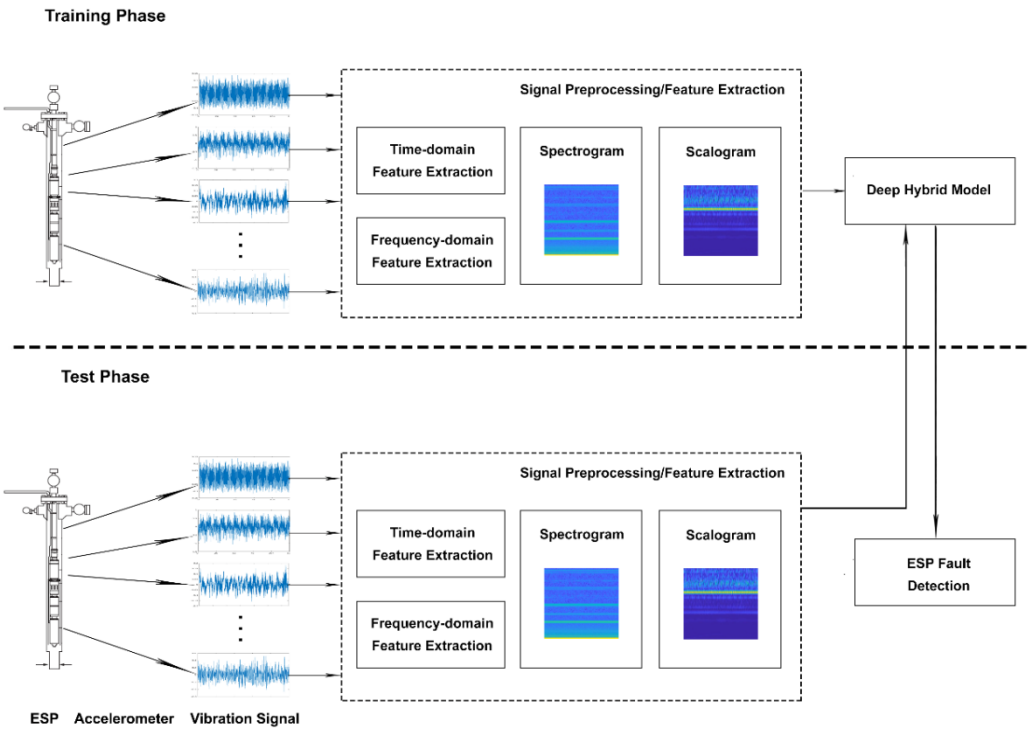


Рис. 4.4.1. Блок-схема предлагаемого подхода

Во-первых, рассчитываются восемь признаков во временной области непосредственно из сигнала вибрации ЭПН и двадцать восемь признаков в частотной области на основе преобразования Фурье. Во-вторых, генерируются скалограмма на основе вейвлет-преобразования и спектрограмма на основе

кратковременного преобразования Фурье. Полученные частотные и временные характеристики и визуальные представления направляются на глубокую гибридную модель, где определяются нормальное состояние ЭПН и его неисправности.

Предлагается использовать простую глубокую нейронную сеть, состоящую из полносвязных слоев для обработки признаков в частотной и временной областях (frequency- and time-domain features, TFDF-DNN), которые рассматриваются как единый вектор. TFDF-DNN состоит из двух скрытых слоев, содержащих 128 и 256 нейронов. Последний полносвязный слой модели состоит из 2500 нейронов.

Предлагаются спектрограммы и скалограммы для обработки изображений с помощью моделей глубоких CNN, которые обозначаются как модель SP-CNN и модель SG-CNN, соответственно. Цветные изображения размером 128x128x3 получаются с помощью предварительной обработки вибрационных сигналов полученных от акселерометров и передаются на модель CNN. Скалограмма позволяет определить различные частотные составляющие вибрационного сигнала. Ее преимущества перед спектрограммой заключаются в том, что она лучше выявляет низкочастотные и быстро меняющиеся составляющие сигнала [266, с. 1-13], что улучшает классификационные характеристики. Поскольку CNN дает плохие результаты с очень большими изображениями, сигнал предварительно разделяется на сегменты.

Входные изображения CNN обрабатываются сверткой. Слой свертки можно рассчитать следующим образом:

$$\boldsymbol{x}_{\boldsymbol{m_k}}^{\boldsymbol{k}} = \sum_{\boldsymbol{m_{k-1}}}^{\boldsymbol{M_{k-1}}} (\boldsymbol{x}_{\boldsymbol{m_{k-1}}}^{\boldsymbol{k-1}} \times \boldsymbol{w}_{\boldsymbol{m_{k-1}}}^{(\boldsymbol{m_k})}) + \boldsymbol{b}^{(\boldsymbol{m_k})}, \qquad (4.4.11)$$

где $\boldsymbol{k}$ – количество слоев, $\boldsymbol{x}_{\boldsymbol{m_{k-1}}}^{\boldsymbol{k-1}}$ – $\boldsymbol{m_{k-1}}$-ая матрица изображений $\boldsymbol{k-1}$, $\boldsymbol{M_{k-1}}$ – количество матриц изображений в $\boldsymbol{k-1}$, $\boldsymbol{w}_{\boldsymbol{m_{k-1}}}^{(\boldsymbol{m_k})}$ – $\boldsymbol{m_{k-1}}$-ый канал $(\boldsymbol{m_k})$-го фильтра в $\boldsymbol{k}$, $\boldsymbol{b}^{(\boldsymbol{m_k})}$ – смещение $\boldsymbol{m_k}$-ого фильтра в $\boldsymbol{k}$ [130, с. 1-13].

Чтобы уменьшить размер признаков предыдущего слоя, используется слой maxpooling.

В качестве функции активации рассматривается ReLU. Она рассчитывается следующим образом:

$$\boldsymbol{f_{ReLU}(x_{m_k}^k) = \max(0, x_{m_k}^{k-1})}. \quad (4.4.12)$$

Затем результаты слоя свертки передаются на слой BatchNormalization (проводится нормализация, чтобы математическое ожидание было равно 0, а дисперсия равна 1). BatchNormalization позволяет ускорить процесс обучения [224, с. 448-456; 212]. Результирующие матрицы изображений сглаживаются.

Модели SP-CNN и SG-CNN состоят из четырех групп слоев свертки и maxpooling. Количество фильтров для первого слоя - 32, для 2-го и 3-го слоев - по 48, для четвертого слоя - 128. Размеры фильтров - 3x3, 3x3, 2x2 и 2x2. Размер слоя maxpooling принимается равным 2x2.

Последний слой моделей на основе CNN включает полносвязнный слой из 1000 нейронов. Затем признаки из TFDF-DNN ($\boldsymbol{F_{TDFD}}$), SP-CNN ($\boldsymbol{F_{SP}}$) и SG-CNN ($\boldsymbol{F_{SG}}$) объединяются для получения итогового вектора признаков, который выражается как

$$\boldsymbol{F = \{F_{TDFD}, F_{SP}, F_{SG}\}}. \quad (4.4.13)$$

Затем они подаются на полносвязный слой.

Прогностические вероятности для всех классов рассчитываются с использованием функции активации softmax:

$$\boldsymbol{p(\hat{x} = j|x) = \frac{e^{x_{m_k}^{k-1}(j)}}{\sum_{s=1}^{S} e^{x_{m_k}^{k-1}(s)}}}, \quad (4.4.14)$$

где $\boldsymbol{S}$ – количество классов.

Структура предлагаемой глубокой модели идентификации аварийного состояния оборудования показана на рис. 4.4.2.

Подробные структуры предложенных одномерных TDFD-DNN и двумерных DCNN показаны в таблицах 4.4.1 и 4.4.2, соответственно. Двумерные модели DCNN используют спектрограмму STFT и скалограмму CWT в качестве входных данных, а модель TDFD-DNN учитывает характеристики во временной и частотной областях.

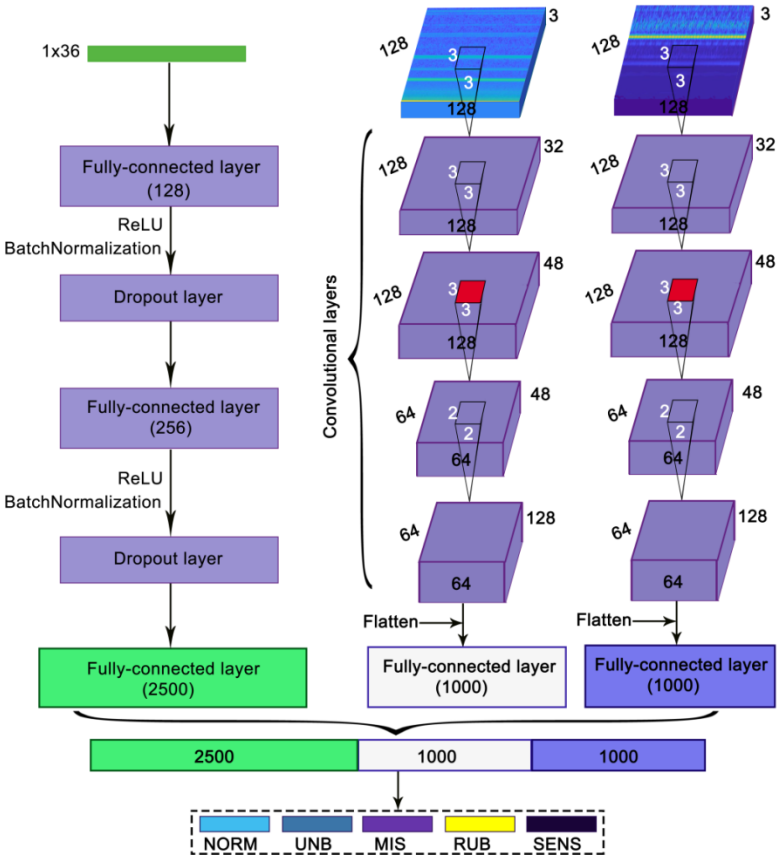


Рис. 4.4.2. Архитектура глубокой гибридной модели для обнаружения неисправностей оборудования [56]

Таблица 4.4.1

Структура предлагаемых двумерных CNN моделей

Слои	Тип	Размер	Ядро	Функция активации	Dropout
1	Input layer	128x128x3	3x3	ReLU	
2	Convolution layer	128x128x32	3x3	ReLU	
3	Convolution layer	128x128x48	3x3	ReLU	0.2
4	MaxPooling layer	2x2			
5	Convolution layer	64x64x48	2x2	ReLU	
6	Convolution layer	64x64x128	2x2	ReLU	0.2
7	Flatten layer	524,288			
8	Fully-connected layer	1000		ReLU	

Параметры предлагаемой модели настраиваются посредством перекрестной проверки. Слои Dropout добавлены для уменьшения переобучения. Эксперимент повторяют десять раз, чтобы уменьшить влияние различных факторов на результаты.

Таблица 4.4.2

Структура предлагаемой TFDF-DNN модели

Слои	Тип	Размер	Функция активации	Dropout
1	Input layer	1x36	ReLU	
2	Fully-connected layer	128	ReLU	0.2
3	Fully-connected layer	256	ReLU	0.2
4	Fully-connected layer	2500	ReLU	

Оптимизатор Adam используется для обновления весов сети (Таблица 4.4.3). Он прост в использовании и обучает глубокие нейронные модели на больших данных [274, с. 1-19]. Для всего процесса обучения оптимальная скорость определяется следующим образом:

$$\boldsymbol{\theta_t = \theta_{t-1} - \alpha \frac{\hat{\mu}_t}{\sqrt{\hat{v}_t}+\varepsilon}}, \tag{4.4.15}$$

$$\begin{matrix} \hat{\mu}_t = \frac{\mu_t}{1-\beta_1^t}, \\ \hat{v}_t = \frac{v_t}{1-\beta_2^t} \end{matrix} \tag{4.4.16}$$

$$\begin{matrix} \mu_t = \beta_1 \mu_{t-1} + (1-\beta_1) g_t \\ v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2, \end{matrix} \tag{4.4.17}$$

где $\boldsymbol{\alpha}$ - скорость обучения, $\boldsymbol{\mu_t}$ - вектор первого момента, $\boldsymbol{v_t}$- вектор второго момента, $\boldsymbol{\beta_1, \beta_2}$- факторы импульса, $\boldsymbol{g_t}$ - градиент для временного шага $\boldsymbol{t}$. Для расчета потерь используется кросс-энтропия [244, с. 1-15].

Таблица 4.4.3

Параметры обучения предлагаемой модели

Параметр	Значение
Оптимизатор	Adam
Размер батча	128
Начальная скорость обучения	0.001
Смещение веса	0.9
Максимальное число эпох	25

Описание набора данных. КФС составляют основу национальной критической инфраструктуры, в частности это добыча и поставки нефти и газа, нефте- и газопроводы. Одним из наиболее эффективных способов искусственного подъема нефти на поверхность является использование систем электрических центробежных погружных насосов (ЭПН). ЭПН представляют

собой сложные подсистемы, обеспечивающие подъем нефти и газа на поверхность на шельфе. Установка и возможная утилизация ЭПН из-за технического обслуживания являются дорогостоящими операциями. Система должна надежно работать после развертывания. Следует избегать удаления неисправного оборудования. Таким образом, важна тщательная оценка надежности. Кроме того, глубоководные работы делают практически невозможным мониторинг системы в режиме реального времени. Эта необходимость мотивирует тщательную проверку оборудования в специальной испытательной среде. Перед установкой система ЭПН тестируется в лабораторных условиях на больших наборах данных. Интеллектуальная диагностическая система помогает специалистам обнаруживать неисправности в оборудовании. Рассмотренный набор экспериментальных данных был получен в лабораториях производителей ЭПН, поставляющих Petrobras, крупнейшей нефтегазовой компании Бразилии [159, с. 67-72]. Исследуемый ЭПН состоит из шести компонентов: двух двигателей, двух протекторов и двух насосов [159, с. 67-72]. Акселерометры прикреплены к каждому компоненту в трех положениях: вверху, посередине и внизу. Они равномерно распределены по ЭПН. Таким образом, 36 (6×3×2) акселерометров соединены попарно с фазовым сдвигом 90 градусов в осевом направлении.

Вибрационные сигналы были получены от каждого акселерометра с частотой дискретизации 4096 Гц и временным интервалом 2.44141E-4 сек. Собранные данные анализируются и маркируются экспертом с использованием преобразования Фурье необработанного сигнала вибрации. Набор данных содержит 9690 помеченных образцов. В таблице 4.4.4 показано процентное соотношение пяти классов, содержащихся в наборе данных.

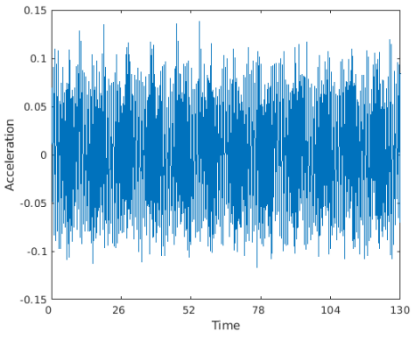
На рис. 4.4.3 и рис. 4.4.4 показаны типичные сигнатуры неисправностей во временной и частотной областях (включая неисправности сенсоров) [56].

Таблица 4.4.4

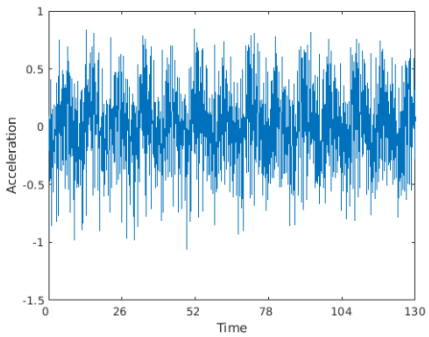
Описание экспериментального набора данных

Класс	Число образцов (%)
Normal	3774 (38.95)
Unbalance	2703 (27.89)
Accelerometer fault	1938 (20.00)
Misalignment	969 (10.00)
Rubbing	255 (2.63)

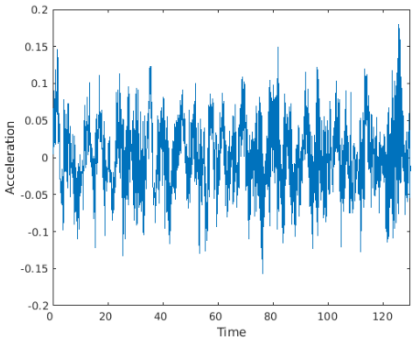
По 1-й и 2-й гармоникам частоты вращения вала [290, с. 843-848] можно определить разбалансировку лопасти насоса и перекос вала насоса.



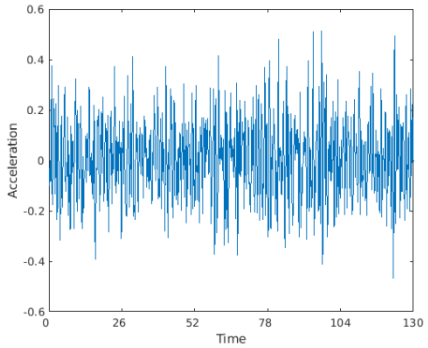
а)



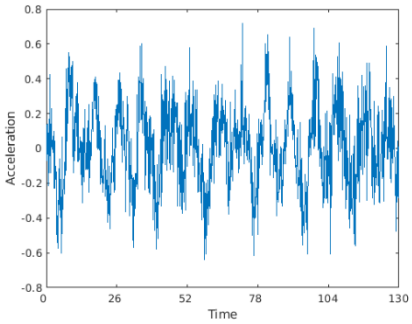
б)



в)



г)



д)

Рис. 4.4.3. Осциллограммы вибрации пяти рассматриваемых категорий

Наличие низкочастотного шума в вибросигнале характеризует наличие трения

[322, с. 121728-121741]. Неисправный сенсор акселерометра – одна из неисправностей [310, с. 415-422].

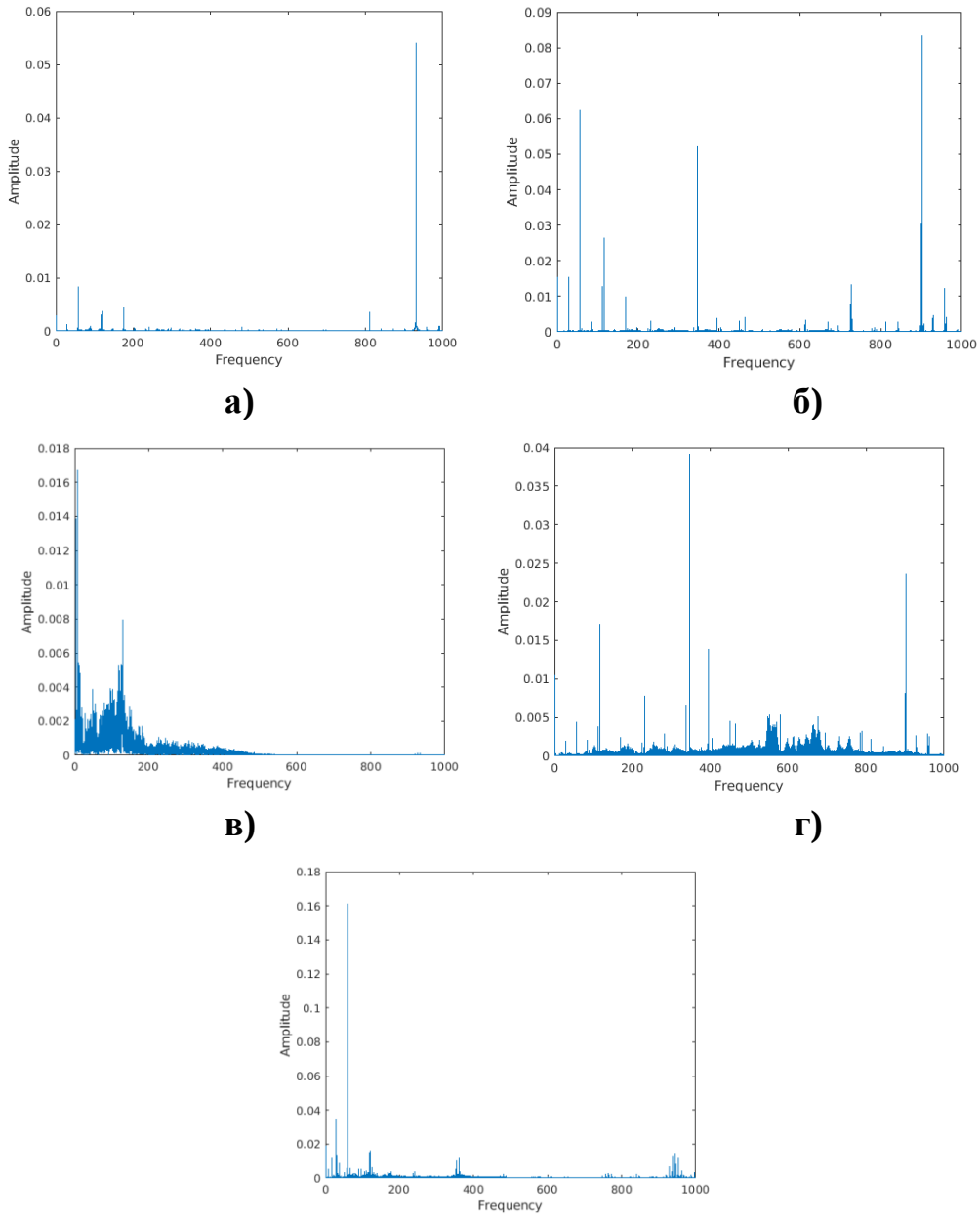


Рис. 4.4.4. Частотный спектр на основе STFT для пяти рассматриваемых классов [56]

Результаты экспериментов. Для объективной оценки работоспособности системы обнаружения отказов ЭПН набор данных разделен на обучающую выборку (80%) и тестовую выборку (20%). Таким образом, для обучения глубокой гибридной модели используется 7752 записи, а для тестирования

системы – 1938 записей. Эту процедуру выполняют двадцать раз. Модель должна различать классы с такими неисправностями, как дисбаланс, несоосность, трение, ошибки сенсоров нормальное состояние системы.

Предлагаемый подход был реализован в Python 2.7.13 с использованием различных библиотек, включая LibROSA, Tensorflow и Keras. Все эксперименты проводились на виртуальной машине Intel Xeon (R), процессор X5670 @ 2,93 ГГц * 4 с 10 ГБ оперативной памяти.

Как показано на рис. 4.4.5, гибридная модель превзошла другие модели. Полученные результаты представлены в таблице 4.4.5.

Таблица 4.4.5

Сравнение результатов классификации

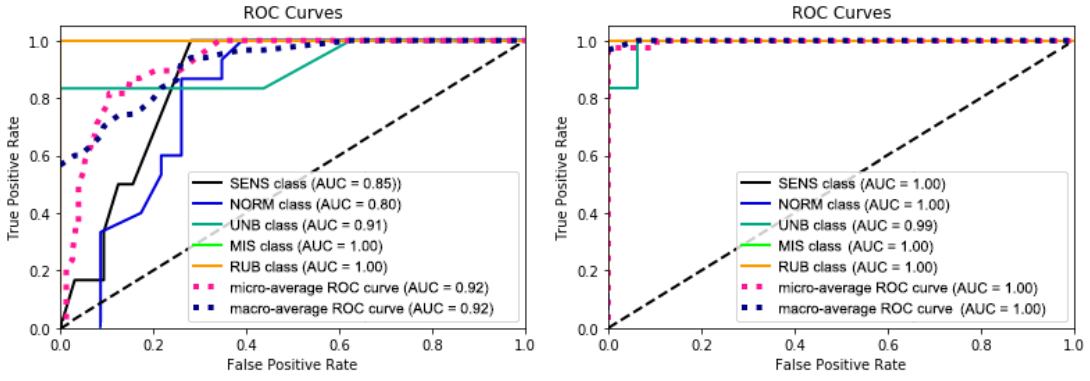
Метод	Метрики оценки	Классы				
		Normal	Unbalance	Misalignment	Rubbing	Faulty sensor
TD-DNN	Precision (%)	0.00 (0.00)	53.57 (2.67)	**100 (0.00)**	75.00 (3.04)	**100 (0.00)**
	Recall (%)	0.00 (0.00)	**100 (0.00)**	33.33 (3.45)	**100 (0.00)**	50.00 (2.14)
	F-measure (%)	0.00 (0.00)	69.77 (1.80)	50.00 (2.10)	85.71 (2.89)	66.67 (1.80)
TDFD-DNN	Precision (%)	**100 (0.00)**	93.75 (1.13)	**100 (0.00)**	**100 (0.00)**	**100 (0.00)**
	Recall (%)	**100 (0.00)**	**100 (0.00)**	83.33 (1.79)	**100 (0.00)**	**100 (0.00)**
	F-measure (%)	**100 (0.00)**	96.77 (0.68)	90.91 (1.01)	**100 (0.00)**	**100 (0.00)**
FD-DNN	Precision (%)	0 (0.00)	68.18 (2.36)	62.50 (2.98)	0 (0.00)	**100 (0.00)**
	Recall (%)	0 (0.00)	**100 (0.00)**	83.33 (1.61)	0 (0.00)	**100 (0.00)**
	F-measure (%)	0 (0.00)	81.08 (3.07)	71.43 (3.55)	0 (0.00)	**100 (0.00)**
SP-CNN	Precision (%)	66.67 (2.73)	58.33 (2.78)	**100 (0.00)**	50.00 (4.35)	**100 (0.00)**
	Recall (%)	33.33 (3.45)	93.33 (0.64)	33.33 (3.45)	33.33 (3.45)	87.50 (1.75)
	F-measure (%)	44.44 (5.02)	71.79 (2.52)	50.00 (3.63)	40.00 (5.87)	93.33 (1.15)
SG-CNN	Precision (%)	**100 (0.00)**	**100 (0.00)**	**100 (0.00)**	75.00 (3.06)	**100 (0.00)**
	Recall (%)	**100 (0.00)**	**100 (0.00)**	83.33 (1.80)	**100 (0.00)**	**100 (0.00)**
	F-measure (%)	**100 (0.00)**	**100 (0.00)**	90.91 (0.98)	85.71 (2.75)	**100 (0.00)**
Предложенный подход	Precision (%)	**100 (0.00)**	99.98 (0.39)	**100 (0.00)**	99.95 (0.40)	**100 (0.00)**
	Recall (%)	**100 (0.00)**	**100 (0.00)**	**99.93 (0.24)**	**100 (0.00)**	**100 (0.00)**
	F-measure (%)	**100 (0.00)**	**100 (0.00)**	**99.96 (0.23)**	**100 (0.00)**	**100 (0.00)**

Видно, что предложенная гибридная модель превосходит отдельные модели классификации (TD-DNN, FD-DNN, TDFD-DNN, SP-CNN и SG-CNN) и достигает среднего уровня точности ~99.98%. Жирный шрифт показывает

наилучшую производительность.

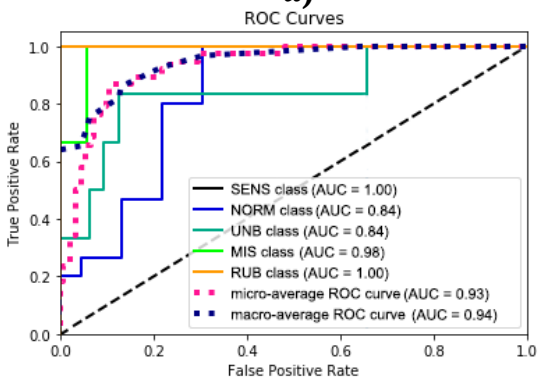
Устойчивость предложенного подхода также оценивалась на основе показателей precision, recall и F-мера. Они показали низкие стандартные отклонения (менее 0.50) и значительное улучшение по сравнению с другими методами. Второй лучший результат показал метод SG-CNN. Предложенный подход сравнивается с методами KNN, LP, SVM и RF.

Анализ графического представления ROC-кривой позволяет оценить качество предложенной глубокой гибридной модели. Экспериментальные результаты показывают, что это лучшая модель для выявления неисправностей ЭПН по рассмотренным признакам, полученным из сигналов вибрации (площадь под ROC-кривой) составила 100% (Рис. 4.4.5). Согласно показателям точности и отзыва, KNN довольно точно распознал класс MIS. SVM показала 100% результаты по MIS, RUB, SENS и NORM классам ЭПН. Класс UNB был точно классифицирован методом RF. Метод LR показывает низкие результаты для классов NORM ЭПН и DIS по сравнению с другими методами.

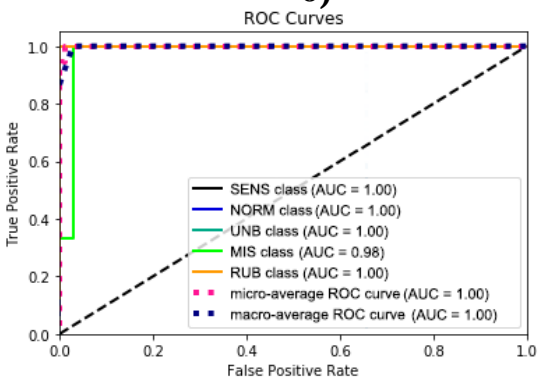


а)

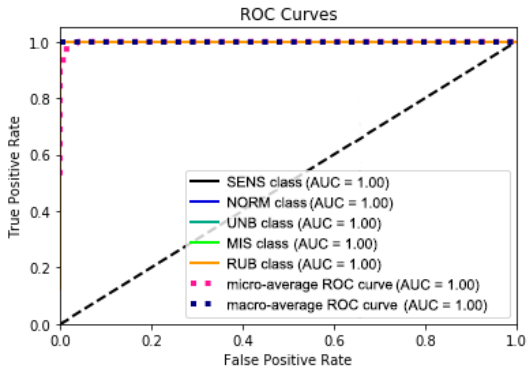
б)



в)



г)



д)

Рис. 4.4.5. ROC-кривые оценки точности классификации рассматриваемых методов для пяти классов: faulty sensor (SENS), normal operational condition (NORM), unbalance (UNB), misalignment (MIS) и rubbing (RUB) [56]

Было проведено сравнение предлагаемого подхода с методами KNN, KNN+, KNN+FS, KNN+EFS [158, с. 1807-1812] и методом, предложенным Oliveira-Santos и соавт. [309, с. 381-396]. В таблице 4.4.6 показаны многообещающие результаты использования глубокой гибридной модели для автоматической диагностики отказов ЭПН. Эксперименты на относительно небольшом наборе данных полученных от акселерометров ЭПН показали достаточно высокую точность предложенного метода, однако обучение модели потребовало времени. Тем не менее, проблему с вычислительной сложностью можно игнорировать, поскольку в реальных КФС системах используется предварительно обученная модель.

Таблица 4.4.6

Сравнение эффективности предлагаемого подхода с другими методами на основе метрики recall

Метод	Классы				
	NORM (%)	UNB (%)	MIS (%)	RUB (%)	SENS (%)
Ансамблевый подход [309]	97.82	97.51	76.00	48.29	95.65
KNN [158]	97.50	89.50	60.00	33.50	87.00
KNN+ [158]	97.50	89.50	68.00	38.00	89.00
KNN+FS [158]	97.50	90.50	71.00	35.50	89.50
KNN+EFS [158]	98.00	93.00	81.00	34.50	91.00
Предложенный подход	**100**	**100**	**99.93**	**100**	**100**

Можно сделать выводы, что в исследовании был представлен усовершенствованный метод обнаружения отказов ОТ КФС. Разработана гибридная модель на основе простой DNN и двух CNN для получения важной информации из признаков вибрационных сигналов и повышения точности классификации отказов. Экспериментальные результаты показали, что применение предложенной модели для обнаружения неисправностей ЭПН, таких как дисбаланс лопастей насоса, несоосность вала и механическое трение, включая обнаружение неисправностей сенсоров, может дать высокие результаты. Глубокая гибридная модель достигла лучших результатов, чем методы KNN, SVM, LR и RF.

Преимущества предложенного подхода состоят в том, что он может быть автоматизирован, применим для анализа больших данных и использован для создания интеллектуальной экспертной системы. Согласно результатам, можно сделать вывод, что глубокая гибридная модель имеет большой потенциал, чтобы стать эффективным и действенным инструментом для диагностики и прогнозирования отказов ОТ КФС и может стать многообещающей альтернативой для интеллектуальных систем технического обслуживания в будущем.

Выводы к главе IV

- Разработан метод обнаружения аномалий в ОТ КФС на основе иерархических скрытых Марковских моделей;
- Разработан метод обнаружения аномалий в акустических сигналах КФС с применением трансферного обучения;
- Предложен метод обнаружения атак на устройства КФС с применением глубокой гибридной модели;
- Разработан метод классификации отказов КФС на основе изображений.

ГЛАВА V. РАЗРАБОТКА МЕТОДОВ ОЦЕНКИ КИБЕРУСТОЙЧИВОСТИ КИБЕРФИЗИЧЕСКИХ СИСТЕМ

Наиболее важной задачей является сохранение способности КФС корректно функционировать даже в условиях деструктивных информационных воздействий, поскольку успешное осуществление кибератак на такие системы может привести к негативным финансовым последствиям и экологическим катастрофам, а также привести к гибели людей [203, с. 269-278]. Поскольку злоумышленники часто ищут и нацеливаются на самое слабое звено – наиболее уязвимый функциональный компонент в КФС – самые уязвимые устройства могут стать хорошей отправной точкой для эффективного исследования киберустойчивости системы. Существующие методы не могут всесторонне анализировать распространенные многоэтапные кибератаки в среде КФС.

В этой главе будут разработаны метод определения критичности уязвимостей функциональных компонентов КФС и метод оценки рисков КФС.

5.1. Метод определения критичности уязвимостей функциональных компонентов киберфизических систем на основе Байесовского графа атак

Интеграция интеллектуальных устройств и технологий в различных отраслях существенно изменила всю технологическую инфраструктуру. КФС распознает окружающую среду с помощью сенсоров, принимает решения в соответствии со своим назначением и обеспечивает корректирующие действия с помощью исполнительных механизмов [184, с. 1454-1467]. Связь между сетями ИТ (информационные технологии) и ОТ (операционные технологии) важна при моделировании КФС систем. Конвергенция ИТ и ОТ растет, что позволяет более эффективно управлять и эксплуатировать КФС. Серьезность отказа компонента КФС влияет на систему в зависимости от количества компонентов и их разнообразия. Это доказывает необходимость измерения критичности уязвимостей компонентов системы.

Своевременное выявление критических недостатков безопасности в КФС позволяет обнаруживать риски и потенциальные угрозы. Для решения этой проблемы создаются модели угроз, позволяющие лучше понять потенциальные уязвимости, которые необходимо учитывать для обеспечения надежности системы. Выбор оптимального решения для оценки критичности функциональных уязвимостей компонентов КФС — сложный процесс, поскольку все уязвимости должны быть идентифицированы, классифицированы и количественно оценены в соответствии с единым подходом в рамках процесса киберустойчивости.

Применение информационных технологий к КФС и их доступность в Интернете делает их привлекательной мишенью для разрушительных кибератак злоумышленников [315, с. 2715-2729]. Статистика за последние несколько лет показывает рост количества кибератак на КФС, при этом в большинстве случаев целью злоумышленников является получение контроля над подсистемой управления.

В настоящее время предложено множество подходов для анализа кибербезопасности КФС, однако корреляции между устройствами и уязвимостями уделяется мало внимания, особенно при выявлении уязвимостей нулевого дня [72, с. 250-260; 81, с. 1-16; 181; 140, с. 41-53; 404, с. 1-5]. Существующие методы не позволяют комплексно анализировать распространенные многоэтапные кибератаки в среде КФС.

Однако решение этой проблемы возможно с использованием байесовского графа атак (БГА), который позволяет оценить критичность уязвимостей компонентов КФС. Он предоставляет информацию о связях между уязвимостями, что является важным фактором. БГА зависит от операционных систем и сетевых протоколов, правил контроля доступа КФС, идентификации программного обеспечения и информации об уязвимостях [287, с. 273–294].

Уровень критичности уязвимостей оценивается для того, чтобы технические специалисты могли принять обоснованное решение об устранении

уязвимостей, выявленных в программном и аппаратном обеспечении по результатам анализа уязвимостей в КФС.

Методы многокритериального принятия решений могут быть использованы в качестве подходящего подхода для решения задачи ранжирования критичности уязвимостей компонентов КФС. Они помогают лицам, принимающим решения, выбирать оптимальные альтернативы на основе заданных критериев. Для этой цели используются TOPSIS [189, с. 1-16], PROMETHEE [26, с. 271-285], VIKOR [311, с. 482-500], ELECTRE [345, с. 1-8] и другие.

PROMETHEE II считается довольно простым среди методов многокритериального анализа и все чаще используется лицами, принимающими решения. Он дает довольно стабильные результаты. Метод основан на попарном сравнении альтернатив, соответствующих каждому критерию [291, с. 47-55].

Целью данного исследования является оценка критичности уязвимостей компонентов КФС на основе БГА [50, с. 1-7; 13, с. 52-54]. Уязвимость количественно оценивается с использованием Общей системы оценки уязвимостей (Common Vulnerability Scoring System, CVSS) и оценок Национальной базы данных уязвимостей (National Vulnerability Database, NVD) [172, с. 1230-1237]. Платформа CVSS сочетается с БГА для обеспечения комплексной оценки критичности уязвимостей. Комбинация сложности эксплойта и оценки воздействия принимается в качестве весов ребер графа. Для определения критичности уязвимостей компонентов КФС используется многокритериальный метод принятия решений PROMETHEE II [369, с. 5301-5308]. Он позволяет ранжировать и выявить наиболее критичные уязвимости рассматриваемой системы.

Основной вклад исследования включает следующее:

1) Представлена модель угроз на примере КФС для транспортировки природного газа по трубопроводу.

2) Создание БГА на основе рассматриваемой структуры КФС с учетом известных уязвимостей ее компонентов.

3) Рассматриваются три сценария кибератак с использованием в качестве примера транспортировку природного газа по трубопроводу.

4) Алгоритм Promethee II применяется для ранжирования и выявления наиболее критических уязвимостей рассматриваемой системы.

5) Сравнение с известными многокритериальными методами принятия решений, такими как TOPSIS, VIKOR и ELECTRE, доказывает эффективность предлагаемого подхода.

Предлагаемый подход. Предлагаемый подход основан на генерации БГА с использованием предопределенной модели угроз КФС. В этом случае узлы БГА представляют собой различные состояния, которых может достичь злоумышленник. Атакующий может осуществить внешнюю атаку на КФС, используя различные сценарии кибератак.

Пусть БГА состоит из множества заданных узлов (устройств) $\{X_1, X_2, \dots, X_n\}$. Ребро (X_i, X_j) существует тогда и только тогда, когда уязвимость позволяет противнику проводить кибератаку из X_i в X_j. Информация на каждом узле X_i $(i = 1,2, \dots, n)$ графа - это $p(X_i|pa_i)$, где pa_i вероятность того, что узел будет скомпрометирован. Оценка вероятности p_{v_i} того, что злоумышленник успешно воспользуется уязвимостью v_i основана на CVSS [163, с. 1-38]. Вероятность того, что противник скомпрометирует узел графа X_i - $p(X_i = T) = pr$, где $pr \in [0,1]$ [133, с. 507-554; 367, с. 507-554].

Входные данные алгоритма генерации БГА включают информацию о топологии сети, компонентах КФС и уязвимостях. Θ содержит информацию о связи компонентов КФС и конфигурации сети для всей критической системы. Он включает в себя различные пути атаки, связанные с разными уязвимостями, а также оценку CVSS, помогающую оценить степень реплизации атаки и ее серьезность. CVSS позволяет получить основной атрибут уязвимости и числовой балл, отражающий ее важность [365, с. 863-878]. Чем выше значение,

тем больше вероятность осуществления кибератаки. Оценка может отражать низкую (0.1–3.9), среднюю (4.0–6.9), высокую (7.0–8.9) или срочную (9.0–10.0) важность [304, с. 3036-3047; 284].

NVD предоставляет CVSS оценки для всех известных уязвимостей [306]. Для CVSS 3.1 метрика возможности использования состоит из четырех основных показателей: вектор атаки *AV*, сложность атаки *AC*, взаимодействие с пользователем *UI* и требуемые привилегии *PR*. Аналогично, метрика воздействия состоит из трех основных показателей: конфиденциальность *IC*, целостность *II* и доступность *IA*. Они рассчитываются по формуле:

$$E_i = 8.22 \times AV_i \times AC_i \times UI_i \times PR_i, \quad (5.1.1)$$

$$I_i = 1 - ((1 - IC_i) \times (1 - II_i) \times (1 - IA_i)). \quad (5.1.2)$$

CVSS вычисляет базовую оценку (base score, BS) уязвимости i как

$$BS_i = \begin{cases} 0, & \text{если } I_i \le 0 \\ rndTo1Dec(\min[(I_i + E_i),10]), & \text{если } Scope \text{ неизменна} \\ rndTo1Dec(\min[1.08 \times (I_i + E_i),10]), & \text{если } Scope \text{ меняется} \end{cases}. \quad (5.1.3)$$

Здесь под Scope (область действия), согласно CVSS, понимается способность уязвимости в одном компоненте ПО влиять на ресурсы, выходящие за рамки его возможностей или привилегий, а $rndTo1Dec$ – операция округления до одного десятичного знака.

Оценка сложности эксплойта по CVSS строится для каждого ребра на пути кибератаки в графе. При этом узлы графа являются компонентами КФС, а ребра – уязвимостями, с помощью которых злоумышленник достигает своей цели. Оценка сложности эксплойта рассчитывается следующим образом [287, с. 273–294]:

$$\xi = 10 - \frac{\varepsilon}{AV}, \quad (5.1.4)$$

где ε - это оценка возможности использования *AV*. Чем меньше ξ, тем легче злоумышленнику скомпрометировать этот узел КФС. Подробности настройки параметров предлагаемой модели угроз приведены в таблице 5.1.1.

Таблица 5.1.1

Значения входных параметров

Параметр	Веса	Значение параметра	CVSS оценка
BS	0.25	Network	0.85
		Local	0.62
γ	0.1		1.0 – 6.0
ε	0.15		1.5 – 4.0
AV	0.1	Network	0.85
		Local	0.55
UI	0.1	None	0.85
		Required	0.62
PR	0.1	None	0.85
		Low	0.62
ξ	0.2		5.29 - 7.27

Узел злоумышленника – это компонент КФС, который указывает на злоумышленника. Алгоритм генерации БГА принимает данный узел в качестве начального узла и генерирует граф Θ.

Для оценки критичности уязвимостей компонентов КФС на основе графа уязвимостей применяется алгоритм PROMETHEE II. γ, BS, ε, AV, ξ, UI, PR выбраны как наиболее информативные показатели для выявления критичности уязвимостей. Положительный поток выигрыша $\phi^{+}(a)$ и отрицательный поток выигрыша $\phi^{-}(a)$ в PROMETHEE II рассчитываются для каждого столбца таблицы уязвимостей компонентов КФС. Критичность уязвимостей компонента КФС ределяется на основе чистого потока ранжирования для всех узлов графа (Алгоритм).

Алгоритм. Оценка критичности уязвимостей функциональных компонентов КФС

Вход: Первоначальный БГА $\Theta = (X, E, W)$

Метрики уязвимостей: γ, BS, ε, AV, ξ, UI, PR

Выход: Значение риска

Шаг 1: Расчет показателей критичности уязвимостей $Z = (BS, \gamma, \varepsilon, AV, \xi, UI, PR)$ для графа Θ.

Шаг 2: Нормализация значений.

Шаг 3: Расчет положительного потока предпочтения $\phi^{+}(a)$ и отрицательного потока предпочтения $\phi^{-}(a)$ для каждого столбца таблицы уязвимостей компонентов КФС:

$$\phi^{+}(a) = \frac{1}{m-1}\sum_{x\in G}\pi(a,x),$$

$$\phi^{-}(a) = \frac{1}{m-1}\sum_{x\in \mathrm{G}}\pi(x,a).$$

Шаг 4: На основании результатов, полученных на шаге 3, рассчитывается поток предпочтения для всех узлов графа Θ:

$$\phi(a) = \phi^{+}(a) - \phi^{-}(a).$$

Шаг 5: Ранжирование критичности уязвимостей компонентов КФС согласно значениям $\phi(a)$.

Графы атак. Графы атак используются для анализа различных уязвимостей КФС. Они обнаруживают потенциальные угрозы для наиболее важных устройств, включая информацию о злоумышленнике, уязвимостях, хостах и т. д. Они показывают, как уязвимости в сети могут быть использованы последовательно или параллельно, показывая различные пути, по которым злоумышленник может добраться до своей целевой жертвы. Графы атак могут быть основаны на состояниях или логических значениях [294, с. 231-244]. В таком графе на основе состояний каждый узел представляет уникальное состояние безопасности. Однако ограничение связано с масштабируемостью. Логические графы атак представляют собой зависимости между кибератаками и условиями безопасности, такими как «И» и «ИЛИ».

БГА – это граф $\Theta = (X, E, W)$, где X – множество узлов, $E \subseteq N \times N$ – множество ребер графа Θ, W – матрица весов графа, где $|N| = n$, а каждая упорядоченная пара (X_i, X_j) из E называется ребром [304, с. 3036-3047]. $W: E \longrightarrow R$ - это функция, которая присваивает значение каждому ребру. Вес W_{ij} на каждом ребре обозначает оценку возможности эксплуатации узла j. Взвешенная матрица смежности, связанная с взвешенным графом Θ,

определяется как квадратная матрица A такого размера $n \times n$, что $[A]_{ij} = W_{ij} > 0$ если $(X_i, X_j) \in E$ и $[A]_{ij} = 0$ во всех других случаях.

Байесовская сеть, примененная к графу атак, преобразует его в БГА. В сети узлы соединены друг с другом через направленные ребра в направленном ациклическом графе [162, с. 75-104; 75, с. 505-541]. Рассмотрен принцип монотонности графа атак, когда злоумышленник, повысивший привилегию, никогда от нее не отказывается.

Таблица условных вероятностей (conditional probability table, CPT) БГА определяется как комбинированный эффект сетевой уязвимости (Frigault et al., 2008). Локальные CPT узлов с логическими условиями «И» и «ИЛИ» рассчитываются в [294, с. 231-244]. Логическое «И» означает, что всех разрешений достаточно для компрометации узла X_i. Тогда как, логическое «ИЛИ» означает, что любого разрешения достаточно для компрометации узла. Непосредственными родительскими узлами X_i являются pa_i, а вероятность того, что уязвимость v_i будет использована, равна p_{v_i}.

$$p(X_i|pa_i) = \begin{cases} 0, & \exists X_j \in pa_i \mid X_j = 0 \\ \prod_{j:X_j} p_{v_j}, & \text{в противном случае} \end{cases}, \quad (5.1.5)$$

$$p(X_i|pa_i) = \begin{cases} 0, & \exists X_j \in pa_i \mid X_j = 0 \\ 1 - \prod_{j:X_j}(1 - p_{v_j}), & \text{в противном случае} \end{cases}. \quad (5.1.6)$$

Оценка критичности уязвимостей компонентов КФС на основе PROMETHEE II. Метод PROMETHEE был предложен Brans и Vincke [111, с. 647-656], который позволяет оценивать набор конечных альтернатив, а затем выбирает наилучшую альтернативу в соответствии с набором противоречивых критериев оценки. PROMETHEE II — самый популярный метод, поскольку он устанавливает полный рейтинг альтернатив [235, с. 717-730; 369, с. 5301-5308]. Он включает в себя следующие шаги.

На первом этапе d_j определяется:

$$d_j(a, b) = g_j(a) - g_j(b), \quad (5.1.7)$$

где $d_j(a,b)$ представляет собой разницу между оценками альтернатив a и b каждого фактора.

На следующем шаге функция предпочтения оценивается как степень предпочтения альтернативы a альтернативе b по j-му критерию в зависимости от $d_j(a,b)$ [305, с. 77-85]:

$$P_j(a,b) = \Psi_j(d_j(a,b)). \quad (5.1.8)$$

Затем оценивается глобальный индекс предпочтений:

$$\pi(a,b) = \sum_{j=1}^{m} P_j(a,b) * W_j, \ \forall a,b \in G, \quad (5.1.9)$$

где $\pi(a,b)$ - взвешенная сумма $P(a,b)$ для каждого критерия C_j, W_j - вес j-го критерия, и G - набор доступных альтернатив.

Далее рассчитываются потоки предпочтения, которые используются для измерения производительности каждой альтернативы. Чистый поток $\phi(a)$ представляет собой разницу между положительным исходящим потоком $\phi^+(a)$ и отрицательным входящим потоком $\phi^-(a)$:

$$\phi(a) = \phi^+(a) - \phi^-(a), \quad (5.1.10)$$

где

$$\phi^+(a) = \frac{1}{m-1}\sum_{x\in G} \pi(a,x), \quad (5.1.11)$$

$$\phi^-(a) = \frac{1}{m-1}\sum_{x\in G} \pi(x,a). \quad (5.1.12)$$

Большие значения $\phi^+(a)$ указывают на явное предпочтение одной альтернативы другим альтернативам.

Кроме того, если $\phi(a) > \phi(b)$, тогда альтернатива a лучше альтернативы b.

Модель угроз. Для иллюстрации потенциальных сценариев кибератак, использующих уязвимости в компонентах КФС, можно рассмотреть пример транспортировки природного газа по трубопроводу как одного из критически важных секторов инфраструктуры.

Природный газ транспортируется по трубопроводам компрессорными станциями. Они состоят из нескольких газовых компрессоров, которые перемещают газ по трубопроводу [190].

После фильтрационной сепарации природный газ подается в компрессорную установку. Системы смазки и охлаждения двигателя компрессора служат для защиты и поддержания скорости потока газа. ОТ рассматриваемой КФС представляют собой программируемые логические контроллеры (programmable logic controllers, PLC) и удаленные терминальные устройства (remote terminal units, RTU), подключающиеся к сенсорам и исполнительным механизмам. Сенсоры передают сигналы на контроллеры, которые отправляют управляющие сигналы на исполнительные механизмы.

Операторы процесса взаимодействуют с системой управления через человеко-машинный интерфейс (human-machine interface, HMI). Архив данных хранит значения параметров процесса. Сервер данных получает информацию от контроллера и передает ее другим контроллерам или HMI. Сеть управления передает инструкции и данные между блоками управления и измерения и устройствами SCADA.

Инженерные рабочие станции могут быть подключены к корпоративной сети или Интернету. Здесь устанавливаются средства разработки программного обеспечения, с помощью которых технический специалист может централизованно вносить изменения и дополнения в конфигурацию системы. Компрессорные станции при транспортировке природного газа обычно контролируются с помощью распределенной системы управления, возможно, сопряженного с отдельной системой безопасности. Станции управляются удаленно через RTU, подключенные к центральной системе SCADA WAN.

После того, как устройства ОТ КФС скомпрометированы, злоумышленники могут использовать различные CVSS для запуска атак на процессы управления, которые влияют на физические операции. Информация об уязвимостях взята из описаний NVD [306]. Злоумышленник использует

подключение ИТ-сети КФС к сети Интернет для выполнения вредоносного кода (Рис. 5.1.1).

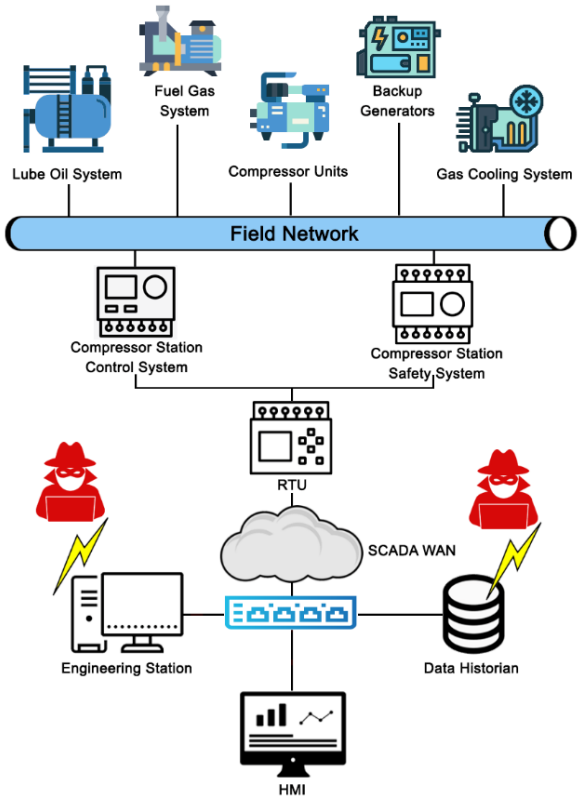


Рис. 5.1.1. Пример конфигурации КФС

Уязвимое программное обеспечение позволяет злоумышленнику повысить привилегии. Это возможно посредством подслушивания, повторного воспроизведения или атак с внедрением ложных пакетов в ИТ.

Сценарии кибератак. На рис. 5.1.2 представлен построенный БГА для рассматриваемой КФС. Предполагается, что целью злоумышленника может быть компрометация архива данных или рабочей станции EWS, которые обычно являются основными целями из-за их взаимосвязанного характера.

Рассмотрено несколько сценариев действий злоумышленника, воспользовавшегося уязвимостями функциональных компонентов КФС:

• Манипулирование элементами управления

В зависимости от конфигурации системы злоумышленник, способный взаимодействовать с WAN SCADA, может использовать CVE-2020-13500 или

CVE-2022-29966 для манипулирования системными настройками, файлами или критическими значениями, связанными с компрессорными станциями.

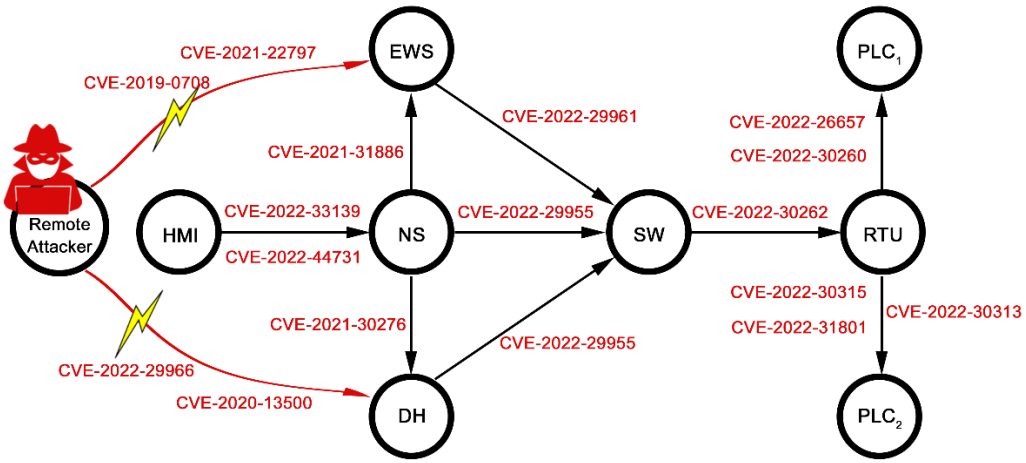


Рис. 5.1.2. Байесовский граф атак [46]

Он может использовать CVE-2022-33139 для обхода аутентификации, а затем манипулировать значениями мониторинга, связанными с компрессорными станциями, чтобы перегружать операторов ложными сигналами тревоги или изменять заданные значения расхода, чтобы нарушить транспортировку природного газа, тем самым расширяя возможности манипуляции.

• Отказ в управлении

Злоумышленник, имеющий возможность взаимодействовать с RTU, может использовать уязвимости CVE-2022-29961 или CVE-2022-29955 для обхода аутентификации в RTU и ввода команд, которые останавливают операцию, а, воспользовавшись уязвимостью CVE-2022-30262, он может лишить операторов возможности управлять и контролировать работу компрессорных станций PLC1 и PLC2.

Альтернативно, злоумышленник, способный взаимодействовать с сетевым коммутатором (NS), может воспользоваться неаутентифицированным соединением (CVE-2021-30276 или CVE-2021-31886) и отправить определенные команды на RTU, что может приостановить работу или

нарушить связь между SCADA WAN и RTU, лишая операторов возможности мониторинга и управления компрессорными станциями.

• Потеря контроля

Для этого злоумышленник взламывает систему управления самой компрессорной станции, используя уязвимости CVE-2022-30262, CVE-2022-26657, CVE-30315, CVE-2022-30260 или CVE-2022-31801. Это позволяет добиться выполнения кода на RTU и без ограничений общаться через различные сетевые интерфейсы. Использование уязвимостей CVE-2022-30313 и CVE-2022-30315 позволяет проникнуть в сеть безопасности. Тем самым злоумышленник производит манипулятивные настройки или даже добивается выполнения кода на PLC2 для отключения систем аварийного отключения и систем пожарной и газовой безопасности.

Первоначальная цель злоумышленника в КФС системах – узлы DH и EWS для возможных путей атаки, а конечная цель злоумышленника – скомпрометировать узлы PLC1 и PLC2 через RTU. В этом случае совместная вероятность для всех узлов БГА будет равна:

$$p(DH, EWS, HMI, NS, SW, RTU, PLC_1, PLC_2) = p(DH)p(EWS)p(NS|HMI) \times p(NS|EWS)p(NS|DH)p(SW|EWS)p(SW|NS) \times p(SW|DH)p(PLC_1|RTU)p(PLC_2|RTU). \quad (5.1.13)$$

Общая схема предлагаемого подхода представлена на рис. 5.1.3. Предлагаемый подход позволяет оценить критичность уязвимостей компонентов системы и может помочь экспертам обеспечить меры безопасности для предотвращения атак или защиты сетевых ресурсов КФС.

Результаты экспериментов. Оценка критичности уязвимостей компонентов КФС проводится на Python 3.7.12 с использованием различных библиотек, в том числе EasyMCDM.

Все эксперименты проводятся на процессоре Intel Xeon(R), процессор X5670 @ 2.93 ГГц * 24 с 24 ГБ оперативной памяти.

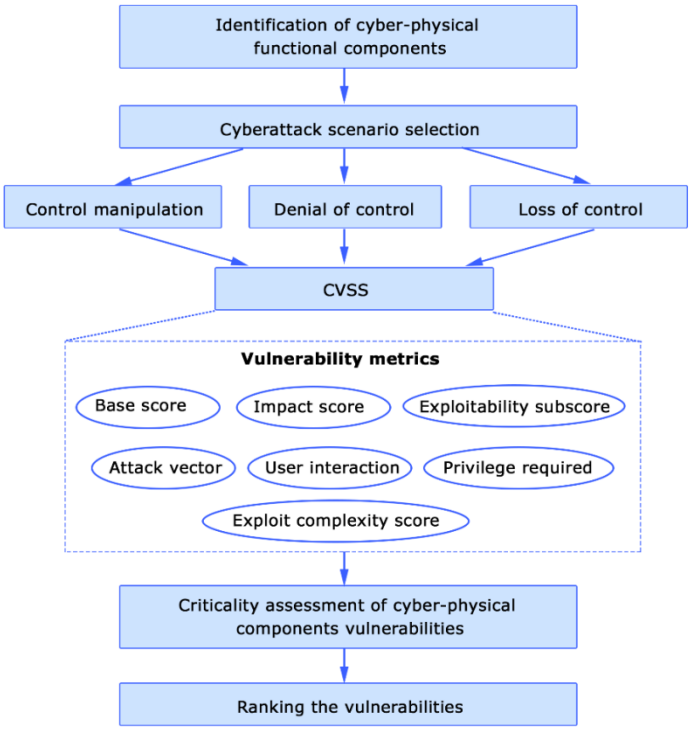


Рис. 5.1.3. Общая схема предлагаемого подхода

Хорошим примером рассматриваемой проблемы является ранжирование рисков уязвимостей компонентов КФС для безопасной и надежной транспортировки природного газа по трубопроводу.

БГА был создан с использованием MulVal для предложенной модели КФС с учетом оценок CVSS, описанных ранее.

Информация о рассматриваемых уязвимостях, включая базовую оценку, оценку воздействия γ, подоценку возможности использования ε, вектор атаки, взаимодействие с пользователем, требуемые привилегии и оценку сложности эксплойта ξ, представлена в таблице 5.1.2.

Как видно из таблицы, в компонентах КФС рассмотрено шестнадцать уязвимостей. Это позволяет визуально исследовать различные уязвимости соответствующих компонентов топологии сети КФС.

Веса на ребрах БГА, полученные на основе топологии сети КФС, указывают на оценку сложности эксплойта и оценку воздействия, связанные с соответствующей функциональной уязвимостью компонента.

Таблица 5.1.2

Информация об уязвимостях

Компонент КФС	Тип	Уязвимость CVE-ID	BS	γ	ε	AV	UI	PR	ξ
PLC_1	Controller	CVE-2022-29957	7.8	5.9	1.8	0.55	0.85	0.62	6.73
		CVE-2022-30260	7.8	5.9	1.8	0.55	0.62	0.85	6.73
PLC_2	Controller	CVE-2022-31801	9.8	5.9	3.9	0.85	0.85	0.85	5.41
		CVE-2022-30313	7.5	3.6	3.9	0.85	0.85	0.85	5.41
		CVE-2022-30315	9.8	5.9	3.9	0.85	0.85	0.85	5.41
RTU	Controller	CVE-2022-30262	7.8	5.9	1.8	0.55	0.85	0.62	6.73
SW	SCADA WAN	CVE-2022-22961	5.3	1.4	3.9	0.85	0.85	0.85	5.41
		CVE-2022-22955	9.8	5.9	3.9	0.85	0.85	0.85	5.41
DH	Data Historian	CVE-2022-29966	8.8	5.9	2.8	0.85	0.85	0.85	6.71
		CVE-2020-13500	9.8	5.9	3.9	0.85	0.85	0.85	5.41
NS	Network Switch	CVE-2021-31886	9.8	5.9	3.9	0.85	0.85	0.85	5.41
		CVE-2021-30276	7.8	5.9	1.8	0.55	0.85	0.62	6.73
HMI	HMI	CVE-2022-33139	9.8	5.9	3.9	0.85	0.85	0.85	5.41
		CVE-2022-44731	5.4	2.5	2.8	0.85	0.85	0.62	6.71
EWS	EWS	CVE-2019-0708	9.8	5.9	3.9	0.85	0.85	0.85	5.41
		CVE-2021-22797	7.8	5.9	1.8	0.55	0.62	0.85	6.73

Например, в RTU есть уязвимость CVE-2022-30262, которой злоумышленник может воспользоваться и получить доступ, тогда ξ и γ примут значения 6.73 и 5.9, соответственно. А в HMI имеется уязвимость CVE-2022-33139. При этом вес ребра графа равен (5.41, 5.9).

Алгоритм PROMETHEE II применяется для оценки критичности уязвимостей компонентов КФС для БГА, рассмотренного на рис. 5.1.2. Для этого рассчитываются значения $\pi(a, b)$, а также положительный исходящий поток $\phi^+(a)$ и отрицательный входящий поток $\phi^-(a)$ для каждой альтернативы (Таблица 5.1.3).

Чистый поток $\phi(a)$ вычисляется для каждого парного сравнения путем вычитания отрицательного входящего потока из положительного исходящего потока.

В таблице 5.1.3 представлены экспериментальные результаты использования этого алгоритма. Максимальные значения коэффициентов выделены жирным шрифтом.

Таблица 5.1.3

Оценка критичности уязвимостей компонентов КФС на основе PROMETHEE II

	PLC_1	PLC_2	RTU	SW	DH	NS	HMI	EWS	$\phi^+(a)$	$\phi(a)$	Ранг
PLC_1	0.000	0.082	0.082	0.165	0.000	0.082	0.165	0.165	0.106	-0.195	4
PLC_2	0.100	0.000	0.000	0.165	0.000	0.000	0.165	0.165	0.085	-0.175	3
RTU	0.100	0.000	0.000	0.165	0.000	0.000	0.165	0.165	0.085	-0.175	3
SW	0.500	0.483	0.483	0.000	0.138	0.483	0.000	0.000	0.298	**0.181**	1
DH	0.309	0.291	0.291	0.163	0.000	0.291	0.163	0.163	0.239	0.179	2
NS	0.100	0.000	0.000	0.165	0.000	0.000	0.165	0.165	0.085	-0.175	3
HMI	0.500	0.483	0.483	0.000	0.138	0.483	0.000	0.000	0.298	**0.181**	1
EWS	0.500	0.483	0.483	0.000	0.138	0.483	0.000	0.000	0.298	**0.181**	1
$\phi^-(a)$	0.301	0.260	0.260	0.117	0.059	0.260	0.117	0.117			

Многокритериальный метод принятия решений Promethee II отображает ранжирование компонентов КФС по их уязвимостям учитывая их важность (Таблица 5.1.3).

Для определения разницы между оценками в степени предпочтения от нуля до единицы использовалась функция предпочтения критерия Гаусса. Ранжирование PROMETHEE II осуществляется путем упорядочивания решений от лучшего к худшему. Концепция с большей значимостью считается более надежной концепцией.

Результаты экспериментов показали, что наиболее критичными являются рабочая станция EWS, HMI и SCADA WAN (SW). Выбор подходящей политики обеспечения кибербезопасности рассматриваемого процесса осуществляется с помощью сравнительного анализа функциональных уязвимостей компонентов КФС.

Основной принцип PROMETHEE II состоит в сравнении альтернатив PLC1, PLC2, RTU, SW, DH, NS, HMI и EWS, которые оцениваются по семи критериям: γ, *BS*, ε, *AV*, ξ, *UI*, *PR* чтобы их максимизировать или

минимизировать. Как показано в таблице 5.1.4, результаты ранжирования рисков с использованием методов PROMETHEE II и TOPSIS [189, с. 1-16] аналогичны.

Таблица 5.1.4

Сравнение предложенного подхода с известными методами

Компонент КФС	TOPSIS	VIKOR	ELECTRE	Предложенный подход
PLC_1	4	4	3	4
PLC_2	3	3	4	3
RTU	3	4	3	3
SW	1	1	1	1
DH	2	2	2	2
NS	1	1	1	1
HMI	1	1	1	1
EWS	1	1	1	1

Более того, на третью и четвертую позиции были помещены два разных варианта решения для методов VIKOR [311, с. 482-500] и ELECTRE [345, с. 1-8].

Подводя итоги, предложен количественный метод оценки критичности уязвимостей компонентов КФС на основе многокритериального метода принятия решений PROMETHEE II. Он позволяет ранжировать и идентифицировать наиболее уязвимые компоненты системы. Предложенный подход оценивался на примере системы КФС для транспортировки природного газа по трубопроводу.

Для этого была представлена модель угроз и рассмотрены три сценария кибератак на эту систему. БГА был создан на основе рассматриваемой структуры КФС с учетом известных уязвимостей ее компонентов. Для эффективной оценки уязвимостей использовались метрики CVSS, такие как базовая оценка, оценка воздействия, подоценка возможности эксплойта, вектор атаки, взаимодействие с пользователем, требуемые привилегии и оценка сложности эксплойта.

В результате удалось выявить наиболее критичные уязвимости функциональных компонентов КФС с помощью PROMETHEE II. Сравнение с методами TOPSIS, VIKOR и ELECTRE доказало эффективность предложенного подхода. Предложенный подход может помочь техническим специалистам принимать более обоснованные решения при ранжировании критических уязвимостей функциональных компонентов КФС для обеспечения мер безопасности и предотвращения кибератак.

Исследования могут быть сосредоточены по таким направлениям как применение глубоких нейронных сетей для изучения сложных КФС систем, выявления уязвимостей нулевого дня и оценки киберустойчивости таких систем.

5.2. Метод оценки рисков киберфизических систем с применением нечеткого интеграла Сугено

Обеспечение защиты от киберфизических атак является серьезной проблемой, требующей методов оценки рисков кибербезопасности, способных исследовать тесные взаимодействия и взаимозависимости между кибер- и физическими компонентами в таких системах [33, с. 72-82; 85, с. 11-32]. Однако существующие методы оценки рисков не учитывают эту специфику КФС.

В настоящее время методы анализа и оценки рисков КФС можно разделить на качественные и количественные. Первые основаны на экспертной оценке для определения вероятности и потенциального воздействия рисков, а также раскрывают природу рисков. При этом количественный анализ рисков использует математические и статистические методы для расчета вероятности и потенциального воздействия рисков.

Однако исследователи отдают предпочтение количественным методам анализа и оценки рисков, поскольку они позволяют более точно оптимизировать ресурсы безопасности.

В отличие существующих подходов, данное исследование направлено на решение конкретной проблемы количественного измерения киберфизической

безопасности КФС с учетом факторов, влияющих как на ее физический слой, так и на киберслой.

Количественная оценка рисков кибербезопасности приведет к выявлению киберуязвимостей и обеспечит наличие плана смягчения последствий для предотвращения угроз безопасности.

Так, Ou и Singhal проанализировали риск системы в контексте кибератаки, используя граф атак без учета критичности узла [312, с. 10-15]. Здесь критичность узла указывает на максимальное количество повреждений, нанесенных системе, когда злоумышленник скомпрометировал узел. Понятие критичности узла было определено с применением схемы атак на ИТ-инфраструктуру на основе предварительной и последующей ассоциации с другими узлами [67, с. 31-43; 354, с. 9-14].

Исследования доказали влияние соседних узлов графа атак на критичность рассматриваемого узла.

Связь ИТ и ОТ сетей важна при моделировании критичности узлов КФС систем. Большинство существующих работ рассматривают критичность узла только в одной среде, т. е. либо в среде ИТ, либо в среде ОТ [77, с. 1-5]. В этом случае предполагается, что сеть однородна, в которой все узлы работают одинаково и настроены на одну и ту же операционную систему.

Применение и преимущества, получаемые от конвергенции ИТ и ОТ, растут и позволяют более эффективно управлять и эксплуатировать КФС. ОТ состоит из множества сенсоров, исполнительных механизмов, ПЛК или удаленных телеметрических устройств, участвующих в производственном процессе. Они включают HMI, системы сигнализации или оповещения, инженерные рабочие станции и т. д. Существующие ИТ-системы и приложения управляют функциями промышленной автоматизации и управления во всей КФС.

Системы и функции на этом уровне взаимодействуют с производственной зоной и обмениваются данными с корпоративными системами и приложениями

для выполнения определенного количества и типов услуг. Серьезность влияния отказа узла на всю КФС зависит от количества услуг узла и их разнообразия. Это доказывает необходимость измерения критичности узлов системы.

В данном исследовании предлагается количественная метрика оценки киберрисков КФС на основе критичности узлов графа атак.

Предлагаемый метод. Структура предлагаемой методологии оценки рисков включает моделирование системы, определение и ранжирование критичности компонентов системы и оценку рисков.

Смоделированный граф атак, нацеленных на устройства КФС, позволяет получить различные измерения для оценки киберрисков. Различные измерения из кибер- и физической среды объединяются в единую количественную оценку, которая используется для диагностики состояния системы.

Значения уязвимостей компонентов в кибер- и физических слоях КФС используются для расчета риска системы на основе нечеткого интеграла Сугено. На основе полученных значений выбираются наиболее критические узлы КФС и прогнозируются возможные пути атаки.

Значение риска может быть статическим [379, с. 78–85], динамическим [182, с. 201–207] и каскадным [102, с. 233-239; 210, с. 2393-2403]. Когда злоумышленник не эксплуатирует уязвимости КФС, оценка риска называется статическим значением риска R0.

Значение R0 показывает, насколько легко справиться с уязвимостями распределительной сети КФС. Динамическое значение риска R (R>R0) указывает на то, что злоумышленник эксплуатировал некоторые уязвимости системы. Цикл оценки риска завершается и сравнивается со статическим значением риска R0. Если R>R0, риск возник, и немедленно принимаются экстренные меры для устранения риска. Кроме того, следующая оценка проводится с определенным интервалом, если риска нет.

Для обеспечения киберустойчивости КФС на основе графа атак необходимо определить критичность ее узлов.

Индексы оценки риска позволяют измерить, насколько уязвима система к кибератакам, и определить расположение компонентов КФС относительно друг друга.

Рассматриваемые индексы охватывают следующие две основные области: (1) ОТ и (2) ИТ. Рассматриваемые среды — (1) физическая и (2) кибер.

Показатели ИТ и ОТ прогнозируют состояние КФС при возникновении нежелательного события. Например, для анализа состояния энергетической системы на основе ветра на предмет отказов оператор должен сначала знать такие параметры, как скорость включения, номинальная скорость, скорость отключения и номинальная мощность на каждом узле системы [132, с. 192-198]. Отказ отдельных компонентов КФС может вызвать неконтролируемые отклонения различных параметров из-за кибератак злоумышленников, что в конечном итоге может привести к разрушению системы [324, с. 1-12; 326, с. 72-130].

На рис. 5.2.1 показан пример схемы КФС, состоящей из 17 узлов: 7 киберузлов и 10 физических узлов. В этом примере сложная система включает два слоя, а именно ИТ и ОТ, и их взаимодействие. На рисунке показаны связи между кибер- и физическими компонентами. Красные стрелки представляют собой путь атаки, по которому злоумышленник может пройти от входного узла до целевого узла атаки. Путь начинается от киберузла системы до целевого узла физического уровня.

CVSS используется как система оценок, которая количественно измеряет киберриск уязвимостей, присутствующих в системе [139].

Граф атак Θ позволяет связывать различные уязвимости. Он определяет потенциальные угрозы для узлов КФС на основе информации об уязвимостях, чтобы представить пути атаки в системе.

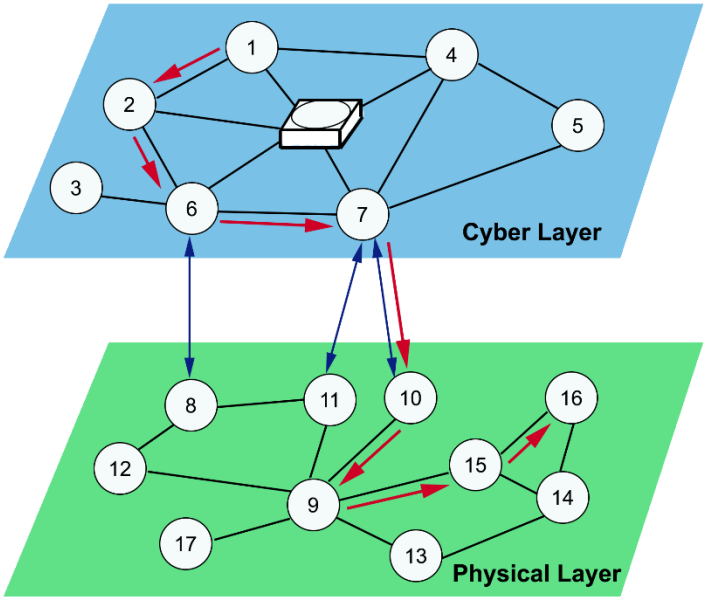


Рис. 5.2.1. Диаграмма прогнозирования пути атаки

Фактически, CVSS оценивает сложность реализации кибератаки, принимая во внимание уязвимости для каждого устройства, которое существует в конкретном узле КФС.

КФС системы имеют разнообразную топологическую структуру. Рассмотрим некоторые индексы для оценки критичности узлов графа атак. Эти индексы предоставляют подробную информацию обо всей сети.

Пусть $G = (\mathcal{N}, \mathcal{E})$ – это граф, где $\mathcal{N} = \{1, \dots, n\}$ – набор узлов (вершин), а $\mathcal{E}$ — набор ребер, $\mathcal{E} = \{(i,j) | i,j \in \mathcal{N}\}$. n представляет собой общее количество узлов в сети.

а) Центральность по близости. Центральность по близости узла измеряет, насколько близко узел (т. е. i) находится ко всем другим узлам, путем вычисления кратчайшей длины пути от одного узла до других узлов в сети. Узлы с высоким показателем центральности по близости имеют большее влияние на другие узлы в сети.

Определение (центральность по близости). Центральность по близости C_i^c для узла i вычисляется следующим образом [44, с. 230-236]:

$$C_i^c = \frac{n-1}{\sum_{\substack{j=1 \\ i \neq j}}^{n} d_{ij}}, \; i = 1, \dots, n, \qquad (5.2.1)$$

где d_{ij} – кратчайшее расстояние от вершины i до вершины j. В графе может быть несколько путей, соединяющих i с j, среди которых d_{ij} – кратчайший путь.

б) Центральность по собственному вектору показывает связь между самой «влиятельной» вершиной графа и соседними вершинами [198, с. 7821-7826; 282, с. 89-119].

Определение (центральность по собственному вектору). Центральность по собственному вектору E_i^c для узла i определяется следующим образом:

$$E_i^c = \frac{1}{\delta_{\max}} \sum_{j=1}^{n} a_{ij} E_j^c, i = 1, \dots, n, \quad (5.2.2)$$

где $\delta_{\max}$ обозначает наибольшее собственное значение матрицы смежности. Матрица смежности представляет собой квадратную матрицу размера $n \times n$, $\mathbb{A} = \|a_{ij}\|_{i,j=1}^{n}$, элементы которой определяются как: $a_{ij} = 1$, когда узел i соединен с узлом j, и $a_{ij} = 0$ в противном случае.

в) Центральность по промежуточности ребра измеряет количество кратчайших путей, проходящих через определенную вершину графа [192, с. 35-41]. Эта теоретико-графовая метрика измеряет, как часто узел выполняет роль «моста» на кратчайших путях между двумя другими узлами.

При переводе сети в теоретико-графовую модель центральность по промежуточности ребра (B_i^c) для узла указывает на возможность атаки, проходящей через этот узел.

Определение (центральность по промежуточности ребра). Центральность по промежуточности ребра B_i^c узла i определяется следующим образом:

$$B_i^c = \sum_{\substack{k,j=1 \\ k \neq j \neq i}}^{n} \frac{\sigma_{kj,i}}{\sigma_{kj}}, i = 1, \dots, n, \quad (5.2.3)$$

где σ_{kj} – общее количество кратчайших путей от исходного узла k до конечного узла j, а $\sigma_{kj,i}$ – количество путей от k до j, проходящих через i.

Кратчайшие пути относятся ко всем кратчайшим путям между каждой парой вершин в графе. Если одна вершина является частью кратчайших путей, то она имеет высокую центральность промежуточности по ребру.

г) Центральность по Катцу – это параметр теории графов, который придает важность узлу с учетом структуры сети и положения узла в сети. Он определяет количество узлов, соединенных через этот путь, а вклад удаленных узлов «штрафуется».

Определение (центральность по Катцу). Центральность по Катцу K_i^c узла i определяется как:

$$K_i^c = \sum_{q=1}^{\infty} \sum_{j=1}^{n} \beta^q (\mathbb{A}^q)_{ji}, i = 1, \dots, n, \quad (5.2.4)$$

где $\beta \in (0,1)$ – коэффициент затухания, т. е. доля участия удаленных вершин, а $(\mathbb{A}^q)_{ji}$ – общее количество связей степени q между узлами i и j.

Вышеуказанные индексы предоставляют подробную информацию обо всех КФС имеющих разнообразную топологическую структуру.

Объединение значений критериев оценки рисков КФС выполняется с использованием нечетких интегралов, которые определяются относительно нечетких мер. Предполагается, что значения нечеткой меры и всех входных параметров изменяются в пределах единичного интервала.

Определение (нечеткая мера). Пусть $N = \{1, \dots, n\}$ конечное множество и $\mu: 2^{\mathcal{N}} \to [0,1]$ – это функция такая, что $\mu(\emptyset) = 0$ и $\mu(\mathcal{N}) = 1$. Если для любых A и B, таких, что $A \subseteq B \subseteq \mathcal{N}$, выполняется условие $\mu(A) \leq \mu(B)$, то нечеткое множество μ называется нечеткой мерой.

Определение (λ-мера Сугено). Пусть $N = \{1, \dots, n\}$ конечное множество и $\lambda \in (-1, +\infty)$. Функция $\mu: 2^{\mathcal{N}} \to [0,1]$ является λ-мерой Сугено, если выполняются следующие условия:

$$\mu(\emptyset) = 0, \quad (5.2.5)$$

$$\mu(\mathcal{N}) = 1, \quad (5.2.6)$$

$$\mu(A) \leq \mu(B), \forall A, B \text{ таких, что } A \subseteq B \subseteq \mathcal{N}, \quad (5.2.7)$$

$$\mu(A \cup B) = \mu(A) + \mu(B) + \lambda\, \mu(A)\, \mu(B), \forall A, B \subseteq \mathcal{N} \text{ с } A \cap B = \emptyset, \quad (5.2.8)$$

где $\emptyset$ — пустое множество.

Уравнения (5.2.5) и (5.2.6) представляют меры пустого множества и комбинации всех множеств соответственно. Уравнение (5.2.7) представляет свойство монотонности. Уравнение (5.2.8) представляет возможные подмножества и объединенные подмножества.

При повторном применении (5.2.8) для каждого $A \subseteq \mathcal{N}$ значение $\mu(A)$ можно вычислить следующим образом [281, с. 1-20]:

$$\mu(A) = \left[\frac{\prod_{i\in A}(1+\lambda\mu(\{i\}))}{\lambda}\right]. \quad (5.2.9)$$

Используя ограничение $\mu(\mathcal{N}) = 1$ (5.2.6) и применяя (5.2.9), значение λ можно вычислить с помощью (5.2.10) [281, с. 1-20; 296, с. 532-549]:

$$\lambda + 1 = \prod_{i=1}^{n}(1 + \lambda\mu_i), \quad (5.2.10)$$

где $\mu_i = \mu(\{i\})$.

Определение (дискретный интеграл Сугено). Пусть μ – это нечеткая мера в $\mathcal{N}$. Дискретный интеграл Сугено функции $x = (x_1, x_2, \dots, x_n)$: $[0,1]^n \to [0,1]$ относительно μ определяется как [281, с. 1-20]:

$$SI_\mu(x) = \max_{1\le i\le n}(\min(x_{\pi(i)}, \mu(\{\pi(1), \pi(2), \dots, \pi(n)\}))) =$$

$$= \max_{1\le i\le n}\{\min\{x_{\pi(i)}, \mu(\{\pi(1)\})\}, \dots, \min\{x_{\pi(n)}, \mu(\{\pi(1), \dots, \pi(n)\})\}\}, \quad (5.2.11)$$

где π – перестановка в $\mathcal{N}$ такая, что $x_{\pi(1)} \le \cdots \le x_{\pi(n)}$.

Основная идея интеграла Сугено основана на взвешенном минимуме и максимуме, что позволяет оценить важность каждой модели с использованием нечетких мер.

Нечеткий интеграл Сугено определяет наивысший уровень сходства между целевыми и прогнозируемыми значениями.

Объединение значений критериев оценки риска КФС выполняется с использованием нечетких интегралов, которые определяются относительно нечетких мер. Предполагается, что значения нечеткой меры и всех входных параметров изменяются в пределах единичного интервала.

В исследовании для оценки рисков путей атаки используется метрика, основанная на нечетком интеграле Сугено [58, с. 779-796]. Риск каждого узла графа атак рассчитывается следующим образом:

$$Risk_i = Probability_i \times ImpactSI_i, i = 1, \dots, n, \quad (5.2.12)$$

где $Probability_i$ – это вероятность доступа к узлу i, которая показывает количество путей атаки и рассчитывается следующим образом:

$$Probability_i = 1 - \prod_{j=1}^{n}(1 - P_j),\ i = 1, \dots, n, \quad (5.2.13)$$

где P_j – это вероятность атаки на узел j.

Здесь P_i рассчитывается как

$$P_i = AV_i \times AC_i \times UI_i \times PR_i,\ i = 1, \dots, n, \quad (5.2.14)$$

где AV_i – вектор атаки, AC_i – сложность атаки, UI_i – взаимодействие с пользователем, PR_i – требование привилегий.

Подверженность риску с использованием нечеткого интеграла Сугено (SI) рассчитывается с использованием индексов B_i^c, C_i^c, E_i^c и K_i^c, а также оценок целостности (I_i), доступности (A_i) и конфиденциальности (C_i), полученных из CVSS v3.1, следующим образом:

$$ImpactSI_i = SI(B_i^c, C_i^c, E_i^c, K_i^c, I_i, A_i, C_i),\ i = 1, \dots, n. \quad (5.2.15)$$

Узлы с высокими значениями $ImpactSI$ считаются более уязвимыми с точки зрения кибербезопасности КФС (Алгоритм).

Алгоритм: Оценка рисков КФС

Вход: Граф атак Θ

Выход: Значение риска R

Шаг 1: Для узла i получить информацию об уязвимости v из базы данных уязвимостей.

Шаг 2: Рассчитать меры центральности с помощью (5.2.1)-(5.2.4) и метрики безопасности.

Шаг 3: Принимая во внимание значения, полученные с помощью (5.2.1)-(5.2.4), рассчитать риск узла i с помощью (5.2.5) в случае успешного использования уязвимости.

Шаг 4: Повторить шаги с 1 по 3 и рассчитать риск рассматриваемого узла i для различных уязвимостей.

Шаг 5: Рассчитать риск КФС системы.

Метрика, основанная на нескольких индексах, лучше характеризует каждый узел системы. В отличие от одного индекса, она более информативна. Ниже приводится пример оценки киберрисков на основе нечеткого интеграла Сугено.

После расчета требуемых коэффициентов они объединяются с помощью многокритериального анализа решений. Нечеткий интеграл Сугено используется для оценки рисков путей атак для обеспечения киберустойчивости КФС.

Например, индексы, используемые в качестве входных данных $ImpactSI_i$ для узла i, могут быть: $x_{1,i} = I_i$, $x_{2,i} = A_i$, $x_{3,i} = C_i$, $x_{4,i} = B_i^c$, $x_{5,i} = C_i^c$, $x_{6,i} = E_i^c$ и $x_{7,i} = K_i^c$. Каждому из этих индексов присваивается соответствующий «экспертный» вес: μ_1, μ_2, μ_3, μ_4, μ_5, μ_6 и μ_7. Далее для простоты второй индекс i переменных x опущен. Предположим, что важность каждого входного индекса выражается через нечеткие плотности $\mu_1 = \mu(x_1) = 0.95$, $\mu_2 = \mu(x_2) = 0.27$, $\mu_3 = \mu(x_3) = 0.25$, $\mu_4 = \mu(x_4) = 0.38$, $\mu_5 = \mu(x_5) = 0.28$, $\mu_6 = \mu(x_6) = 0.19$ и $\mu_7 = \mu(x_7) = 0.12$.

Численное значение параметра λ можно рассчитать как решение следующего уравнения:

$$1 + \lambda = (1 + \lambda\mu_1)(1 + \lambda\mu_2)(1 + \lambda\mu_3)(1 + \lambda\mu_4)(1 + \lambda\mu_5)(1 + \lambda\mu_6)(1 + \lambda\mu_7). \quad (5.2.16)$$

В результате было получено решение $\lambda = -0.9893$. А нечеткие меры Сугено также включают $\mu(x_1, x_2) = 0.9662$, …, $\mu(x_3, x_4) = 0.5360$, …, $\mu(x_6, x_7) = 0.2874$, …, $\mu(x_1, x_2, x_3) = 0.9772$, …, $\mu(x_1, x_4, x_5) = 0.9834$, …, $\mu(x_3, x_4, x_5, x_6) = 0.7320$, …, $\mu(x_1, x_2, x_3, x_4, x_5, x_6, x_7) = 1$. Далее полученные значения подвергаются фазификации на основе гауссовой функции принадлежности. В этом случае нечеткий интеграл Сугено равен 0.9382.

Вводится переменная R0 – пороговое значение, при котором узел КФС с *Risk* выше этого порога считается «неустойчивым». На практике это значение должны определять опытные эксперты.

Если узел «неустойчив», немедленно принимаются экстренные меры по устранению риска. Если *Risk* ниже R0, то узел считается «устойчивым», и риска нет. В этом случае следующая оценка рисков может быть выполнена через определенный интервал.

Окончательное решение *Risk* делает систему более киберустойчивой с точки зрения киберфизической безопасности. На рис. 5.2.2 представлена общая схема предлагаемого подхода к оценке рисков КФС для достижения киберустойчивой работы системы.

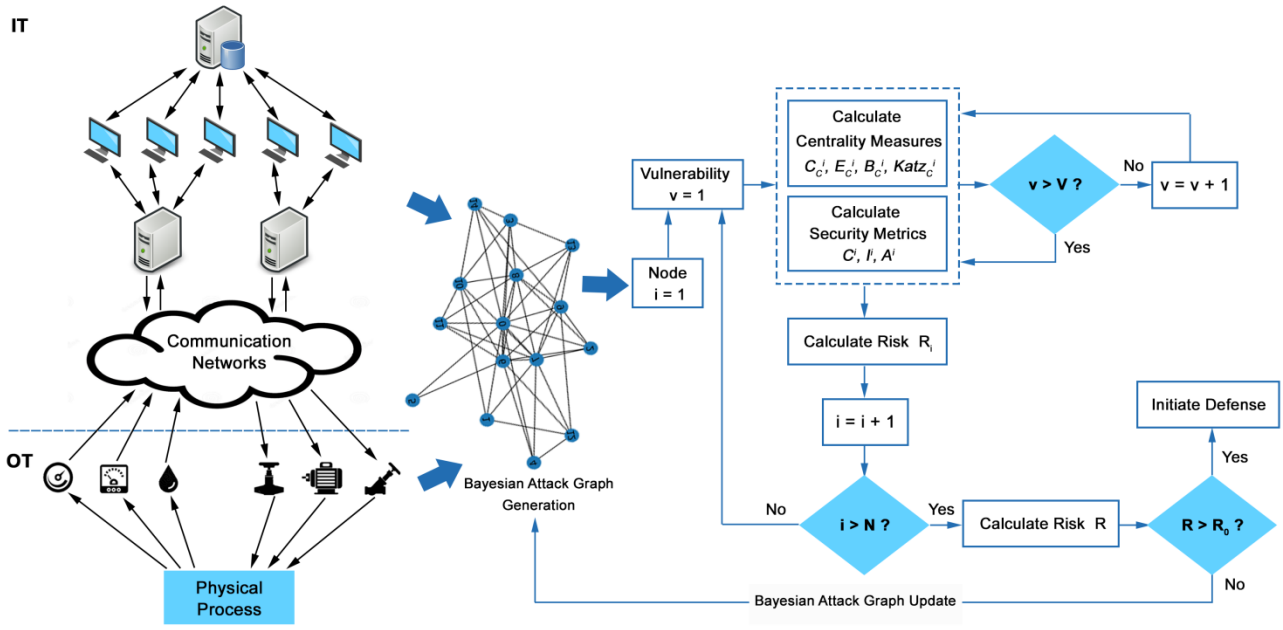


Рис. 5.2.2. Общая схема предложенного подхода оценки рисков КФС

Пример КФС. Энергия, полученная на основе ветра, является одним из критических секторов инфраструктуры (Рис. 5.2.3). Ветровые турбины широко используются на различных объектах: предприятиях, в домохозяйствах, частных домах и т. д. Потоки ветра вращают лопасти ветряной турбины, приводя ее в движение [190]. Чем сильнее ветер, тем больше вырабатывается энергии. Это вращение запускает турбину, которая также начинает вращаться.

Ветровые турбины – это устройства, преобразующие энергию ветра в электрическую энергию. Энергия передается по валу ротора, который соединен с редуктором, который приводит в действие электрогенератор.

Турбина состоит из системы охлаждения, системы мониторинга состояния и флюгера. Они служат входными данными для контроллера, который определяет положение лопастей и ротора. Система управления батареями контролирует и управляет несколькими аккумуляторными батареями и обеспечивает стабилизацию сети.

Значения параметров процесса, полученные в ходе мониторинга, хранятся в архиве данных. HMI обеспечивает взаимодействие операторов процесса с системой управления (Рис. 5.2.3).

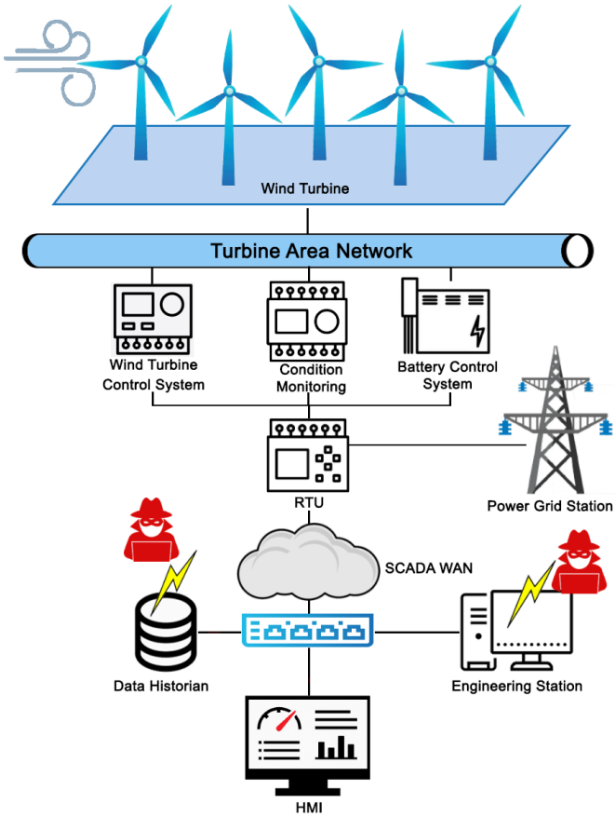


Рис. 5.2.3. Примеры сценариев кибератак на КФС

Инженерные рабочие станции содержат средства разработки программного обеспечения, с помощью которых специалист может вносить изменения и дополнения в конфигурацию системы через корпоративную сеть или Интернет. Подключение через RTU (удаленный терминал) связывает

ветропарк с центральной системой SCADA (диспетчерское управление и сбор данных).

Предположим, что злоумышленник воспользовался уязвимостями компонентов КФС и реализовал следующие сценарии кибератак (Рис. 5.2.3):

- *Манипуляция и отказ в управлении.* Злоумышленник, способный взаимодействовать с сервером SCADA, может использовать уязвимость CVE-2019-14925 для манипулирования системными конфигурациями, файлами или критическими значениями, связанными с работой ветропарка. Уязвимость может привести к несанкционированному доступу к конфиденциальным данным, включая имена пользователей, пароли и другую конфиденциальную информацию.

Злоумышленник также может злоупотребить тем фактом, что соединения не аутентифицированы (CVE-2021-27395), что позволит несанкционированно манипулировать данными и выдавать команды RTU. Это может остановить выполнение логических задач и нарушить связь между SCADA и RTU. Это не позволит операторам контролировать компрессорные станции.

- *Потеря управления.* Чтобы скомпрометировать учетные данные и получить доступ к сети ветропарка, злоумышленник использует уязвимости CVE-2019-9013, CVE-2022-1159 и CVE-2021-22797 для выполнения кода на RTU.

Это позволит перейти к управлению отдельными турбинами. Злоумышленник может проникнуть во внутреннюю сеть системы управления, используя уязвимость CVE-2018-5452, чтобы манипулировать конфигурацией системы, рабочими параметрами и прошивкой контроллера.

Это может привести к отключению функции защиты от превышения скорости, встроенной в RTU, и отключению нагрузки, что приведет к отключению турбины.

Используя CVE-2020-7566, злоумышленник может скомпрометировать данные и отключить системы мониторинга работоспособности, которые обеспечивали

бы раннее предупреждение об опасности. Кроме того, они могут нацелиться на ПЛК и использовать CVE-2020-6992 для компрометации учетных данных и получения кода выполнения на ПЛК. С помощью этого злоумышленник может повлиять на функции управления батареей, что приведет к простою системы и потенциальной дестабилизации КФС.

Результаты экспериментов. Представлены результаты оценки предлагаемого подхода на основе графа атак с использованием предопределенной модели уязвимости КФС.

Злоумышленник — это хост, на котором находится нарушитель во внешней сетевой среде. Рассматриваемая модель КФС состоит из двух слоев и 17 узлов. Конечной целью злоумышленника является компрометация узлов ПЛК и RTU. В качестве угрозы рассматривалась атака типа «отказ в обслуживании» (DoS).

Оценка рисков КФС проводилась на Python 3.7.12 с использованием различных библиотек, включая модуль NetworkX. Все эксперименты проводились на процессоре Intel Xeon(R) X5670 @ 2.93GHz*24 с 24GB RAM.

Для каждого узла i системы были рассчитаны значения индексов C_i^c, E_i^c, K_i^c, B_i^c, а также целостности (I_i), доступности (A_i), конфиденциальности (C_i), полученные на основе CVSS v3.1. Это позволило выявить критические узлы системы.

Индексам были присвоены «экспертные» веса. Чем выше вес, тем выше «информативность» индекса.

Информация о рассматриваемых уязвимостях, включая базовую оценку *BS*, оценку воздействия *IS*, возможность использования ζ, вектор доступа, *I, A, C, UI, PR* и *AC*, приведена в Таблице 5.2.1.

В таблице 5.2.2 приведены значения индексов для всех 17 узлов рассматриваемой КФС.

Таблица 5.2.1

Информация об уязвимостях

Узел	Уязвимость CVE-ID	*BS*	*IS*	ζ	*AV*	*C*	*I*	*A*	*UI*	*PR*	*AC*
1	CVE- 2021-41773	7.5	3.6	3.9	0.85	0.56	0	0	0.85	0.85	0.77
2	CVE-2022-22720	9.8	5.9	3.9	0.85	0.56	0.56	0.56	0.85	0.85	0.77
3	CVE-2022-30522	7.5	3.6	3.9	0.85	0	0	0.56	0.85	0.85	0.77
4	CVE-2014-7844	7.8	5.9	1.8	0.55	0.56	0.56	0.56	0.85	0.85	0.77
5	CVE-2019-9557	6.1	2.7	2.8	0.85	0.22	0.22	0	0.62	0.85	0.77
6	CVE-2020-2512	5.9	3.6	2.2	0.85	0	0	0.56	0.85	0.85	0.44
7	CVE-2020-24673	9.8	5.9	3.9	0.85	0.56	0.56	0.56	0.85	0.85	0.77
8	CVE-2020-6992	6.7	5.9	0.8	0.55	0.56	0.56	0.56	0.85	0.27	0.77
9	CVE-2021-27395	8.1	5.2	2.8	0.85	0	0.56	0.56	0.85	0.62	0.77
10	CVE-2020-3960	8.4	5.8	2.0	0.55	0.56	0	0.56	0.85	0.62	0.77
11	CVE-2021-22797	7.8	5.9	1.8	0.55	0.56	0.56	0.56	0.62	0.85	0.77
12	CVE-2019-14925	6.5	3.6	2.8	0.85	0.56	0	0	0.85	0.62	0.77
13	CVE-2019-9013	8.8	5.9	2.8	0.62	0.56	0.56	0.56	0.85	0.85	0.77
14	CVE-2022-1159	7.2	5.9	1.2	0.85	0.56	0.56	0.56	0.85	0.27	0.77
15	CVE-2020-7566	7.3	5.2	2.1	0.62	0.56	0.56	0	0.62	0.85	0.77
16	CVE-2018-5452	7.5	3.6	3.9	0.85	0	0	0.56	0.85	0.85	0.77
17	CVE-2023-0286	7.4	5.2	2.2	0.85	0.56	0	0.56	0.85	0.85	0.44

Эти значения позволяют оценить критичность устройств системы на основе значений воздействия и вероятности.

Таблица 5.2.2

Индексы оценки критичности узлов КФС

Узел	B_i^c	C_i^c	E_i^c	K_i^c
1	0.0125	0.1667	0.0777	0.2411
2	0.0422	0.1667	0.0777	0.2411
3	0.0125	0.1250	0.0530	0.2193
4	0.0125	0.1000	0.0362	0.2172
5	0.0042	0.1667	0.0777	0.2411
6	0.0453	0.3125	0.2199	0.3112
7	0.2060	0.3214	0.3169	0.3182
8	0.0190	0.2045	0.1500	0.2264
9	0.2500	0.2500	0.3650	0.2429
10	0.0943	0.2188	0.2162	0.2271
11	0.1390	0.2667	0.3186	0.2497
12	0.0000	0.2101	0.2490	0.2195
13	0.0229	0.2101	0.2490	0.2195
14	0.0042	0.2156	0.3398	0.2391
15	0.0688	0.2101	0.2490	0.2195
16	0.0000	0.2402	0.4018	0.2411
17	0.0000	0.2101	0.2490	0.2195

Предложенный подход был выбран для оценки серьезности уязвимости по критериям:

$$R = \begin{cases} \text{critical}, & v \in [5, 10] \\ \text{high}, & v \in [3, 5) \\ \text{medium}, & v \in [2, 3) \\ \text{low}, & v \in [0, 2) \end{cases}. \quad (5.2.17)$$

Были рассмотрены различные пути атак, направленные на возможные узлы графа для изменения их состояний. Для краткости рассматриваются три цели атак в моделируемой сети: узлы 17, 16 и 12. В таблицах 5.2.3, 5.2.4 и 5.2.5 показаны рассмотренные пути атак и серьезность уязвимостей, которые могут привести к нежелательным событиям.

Итак, существуют пять возможных путей атак на узлы 17 и 12 (Таблица 5.2.3 и Таблица 5.2.5).

Согласно таблице 5.2.4, наибольшая атака состояла из шести шагов для узла 16. Были получены пять путей атак для каждого из рассматриваемых узлов.

Таблица 5.2.3

Оценка рисков путей атаки на узел #17

Путь	Описание пути атаки	Вероятность	Риск пути	Серьезность уязвимости
P1	$X_1 \to X_2 \to X_7 \to X_{10} \to X_8 \to X_{17}$	0.8968	3.1415	High
P2	$X_1 \to X_6 \to X_7 \to X_{10} \to X_8 \to X_{17}$	0.8571	2.3053	Medium
P3	$X_1 \to X_2 \to X_7 \to X_{11} \to X_8 \to X_{17}$	0.7318	2.4721	Medium
P4	$X_1 \to X_6 \to X_7 \to X_{11} \to X_8 \to X_{17}$	0.8571	2.1983	Medium
P5	$X_1 \to X_6 \to X_9 \to X_{11} \to X_8 \to X_{17}$	0.7553	1.9372	Low

Таблица 5.2.4

Оценка рисков путей атаки на узел #16

Путь	Описание пути атаки	Вероятность	Риск пути	Серьезность уязвимости
P1	$X_1 \to X_6 \to X_7 \to X_{10} \to X_8 \to X_{15} \to X_{16}$	0.9228	2.3662	Medium
P2	$X_1 \to X_2 \to X_7 \to X_{11} \to X_8 \to X_{15} \to X_{16}$	0.9149	2.9764	High
P3	$X_1 \to X_2 \to X_7 \to X_{10} \to X_8 \to X_{15} \to X_{16}$	0.9442	3.1896	High
P4	$X_1 \to X_6 \to X_7 \to X_{11} \to X_8 \to X_{15} \to X_{16}$	0.9228	2.2515	Medium
P5	$X_1 \to X_6 \to X_9 \to X_{11} \to X_8 \to X_{15} \to X_{16}$	0.8677	2.1171	Medium

Таблица 5.2.5

Оценка рисков путей атаки на узел #12

Путь	Описание пути атаки	Вероятность	Риск пути	Серьезность уязвимости
P1	$X_1 \to X_2 \to X_7 \to X_{10} \to X_8 \to X_{12}$	0.9084	2.3299	Medium
P2	$X_1 \to X_6 \to X_7 \to X_{10} \to X_8 \to X_{12}$	0.8725	1.5228	Low
P3	$X_1 \to X_2 \to X_7 \to X_{11} \to X_8 \to X_{12}$	0.9084	2.2164	Medium
P4	$X_1 \to X_6 \to X_7 \to X_{11} \to X_8 \to X_{12}$	0.8725	1.4192	Low
P5	$X_1 \to X_6 \to X_9 \to X_{11} \to X_8 \to X_{12}$	0.7804	1.2694	Low

Таким образом, наиболее вероятный путь атаки лежит через узлы X_1, X_2, X_7, X_{10}, за которыми следует атака на X_8 и X_{15}, и, наконец, атакуется узел X_{16} в P3. Поэтому уязвимости рассматриваемых узлов являются важнейшими объектами КФС, на которых необходимо сосредоточить внимание экспертов.

На рис. 5.2.4 показаны значения *P* и *ImpactSI* для каждого пути атаки на систему. Сравнение этих оценок упрощает выявление наиболее вероятных путей атаки, которые могут привести к нежелательным событиям. Значения P для пути атаки, нацеленной на конечные узлы 17 и 12, выделены пурпурным и голубым цветом, соответственно. Тогда как значения *ImpactSI* для пути атаки на узлы 17 и 12 обозначены красным и синим цветом, соответственно.

Преимущество предлагаемого метода более наглядно доказывается при использовании нескольких индексов при оценке рисков компонентов КФС. Поэтому были рассмотрены три наиболее критичных узла системы. Использование нечеткого интеграла Сугено позволило наиболее точно определить риск пути атаки на основе вероятности и воздействия, направленного на наиболее уязвимые компоненты КФС. Это поможет лицам, принимающим решения, выделять ресурсы для ограничения доступа к определенным компонентам системы, которые злоумышленники могут использовать для атаки на критическую инфраструктуру.

Таким образом, для решения задач обеспечения киберустойчивости системы была предложена количественная метрика оценки киберрисков КФС на основе нечеткого интеграла Сугено.

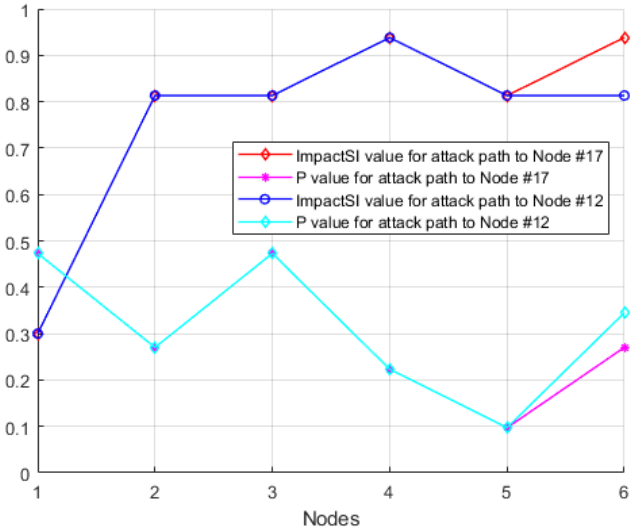


Рис. 5.2.4. Распределение значений ImpactSI и P по рассмотренным путям атак

Она предназначена для включения различных типов критических индексов, которые влияют на киберустойчивость и работу КФС, обеспечивая при этом количественную оценку кибер- и физического состояния системы. Эта метрика учитывает уязвимости в устройствах ИТ и ОТ, развернутых на физическом и киберуровне КФС. Экспериментальные результаты, основанные на DoS-атаке, показали эффективность оценки рисков в узлах рассматриваемой системы.

Подводя итоги данного исследования, был проведен анализ путей кибератаки на КФС с учетом критичности ее функциональных компонентов на основе теории графов. Это позволяет проводить осознанный контроль над киберустойчивостью КФС, помогает идентифицировать угрозы на основе уязвимостей и снижает риск, поддерживая систему в рабочем состоянии в случае кибератаки. Был предложен метод количественной оценки рисков путей кибератаки на КФС с использованием метрики, основанной на нечетком интеграле Сугено. В качестве индексов оценки уязвимостей ИТ и ОТ рассматривались центральность по близости, центральность по собственному вектору, центральность по промежуточности ребра, центральность по Катцу, а также целостность, доступность и конфиденциальность, приводящие к кибератакам на КФС системы.

Предложенная метрика учитывает взаимосвязь между уязвимостями и влиянием кибератак на физический уровень рассмотренной системы. Применена модель угроз на примере КФС для генерации энергии ветра.

Направления исследований включают:

• *Внедрение цифрового двойника.* В основе цифрового двойника лежит разработка высокоточных прототипов промышленных машин и компонентов на основе данных. Эта система будет работать в режиме реального времени и позволит проводить проактивную проверку.

• *Разработка архитектуры Zero Trust*. Она основана на данных для контроля безопасности между активами, системами и пользователями. Преимуществами Zero Trust являются оптимизированный стек безопасности, снижение транзакционных издержек и более гибкая адаптация сотрудников и поставщиков.

• *Цифровая иммунная система.* Она позволит развить иммунитет КФС к действиям злоумышленников, аналогичный иммунной системе защиты живого организма. Система разрабатывает и накапливает контрмеры против неизвестных кибератак, обнаруживает нежелательные эффекты и частично или полностью восстанавливает вычислительные процессы.

• *Смягчение каскадных эффектов, связанных с кибератаками*. Каскадные эффекты, приводящие к сбоям в одной инфраструктуре, распространяются на многие другие инфраструктуры с учетом зависимостей компонентов, уязвимостей и существующих мер. Исследования и разработка инструментов для смягчения каскадных эффектов позволят проводить ранний анализ рисков КФС.

Предложенный метод позволит определить возможный риск системы и обнаружить конечную цель на пути кибератаки, что позволит эксперту эффективнее принять окончательное решение. Результаты исследования будут полезны при разработке подходов с использованием искусственного интеллекта для обеспечения кибербезопасности и киберустойчивости КФС.

Выводы к главе V

- Разработан метод определения критичности уязвимостей функциональных компонентов КФС на основе Байесовского графа атак и рассмотрена возможность релизации для системы транспортировки природного газа по трубопроводу;
- Разработан метод оценки рисков КФС систем с применением нечеткого интеграла Сугено и рассмотрена возможность релизации для системы генерации энергии ветра.

ЗАКЛЮЧЕНИЕ

Основные научно-теоретические и практические результаты, полученные при решении задач в рамках диссертационной работы, заключаются в следующем:

1. Проанализированы существующие методы, выявлены актуальные

проблемы для обеспечения киберустойчивости КФС: разработка методов и алгоритмов интеллектуальной обработки больших данных, разработка методов и алгоритмов обеспечения кибербезопасности ИТ и защищенности ОТ КФС, разработка метода определения критичности уязвимостей функциональных компонентов КФС и метода оценки рисков КФС. Представлена постановка научной задачи исследования. Разработана трехуровневая концептуальная модель для решения ряда задач по обеспечению киберустойчивости КФС.

2. Разработаны алгоритмы, основанные на кластеризации с учетом компактности и разделимости кластеров, для обнаружения выбросов в больших данных на основе метода k-средних с целью снижения риска блокировки КФС.
3. Предложен алгоритм анализа больших данных КФС на основе взвешенной кластеризации, где веса, полученные путем суммирования весов каждой точки из набора данных, были назначены кластерам, что точнее обнаруживает аномалии в КФС по сравнению с алгоритмом k-средних.
4. Разработан метод параллельной обработки больших данных для снижения риска выхода КФС из строя, который сократил время кластеризации по сравнению с алгоритмом k-средних.
5. Предложен метод кластерного анализа больших данных КФС на основе консенсусного ансамбля с использованием квадрата евклидовой метрики.
6. Разработан метод классификации кибератак на КФС с применением экстремального машинного обучения и генетического алгоритма поведения «светлячков».
7. Разработан метод обнаружения DoS атак на КФС с применением ансамбля из пяти классификаторов, где полученная метка класса соответствует максимальному значению среди всех оценок, полученных методами классификации для данной точки [10, 12].

8. Предложен алгоритм обнаружения вредоносного ПО в КФС на основе изображений с применением глубоких нейронных сетей.
9. Разработан метод обнаружения аномалий в ОТ КФС на основе иерархических скрытых Марковских моделей для системы водоочистки.
10. Разработан метод обнаружения аномалий в акустических сигналах КФС с применением трансферного обучения и метода экстремального градиентного бустинга [26].
11. Предложен метод обнаружения атак на устройства КФС с применением глубокой гибридной модели для системы водоочистки.
12. Разработан метод классификации отказов КФС на основе изображений с применением глубоких нейронных сетей на примере нефтяного погружного насоса.
13. Разработан метод определения критичности уязвимостей функциональных компонентов КФС на основе Байесовского графа атак для транспортировки природного газа по трубопроводу.
14. Разработан метод оценки рисков КФС с применением нечеткого интеграла Сугено на примере системы генерации энергии ветра.

Результаты, полученные при экспериментальных исследованиях, показали преимущества разработанных методов и доказали перспективы выбранных подходов.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Alıquliyev, R., Hacırəhimova, M. Big data analitika əsasında informasiya təhlükəsizliyi obyektində anomaliyaların aşkarlanması modeli // “İnformasiya təhlükəsizliyinin aktual problemləri”, III respublika elmi-praktiki seminarı, Bakı: Azərbaycan, – 2017, – s. 96-99.

2. “Azərbaycan Respublikasında Rəqəmsal İnkişaf Konsepsiyası” Azərbaycan Respublikasi Prezidentinin Fərmani // 16 yanvar 2025-ci ildə təsdiq edilmişdir. – Bakı: Qanun, – 2025, – 25 s.
3. Azərbaycan Respublikasının informasiya təhlükəsizliyi və kibertəhlükəsizliyə dair 2023–2027-ci illər üçün Strategiyası”nın təsdiq edilməsi haqqında Azərbaycan Respublikası Prezidentinin Sərəncamı // 28 avqust 2023-cü ildə təsdiq edilmişdir. – Bakı: Qanun, – 2024, – 31 s.
4. Ələkbərov R.Q., Bulud əsaslı elektron dövlət sistemi: mövcud vəziyyəti, problemləri və təhlükəsizlik məsələləri, 2021, №1, s. 18-31.
5. Əliyev, Ə.Q. İnformasiya iqtısadiyyatı sektorlarının və innovativ strukturlarının inkişafına dördüncü sənaye inqilabının təsirinin analizi / Ə.Q.Əliyev, R.O. Şahverdiyeva // İnformasiya cəmiyyəti problemləri, – 2018, 2, – s. 52–64.
6. Fətəliyev, T.X, Mehdiyev, Ş.A. Əşyaların internetinin neft-qaz kompleksində tətbiqi haqqında // İnformasiya sistemləri və texnologiyalar nailiyyətlər və perspektivlər beynəlxalq elmi konfransı, – Sumqayıt: Sumqayıt Dövlət Universiteti nəşriyyatı, – 15-16 noyabr, – 2018, – s. 112-115.
7. Fətəliyev, T.X. Neft-qaz kompleksinin mövcud texnoloji problemlərinin ikt yönümlü innovativ texnolojiyalar əsasında həlli yolları / T.X.Fətəliyev, Ş.A.Mehdiyev // İnformasiya Texnologiyaları Problemləri, – Bakı: İnformasiya Texnologiyaları İnstitutunun "İnformasiya Texnologiyaları" nəşriyyatı, – 2019. No1, – s. 98-108.
8. Həşimov, M.A. “Ağilli şəhər” konsepsiyasi: mövcud vəziyyəti, tətbiq sahələri və problemləri // – Bakı: İnformasiya cəmiyyəti problemləri, – 2021. №2, – s. 95–107.
9. “İnformasiya, informasiyalaşdırma və informasiyanın mühafizəsi haqqında” Azərbaycan Respublikasının Qanunu // 19 iyun 1998-ci ildə qəbul edilmişdir (27 may 2022-ci il tarixdə olan dəyişiklik və əlavələr). – Bakı: Qanun, – 2022, – 9 s.

10. “İnformasiya, informasiyalaşdırma və informasiyanın mühafizəsi haqqında” Azərbaycan Respublikasının Qanunu // 19 iyun 1998-ci ildə qəbul edilmişdir (29 noyabr 2024-cü il tarixdə olan dəyişiklik və əlavələr). – Bakı: Qanun, – 2024, – 2 s.
11. “Kritik informasiya infrastrukturunun təhlükəsizliyinin təmin edilməsi sahəsində bəzi tədbirlər haqqında” Azərbaycan Respublikasi Prezidentinin Fərmani // 17 aprel 2021-ci ildə qəbul edilmişdir. – Bakı: “Azərbaycan” qəzeti, – 2021, – 4 s.
12. Алгулиев, Р.М., Алыгулиев, Р.М., Имамвердиев, Я.Н., Сухостат, Л.В. Обнаружение DoS атак с применением ансамбля классификаторов // “İnformasiya təhlükəsizliyinin multidissiplinar problemləri” III respublika elmi-praktiki konfransı. Баку, Азербайджан, – 2017, – с. 12-18.
13. Алгулиев, Р.М., Алыгулиев, Р.М., Сухостат, Л.В. Оценка критичности киберфизических систем на основе графа атак // XVII Internatıonal scientific conference “Recognition - 2023”. Курск, Россия, – 2023, – с. 52-54.
14. Алгулиев, Р.М., Имамвердиев, Я.Н., Сухостат Л.В. Киберфизические системы: основные понятия и вопросы обеспечения безопасности / Р.М.Алгулиев, Я.Н.Имамвердиев, Л.В.Сухостат // Информационные технологии, – 2017, 7, – с. 517–528.
15. Алгулиев, Р.М., Имамвердиев, Я.Н., Сухостат, Л.В. Обеспечение информационной безопасности киберфизических систем // “Proqram mühəndisliyinin aktual elmi-praktiki problemləri” konfransı. Баку, Азербайджан, – 2017, – с. 40-45.
16. Алгулиев, Р.М., Имамвердиев, Я.Н., Шыхалиев, Р.Г., Сухостат, Л.В. Об одном подходе по выявлению кибератак на киберфизические системы с применением глубокой гибридной модели // «GLOBAL CYBER SECURITY FORUM 2019». Харьков, Украина, 2019, с. 27-28.
17. Алыгулиев, Р.М., Имамвердиев, Я.Н., Сухостат, Л.В. Оптимизационный подход к обнаружению аномалий в Big data // XIII Международная научно-

техническая конференция «Распознавание-2017». Курск, Россия, – 2017, – с. 38-40.

18. Имамвердиев, Я.Н. Обнаружение аномалий в сетевом трафике на основе информативных признаков / Я.Н.Имамвердиев, Л.В.Сухостат // Radio Electronics, Computer Science, Control, – 2017, 3, – с. 113-120.

19. Имамвердиев, Я.Н., Сухостат, Л.В. Вопросы безопасности киберфизических систем // “Riyaziyyatın tətbiqi məsələləri və yeni informasiya texnologiyaları” üzrə III respublika elmi-praktiki konfransı. Сумгаит, Азербайджан, – 2016, – с. 257-259.

20. Имамвердиев, Я.Н., Сухостат, Л.В. Вопросы применения методов машинного обучения для решения проблем информационной безопасности // "Big data: imkanları, multidissiplinar problemləri və perspektivləri" I respublika elmi-praktiki konfransı. Баку, Азербайджан, – 2016, – с. 128-131.

21. Сухостат, Л.В. Вопросы защиты персональных данных в киберфизических системах // “İnformasiya təhlükəsizliyinin aktual multidissiplinar elmi-praktiki problemləri” V respublika konfransı. Баку, Азербайджан, – 2019, – с. 92-96.

22. Сухостат, Л.В. Об одном подходе по обнаружения аномалий в киберфизических системах на основе акустических сигналов // International Conference on “İnformation security: problems and prospects”. Баку, Азербайджан, – 2021, – с. 115-118.

23. Сухостат, Л.В. Обзор некоторых решений безопасности современных АСУ ТП / Л.В.Сухостат // Телекоммуникации, – 2022, 2, – с. 14-23.

24. Сухостат, Л.В. Обнаружение атак на киберфизические системы на основе глубокого обучения // “İnformasiya təhlükəsizliyinin aktual multidissiplinar elmi-praktiki problemləri” IV respublika konfransı. Баку, Азербайджан, – 2018, – с. 42-46.

25. Abbasi, S. OutlierNets: highly compact deep autoencoder network architectures for on-device acoustic anomaly detection / S.Abbasi, M.Famouri, M.J.Shafiee [et al.] // Sensors, – 2021, 21 (14), – p. 1-12.
26. Abdullah, L. Application of PROMETHEE method for green supplier selection: a comparative result based on preference functions / L.Abdullah, W.Chan, A.Afshari // Journal of Industrial Engineering International, – 2019, 15, – p. 271–285.
27. Abdullayeva, F.J. Cyber resilience and cyber security issues of intelligent cloud computing systems // Results in Control and Optimization, – 2023, 12, – p. 1-16.
28. Aboubi, Y. BAT-CLARA: BAT-inspired algorithm for clustering large applications / Y.Aboubi, H.Drias, N.Kamel // IFAC-PapersOnLine, – 2016, 49 (12), – p. 243–248.
29. Aburomman, A.A. A novel SVM-kNN-PSO ensemble method for intrusion detection system / A.A.Aburomman, M.B.I.Reaz // Applied Soft Computing, – 2016, 38, – p. 360–372.
30. Aburomman, A.A. A novel weighted support vector machines multiclass classifier based on differential evolution for intrusion detection systems / A.A.Aburomman, M.B.I.Reaz // Information Sciences, – 2017, 414, – p. 225-246.
31. Abusnaina, A. DL-FHMC: deep learning-based fine-grained hierarchical learning approach for robust malware classification / A.Abusnaina, M.Abuhamad, H.Alasmary [et al.] // IEEE Transactions on Dependable and Secure Computing, – 2021, 19 (5), – p. 3432-3447.
32. Acharya, N. Development of an artificial neural network based multimodel ensemble to estimate the northeast monsoon rainfall over south peninsular India: an application of extreme learning machine / N.Acharya, N.A.Shrivastava, B.Panigrahi [et al.] // Climate Dynamics, – 2014, 43 (5), – p. 1303–1310.
33. Adamsky, F. Integrated protection of industrial control systems from cyber-attacks: the ATENA approach / F.Adamsky, M.Aubigny, F.Battisti [et al.] // International Journal of Critical Infrastructure Protection, – 2018, 21, – p. 72-82.

34. Aggarwal, C.C. Data clustering: algorithms and applications / C.C.Aggarwal, C.K.Reddy. – New York: CRC Press, – 2014. – 652 p.

35. Aggarwal, P. Analysis of KDD dataset attributes-class wise for intrusion detection / P.Aggarwal, S.K.Sharma // Procedia Computer Science, – 2015, 57, – p. 842–851.

36. Ahmed, C.M. Challenges and opportunities in CPS security: a physics-based perspective / C.M.Ahmed, J.Zhou // IEEE Security & Privacy, – 2020, 18 (6), – p. 14-22.

37. Ahmed, C.M., Murguia, C., Ruths, J. Model-based attack detection scheme for smart water distribution networks // 17th ACM ASIA Conference on Computer and Communications Security. Abu Dhabi, UAE, – 2017, – p. 101–113.

38. Ahmed, M. A survey of network anomaly detection techniques / M.Ahmed, A.Mahmood, J.Hu // Journal of Network and Computer Applications, – 2016, 60, – p. 19–31.

39. Ahmed, M. An inception V3 approach for malware classification using machine learning and transfer learning / M.Ahmed, N.Afreen, M.Ahmed [et al.] // International Journal of Intelligent Networks, – 2023, 4, – p. 11-18.

40. Ahmed, M., Naser, A. A novel approach for outlier detection and clustering improvement // IEEE Conference on Industrial Electronics and Applications (ICIEA). Melbourne, Australia, – 2013, – p. 577–582.

41. Ahmed, M.U. EER-ASSL: combining rollback learning and deep learning for rapid adaptive object detection / M.U.Ahmed, Y.H.Kim, P.K.Rhee // KSII Transactions on Internet and Information Systems, – 2020, 14 (12), – p. 4776-4794.

42. Aida-Zade, K., Mustafayev, E., Rustamov, S. Comparison of deep learning in neural networks on CPU and GPU-based frameworks // 2017 IEEE 11th International Conference on Application of Information and Communication Technologies (AICT). Russia, Moscow, – 2017, – p. 1-4.

43. Akbar, S. Intrusion detection system methodologies based on data analysis / S.Akbar, R.K.Nageswara, J.A.Chandulal // International Journal of Computer Applications, – 2010, 5 (2), – p. 10–20.

44. Akbarzadeh, A., Katsikas, S. Identifying critical components in large scale cyber physical systems // IEEE/ACM 42nd International Conference on Software Engineering Workshops (ICSEW'20). New York, USA, – 2020, – p. 230–236.

45. Alakbarov, R.G. Application and security issues of internet of things in oil-gas industry / R.G.Alakbarov, M.A.Hashimov // International Journal of Education and Management Engineering, – 2018, 8 (6), – p. 24-36.

46. Alasmary, H., Abusnaina, A., Jang, R., Abuhamad, M., Anwar, A., Nyang, D., Mohaisen, D. Soteria: detecting adversarial examples in control flow graph-based malware classifiers // IEEE International Conference on Distributed Computing Systems (ICDCS). Singapore, Singapore, – 2020, – p. 888-898.

47. Alguliev, R.M. pSum-SaDE: A modified p-median problem and self-adaptive differential evolution algorithm for text summarization / R.M.Alguliev, R.M.Aliguliyev, C.A.Mehdiyev // Applied Computational Intelligence and Soft Computing, – 2011, 2011, – p. 1–13.

48. Alguliev, R.M. Weighted consensus index for assessment of the scientific performance of researchers / R.M.Alguliev, R.M.Aliguliyev, T.Kh.Fataliyev [et al.] // COLLNET Journal of Scientometrics and Information Management, – 2014, 8, – p. 371-400.

49. Alguliyev, R. An anomaly detection based on optimization / R.Alguliyev, R.Aliguliyev, Y.Imamverdiyev, L.Sukhostat // International Journal of Intelligent Systems and Applications, – 2017, 9 (12), – p. 87-96.

50. Alguliyev, R. An approach for assessing the functional vulnerabilities criticality of CPS components / R.Alguliyev, R.Aliguliyev, L.Sukhostat // Cyber Security and Applications, – 2025, 3, – p. 1-7.

51. Alguliyev, R. Cyber-physical systems and their security issues / R.Alguliyev, Y.Imamverdiyev, L.Sukhostat // Computers in Industry, – 2018, 100, – p. 212-223.

52. Alguliyev, R. Radon transform based malware classification in cyber-physical system using deep learning / R.Alguliyev, R.Aliguliyev, L.Sukhostat // Results in Control and Optimization, – 2024, 14, – p. 1-14.
53. Alguliyev, R. Weighted clustering for anomaly detection in big data / R.Alguliyev, R.Aliguliyev, Y.Imamverdiyev, L.Sukhostat // Statistics, Optimization & Information Computing, – 2018, 6 (2), – p. 178–188.
54. Alguliyev, R.M. An improved ensemble approach for DoS attacks detection / R.M.Alguliyev, R.M.Aliguliyev, Y.N.Imamverdiyev, L.V.Sukhostat // Radio Electronics, Computer Science, Control, – 2018, 2 (45), – p. 73-82.
55. Alguliyev, R.M. Anomaly detection in Big data based on clustering / R.M.Alguliyev, R.M.Aliguliyev, Y.N.Imamverdiyev, L.V.Sukhostat // Statistics, Optimization and Information Computing, – 2017, 5 (4), – p. 325-340.
56. Alguliyev, R.M. Efficient algorithm for big data clustering on single machine / R.M.Alguliyev, R.M.Aliguliyev, L.V.Sukhostat // CAAI Transactions on Intelligence Technology, – 2020, 5 (1), – p. 9-14.
57. Alguliyev, R.M. Hybrid DeepGCL model for cyber-attacks detection on cyber-physical systems / R.M.Alguliyev, Y.N.Imamverdiyev, L.V.Sukhostat // Neural Computing and Applications, – 2021, 33 (16), – p. 10211-10226.
58. Alguliyev, R.M. Method for quantitative risk assessment of cyber-physical systems based on vulnerability analysis / R.M.Alguliyev, R.M.Aliguliyev, L.V.Sukhostat // Kybernetika, – 2024, 60 (6), – p. 779-796.
59. Alguliyev, R.M. Parallel batch k-means for big data clustering / R.M.Alguliyev, R.M.Aliguliyev, L.V.Sukhostat // Computers and Industrial Engineering, – 2021, 152, – p. 1-11.
60. Alguliyev, R.M. Weighted consensus clustering and its application to Big data / R.M.Alguliyev, R.M.Aliguliyev, L.V.Sukhostat // Expert Systems with Applications, – 2020, 150, – p. 1-15.

61. Alguliyev, R.M. Intelligent diagnosis of petroleum equipment faults using a deep hybrid model / R.M.Alguliyev, Y.N.Imamverdiyev, L.V.Sukhostat // SN Applied Sciences, – 2020, 2 (5), – p. 1-16.

62. Alguliyev, R.M., Aliguliyev, R.M., Bagirov, A., Karimov, R. Batch clustering algorithm for Big data sets // IEEE International Conference on Application of Information and Communication Technologies (AICT). Baku, Azerbaijan, – 2016, – p. 79-84.

63. Alguliyev, R.M., Aliguliyev, R.M., Sukhostat, L.V. Consensus clustering by weight optimization of input partitions // IEEE International Conference “Application of Information and Communication Technologies (AICT’2019). Baku, Azerbaijan, – 2019, – p. 1-4.

64. Alguliyev, R.M., Aliguliyev, R.M., Sukhostat, L.V. Purity-based consensus clustering for anomaly detection in Big data // XIV Internatıonal scientific conference “Recognition - 2018”. Kursk, Russia, – 2018, – p. 17-20.

65. Alguliyev, R.M., Imamverdiyev, Y.N., Sukhostat, L.V. Diagnostics of DoS attacks on cyber-physical systems based on hierarchical hidden Markov models // XIII International Conference for young researchers "TECHNICAL SCIENCES. INDUSTRIAL MANAGEMENT". Borovets, Bulgaria, – 2020, – p. 39-41.

66. Alguliyev, R.M., Imamverdiyev, Y.N., Sukhostat, L.V. Detection of cyber-attacks on cyber-physical systems using deep neural network. XV International scientific conference “Recognition - 2019”. Kursk, Russia, – 2019, – p. 18-20.

67. Alhomidi, M. Attack graph-based risk assessment and optimisation approach / M.Alhomidi, M.Reed // International Journal of Network Security & Its Applications, – 2014, 6 (3), – p. 31-43.

68. Ali, M. MALGRA: Machine learning and n-gram malware feature extraction and detection system / M.Ali, S.Shiaeles, G.Bendiab [et al.] // Electronics, – 2020, 9 (11), – p. 1-21.

69. Aliev, T.A., Musaeva, N.F., Nusratov, O.Q., Rzayev, A.G., Sattarova, U.E. Models for indicating the period of failure of ındustrial objects // 4th World Conference on Soft Computing. Berkeley, USA, – 2016, – p. 389–405.

70. Aliguliyev, R.M. Performance evaluation of density-based clustering methods // Information Sciences, – 2009, 179, – p. 3583-3602.

71. Aliguliyev, R.M., Imamverdiyev, Y.N., Hajirahimova, M.S. Multidisciplinary problems of big data in information security // International Scientific and Practical Conference on Information Security and Computer Technologies. Kropyvnytskyi, Ukraine, – 2017, – p. 10-11.

72. Allen, T.T. Timely decision analysis enabled by efficient social media modeling / T.T.Allen, Z.Sui, N.L.Parker // Decision Analysis, – 2017, 14 (4), – p. 250–260.

73. Almalawi, A. An unsupervised anomaly-based detection approach for integrity attacks on SCADA systems / A.Almalawi, X.Yu, Z.Tari [et al.] // Computers & Security, – 2014, 46, – p. 94-110.

74. Anceaume, E., Busnel, Y., Merrer, E.L., Ludinard, R., Marchand, J., Sericola, B. Anomaly characterization in large scale networks // Annual IEEE/IFIP International Conference on Dependable Systems and Networks. Atlanta, USA, – 2014, – p. 68–79.

75. Andersson, S.A. A characterization of Markov equivalence classes for acyclic digraphs / S.A.Andersson, D.Madigan, M.D.Perlman // Annals of Statistics, – 1997, 25, – p. 505-541.

76. Ani, U.P.D. Design considerations for building credible security testbeds: Perspectives from industrial control system use cases / U.P.D.Ani, J.M.Watson, B.Green [et al.] // Journal of Cyber Security Technology, – 2021, 5 (2), – p. 71-119.

77. Anikin, I.V. Information security risk assessment and management method in computer networks // International Siberian Conference on Control and Communications. Omsk, Russia, – 2015, – p. 1-5.

78. Ankerst, M. OPTICS: Ordering points to identify the clustering structure / M.Ankerst, M.M.Breunig, H.-P.Kriegel // ACM SIGMOD Record, – 1999, 28 (2), – p. 49–60.

79. Arghandeh, R. On the definition of cyber-physical resilience in power systems / R. Arghandeh, A. Von Meier, L. Mehrmanesh [et al.] // Renewable and Sustainable Energy Reviews, – 2016, 58, – p. 1060–1069.

80. Arthur, D., Vassilvitskii, S. K-Means++: the advantages of careful seeding // Annual ACM-SIAM symposium on Discrete algorithms. New Orleans, Louisiana, USA, – 2007, – p. 1027–1035.

81. Asghar, M.R. Cybersecurity in industrial control systems: Issues, technologies, and challenges / M.R.Asghar, Q.Hu, S.Zeadally // Computer Networks, – 2019, 165, – p. 1-16.

82. Ashibani, Y. Cyber-physical systems security: analysis, challenges and solutions / Y.Ashibani, Q.H.Mahmoud // Computers & Security, – 2017, 68, – p. 81-97.

83. Ashok, A. Cyber-physical security of wide-area monitoring, protection and control in a smart grid environment / A.Ashok, A.Hahn, M.Govindarasu // Journal of Advanced Research, – 2014, 5, – p. 481–489.

84. Atzori, M. Deep learning with convolutional neural networks applied to electromyography data: a resource for the classification of movements for prosthetic hands / M. Atzori, M. Cognolato, H. Müller // Frontiers in Neurorobotics, – 2016, 10 (9), – p. 1–10.

85. Avizienis, A. Basic concepts and taxonomy of dependable and secure computing / A.Avizienis, J.-C.Laprie, B.Randell [et al.] // IEEE Transactions on Dependable and Secure Computing, – 2004, 1 (1), – p. 11–32.

86. Ayodeji, A. A new perspective towards the development of robust data-driven intrusion detection for industrial control systems / A.Ayodeji, Y.Liu, N.Chao // Nuclear Engineering and Technology, – 2020, 52 (12), – p. 2687-2698.

87. Baheti, R. Cyber-physical systems. The impact of control technology / R.Baheti, H.Gill // IEEE Control Systems Society, – 2011, 12, – p. 161–166.

88. Bahmani, B. Scalable k-means++ / B.Bahmani, B.Moseley, A.Vattani [et al.] // Proceedings of the VLDB Endowment, – 2012, 5 (7), – p. 622–633

89. Banati, H. Fire Fly based feature selection approach / H.Banati, M.Bajaj // IJCSI International Journal of Computer Science Issues, – 2011, 8 (4), – p. 473-480.

90. Banerje, A. Frequency-sensitive competitive learning for scalable balanced clustering on high-dimensional hyperspheres / A.Banerje, J.Ghosh // IEEE Transactions on Neural Networks, – 2004, 15, – p. 702–719.

91. Barbara, D. Requirements for clustering data streams // ACM SIGKDD Explorations Newsletter, – 2003, 3, – p. 23-27.

92. Barnum, S. Roundtable: reliability of embedded and cyber-physical systems / S.Barnum, S.Sastry, J.A.Stankovic // IEEE Security & Privacy, – 2010, 8 (5), – p. 27–32.

93. Baum, L. Statistical inference for probabilistic functions of finite state Markov Chains / L.Baum, T.Petrie // The Annals of Mathematical Statistics, – 1966, 37, – p. 1554–1563.

94. Beaver, J.M., Borges-Hink, R.C., Buckner, M.A. An evaluation of machine learning methods to detect malicious SCADA communications // International Conference on Machine Learning and Applications (ICMLA), Miami, USA, – 2013, – p. 54-59.

95. Belayneh, A. Standard precipitation index drought forecasting using neural networks, wavelet neural networks, and support vector regression / A.Belayneh, J.Adamowski // Applied Computational Intelligence and Soft Computing, – 2012, 2012, – p. 1–13.

96. Belguendouz, H., Guerid, H., Kaddour, M. Static classification of IoT malware using grayscale image representation and lightweight convolutional neural networks // Proceedings of the IEEE 5th International Conference on Advanced

Communication Technologies and Networking. Marrakech, Morocco, – 2022, – p. 1-8.

97. Berikov, V. Ensemble clustering based on weighted coassociation matrices: Error bound and convergence properties / V.Berikov, I.Pestunov // Pattern Recognition, – 2017, 63, – p. 427–436.

98. Berry, M.W. Lecture notes in data mining / M.W.Berry, M.Browne. – Singapore: World Scientific, – 2006. – 236 p.

99. Bezdek, J.C. Some new indexes of cluster validity / J.C.Bezdek, N.R.Pal // IEEE Transactions on Systems, Man and Cybernetics – Part B: Cybernetics, – 1998, 28 (3), – p. 301–315.

100. Bhamare, D. Cybersecurity for industrial control systems: a survey / D.Bhamare, M.Zolanvari, A.Erbad [et al.] // Computers & Security, – 2020, 89, – p. 1-12.

101. Bhattacharyya, D. Network anomaly detection: a machine learning perspective / D.Bhattacharyya, J.Kalita – New York: Chapman and Hall/CRC, – 2013. – 336 p.

102. Bhuiyan, M.Z.A. Review of static risk-based security assessment in power system / M.Z.A.Bhuiyan, G.J.Anders, J.Philhower, S.Du // IET Cyber-Physical Systems: Theory & Applications, – 2019, 4 (3), – p. 233-239.

103. Blei, D.M. Latent Dirichlet allocation / D.M.Blei, A.Y.Ng, M.I.Jordan // Journal of Machine Learning Research, – 2003, 3, – p. 993–1022.

104. Bouaguel, W., Mufti, B., Limam, M. A fusion approach based on wrapper and filter feature selection methods using majority vote and feature weighting // International Conference on Computer Applications Technology. Sousse, Tunisia, – 2013, – p. 1–6.

105. Boutin, F., Hascoet, M. Cluster validity indices for graph partitioning // IEEE International Conference on Information Visualization. London, UK, – 2004, – p. 376–381.

106. Boutsidis, C., Drineas, P., Mahoney, M.W. Unsupervised feature selection for the k-means clustering problem // International Conference on Neural Information Processing Systems (NIPS). Vancouver, Canada, – 2009, – p. 153-161.
107. Boyer, S.A. SCADA: Supervisory Control and Data Acquisition / S.A. Boyer, – Research Triangle Park: International Society of Automation, – 2010. – 257 p.
108. Boyes, H. The Industrial Internet of Things (IIoT): an analysis framework / H.Boyes, B.Hallaq, J.Cunningham, T.Watson // Computers in Industry, – 2018, 101, – p. 1-12.
109. Bracewell, R. The Fourier transform and ıts applications / R.Bracewell. – Boston: McGraw-Hill, – 2000. – 640 p.
110. Brandt, A. Noise and vibration analysis: signal analysis and experimental procedures / A.Brandt. – Chichester: John Wiley and Sons, – 2011. – 464 p.
111. Brans, J.P. A preference ranking organisation method. The PROMETHEE method for MCDM / J.P.Brans, Ph.Vincke // Management Science, – 1985, 31 (6), – p. 647–656.
112. Breiman, L. Random forests // Machine Learning, – 2001, 45 (1), – p. 5–32.
113. Breunig, M.M. LOF: identifying density-based local outliers / M.M.Breunig, H.P.Kriegel, R.T.Ng [et al.] // ACM Sigmod Record, – 2000, 29 (2), – p. 93-104.
114. Brooks, Z. Hacking Driverless Vehicles: [Electronic resource] / DEFCON 21. – August 1-4, 2013. URL: https://www.defcon.org/images/defcon-21/dc-21-presentations/Zoz/DEFCON-21-Zoz-Hacking-Driverless-Vehicles.pdf
115. Burnaev, E., Smolyakov, D. One-class SVM with privileged information and its application to malware detection // IEEE 16th International Conference on Data Mining Workshops (ICDMW). Barcelona, Spain, – 2016, – p. 273-280.
116. Calinski, T. A dendrite method for cluster analysis / T.Calinski, J.Harabasz // Communications in Statistics, – 1974, 3 (1), – p. 1–27.
117. Canizo, M. Multi-head CNN–RNN for multi-time series anomaly detection: an industrial case study / M.Canizo, I.Triguero, A.Conde [et al.] // Neurocomputing, – 2019, 363, – p. 246-260.

118. Capo, M. An efficient approximation to the k-means clustering for massive data / M.Capo, A.Pérez, J.A.Lozano // Knowledge-Based Systems, –2017, 117, – p. 56-69.

119. Cardenas, A.A., Amin, S., Sastry, S. Research challenges for the security of control systems // USENIX Workshop on Hot Topics in Security. San Jose, USA, – 2008, – p. 1-6.

120. Cardenas, A.A., Amin, S., Sinopoli, B., Giani, A., Perrig, A., Sastry, S. Challenges for securing cyber physical systems // Workshop on Future Directions in Cyber-physical Systems Security. Newark, USA, – 2009, – p. 1-7.

121. Carletti, V., Greco, A., Saggese, A., Vento, M. Robustness evaluation of convolutional neural networks for malware classification / Proceedings of the Italian Conference on Cybersecurity (ITASEC). Bari, Italy, – 2021, – p. 414–423.

122. Catania, C.A. Automatic network intrusion detection: current techniques and open issues / C.A.Catania, C.G.Garino // Computers and Electrical Engineering, – 2012, 38 (5), – p. 1062–1072.

123. Chalamasetty, G.K., Mandal, P., Tseng, T.L. Secure SCADA communication network for detecting and preventing cyber-attacks on power systems // Clemson University Power Systems Conference (PSC). Clemson, USA, – 2016, – p. 1-7.

124. Chandola, V. Anomaly detection: a survey / V.Chandola, A.Banerjee, V.Kumar // ACM Computing Surveys, – 2009, 41 (3), – p. 1–58.

125. Chandrika, J. Dynamic clustering of high speed data streams / J.Chandrika, K.R.Ananda Kumar // International Journal of Computer Science Issues, – 2012, 9 (1), – p. 224-228.

126. Chen, P.-Y. Information fusion to defend intentional attack in Internet of Things / P.-Y.Chen, S.-M.Cheng, K.-C.Chen // IEEE Internet of Things Journal, – 2014, 1 (4), – p. 337–348.

127. Chen, R. Automated labeling and learning for physical layer authentication against clone node and Sybil attacks in industrial wireless edge networks / R.Chen,

Z.Pang, H.Wen [et al.] // IEEE Transactions on Industrial Informatics, – 2021, 17 (3), – p. 2041-2051.

128. Chen, T., Guestrin, C. XGBoost: a scalable tree boosting system // ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, USA, – 2016, – p. 785-794.

129. Chen, Y. Privacy-preserving image multi-classification deep learning model in robot system of industrial IoT / Y.Chen, Y.Ping, Z.Zhang [et al.] // Neural Computing and Applications, – 2021, 33, – p. 4677–4694.

130. Cheng, C. Deep convolutional neural network-based in-process tool condition monitoring in abrasive belt grinding / C.Cheng, J.Li, Y.Liu [et al.] // Computers in Industry, – 2019, 106, – p. 1-13.

131. Cheng, P. Cyber security for industrial control systems: from the viewpoint of close-loop / P.Cheng, H.Zhang, J.Chen. – USA: CRC Press, – 2016. – 325 p.

132. Chermitti, A. Assessment parameters and matching between the sites and wind turbines / A.Chermitti, M.Bencherif, Z.Nakoul [et al.] // Physics Procedia, – 2014, 55, – p. 192-198.

133. Chickering, D.M. Optimal structure identification with greedy search // The Journal of Machine Learning Research, – 2002, 3, – p. 507-554.

134. Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation // International Conference on Empirical Methods in Natural Language (EMNLP). Doha, Qatar, – 2014, – p. 1724–1734.

135. Chollet, F. Xception: deep learning with depthwise separable convolutions // IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, – 2017, – p. 1800-1807.

136. Chow, C. Approximating discrete probability distributions with dependence trees / C.Chow, C.Liu // IEEE Transactions on Information Theory, – 1968, 14 (3), – p. 462–467.

137. Coelho, G., Pereira, P., Matos, L., Ribeiro, A., Nunes, E., Ferreira, A., Cortez, P., Pilastri, A. Deep dense and convolutional autoencoders for machine acoustic anomaly detection // IFIP International Conference on Artificial Intelligence Applications and Innovations. Hersonissos, Greece, – 2021, – p. 337-348.
138. Collins, S. Stuxnet: the emergence of a new cyber weapon and its implications / S.Collins, S.McCombie // Journal of Policing, Intelligence and Counter Terrorism, – 2012, 7 (1), – p. 80–91.
139. Common Vulnerability Scoring System version 3.1: [Electronic resource] / FIRST.Org, Inc., – 2019. URL: https://www.first.org/cvss/v3-1/cvss-v31-specification_r1.pdf
140. Complex Systems Design & Management Asia / M. A.Cardin, D.Krob, P.Lui [et al.]. – Cham: Springer, – 2015. – 176 p.
141. Converged Plantwide Ethernet (CPwE) design and implementation guide: [Electronic resource] / Cisco and Rockwell Automation. – September 9, 2011. URL: https://literature.rockwellautomation.com/idc/groups/literature/documents/td/enet-td001_-en-p.pdf
142. Cooper, G.F. A bayesian method for the induction of probabilistic networks from data / G.F.Cooper, E.Herskovits // Machine Learning, – 1992, 9, – p. 309-347.
143. Cortes, C. Support-vector networks / C.Cortes, V.Vapnik // Machine Learning, – 1995, 20, – p. 21-27.
144. Cover, T.M. Nearest neighbor pattern classification / T.M.Cover, P.E.Hart // IEEE Transactions on Information Theory, – 1968, 13, – p. 21-27.
145. Critical information infrastructures security. CRITIS 2016. Lecture notes in computer science / G.Havarneanu, R.Setola, H.Nassopoulos [et al.]. – Cham: Springer, – 2017. – 348 p.
146. Cruz, T. A cybersecurity detection framework for supervisory control and data acquisition systems / T.Cruz, L.Rosa, J.Proenca [et al.] // IEEE Transactions on Industrial Informatics, – 2016, 12, – p. 2236-2246.

147. Cui, Z. Detection of malicious code variants based on deep learning / Z.Cui, F.Xue, X.Cai // IEEE Transactions on Industrial Informatics, – 2018, 14 (7), – p. 3187-3196.

148. Cuomo, S. A GPU-accelerated parallel k-means algorithm / S.Cuomo, V.De Angelis, G.Farina [et al.] // Computers & Electrical Engineering, – 2019, 75, – p. 262–274.

149. Cyber Resilience Act: [Electronic resource] / European Union. – March 12, 2024. URL: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0130_EN.html

150. Damiano, L. A brief technical introduction to Hidden Markov Models: [Electronic resource] / L.Damiano, B.Peterson, M.Weylandt. – August 29, 2017. URL: https://luisdamiano.github.io/gsoc17/hmm_techreview.pdf

151. Damodaran, A. A comparison of static, dynamic, and hybrid analysis for malware detection / A.Damodaran, F.D.Troia, C.A.Visaggio, C.A., [et al.] // Journal of Computer Virology and Hacking Techniques, 2017, 13, p. 1–12.

152. Dave, R. Robust clustering methods: a unified view / R.Dave, R.Krishnapuram // IEEE Transactions on Fuzzy Systems, – 1997, 5 (2), – p. 270-293.

153. Davies, D.L. A cluster separation measure / D.L.Davies, D.W.Bouldin // IEEE Transactions on Pattern Analysis and Machine Intelligence, – 1979, PAMI-1 (2), – p. 224–227.

154. Davies, R., The Internet of Things: Opportunities and challenges: [Electronic resource] / European Parliamentary Research Service. – May 19, 2015. URL: http://www.europarl.europa.eu/RegData/etudes/BRIE/2015/557012/EPRS_BRI(2015)557012_EN.pdf

155. Davis, J.J. Data preprocessing for anomaly based network intrusion detection: a review / J.J.Davis, A.J.Clark // Computers & Security, – 2011, 30, – p. 353–375.

156. Davis, S. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences / S.Davis, P.Mermelstein // Readings in speech recognition, – 1990, 28 (4), – p. 65-74.

157. Dazhuang, H., Lobov, A., Moctezumas, L.E.G.A. An approach to use PERA in enterprise modeling for industrial systems // Annual Conference on IEEE Industrial Electronics Society. Montreal, Canada, – 2012, – p. 4196-4203.
158. de Assis Boldt, F., Rauber, T.W., Oliveira-Santos, T., Rodrigues, A., Varejão, F.M., Ribeiro, M.P. Binary feature selection classifier ensemble for fault diagnosis of submersible motor pump // IEEE International Symposium on Industrial Electronics (ISIE). Edinburgh, UK, – 2017, – p. 1807-1812.
159. de Assis Boldt, F., Rauber, T.W., Varejão, F.M., Ribeiro, M.P. Performance analysis of extreme learning machine for automatic diagnosis of electrical submersible pump conditions // IEEE International Conference on Industrial Informatics. Porto Alegre, Brazil, – 2014, – p. 67-72.
160. De Oliveira, J.V. Particle swarm clustering in clustering ensembles: exploiting pruning and alignment free consensus / J.V.De Oliveira, A.Szabo, L.N.de Castro // Applied Soft Computing, – 2017, 55, – p. 141–153.
161. De Porcellinis, S. Modelling critical infrastructure via a mixed holistic reductionistic approach / S.De Porcellinis, S.Panzieri, R.Setola // International Journal of Critical Infrastructures, – 2009, 5 (1-2), – p. 86–99.
162. Dechter, R. Bucket elimination: A unifying framework for probabilistic inference // Learning in Graphical Models. NATO ASI Series, – 1998, – p. 75-104.
163. Dempster, A. Maximum likelihood from incomplete data via the EM algorithm / A.Dempster, N.Laird, D.Rubin // Journal of the Royal Statistical Society. Series B, – 1977, 39 (1), – p. 1-38.
164. Deo, R.C. Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern / R.C. Deo, M. Şahin // Australia Atmospheric Research, – 2015, 153, – p. 512–525.
165. Derler, P. Modeling cyber-physical systems / P.Derler, E.A.Lee, A.Sangiovanni-Vincentelli // Proceedings of IEEE, – 2012, 100 (1), – p. 1–28.

166. Difallah, D. Scalable anomaly detection for smart city infrastructure networks / D.Difallah, P.Cudre-Mauroux, S.McKenna // IEEE Internet Computing, – 2013, 17 (6), – p. 39–47.

167. Ding, D. A survey on security control and attack detection for industrial cyber-physical systems / D.Ding, Q.-L.Han, Y.Xiang [et al.] // Neurocomputing, – 2018, 275 (1), – p. 1674-1683.

168. Dinh, T.N., Shen, Y., Thai, M.T. An efficient spectral bound for link vulnerability assessment in large-scale networks // IEEE Military Communications Conference, – 2012, – p. 1-6.

169. Djouadi, S.M., Melin, A.M., Ferragut, E.M., Laska, J.A., Dong, J. Finite energy and bounded actuator attacks on cyber-physical systems // IEEE European Control Conference. Linz, Austria, – 2015, – p. 3659-3664.

170. Dong, Y., Wang, R., He, J. Real-time network intrusion detection system based on deep learning // Proceedings of IEEE 10th International Conference on Software Engineering and Service Science (ICSESS). Beijing, China, – 2019, – p. 1-4.

171. Dua, S. Data mining and machine learning in cybersecurity / S. Dua, X. Du. – Boca Raton, FL: CRC Press, – 2011. – 256 p.

172. Dudenhoeffer, D.D., Permann, M.R., Woolsey, S., Timpany, R., Miller, C., McDermott, A., Manic, M. Interdependency modeling and emergency response // Summer Computer Simulation Conference. San Diego, USA, – 2007, – p. 1230–1237.

173. Eastman, R., Versace, M., Webber, A. Big data and predictive analytics: on the cybersecurity front line: [Electronic resource] / SAS Institute. – February 2015. URL: https://v2.itweb.co.za/whitepaper/Whitepaper_SAS_Cyber_Security.pdf

174. Eggermont, J., Kok, J.N., Kosters, W.A. Genetic programming for data classification: partitioning the search space // ACM Symposium on Applied Computing. Nicosia, Cyprus, – 2004, – p. 1001–1005.

175. Emmanuel, M. Communication technologies for smart grid applications: a survey / M.Emmanuel, R.Rayudu // Journal of Network and Computer Applications, – 2016, 74, – p. 133-148.

176. Ertoz, L., Steinbach, M., Kumar, V. A new shared nearest neighbor clustering algorithm and its applications // Workshop on clustering high dimensional data and its applications, SIAM International Conference on Data Mining. Arlington, USA, – 2002, – p. 1-15.

177. Esposito, C., Gouglidis, A., Hutchison, D., Gurtov, A., Helvik, B., Heegaard, P., Rizzo, G., Rak, J. On the disaster resiliency within the context of 5G networks: the RECODIS experience // European Conference on Networks and Communications, Ljubljana, Slovenia, – 2018, – p. 1-4.

178. Ester, M., Kriegel, H. P., Sander, J., Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise // International Conference on Knowledge Discovery and Data Mining, Portland. Oregon, USA, – 1996, – p. 226–231.

179. Eusgeld, I. Dependability Metrics. Lecture Notes in Computer Science / I. Eusgeld, F.C. Freiling, R. Reussner. – Berlin: Springer, – 2008. – 305 p.

180. Eyal, I. Distributed data clustering in sensor networks / I. Eyal, I. Keidar, R. Rom // Distributed Computing, – 2011, 24, – p. 207-222.

181. Falco J., Wavering, A., Proctor, F. IT security for industrial control systems: [Electronic resource] / US Department of Commerce, National Institute of Standards and Technology, – Gaithersburg, 2015. URL: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=821684

182. Fang, D.Z. Improved hybrid approach to transient stability assessment / D.Z.Fang, A.K.David, C.Kai, C.Yunli // IEE Proceedings - Generation, Transmission and Distribution, – 2005, 152 (2), – p. 201-207.

183. Farabet, C. Learning hierarchical features for scene labeling / C.Farabet, C.Couprie, L.Najman // IEEE Transactions on Pattern Analysis and Machine Intelligence, – 2013, 35 (8), – p. 1915–1929.

184. Fawzi, H. Secure estimation and control for cyber-physical systems under adversarial attacks / H.Fawzi, P.Tabuada, S.Diggavi // IEEE Transactions on Automatic Control, – 2014, 59 (6), – p. 1454-1467.

185. Ferri, C. An experimental comparison of performance measures for classification / C. Ferri, J. Hernández-Orallo, R. Modroiu // Pattern Recognition Letters, – 2009, 30 (1), – p. 27–38.

186. Filonov, P., Lavrentyev, A., Vorontsov, A. Multivariate industrial time series with cyber-attack simulation: fault detection using an LSTM-based predictive data model: [Electronic resource] / Cornell University. – 2016. URL: https://arxiv.org/pdf/1612.06676

187. Fine, S. The hierarchical hidden Markov model: analysis and applications / S.Fine, Y.Singer, N.Tishby // Machine Learning, – 1998, 32, – p. 41–62.

188. Fiorea, U. Network anomaly detection with the restricted Boltzmann machine / U.Fiorea, F.Palmierib, A.Castiglionec [et al.] // Neurocomputing, – 2013, 122, – p. 13–23.

189. Forcina, A. Exploring Industry 4.0 technologies to improve manufacturing enterprise safety management: A TOPSIS-based decision support system and real case study / A.Forcina, L.Silvestri, F.De Felice [et al.] // Safety Science, – 2023, 169, – p. 1-16.

190. Forescout Vedere Labs. OT:ICEFALL: The legacy of “insecure by design” and its implications for certifications and risk management: [Electronic resource] / Forescout Technologies, Inc., – 2022. URL: from https://www.forescout.com/resources/ot-icefall-report/

191. Fovino, I., Carcano, A., De Lacheze Murel, T., Trombetta, A., Masera, M. Modbus/DNP3 state-based intrusion detection system // IEEE International Conference on Advanced Information Networking and Applications (AINA), – Perth: IEEE, – 2010, – p. 729-736.

192. Freeman, L.C. A set of measures of centrality based on betweenness // – Washington: Sociometry, – 1977, 40, – p. 35-41.

193. Freitas, S., Duggal, R., Chau, D.H. Malnet: A large-scale cybersecurity image database of malicious software // Proceedings of the 31st ACM International Conference on Information & Knowledge Management (CIKM), – Atlanta: ACM, – 17-21 October, – 2022, – p. 3948–3952.

194. Frigault, M., Wang, L., Singhal, A., Jajodia, S. Measuring network security using dynamic Bayesian network // ACM Workshop on Quality of Protection, – Alexandria, USA: ACM, – 27 October, – 2008, – p. 1-7.

195. Gastaldi, X. Shake-shake regularization of 3-branch residual networks // International Conference on Learning Representations (ICLR). Toulon, France, – 2017, – p. 1–5.

196. Gates, C., Taylor, C. Challenging the anomaly detection paradigm: a provocative discussion // Workshop on New Security Paradigms, – Schloss Dagstuhl: ACM, – 19-22 September, – 2006, – p. 21–29.

197. Giannella, C., Sayrafi, B. An information theoretic histogram for single dimensional selectivity estimation // ACM Symposium on Applied computing (SAC). Santa Fe, New Mexico, – 2005, – p. 676-677.

198. Girvan, M. Community structure in social and biological networks / M.Girvan, M.E.Newman // Proceedings of the National Academy of Sciences, – 2002, 99 (12), – p. 7821-7826.

199. Gouglidis, A., Green, B., Busby, J.S., Rouncefield, M.F., Hutchison, D., Schauer, S. Threat awareness for critical infrastructures resilience // IEEE International Workshop on Resilient Networks Design and Modeling (RNDM). Halmstad, Sweden, – 2016, – p. 196-202.

200. Gouglidis, A., Hu, V.C., Busby, J.S., Hutchison, D. Verification of resilience policies that assist attribute based access control // ACM Workshop on Attribute-Based Access Control. Arizona, USA, – 2017, – p. 43–52.

201. Grollmisch, S., Johnson, D., Abeßer, J., Lukashevich, H. IAEO3 – combining OpenL3 embeddings and interpolation autoencoder for anomalous sound detection //

Workshop on Detection and Classification of Acoustic Scenes and Events. Tokyo, Japan, – 2020, – p. 1-3.

202. Gugulothu, N., Malhotra, P., Vig, L., Shroff, G. Sparse neural networks for anomaly detection in high-dimensional time series // Workshop on AI for internet of things (AI4IOT'18). Stockholm, Sweden, – 2018, – p. 1–7.

203. Guo, C., Ma, Y., Yang, B., Jensen, C., Kaul, M. EcoMark: evaluating models of vehicular environmental impact // ACM International Conference on Advances in Geographic Information Systems. California, USA, – 2012, – p. 269-278.

204. Hahn, A. A multi-layered and kill-chain based security analysis framework for cyber-physical systems / A.Hahn, R.K.Thomas, I.Lozano [et al.] // International Journal of Critical Infrastructure Protection, – 2015, 11, – p. 39–50.

205. Hanson, S.J., Pratt, L.Y. Comparing biases for minimal network construction with back-propagation. 1st International Conference on Neural Information Processing Systems. Colorado, USA, – 1988, – p. 177–185.

206. Hassanzadeh, A. A review of cybersecurity incidents in the water sector / A.Hassanzadeh, A.Rasekh, S.Galelli [et al.] // Journal of Environmental Engineering, – 2020, 146 (5), – p. 1-15.

207. Hassanzadeh, A., Modi, S., Mulchandani, S. Towards effective security control assignment in the Industrial Internet of Things // IEEE World Forum on Internet of Things (WF-IoT). Milan, Italy, – 2015, – p. 1-6.

208. He, H., Maple, C., Watson, T., Tiwari, A., Mehnen, J., Jin, Y., Gabrys, B. The security challenges in the IoT enabled cyberphysical systems and opportunities for evolutionary computing & other computational intelligence // IEEE Congress on Evolutionary Computation. Vancouver, Canada, – 2016a, – p. 1015-1021.

209. He, K., Zhang, X., Ren, S., Sun, J. Deep residual learning for image recognition // IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, –2016b, – p. 770-778.

210. Henneaux, P., Labeau P.E., Maun J.C., Haarla L. A two-level probabilistic risk assessment of cascading outages / P.Henneaux, P.E.Labeau, J.C.Maun, L.Haarla // IEEE EEE Transactions on Power Systems, – 2015, 31 (3), – p. 2393–2403.

211. Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. Improving neural networks by preventing co-adaptation of feature detectors: [Electronic resource] / Cornell University. – 2012. URL: https://arxiv.org/abs/1207.0580

212. Hollander, M. Nonparametric statistical methods / M.Hollander, D.A.Wolfe, E.Chicken, – New Jersey: John Wiley & Sons, – 2015. – 828 p.

213. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H. MobileNets: efficient convolutional neural networks for mobile vision applications: [Electronic resource] / arXiv preprint. – 2017. URL: https://arxiv.org/pdf/1704.04861

214. Hu, Y. A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition / Y.Hu, Y.Wong, W.Wei [et al.] // PLoS ONE, – 2018, 13 (10), – p. 1–18.

215. Huang, G. Trends in extreme learning machines: a review / G.Huang, G.-B.Huang, S.Song [et al.] // Neural Networks, – 2015, 61, – p. 32–48.

216. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q. Densely connected convolutional networks // IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, – 2017, – p. 2261-2269.

217. Huang, G.-B. Convex incremental extreme learning machine / G.-B.Huang, L.Chen // Neurocomputing, – 2007, 70, – p. 3056–3062.

218. Huang, G.-B. Extreme learning machine: theory and applications / G.-B.Huang, Q.-Y.Zhu, C.-K.Siew // Neurocomputing, – 2006, 70, – p. 489–501.

219. Humayun, M. Cyber security threats and vulnerabilities: a systematic mapping study / M. Humayun, M. Niazi, N. Jhanjhi [et al.] // Arabian Journal for Science and Engineering, – 2020, 45, – p. 3171–3189.

220. Ibrahimov, B.G. Research on the performance of intelligent communication networks / B.G.Ibrahimov, M.Q.Hasanov, R.T.Humbatov [et al.] // IFAC-PapersOnLine, – 2018, 51 (30), – p. 768-770.

221. Imamverdiyev, Y.N., Sukhostat, L.V. Anomaly detection in network traffic using extreme learning machine // IEEE International Conference on Application of Information and Communication Technologies (AICT'2016). Baku, Azerbaijan, – 2016, – p. 418-421.

222. Inoue, J., Yamagata, Y., Chen, Y., Poskitt, C.M., Sun, J. Anomaly detection for a water treatment system using unsupervised machine learning // IEEE International Conference on Data Mining Workshops (ICDMW). New Orleans, USA, – 2017, – p. 1058–1065.

223. Intelligent Data Engineering and Automated Learning – IDEAL 2016. Lecture Notes in Computer Science / H.Yin, Y.Gao, B.Li [et al.]. Cham: Springer, – 2016. – 647 p.

224. Ioffe, S., Szegedy, C. Batch normalization: accelerating deep network training by reducing internal covariate shift // International Conference on Machine Learning. Lille, France, – 2015, – p. 448-456.

225. ISA99, Industrial Automation and Control Systems Security: [Electronic resource] / International Society of Automation. – 2004. URL: https://www.isa.org/standards-and-publications/isa-standards/isa-standards-committees/isa99

226. Islam, R., Islam, M.M.M., Kim, J.-M. Feature selection techniques for increasing reliability of fault diagnosis of bearings // IEEE International Conference on Electrical and Computer Engineering. Dhaka, Bangladesh, – 2016, – p. 396-399.

227. ISO 22301, Security and resilience – Business continuity management systems – Requirements: [Electronic resource] / International Organization for Standardization. – 2019. URL: https://www.iso.org/obp/ui/#iso:std:iso:22301:ed-2:v1:en

228. Iturbe, M. Towards large-scale, heterogeneous anomaly detection systems in industrial networks: a survey of current trends / M.Iturbe, I.Garitano, U.Zurutuza [et al.] // Security and Communivation Networks, – 2017, 2017, – p. 1-17.
229. Jaatun, M.G. Cloud Computing – CloudCom 2009. Lecture Notes in Computer Science / M.G. Jaatun, G. Zhao, C. Rong. – Berlin: Springer, – 2009. – 707 p.
230. Janssens, O. Convolutional neural network based fault detection for rotating machinery / O. Janssens, V. Slavkovikj, B. Vervisch [et al.] // Journal of Sound and Vibration, – 2016, 377, – p. 331–345.
231. Jara, A.J. Big data for cyber physical systems—an analysis of challenges, solutions and opportunities // IEEE International Conference on innovative mobile and Internet services in ubiquitous computing. Birmingham, UK, – 2014, – p. 376-380.
232. Jarvis, R.A. Clustering using a similarity measure based on shared nearest neighbors / R.A.Jarvis, E.A.Patrick // IEEE Transactions on Computers, – 1973, C-22 (11), – p. 1025–1034.
233. Jin, S. Practical speed measurement for an intelligent vehicle based on double radon transform in urban traffic scenarios / S.Jin, J.Yin, M.Tian [et al.] // Measurement Science and Technology, – 2020, 32(2), – p. 1-12.
234. Johansson, K.H. Guest editorial special issue on control of cyber-physical systems / K.H.Johansson, G.J.Pappas, P.Tabuada [et al.] // IEEE Transactions on Automatic Control, – 2014, 59 (12), – p. 3120-3121.
235. Kabir, G. Power substation location selection using fuzzy analytic hierarchy process and PROMETHEE: A case study from Bangladesh / G.Kabir, R.S.Sumi // Energy, – 2014, 72, – p. 717-730.
236. Kanungo, T. An efficient k-means clustering algorithm: Analysis and implementation / T.Kanungo, D.M.Mount, N.S.Netanyahu [et al.] // IEEE Transactions on Pattern Analysis and Machine Intelligence, – 2002, 24 (7), – p. 881-892.

237. Karanja, E.M. Analysis of Internet of things malware using image texture features and machine learning techniques / E.M.Karanja, S.Masupe, M.G.Jeffrey // Internet of Things, – 2020, 9, – p. 1-15.

238. Kassani, S.H. Diabetic retinopathy classification using a modified Xception architecture / S.H.Kassani, P.H.Kassani, R.Khazaeinezhad [et al.] // IEEE International Symposium on Signal Processing and Information Technology. Ajman, United Arab Emirates, – 2019, – p. 1-6.

239. Kaufman, L. Finding groups in data: An introduction to cluster analysis / L.Kaufman, P.J.Rousseeuw, – New York: John While & Sons, – 1990. – 342 p.

240. Kaur, P. Big data and machine learning-based secure healthcare framework / P.Kaur, M.Sharma, M.Mittal // Procedia Computer Science, – 2018, 132, – p. 1049-1059.

241. Kayan, H. Cybersecurity of industrial cyber-physical systems: a review / H.Kayan, M.Nunes, O.Rana [et al.] // ACM Computing Surveys, – 2022, 54 (11s), – p. 1-35.

242. Khan, R.U. Analysis of ResNet and GoogleNet models for malware detection / R.U.Khan, X.Zhang, R.Kumar // Journal of Computer Virology and Hacking Techniques, – 2019, 15, – p. 29–37.

243. Kim, M.J. Development of real-time Big data analysis system and a case study on the application of information in a medical institution / M.J.Kim, Y.S.Yu // International Journal of Software Engineering and Its Applications, – 2015, 9 (7), – p. 93-102.

244. Kingma, D.P., Ba, J. Adam: A method for stochastic optimization // International Conference on Learning Representations. Banff, Canada, – 2014, – p. 1-15.

245. Kiselak, J. “SPOCU”: scaled polynomial constant unit activation function / J.Kiselak, Y.Lu, J.Svihra [et al.] // Neural Computing and Applications, – 2021, 33, – p. 3385–3401.

246. Kotenko, I.V. Parallel big data processing system for security monitoring in Internet of Things networks / I.V.Kotenko, I.Saenko, A.Kushnerevich // Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, – 2017, 8 (4), – p. 60-74.

247. Kott, A. Cyber Resilience of Systems and Networks / A.Kott, I.Linkov, – USA: Springer, – 2018. – 488 p.

248. Kou, Y., Lu, C., Sirwongwattana, S., Huang, Y. Survey of fraud detection techniques // IEEE ICNSC Conference. Taipei, Taiwan, – 2004, – p. 749-754.

249. Koubaa, A., Andersson, B. A vision of cyber-physical Internet // International Workshop on Real-Time Networks. Dublin, Ireland, – 2009, – p. 1-6.

250. Kravchik, M., Shabtai, A. Detecting cyber attacks in industrial control systems using convolutional neural networks // Workshop on Cyber-Physical Systems Security and PrivaCy. New York, USA, – 2018, – p. 72-83.

251. Krizhevsky, A. Imagenet classification with deep convolutional networks / A.Krizhevsky, I.Sutskever, G.Hinton // Communications of the ACM, – 2012, 60 (6), – p. 84-90.

252. Krotofil, M. Vulnerabilities of cyber-physical systems to stale data-determining the optimal time to launch attacks / M.Krotofil, A.A.Cardenas, J.Larsen // International Journal of Critical Infrastructure Protection, – 2014, 7, – p. 213–232.

253. Krotofil, M., Larsen, J., Gollmann, D. The process matters: ensuring data veracity in cyber-physical systems // ACM Symposium on Information, Computer and Communications Security, Singapore: Republic of Singapore. – 2015, – p. 133–144.

254. Krutz, R.L. Securing SCADA systems / R.L.Krutz. – USA: John Wiley & Sons, – 2005. – 240 p.

255. Kumar, S. MCFT-CNN: malware classification with fine-tune convolution neural networks using traditional and transfer learning in Internet of Things // Future Generation Computer Systems, – 2021, 125, – p. 334–351.

256. Lachtar, N. Toward mobile malware detection through convolutional neural networks / N.Lachtar, D.Ibdah, A.Bacha // IEEE Embedded Systems Letters, – 2021, 13 (3), – p. 134–137.

257. Langner R. Stuxnet: dissecting a cyberwarfare weapon / R.Langner // IEEE Security & Privacy, – 2011, 9 (3), – p. 49–51.

258. Lazarevic, A., Ertoz, L., Kumar, V., Ozgur, A., Srivastava, J. A comparative study of anomaly detection schemes in network intrusion detection // SIAM Conference on Data Mining. San Francisco, USA, – 2003, – p. 25-36.

259. LeCun, Y. Deep learning / Y.LeCun, Y.Bengio, G.Hinton // Nature, – 2015, 521, – p. 436–444.

260. Lee, E.A. Cyber physical systems: design challenges // International Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing. Orlando, USA, – 2008, – p. 363-369.

261. Li, B. Hearing loss classification via AlexNet and extreme learning machine // International Journal of Cognitive Computing in Engineering, – 2021, 2, – p. 144–153.

262. Li, D., Chen, D., Jin, B., Shi, L., Goh, J., Ng, S.K. MAD-GAN: multivariate anomaly detection for time series data with generative adversarial networks // International Conference on Artificial Neural Networks. Munich, Germany, – 2019, – p. 703–716.

263. Li, G. A CNN model based on innovative expansion operation improving the fault diagnosis accuracy of drilling pump fluid end / G.Li, J.Hu, D.Shan [et al.] // Mechanical Systems and Signal Processing, – 2023, 187, – p. 1-15.

264. Li, H. Bearing fault diagnosis with a feature fusion method based on an ensemble convolutional neural network and deep neural network / H.Li, J.Huang, S.Ji // Sensors, – 2019, 19 (9), – p. 1-18.

265. Li, H. Multicast routing for decentralized control of cyber physical systems with an application in smart grid / H.Li, L.Lai, H.V.Poor // IEEE Journal on Selected Areas in Communications, – 2012, 30, – p. 1097–1107.

266. Liang, Y. Photoplethysmography and deep learning: enhancing hypertension risk stratification / Y.Liang, Z.Chen, R.Ward [et al.] // Biosensors, – 2018, 8(4), – p. 1-13.

267. Lichman, M. UCI Machine Learning Repository: [Electronic resource] / University of California. – 2013. URL: http://archive.ics.uci.edu/ml

268. Lin, Q., Adepu, S., Verwer, S., Mathur, A. TABOR: a graphical model-based approach for anomaly detection in industrial control systems // Asia Conference on Computer and Communications Security (ASIACCS'18). Incheon, Republic of Korea, – 2018, – p. 525–536.

269. Lin, S.-W., Miller, B., Durand, J., Bleakley, G., Chigani, A., Martin, R., Murphy, B., Crawford, M. The Industrial Internet of Things Volume G1: Reference Architecture: [Electronic resource] / Industrial Internet Consortium, – 2019. URL: https://www.iiconsortium.org/pdf/IIRA-v1.9.pdf

270. Lin, S.-W., Murphy, B., Clauer, E., Loewen, U., Neubert, R., Bachmann, G., Pai, M., Hankel, M. IIoT Architecture Alignment and Interoperability. An Industrial Internet Consortium and Plattform Industrie 4.0 Joint Whitepaper: [Electronic resource] / Industrial Internet Consortium, – 2017. URL: https://www.iiconsortium.org/pdf/JTG2_Whitepaper_final_20171205.pdf

271. Liu, C. Cyber-physical machine tool–the era of machine tool 4.0 / C.Liu, X.Xu // Procedia CIRP, – 2017, 63, – p. 70–75.

272. Liu, C.H. Cyber physical systems: architectures, protocols and applications / C.H.Liu, Y.Zhang. – Florida: CRC Press, – 2016. – 272 p.

273. Lo, W.W., Yang, X., Wang, Y. An Xception convolutional neural network for malware classification with transfer learning // IEEE IFIP International Conference on New Technologies, Mobility and Security. Canary Islands, Spain, – 2019, – p. 1-5.

274. Loshchilov, I., Hutter, F. Decoupled weight decay regularization // International Conference on Learning Representations (ICLR). New Orleans, USA, – 2019, – p. 1-19.

275. Luiijf, E. Critical information infrastructures security. CRITIS 2018. Lecture notes in computer science / E.Luiijf, I.Žutautaitė, B.M.Hämmerli. – Cham: Springer, – 2018. – 210 p.

276. Lun, Y.Z. State of the art of cyber-physical systems security: an automatic control perspective / Y.Z.Lun, A.D'Innocenzo, F.Smarra [et al.] // Journal of Systems and Software, – 2019, 149, – p. 174–216.

277. Machuca, C.M., Secci, S., Vizarreta, P., Kuipers, F., Gouglidis, A., Hutchison, D., Jouet, S., Pezaros, D., Elmokashfi, A., Heegaard, P., Ristov, S. Technology-related disasters: a survey towards disaster-resilient software defined networks // IEEE International Workshop on Resilient Networks Design and Modeling (RNDM). Halmstad, Sweden, – 2016, – p. 35-42.

278. Malchiodi, D., Bassis, S., Valerio, L. Discovering regression data quality through clustering methods // Italian Workshop on Neural Networks, – 2008, – p. 76–85.

279. Mammadova, M. The problems of information security of electronic personal health data // IEEE International Conference on Information Technology in Medicine and Education (ITME). Huangshan, China, – 2015, – p. 678-682.

280. Mammadova, M., Ahmadova, A. Formation of unified digital health information space in Healthcare 4.0 environment and interoperability issues // IEEE International Conference on Application of Information and Communication Technologies (AICT). Washington DC, USA, – 2022, – p. 1-6.

281. Martínez, G.E. General type-2 fuzzy Sugeno integral for edge detection / G.E.Martínez, C.I.Gonzalez, O.Mendoza [et al.] // Journal of Imaging, – 2019, 5 (8), – p. 1-20.

282. Mason, O. Graph theory and networks in biology / O.Mason, M.Verwoerd // IET Systems Biology, – 2007, 1(2), – p. 89–119.

283. McHugh, J. Testing intrusion detection systems: a critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by Lincoln

laboratory // ACM Transactions on Information and System Security, – 2000, 3 (4), – p. 262–294.

284. Mell, P., Scarfone, K., Romanosky, S. A Complete Guide to the Common Vulnerability Scoring System Version 2.0: [Electronic resource] / FIRST - Forum of Incident Response and Security Teams, – 2007. URL: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=51198

285. Mirkin, B. Mathematical classification and clustering / B.Mikrin. – New York: Springer, – 2011, – 448 p.

286. Mitchell, R. Effect of intrusion detection and response on reliability of cyber physical systems / R.Mitchell, I.R.Chen // IEEE Transactions on Reliability, – 2013, 62 (1), – p. 199–210.

287. Modeling and design of secure Internet of Things / A.Kott, S.Shetty, L.L.Njilla [et al.]. – Hoboken: Wiley-IEEE Press, – 2020. – 704 p.

288. Mohammad, R., Thabtah, F.A., McCluskey, T.L. Phishing websites dataset: [Electronic resource] / University of Huddersfield. – 2015. URL: https://archive.ics.uci.edu/ml/datasets/Phishing+Websites

289. Molina, L.C. Feature selection algorithms: a survey and experimental evaluation / L.C.Molina, L.Belanche, A.Nebot // IEEE International Conference on Data Mining. Maebashi City, Japan, – 2002, – p. 306–313.

290. Monjardim, G.E., Rodrigues, A., Varejao, F.M., Souza, V.E.S., Ribeiro, M.P. A domain-specific language for fault diagnosis in electrical submersible pumps // IEEE International Conference on Industrial Informatics (INDIN). Porto, Portugal, – 2018, – p. 843-848.

291. Moradpour, S. Using hybrid fuzzy PROMETHEE II and fuzzy binary goal programming for risk ranking: a case study of highway construction projects / S.Moradpour, S.Ebrahimnejad, E.Mehdizadeh [et al.] // Journal of Optimization in Industrial Engineering, – 2011, 9, – p. 47-55.

292. Morita, K. Anomalous sound detection by using local outlier factor and Gaussian mixture model / K.Morita, T.Yano, K.Q.Tran // Workshop on Detection and Classification of Acoustic Scenes and Events. Tokyo, Japan, – 2020, – p. 2-4.
293. Müller, R., Ritz, F., Illium, S., Linnhoff-Popien, C. Acoustic anomaly detection for machine sounds based on image transfer learning // International Conference on Agents and Artificial Intelligence. Vienna, Austria, – 2021, – p. 49-56.
294. Muñoz-González, L. Exact inference techniques for the analysis of Bayesian attack graphs / L.Muñoz-González, D.Sgandurra, M.Barrère [et al.] // IEEE Transactions on Dependable and Secure Computing, – 2017, 16 (2), – p. 231-244.
295. Munshi, A. A. Big data framework for analytics in smart grids / A.A.Munshi, Y.A.Mohamed // Electric Power Systems Research, – 2017, 151, – p. 369–380.
296. Murofushi, T. A theory of fuzzy measures. Representation, the Choquet integral and null sets / T.Murofushi, M.Sugeno // Journal of Mathematical Analysis and Applications, – 1991, 159 (2), – p. 532-549.
297. Naeem, H. Malware detection in industrial internet of things based on hybrid image visualization and deep learning model / H.Naeem, F.Ullah, M.R.Naeem [et al.] // Ad Hoc Networks, – 2020, 105, – p. 1-12.
298. Nallaivarothayan, H., Ryan, D., Denman, S., Sridharan, S., Fookes, C. An evaluation of different features and learning models for anomalous event detection // International Conference on Digital Image Computing: Techniques and Applications. Hobart, Australia, – 2013, – p. 1–8.
299. Nataraj, L., Yegneswaran, V., Porras, P., Zhang, J. A comparative assessment of malware classification using binary texture analysis and dynamic analysis // ACM Workshop on Security and Artificial Intelligence. NY, USA, – 2011, – p. 21–30.
300. Neuman, C., Tan, K. Mediating cyber and physical threat propagation in secure Smart grid architectures // IEEE International Conference on Smart Grid Communications. Brussels, Belgium, – 2011, – p. 238-243.
301. Ngo, Q. A survey of IoT malware and detection methods based on static features / Q.Ngo, H.Nguyen, V.Le [et al.] // ICT Express, – 2020, 6 (4), – p. 280-286.

302. Nguyen, N., Caruana, R. Consensus clusterings // IEEE International Conference on Data Mining. Omaha, USA, – 2007, – p. 607–612.
303. Nguyen, N.T. Advances in multimedia and network information system technologies / N.T.Nguyen, A.Zgrzywa, A.Czyzewski. – Berlin: Springer-Verlag, – 2010. – 318 p.
304. Nicol, D.M., Mallapura, V. Modeling and analysis of stepping stone attacks // Winter Simulation Conference. Savannah, USA, – 2014, – p. 3036–3047.
305. Nikouei, M.A. The PROMETHEE multiple criteria decision making analysis for selecting the best membrane prepared from sulfonated poly (ether ketone)s and poly (ether sulfone)s for proton exchange membrane fuel cell / M.A.Nikouei, M.Oroujzadeh, S.Mehdipour-Ataei // Energy, – 2017, 119, – p. 77-85.
306. NIST National vulnerability database: [Electronic resource] / National Institute of Standards and Technology, – 2024. URL: https://nvd.nist.gov/vuln/data-feeds
307. Obregon, L. Secure Architecture for Industrial Control Systems: [Electronic resource] / SANS Institute. – 2015. URL: https://www.sans.org/reading-room/whitepapers/ICS/securearchitecture-industrial-control-systems-36327
308. Offshore pipelines: design, installation, and maintenance / B.Guo, S.Song, A.Ghalambor [et al.]. – Waltham: Gulf Professional Publishing, – 2014. – 384 p.
309. Oliveira-Santos, T. Combining classifiers with decision templates for automatic fault diagnosis of electrical submersible pumps / T.Oliveira-Santos, A.Rodrigues, V.F.Rocha [et al.] // Integrated Computer-Aided Engineering, – 2018, 25 (4), – p. 381-396.
310. Oliveira-Santos, T., Rauber, T.W., Varejão, F.M., Martinuzzo, L., Oliveira, W., Pellegrini Ribeiro, M., Rodrigues, A. Submersible motor pump fault diagnosis system: a comparative study of classification methods // IEEE International Conference on Tools with Artificial Intelligence (ICTAI). San Jose, USA, – 2016, – p. 415-422.

311. Ou Yang, Y. A VIKOR technique based on DEMATEL and ANP for information security risk control assessment / Y.Ou Yang, H.Shieh, G.Tzeng // Information Sciences, – 2013, 232, – p. 482-500.

312. Ou, X. Quantitative security risk assessment of enterprise networks / Ou, X., Singhal, A. – New York: Springer, – 2011, – 28 p.

313. Panteli, M. The GRID: stronger, bigger, smarter? Presenting a conceptual framework of power system resilience / M.Panteli, P.Mancarella // IEEE Power Energy Magazine, – 2015, 13 (3), – p. 58–66.

314. Parker, J.K. Accelerating fuzzy-c means using an estimated subsample size / J.K.Parker, L.O.Hall // IEEE Transactions on Fuzzy Systems, – 2014, 22 (5), – p. 1229–1244.

315. Pasqualetti, F. Attack detection and identification in cyber-physical systems / F.Pasqualetti, F.Dorfler, F.Bullo // IEEE Transactions on Automatic Control, – 2013, 58 (11), – p. 2715-2729.

316. Patrikainen, A. Comparing subspace clusterings / A.Patrikainen, M.Meila // IEEE Transactions on Knowledge and Data Engineering, – 2006, 18 (7), – p. 902–916.

317. Peng, Z. Time–frequency data fusion technique with application to vibration signal analysis / Z.Peng, W.Zhang, Z.Lang [et al.] // Mechanical Systems and Signal Processing, – 2011, 29, – p. 164–173.

318. Philip, S.Y. Machine learning in cyber trust / S.Y.Philip, J.P.T.Jeffrey. –New York: Springer-Verlag, – 2009, – 362 p.

319. Piqueira, J.R.C., Cabrera, M. A., C. M. Batistela, Malware propagation in clustered computer networks / J.R.C.Piqueira, M.A.Cabrera, C.M.Batistela // Physica A, – 2021, 573, – p. 1-15.

320. Pivoto, D.G. Cyber-physical systems architectures for industrial internet of things applications in Industry 4.0: A literature review / D.G.Pivoto, L.F.de Almeida, R.da Rosa Righi [et al.] // Journal of Manufacturing Systems, – 2021, 58, – p. 176-192.

321. Popov, O.B. Cybersecurity for Critical Infrastructure Protection via Reflection of Industrial Control Systems, NATO Science for Peace and Security Series D: Information and Communication Security / O.B.Popov, L.Sukhostat // Amsterdam: IOS Press, – 2022. – 175 p.

322. Prosvirin, A.E. An improved algorithm for selecting IMF components in ensemble empirical mode decomposition for domain of rub-impact fault diagnosis / A.E.Prosvirin, M.M.M.Islam, J.-M.Kim // IEEE Access, – 2019, 7, – p. 121728-121741.

323. Purohit, H., Tanabe, R., Ichige, K., Endo, T., Nikaido, Y., Suefusa, K., Kawaguchi, Y. MIMII dataset: sound dataset for malfunctioning industrial machine investigation and inspection // Workshop on Detection and Classification of Acoustic Scenes and Events. New York, USA, – 2019, – p. 209–213.

324. Qu, Z. Electric power cyber-physical systems vulnerability assessment under cyber-attack / Z.Qu, W.Sun, J.Dong [et al.] // Frontiers in Energy Research, – 2023, 10, – p. 1-12.

325. Quatieri, T.F. Discrete Time Speech Signal Processing: Principles and Practice / T.F.Quatieri. – New Jersey: Prentice Hall Press, – 2001. – 802 p.

326. Rahman, I. Hybrid bio-inspired computational intelligence techniques for solving power system optimization problems: A comprehensive survey / I.Rahman, J.Mohamad-Saleh // Applied Soft Computing, – 2018, 69, – p. 72-130.

327. Rass, S. Game Theory for Security and Risk Management. Static & Dynamic Game Theory: Foundations & Applications / S.Rass, S.Schauer. – Birkhäuser: Springer, – 2018. – 418 p.

328. Reddi, S.J., Kale, S., Kumar, S. On the convergence of Adam and beyond // International Conference on Learning Representations (ICLR). Vancouver, Canada, – 2018, – p. 1–23.

329. Roberson, D. Improving grid resilience using high-voltage DC: strengthening the security of power system stability / D.Roberson, H.C.Kim, B.Chen [et al.] // IEEE Power Energy Magazine, – 2019, 17 (3), – p. 38–47.

330. Ronen, R., Radu, M., Feuerstein, C., Yom-Tov, E., Ahmadi, M. Microsoft Malware Classification Challenge: [Electronic resource] / Cornell University. – 2018. URL: https://www.kaggle.com/c/malware-classification

331. Rosa, L. Intrusion and anomaly detection for the next-generation of industrial automation and control systems / L.Rosa, T.Cruz, M.B.de Freitas [et al.] // Future Generation Computer Systems, 2021, 119, p. 50-67.

332. Rose, J.R. IDERES: intrusion detection and response system using machine learning and attack graphs / J.R.Rose, M.Swann, K.P.Grammatikakis [et al.] // Journal of Systems Architecture, – 2022, 131, – p. 1-12.

333. Roseline, S.A. Intelligent vision-based malware detection and classification using deep random forest paradigm / S.A.Roseline, S.Geetha, S.Kadry [et al.] / IEEE Access, – 2020, 8, – p. 206303–206324.

334. Rosenberg, A., Hirschberg, J. V-Measure: A conditional entropy-based external cluster evaluation measure // Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Prague, Czech Republic, – 2007, – p. 410–420.

335. Rubinov, A.M. Classes and clusters in data analysis / A.M.Rubinov, N.V.Soukhorukova, J.Ugon // European Journal of Operational Research, – 2006, 173 (3), – p. 849–865.

336. Sánchez-Monedero, J. Simultaneous modelling of rainfall occurrence and amount using a hierarchical nominal-ordinal support vector classifier / J.Sánchez-Monedero, S.Salcedo-Sanz, P.Gutiérrez [et al.] // Engineering Applications of Artificial Intelligence, – 2014, 34, – p. 199–207.

337. Sardar, T.H. An analysis of MapReduce efficiency in document clustering using parallel K-means algorithm / T.H.Sardar, Z.Ansari // Future Computing and Informatics Journal, – 2018, 3, – p. 200–209.

338. Seneviratne, S. Self-supervised vision transformers for malware detection / S.Seneviratne, R.Shariffdeen, S.Rasnayaka [et al.] // IEEE Access, – 2022, 10, – p. 103121-103135.

339. Shang, W. Intrusion detection algorithm based on OCSVM in industrial control system / W.Shang, P.Zeng, M.Wan [et al.] // Security and Communication Networks, – 2016, 9 (10), – p. 1040-1049.

340. Shankarapani, M.K. Malware detection using assembly and API call sequences / M.K.Shankarapani, S.Ramamoorthy, R.S.Movva [et al.] // Journal in Computer Virology, – 2011, 2 (7), – p. 107–119.

341. Sheth, A. Physical-cyber-social computing: an early 21st century approach / A.Sheth, P.Anantharam, C.Henson // IEEE Intelligent Systems, – 2013, 28 (1), – p. 78–82.

342. Shin, J. Cyber-physical attack detection and recovery based on RNN in automotive brake systems / J.Shin, Y.Baek, J.Lee // Applied Sciences, – 2019, 9 (1), – p. 1–21.

343. Siaterlis, C. Cyber-physical testbeds / C.Siaterlis, B.Genge // Communications of the ACM, – 2014, 57(6), – p. 64–73.

344. Sicari, S. Security, privacy and trust in Internet of Things: the road ahead / S.Sicari, A.Rizzardi, L.A.Grieco [et al.] // Computer Networks, – 2015, 76, – p. 146–164.

345. Singh, D.K. Intrusion response prioritization based on fuzzy ELECTRE multiple criteria decision making technique / D.K.Singh, P.Kaushik // Journal of Information Security and Applications, – 2019, 48, – p. 1-8.

346. Singhal, A. Data Warehousing and Data Mining Techniques for Cyber Security / A.Singhal. – New York: Springer, – 2007. – 159 p.

347. Siroky, D. Navigating Random Forests and related advances in algorithmic modeling / D.Siroky // Statistics Surveys, – 2009, 3, – p. 147–163.

348. Smith, R. A decoupled feedback structure for covertly appropriating networked control systems / R. Smith // IFAC Proceedings Volumes, – 2011, 44 (1), – p. 90-95.

349. Sobhrajan, P. Comparative study of abstraction in cyber physical system / P.Sobhrajan, S.Y.Nikam // International Journal of Computer Science and Information Technologies, – 2014, 5 (1), – p. 466–469.

350. Song, Z. Abnormal detection method of industrial control system based on behavior model / Z.Song, Z.Liu // Computers & Security, – 2019, 84, – p. 166-178.
351. Stankovic, J.A., Research directions for the Internet of Things / J.A.Stankovic // IEEE Internet of Things Journal, 2014, 1 (1), p. 3–9.
352. Stellios, I. Risk assessment for IoT-enabled cyber-physical systems / I.Stellios, P.Kotzanikolaou, M.Psarakis [et al.] // Advances in Core Computer Science-Based Technologies, – 2021, 14, – p. 157–173.
353. Stisen, A., Blunck, H., Bhattacharya, S., Prentow, T.S., Kjærgaard, M.B., Dey, A., Sonne, T., Jensen, M.M. Smart devices are different: assessing and mitigating mobile sensing heterogeneities for activity recognition // ACM Conference on Embedded Networked Sensor Systems. Seoul, South Korea, – 2015, – p. 127-140.
354. Suh-Lee, C., Jo, J. Quantifying security risk by measuring network risk conditions, IEEE/ACIS International Conference on Computerand Information Science. Las Vegas, USA, – 2015, – p. 9–14.
355. Sukhostat, L.V. An intelligent model based on deep transfer learning for detecting anomalies in cyber-physical systems / L.V.Sukhostat // Radio Electronics, Computer Science, Control, – 2021, 3, – p. 124-132.
356. Sundaram, S., Pajic, M., Hadjicostis, C., Mangharam, R., Pappas, G. The wireless control network: monitoring for malicious behaviour / // IEEE Conference on Decision and Control. Atlanta, USA, – 2010, – p. 5979-5984.
357. Sundermeyer, M., Schlüter, R., Ney, H. LSTM neural networks for language modeling // Proceedings of the 13th Annual onference of the International Speech Communication Association. Portland, USA, – 2012, – p. 194–197.
358. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A. Going deeper with convolutions // IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, – 2015, – p. 1-9.
359. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z. Rethinking the inception architecture for computer vision // IEEE Conference on Computer Vision and Pattern Recognitionю Las Vegas, USA, – 2016, – p. 2818-2826.

360. Tang, R. Clustering big IoT data by metaheuristic optimized mini-batch and parallel partition-based DGC in Hadoop / R.Tang, S.Fong // Future Generation Computer Systems, – 2018, 86, – p. 1395–1412.

361. Tavallaee, M., Bagheri, E., Lu, W., Ghorbani, A. A detailed analysis of the KDD CUP 99 Data Set // IEEE Symposium on Computational Intelligence for Security and Defense Applications. Ottawa, Canada. – 2009. – p. 53–58.

362. Teixeira, A.M.H., Pérez, D., Sandberg, H., Johansson, K.H. Attack models and scenarios for networked control systems // 1st International Conference on High Confidence Networked Systems (HiCoNS'12). Beijing, China, – 2012, – p. 55-64.

363. Thompson, S.K. Sample size for estimating multinomial proportions / S.K.Thompson // The American Statistician, – 1987, 41(1), – p. 42–46.

364. Tzanetakis, G. Musical genre classification of audio signal / G.Tzanetakis, P.Cook // IEEE Transactions on Speech and Audio Processing, – 2002, 10 (3), – p. 293–302.

365. Ur-Rehman, A. Vulnerability modelling for hybrid industrial control system networks / A.Ur-Rehman, I.Gondal, J.Kamruzzaman [et al.] // Journal of Grid Computing, – 2020, 18, – p. 863–878.

366. Uslar, M., Rosinger, C., Schlegel, S. Security by design for the smart grid: combining the SGAM and NISTIR 7628 // IEEE International Computer Software and Applications Conference Workshops. Vasteras, Sweden, – 2014, – p. 110-115.

367. Vafaee, F. Learning the structure of large-scale bayesian networks using genetic algorithm // Conference on Genetic and Evolutionary Computation. Vancouver, Canada, – 2014, – p. 507-554.

368. Vega-Pons, S. A survey of clustering ensemble algorithms / S.Vega-Pons, J.Ruiz-Shulcloper // International Journal of Pattern Recognition and Artificial Intelligence, – 2011, 25 (3), – p. 337–372.

369. Vinodh, S. PROMETHEE based sustainable concept selection / S.Vinodh, R.J.Girubha // Applied Mathematical Modelling, – 2012, 36 (11), – p. 5301-5308.

370. VirusTotal: [Electronic resource] / Hispasec Sistemas, – 2023. URL: https://www.virustotal.com

371. Wang, E.K., Ye, Y., Xu, X., Yiu, S.M., Hui, L.C.K., Chow, K.P. Security issues and challenges for cyber physical system // IEEE/ACM International Conference on Cyber, Physical and Social Computing. Hangzhou, China, – 2010, – p. 733-738.

372. Wang, H. A feature extraction method based on information theory for fault diagnosis of reciprocating machinery / H.Wang, P.A.Chen // Sensors, – 2009, 9 (4), – p. 2415–2436.

373. Wang, J. Generalization of deep neural networks for imbalanced fault classification of machinery using generative adversarial networks / J.Wang, S.Li, B.Han [et al.] // IEEE Access, – 2019, 7, – p. 111168–111180.

374. Wang, L. Cloud-Based cyber-physical systems in manufacturing / Wang L., X.V. Wang. – London: Springer, – 2018, – 404 p.

375. Wei, D., Ji, K. Resilient industrial control system (RICS): concepts, formulation, metrics, and insights // IEEE International Symposium on Resilient Control Systems. Idaho Falls, USA, – 2010, – p. 15–22.

376. Williams, T.J. A reference model for computer integrated manufacturing from the viewpoint of industrial automation / T.J.Williams // IFAC Proceedings Volumes, – 1990, 23 (8), – p. 281–291.

377. World Economic Forum (WEF) Launches Centre for Fourth Industrial Revolution in Azerbaijan: [Electronic resource] / World Economic Forum, – 2022. URL: https://www.weforum.org/press/2022/02/world-economic-forum-launches-centre-for-fourth-industrial-revolution-in-azerbaijan/

378. Xia, Z. Spectral regression based fault feature extraction for bearing accelerometer sensor signals / Z.Xia, S.Xia, L.Wan [et al.] // Sensors, – 2012, 12 (10), – p. 13694–13719.

379. Xiao, F. Power system risk assessment and control in a multobjective framework / F.Xiao, J.D.McCalley // IEEE Transactions on Power Systems, – 2009, 24 (1), – p. 78–85.

380. Xiao, M. Image-based malware classification using section distribution information / M.Xiao, C.Guo, G.Shen [et al.] // Computers & Security, – 2021, 110, – p. 1-14.

381. Xiao, Y. Hybrid LSTM neural network for short-term traffic flow prediction / Y.Xiao, Y.Yin // Information, – 2019, 10 (105), – p. 1–22.

382. Xie, M. Anomaly detection in wireless sensor networks: a survey / M.Xie, S.Han, B.Tian, S. Parvin // Journal of Network and Computer Applications, – 2011, 34, – p. 1302–1325.

383. Xin, S. Cyber-physical modeling and cybercontingency assessment of hierarchical control systems / S.Xin, Q.Guo, H.Sun [et al.] // IEEE Transactions on Smart Grid, – 2015, 6 (5), – p. 2375–2385.

384. Xinlan, Z., Zhifang, H., Guangfu, W., Xin, Z. Information security risk assessment methodology research: group decision making and analytic hierarachy process // Second World Congress on Software Engineering. Wuhan, China, – 2010, – p. 157-160.

385. Xu, D. Automated security test generation with formal threat models / D.Xu, M.Tu, M.Sanford [et al.] // IEEE Transactions on Dependable and Secure Computing, – 2012, 9 (4), – p. 525–539.

386. Xu, G. Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning / G.Xu, M.Liu, Z.Jiang [et al.] // Sensors, – 2019, 19 (5), – p. 1-21.

387. Xu, Y. Household Service Robotics. Intelligent Systems Series / Y.Xu, H.Qian, X.Wu. – Cambridge: Academic Press, – 2015. – 564 p.

388. Yadav, P. A two-stage deep learning framework for image-based android malware detection and variant classification / P.Yadav, N.Menon, V.Ravi [et al.] // Computational Intelligence, – 2022, 38 (5), – p. 1748-1771.

389. Yan, H. Automatic malware classification via PRICoLBP / H.Yan, H.Zhou, H.Zhang // Chinese Journal of Electronics, – 2018, 27 (4), – p. 852-859.

390. Yang, B. Malicious software spread modeling and control in cyber-physical systems / B.Yang, Z.Yu, Y.Cai // Knowledge-Based Systems, – 2022, 248, – p. 1-14.

391. Yang, W., Haider, S.N., Zou, J., Zhao, Q. Industrial Big data platform based on open source software // Proceedings of the International Conference on Computer Networks and Communication Technology (CNCT). Xiamen, China, – 2016, – p. 649-658.

392. Yang, X.-S. Firefly algorithms for multimodal optimization // Stochastic Algorithms: Foundations and Applications, – 2009, 5792, – p. 169–178.

393. Yu, J. Bearing performance degradation assessment using locality preserving projections and Gaussian mixture models // Mechanical Systems and Signal Processing, – 2011, 25 (7), – p. 2573-2588.

394. Yu, Z. SEI2RS malware propagation model considering two infection rates in cyber–physical systems / Z.Yu, H.Gao, D.Wang [et al.] // Physica A, – 2022, 597, – p. 1-12.

395. Zeadally, S. Cyber-Physical system design with sensor networking technologies / S.Zeadally, N.Jabeur. – London: The Institution of Engineering and Technology, – 2016. – 340 p.

396. Zegeye, W.K. Multi-layer hidden Markov model based intrusion detection system / W.K.Zegeye, R.A.Dean, F. Moazzami // Machine Learning and Knowledge Extraction, – 2019, 1 (1), – p. 265–286.

397. Zeng, J. A survey: cyber-physical-social systems and their system-level design methodology / J.Zeng, L.T.Yang, M.Lin [et al.] // Future Generation Computer Systems, – 2016, 105, – p. 1028-1042.

398. Zhang, F. Matrix Theory: Basic Results and Techniques. – New York: Springer, – 2011. – 399 p.

399. Zhang, M.-L. A k-nearest neighbor based algorithm for multi-label classification / M.-L.Zhang, Z.-H.Zhou // Proceedings of IEEE International Conference on Granular Computing. Beijing, China. – 2005, – p. 718–721.

400. Zhang, Z. MIDCAN: A multiple input deep convolutional attention network for COVID-19 diagnosis based on chest CT and chest X-ray / Z.Zhang, X.Zhang // Pattern Recognition Letters, – 2021, 150, – p. 8-16.

401. Zhao, H. A multi-classification method of improved SVM-based information fusion for traffic parameters forecasting / H.Zhao, D.Sun, M.Zhao [et al.] // Promet-Traffic&Transportation, – 2016, 28 (2), – p. 117–124.

402. Zhao, R. Deep learning and its applications to machine health monitoring / R.Zhao, R.Yan, Z.Chen [et al.] // Mechanical Systems and Signal Processing, – 2019, 115, – p. 213–237.

403. Zhong, F. Malware-on-the-Brain: Illuminating malware byte codes with images for malware classification / F.Zhong, Z.Chen, M.Xu [et al.] // IEEE Transactions on Computers, – 2022, 72 (2), – p. 438 – 451.

404. Zhou, X., Xu, Z., Wang, L., Chen, K., Chen, C., Zhang, W. Kill chain for industrial control system. International Conference on Smart Materials // Intelligent Manufacturing and Automation. Nanjing, China, – 2018, – p. 1-5.

405. Ziou, D. Scale space Radon transform / D.Ziou, N.Nacereddine, A.B.Goumeidane // IET Image Processing, – 2021, 15 (9), – p. 2097-2111.

406. Zobel, C.W. Comparative visualization of predicted disaster resilience // International ISCRAM Conference. Seattle, USA, – 2010, – p. 1-6.

407. Zobel, C.W. Quantitatively representing nonlinear disaster recovery / C.W.Zobel // Decision Sciences, – 2014, 45 (6), – p. 1053-1082.

408. Zobel, C.W. Representing perceived tradeoffs in defining disaster resilience / C.W. Zobel // Decision Support Systems, – 2011, 50 (2), – p. 394–403.

409. Zygouras, N., Zacheilas, N., Kalogeraki, V., Kinane, D., Gunopulos, D. Insights on a scalable and dynamic traffic management system // Proceedings of the 18th International Conference on Extending Database Technology. Brussels, Belgium, – 2015, – p. 653-664.

Список сокращений и условных обозначений

КФС – киберфизическая система

ИТ – информационные технологии

ОТ – операционные технологии

БГА – байесовский граф атак

CPS – Cyber-Physical System – киберфизическая система

SCADA – Supervisory Control and Data Acquisition – диспетчерское управление и сбор данных

NIST – National Institute of Standards and Technology – Национальный институт стандартов и технологий

IoT – Internet of Things – Интернет вещей

ICS – Industrial Control System – промышленная система управления

SVM – support vector machines – машины опорных векторов

MFCC – mel-frequency cepstral coefficients – мел-частотные кепстральные коэффициенты

HMM – Hidden Markov Models – скрытые модели Маркова

HHMM – Hierarchical Hidden Markov Models – иерархические скрытые модели Маркова

GMM – Gausian Mixture Models – смешанные гауссовы модели

ESP – Electric Submersible Pump – погружной насос центробежного типа

CWT – Continuous Wavelet Transform – непрерывное вейвлет преобразование

EER – equal error rate – равный уровень ошибок

FAR – false acceptance rate – ошибка ложного принятия

FFT – fast Fourier transform – быстрое преобразование Фурье

ISO – International Organization for Standardization – Международная организация по стандартизации

IEC – International Electrotechnical Commission – Международная электротехническая комиссия

PLC – Programmable logic controller – Программируемый логический

контроллер

CVSS – Common Vulnerability Scoring System – Общая система оценки уязвимостей

DBSCAN – density-based clustering algorithm – алгоритм кластеризации на основе плотности

DBN – deep belief network – сети глубокого доверия

RNN – recurrent neural network – рекуррентные нейронные сети

DNN – Deep Neural Network – глубокая нейронная сеть

CNN – convolutional neural network – сверточная нейронная сеть

LSTM – long short-term memory – сети с долгой краткосрочной памятью

MFTP – Machinery Failure Prevention Technology – технология предотвращения отказов оборудования

DoS – Denial-of-Service – отказ в обслуживании

BAG – Bayesian attack graph – байесовский граф атак

DDoS – distributed denial-of-service – распределенная атака типа «отказ в обслуживании»

RF – random forest – случайный лес

BN – Naïve Bayes – наивный Байес

DT – decision tree – деревья решений

KNN – k-nearest neighbors algorithm – алгоритм к-ближайших соседей

ANN – artificial neural network – искусствнная нейронная сеть

RBF – radial basis function – радиально-базисная функция

XGBoost – extreme gradient boosting – экстремальный градиентный бустинг

AUC – Area under ROC Curve – площадь под ROC-кривой