ISSN 2411-3336; e-ISSN 2541-9404

ЗАПИСКИ ГОРНОГО ИНСТИТУТА

Journal of Mining Institute

Сайт журнала: pmi.spmi.ru

Научная статья

Выявление аномалий в технологическом процессе очистки сточных вод для оценки рисков киберустойчивости

Е.С.Новикова[1], **Е.В.Федорченко**[1]✉, **М.А.Бухтияров**[2], **И.Б.Саенко**[1]

[1] *Санкт-Петербургский Федеральный исследовательский центр РАН, Санкт-Петербург, Россия*

[2] *ООО «Webim», Москва, Россия*

Как цитировать эту статью: Новикова Е.С., Федорченко Е.В., Бухтияров М.А., Саенко И.Б. Выявление аномалий в технологическом процессе очистки сточных вод для оценки рисков киберустойчивости // Записки Горного института. 2024. Т. 267. С. 488-500. EDN TBPPHN

Аннотация. Своевременное выявление и предотвращение нарушений в технологическом процессе сточных вод в результате реализации угроз разной природы является актуальной задачей. Современные системы снабжены большим количеством технологических датчиков. Данные этих датчиков могут использоваться для выявления аномалий в технологическом процессе. Их своевременное выявление, прогнозирование и обработка обеспечит непрерывность и отказоустойчивость технологического процесса. Цель исследования – повышение точности обнаружения таких аномалий. Предлагается методика выявления и последующей оценки рисков киберустойчивости технологического процесса очистки сточных вод, включающая оригинальное формирование обучающих наборов данных и выявление аномалий на основе методов глубокого обучения. Наличие обучающих наборов данных – необходимое условие эффективной работы методики. Отличительная особенность методики выявления аномалий – новый метод обработки данных технологических датчиков, который позволяет использовать вычислительно эффективные аналитические модели с высокой точностью обнаружения аномалий и превосходит результаты ранее опубликованных методов.

Ключевые слова: системы водоочистных сооружений; промышленные киберфизические системы; киберустойчивость; риски; выявление аномалий; обучающие наборы; тестовый стенд

Благодарность. Работа выполнена при поддержке гранта Российского научного фонда № 23-11-20024 и Санкт-Петербургского научного фонда.

Поступила: 07.03.2024 ***Принята: 14.06.2024*** ***Онлайн: 04.07.2024*** ***Опубликована: 04.07.2024***

Введение. Системы комплексных очистных сооружений относятся к объектам критической инфраструктуры, от устойчивого функционирования которых зависит безопасность населения. Нарушения в технологическом процессе водоочистных сооружений могут привести к непоправимым последствиям для здоровья и экологии [1, 2].

Водоочистные сооружения снабжены автоматизированными системами управления для мониторинга и своевременного управления процессами очистки воды [3-5]. Внедрение таких систем приводит к возникновению новых рисков нарушения киберустойчивости в результате кибератак. Так, в 2000 г. была проведена кибератака на водоочистные сооружения в Маручи (Австралия). В результате за три месяца в дождевую канализацию был сброшен один миллион литров неочищенных сточных вод. В последние годы количество киберугроз, направленных на системы очистки воды, растет. В 2021 г. целью кибератаки стала водоочистная станция Oldsmar в США, в результате которой злоумышленнику удалось повысить уровень гидроксида натрия в воде[1].

[1] 21-015 Detectives Investigate Computer Software Intrusion at Oldsmar's Water Treatment Plant. URL: https://pcsoweb.com/21-015-detectives-investigate-computer-software-intrusion-at-oldsmar%E2%80%99s-water-treatment-plant (дата обращения 07.03.2024).

Записки Горного института. 2024. Т. 267. С. 488-500

В 2022 г. злоумышленники реализовали киберугрозу против компании South Staffs Water в Великобритании с использованием программы-вымогателя Clop[2].

Для мониторинга управления процессами очистки воды автоматизированные системы управления собирают и анализируют различные технологические данные. Эти данные могут быть использованы для выявления аномалий, возникающих в результате реализации угроз, прогнозирования их появления и оценки рисков нарушения киберустойчивости системы. Своевременное выявление, прогнозирование и обработка таких аномалий и связанных с ними рисков обеспечит непрерывность и отказоустойчивость технологического процесса [6-8].

В последнее время было предложено большое количество различных методов обнаружения аномалий в функционировании киберфизических объектов [9], и основной фокус в научных исследованиях делается на применение глубоких нейронных сетей, что связано с их способностью моделировать сложные нелинейные зависимости между различными параметрами объекта, выявлять временные и пространственные паттерны в данных [10]. В частности, в статье [11] для выявления аномалий в данных от систем водоподготовки и водоочистных сооружений предложена модель вариационного автоэнкодера MTS-DVGAN на основе двух сетей долгой краткосрочной памяти (LSTM-сетей). Выполненные эксперименты показали, что применение такой модели позволяет с высоким уровнем эффективности (до 97,84 %) выявлять отклонения в функционировании системы водоподготовки. В исследовании [12] представлена модель MTAD-CAN, состоящая из автоэнкодера и двух декодеров с механизмом сопряженного внимания (coupled attention), предназначенного для извлечения как временных зависимостей между параметрами многомерного временного ряда, так и корреляционных связей между самими параметрами. Эксперименты с данными от системы водоочистных сооружения показали точность обнаружения аномалий до 92 %. Исследованы методы спектрального анализа временных для выявления дефектов в оборудовании технологических процессов, которые могут привести к нарушению киберустойчивости системы [13], и предложено решение, сочетающее методы визуального анализа данных и машинного обучения [14].

Выявленные аномалии могут применяться для оценки и прогнозирования рисков киберустойчивости технологического процесса [15-17]. В качестве входных данных зачастую используются журналы событий и сетевой трафик, а также данные мониторинга физических датчиков [18-20]. Для анализа и прогнозирования рисков киберустойчивости необходимо учесть вероятность успешной реализации угроз и возможный ущерб в случае их успешной реализации [21]. Для определения вероятности успешной реализации угроз используют различные модели угроз [22-24], для определения которых необходимо вначале определить модель анализируемой системы или технологического процесса.

Модели угроз могут быть представлены в виде графов и марковских цепей, а также построены с использованием методов машинного обучения для выявления аномалий. Для оценивания могут применяться табличные методы, методы на основе графов и теории вероятностей. Одной из существующих проблем является анализ ущерба в случае реализации угроз на системы водоочистки (или другие системы автоматизированного управления) как одного из компонентов риска [25]. С этой целью может использоваться экспертная оценка потенциального ущерба либо моделирование распространения загрязнений и учет стоимости их устранения.

Для определения риска киберустойчивости применяются интегральные метрики уровня риска как качественные, так и количественные. При формировании интегральных метрик может использоваться табличный подход, весовая функция, минимаксный подход и др. При этом одна из задач состоит в определении метрик, входящих в интегральную оценку, разработке методики их интеграции в оценку риска и анализе чувствительности предложенной интегральной оценки.

[2] South Staffs Water is victim of botched Clop attack. URL: https://www.computerweekly.com/news/252523856/South-Staffs-Water-is-victim-of-botched-Clop-attack (дата обращения 07.03.2024).

Модели угроз, построенные с использованием методов машинного обучения, позволяют с высокой точностью выявлять и прогнозировать аномалии, являющиеся результатом реализации разного рода угроз. Однако их применение связано с решением двух ключевых практических задач: созданием подготовленного набора данных для обучения модели анализа и наличием вычислительных ресурсов. Подготовленные данные – реалистичные структурированные данные, которые имеют разметку в виде описания состояния объекта в различные моменты времени. Анализ наборов данных, находящихся в открытом доступе, показал, что подобных наборов крайне мало [26, 27]. Кроме того, представленные в них аномалии являются тривиальными, и их выявление не требует применения методов машинного обучения [28]. Одной из возможных причин отсутствия достоверных, размеченных наборов данных, описывающих функционирование киберфизических систем, является отсутствие единой методологии их создания. В настоящей работе предлагается методика формирования наборов данных, моделирующих функционирование очистных сооружений на уровне технологического процесса. Она регламентирует этапы, начиная от выбора технологического процесса, заканчивая спецификацией модели нарушителя и возможных деструктивных воздействий.

Применение методов обнаружения аномалий на основе глубокого обучения предъявляет высокие требования к вычислительным ресурсам, что на практике не всегда возможно. Таким образом, актуальна задача разработки методов выявления и прогнозирования аномалий, оптимизированных под устройства с ограниченными вычислительными ресурсами, например промышленные микрокомпьютеры, которые не оборудованы графическими ускорителями и при этом обеспечивают эффективность обнаружения аномалий, сравнимую с классическими моделями глубокого обучения. В настоящей работе такая задача решается использованием специального преобразования вектора данных в изображение, что позволяет применять «легковесные» сверточные нейронные сети небольшой глубины.

Цель исследования – повышение точности обнаружения аномалий в технологическом процессе очистки сточных вод для оценки рисков киберустойчивости с учетом ограничений по используемым вычислительным ресурсам.

Задачи исследования включают разработку методики обнаружения и оценки рисков в технологическом процессе очистки сточных вод на основе машинного обучения; разработку методики формирования наборов данных, моделирующих функционирование очистных сооружений на уровне технологического процесса; разработку и тестирование методики выявления аномалий в потоке данных от технологического оборудования в режиме реального времени; разработку метода преобразования входного вектора данных в изображения.

Разработана методика обнаружения и оценки рисков в технологическом процессе очистки сточных вод на основе машинного обучения. Такая методика включает этапы формирования обучающего набора данных, выявления аномалий в потоке данных от технологического оборудования в режиме реального времени и формирования динамических оценок рисков с учетом обнаруженных аномалий. Для выявления аномалий в технологическом процессе предложен метод преобразования входного вектора данных в изображения, который позволяет использовать сверточные нейронные сети небольшой глубины.

Методы. Предлагается методика выявления и оценки рисков киберустойчивости в технологическом процессе очистки сточных вод на основе машинного обучения. Она включает методику формирования обучающего набора данных, необходимого для выявления аномалий с использованием метода машинного обучения, и методику анализа данных от датчиков технологического процесса сточных вод.

Методика выявления и оценки рисков киберустойчивости в технологическом процессе очистки сточных вод. Основана на анализе данных, получаемых от автоматизированной системы управления процессом очистки сточных вод (рис.1). Методика включает два режима работы – режим проектирования и эксплуатации – и три основных этапа – формирование набора данных, выявление аномалий и оценка рисков киберустойчивости. Этап формирования набора данных

Записки Горного института. 2024. Т. 267. С. 488-500

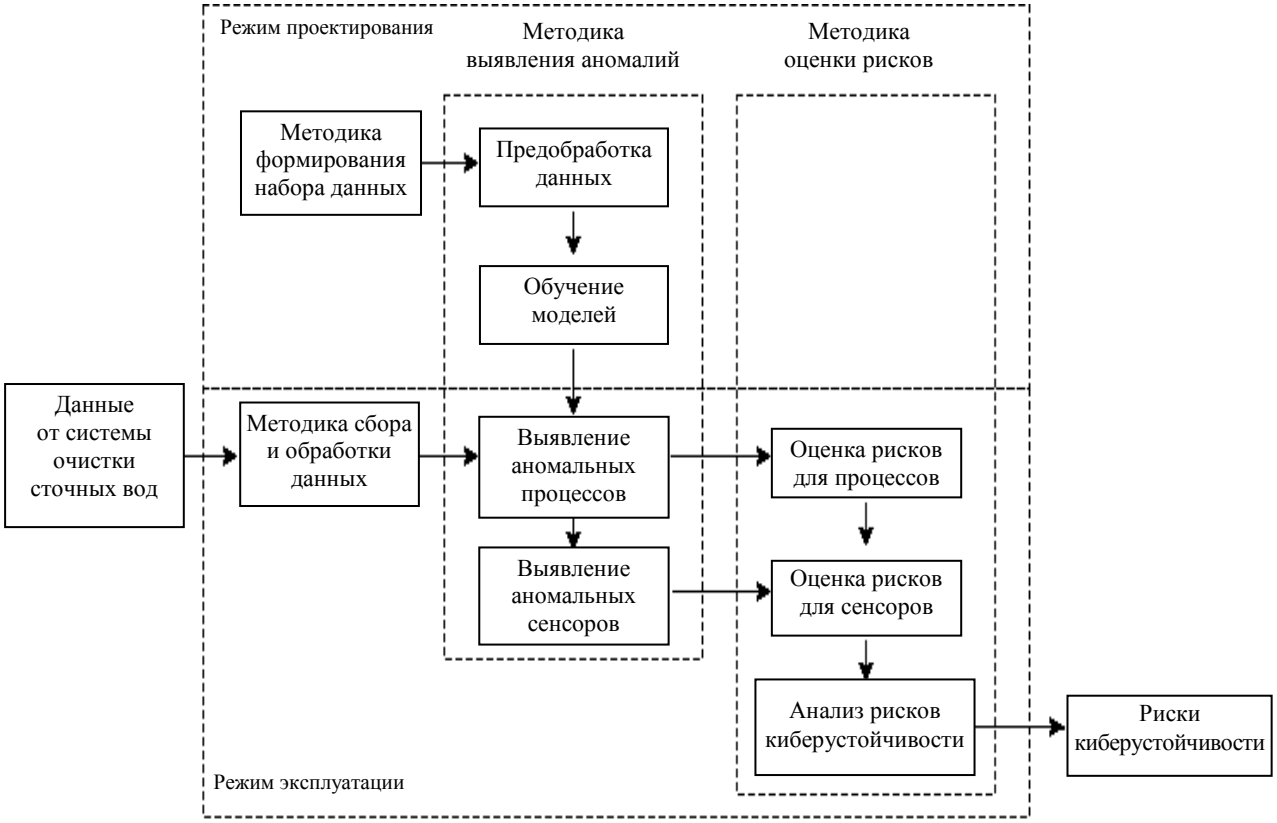

Рис.1. Схема методики выявления и оценки рисков в технологическом процессе очистки сточных вод

представлен методикой формирования обучающего набора данных. Выходными данными методики является набор входных данных для работы следующего этапа – обучения моделей для обнаружения аномалий в технологическом процессе в режиме проектирования, представленного методикой выявления аномалий. В режиме эксплуатации обученные модели и данные от сенсоров/актуаторов системы управления технологическим процессом используются для обнаружения аномалий. Выходными данными методики являются выявленные аномальные процессы и сенсоры/актуаторы. В свою очередь эти данные – входные для следующего этапа – оценка рисков киберустойчивости, представленного методикой оценки рисков киберустойчивости технологического процесса очистки сточных вод.

Методика формирования обучающего набора данных. Разработанная методика учитывает требования, сформулированные в [27-29]:

• Включение данных как от датчиков технологического процесса, так и коммуникационно-вычислительной инфраструктуры системы, т.е. данные сетевого трафика, журналы автоматизированных систем управления и т.д.

• Наличие разметки и структурированной схемы аннотаций, поясняющей разметку и включающей информацию об аномалиях в технологическом процессе, в том числе возможную причину – преднамеренное или непреднамеренное воздействие на процесс.

• Максимальное приближение к реальным данным, а имеющиеся ограничения, в силу использования упрощенных математических моделей или ограниченных возможностей используемых аппаратно-технических средств, должны быть задокументированы.

Таким образом, методика генерации наборов данных, моделирующих функционирование технологического процесса, состоит из следующих этапов:

• определение и спецификация технологического процесса;

• определение типа тестового стенда и его реализация;

• генерация данных, соответствующих нормальному функционированию системы;

• разработка модели угроз непрерывности и отказоустойчивости рассматриваемого технологического процесса с учетом возможных последствий их нарушения;

• разработка сценариев реализации угроз с учетом используемого технологического стека для моделирования технологического процесса;

• реализация сценариев угроз и сбор данных о поведении технологического процесса;

• оценка и валидация сформированного набора данных.

Методика определяет вход и выход выполнения каждого этапа, в том числе документально, поскольку каждый этап зависит от результата выполнения предыдущего. Кроме того, это позволяет при необходимости воспроизвести моделируемый процесс, валидировать полученные данные и объяснять полученные результаты. Например, в ходе выполнения первого этапа работы методики формируются формализованная модель технологического процесса с заданным множеством значимых параметров, влияющих на непрерывность и отказоустойчивость технологического процесса, и технологическая схема. По результатам выполнения данного этапа определяется тип создаваемого тестового стенда. Согласно статье [27], существуют три типа тестовых стендов: программный (виртуальный), аппаратный и гибридный. В программном стенде применяются только программные средства для моделирования технологического процесса. Такие стенды обладают низкой стоимостью реализации и высоким уровнем воспроизводимости. В некоторых случаях, например при моделировании процессов, выполняемых в опасных условиях, использование таких стендов является единственно возможным решением. Однако математическое моделирование многих процессов является крайне сложной задачей, например при описании флотационного процесса очистки воды необходимо построить гидродинамическую модель процесса с большим числом управляющих воздействий и контролируемых переменных с учетом физико-химического взаимодействия веществ [30, 31], поэтому результаты их использования могут быть менее точными и достоверными. Аппаратные стенды строятся с применением специализированного технологического оборудования и программных средств, поэтому данные, получаемые с их помощью, достоверны. Они отражают возможные задержки и неточности в данных, возникающие при использовании физических устройств и датчиков. Существенный недостаток – стоимость разработки таких стендов и, как следствие, низкий уровень воспроизводимости. Гибридные тестовые стенды являются компромиссом между программными и аппаратными стендами. Для их реализации часть технологических процессов моделируется программно, часть – с помощью технологического оборудования. Таким образом, на втором этапе с учетом выбранного типа стенда прорабатывают особенности его построения, а также определяют формат собираемых данных и интервал их получения.

Результаты первого этапа также влияют на модель угроз и сценариев их выполнения. Например, для флотационного процесса очистки воды могут быть определены следующие возможные деструктивные цели: остановка смешивания сырой воды; повреждение расходомера сырой воды, насоса для подачи реагентов, датчиков кислотности; отсутствие коагулянта в резервуаре, флокулянта в баке, соды в резервуаре и т.д. Угрозы могут быть реализованы на физическом, информационно-коммуникационном уровне. К физическому уровню относятся физические устройства, например датчики потока воды, потока воздуха, воды, цифровые преобразователи, контроллеры и телекоммуникационное оборудование. К информационно-коммуникационному уровню – информационные потоки от датчиков к контроллерам, от контроллеров к SCADA-системе и т.д. К логическим компонентам – микропрограммы и программное обеспечение.

На заключительном этапе происходит оценка и валидация сформированного набора данных. Они включают статистический анализ полученных данных, а также при наличии реальных данных – выполнение сравнительного анализа сформированных и реальных данных путем оценки разницы между распределениями вероятностей реальных данных и синтетических.

Собранный в результате набор данных может использоваться для анализа данных технологического процесса.

Записки Горного института. 2024. Т. 267. С. 488-500

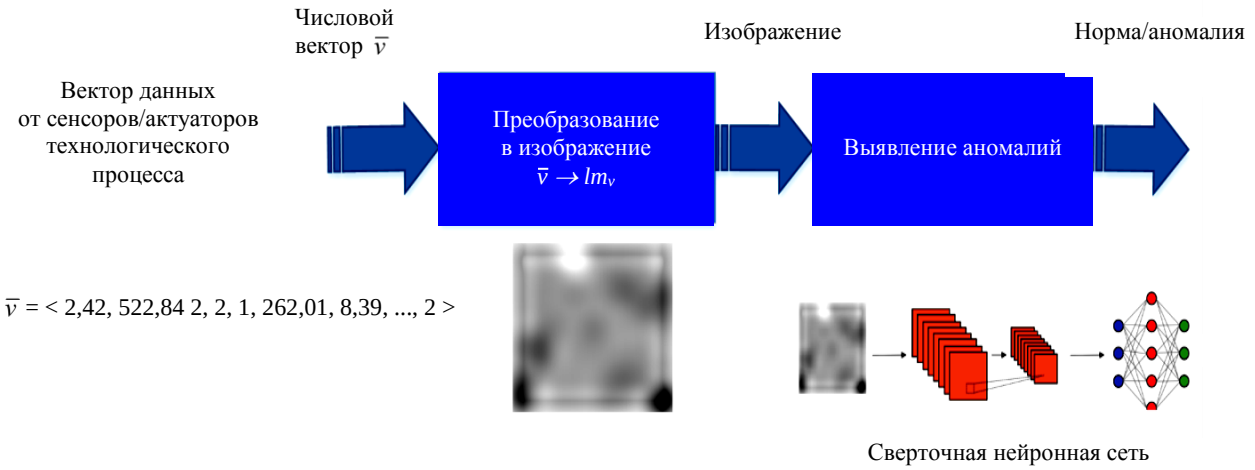

Рис.2. Схема процесса обнаружения аномалий в технологических процессах

Методика выявления аномалий, основанная на преобразовании табличных данных в изображение. В основе предложенной методики лежит идея использования сверточных нейронных сетей, которые хорошо извлекают пространственные связи между атрибутами. Их применение требует преобразования вектора входных данных к двумерной матрице, поскольку операция двумерной свертки лучше извлекает пространственные взаимосвязи. Пусть $\bar{v} = \{v_0, v_1, \ldots v_k\}$ – входной вектор значений от k анализируемых датчиков и актуаторов системы, тогда предлагаемая методика состоит из следующих этапов:

• преобразование каждого одномерного вектора признаков $\bar{v}$ в двумерную матрицу Im_v, обычно рассматриваемую как полутоновое изображение;

• обнаружение аномалий с использованием сверточной нейронной сети (CNN).

Ключевые этапы методики представлены на рис.2.

Для построения изображения на основе одномерного вектора $\bar{v}$ выполняются следующие шаги:

• начальная предобработка значений признаков, которая заключается в их нормализации в диапазоне [0,1] для числовых атрибутов или кодировании значений в формате «one-hot» для категориальных атрибутов;

• предобработанные значения признаков интерпретируются как 8-битные значения для полутонового изображения;

• компоновка пикселей по заданному алгоритму в изображение.

Для определения координат атрибутов-пикселей могут быть использованы алгоритмы прямой компоновки и на основе нелинейного преобразования, зависящего от подобия атрибутов [32-34].

Алгоритм прямой компоновки пикселей является наиболее распространенным подходом к созданию изображения. Ключевым моментом в его работе является определение размерности генерируемого изображения. Обычно входом для сверточных нейронных сетей служат изображения размером $n \times n$, поэтому для определения размера изображения используется следующий подход. Пусть N – количество числовых анализируемых признаков, а M – число значений, которые могут принимать все категориальные атрибуты, тогда $n = \text{ceil}(N + M)/2$, где ceil – функция округления в большую сторону до ближайшего целого числа. Далее изображение генерируется построчно, каждая строка формируется последовательно. Неиспользуемые пиксели обычно заполняются значением 0x00 (черный цвет). Таким образом, при прямой компоновке пикселей порядок признаков в векторе и форма изображения определяют положение соответствующего пикселя на изображении, в результате чего соседние пиксели могут не иметь корреляции друг с другом.

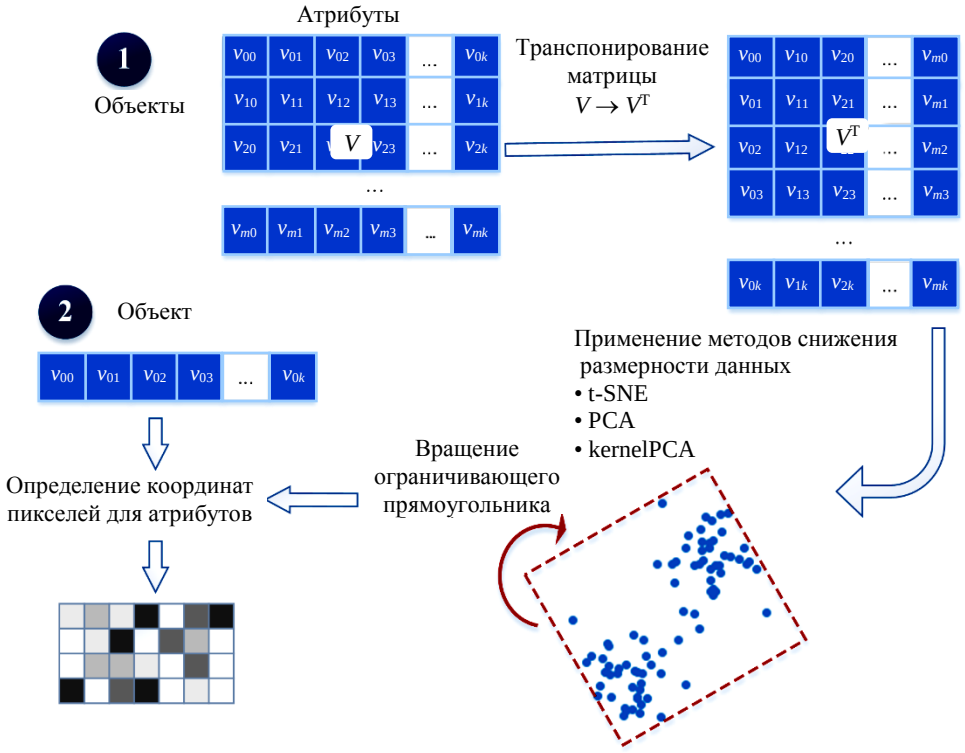

Рис.3. Схема определения координат атрибутов с помощью нелинейного преобразования

Использование алгоритмов на основе нелинейного преобразования, таких как DeepInsight [33, 35], позволяет учесть взаимосвязи между признаками и размещать похожие атрибуты близко друг к другу. Основная идея – использование методов проецирования многомерных данных в пространство меньшей размерности для определения положения данного признака на двумерной плоскости. Схема этого подхода представлена на рис.3.

Пусть $V = \{\overline{v}_i\}_{i=0}^{m}$ – исходный (обучающий) набор данных, представленный в виде матрицы, где m строк соответствуют m векторам с k атрибутами. Тогда процедура создания изображения с использованием алгоритма DeepInsight включает следующие шаги:

• Транспонирование матрицы V так, чтобы каждый признак был представлен вектором из m элементов, т.е. соответствующей строкой транспонированной матрицы V^T.

• Применение метода снижения размерности данных для отображения каждого признака на двумерную плоскость. На этом шаге используются нелинейные методы сокращения размерности, такие как kernelPCA, t-SNE [36]. Полученные проекции определяют положение признаков, но не их значения.

• Построение наименьшего прямоугольника, который охватывает все проекции атрибутов, и вращение их координат в декартовом пространстве, чтобы получить координаты пикселей для признаков.

• Сопоставление атрибутов координатам пикселей в полученной проекции.

Преобразование табличных данных в матричную структуру (изображение) характеризуется дополнительными вычислительными и временными затратами, необходимыми для определения расположения признаков на сетке изображения. Однако данное преобразование выполняется только на этапе обучения и не влияет на вычислительные и временные характеристики процесса обнаружения аномалий в режиме реального времени.

Для обнаружения аномалий было предложено использовать двуслойную сверточную нейронную сеть, структура которой представлена на рис.4. Она состоит из двух сверточных и двух полносвязных слоев. В качестве функции активации в подвыборочных слоях используется функция ReLU, а в последнем полносвязном слое – сигмоидная функция. При обучении применяется оптимизатор Adam, а в качестве функции потерь – логарифмическая функция потерь.

Записки Горного института. 2024. Т. 267. С. 488-500

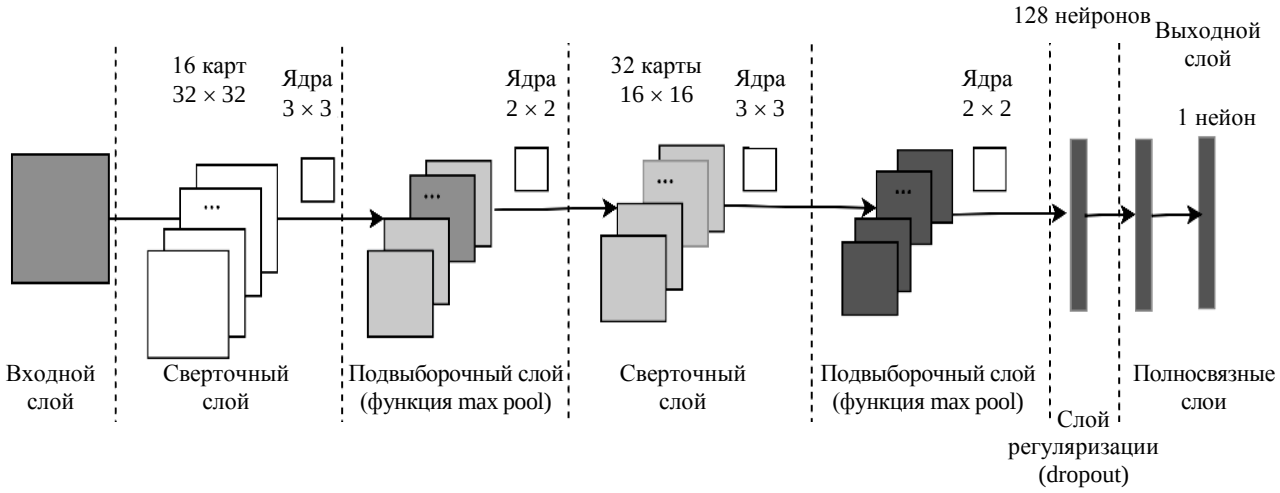

Рис.4. Структура сверточной нейронной сети, используемой для обнаружения аномалий

Методика оценки рисков киберустойчивости технологического процесса очистки сточных вод. Выявленные аномалии в процессах очистки сточных вод могут привести к нарушению киберустойчивости. Поэтому они применяются для оценки ее рисков. При этом само по себе наличие отдельной аномалии не свидетельствует о реализации рисков киберустойчивости, оценка риска растет с ростом количества аномалий. Отметим, что выделяются риски для процессов и отдельных сенсоров/актуаторов, измеряющих различные параметры процесса. Риски киберустойчивости определяются с учетом критичности процессов и соответственно ущерба, который может быть нанесен технологическому процессу очистки воды в случае нарушения киберустойчивости процесса.

Входными данными методики являются выявленные аномалии, а также описание процессов очистки сточных вод, их критичности и сенсоров/актуаторов, измеряющих различные параметры процессов. Для оценки рисков киберустойчивости разработан алгоритм. Он основан на следующих предположениях: уровень риска выше для процессов, которые чаще демонстрируют аномальное поведение; уровень риска выше для более критичных процессов. Кроме того, вводится модель анализируемой системы:

$$M = < P_i: R_i, Pr_i, Cr_i, S^i >,$$

где P_i – процесс очистки сточных вод, $i \in \{1{:}n\}$; n – количество процессов, реализующих очистку сточных вод; S^i – множество сенсоров/актуаторов; Pr_i – вероятность аномалии в процессе P_i; Cr_i – критичность процесса P_i, $Cr_i \in \{1,2,3\}$, определяется экспертно на шкале: 1 – низкая, 2 – высокая, 3 – критичная; R_i – риск, $R_i \in \{0{:}3\}$.

Риск определяется комбинацией критичности процесса и вероятностью реализации киберугроз, которая зависит от количества выявленных аномалий. Отметим, что каждая аномалия также имеет вероятность, связанную с точностью работы алгоритмов машинного обучения. Входные данные алгоритма – общее количество записей, содержащих показания сенсоров/актуаторов технологического процесса, собранных за весь период мониторинга; количество последовательных записей, выделяемых для анализа и называемых партией, m; общее количество партий за весь период мониторинга n. Уровень риска R определяется на шкале [0,6], изначально $R = 0$. Алгоритм вычисления уровня риска:

1. $Pr = 0$ // исходная вероятность реализации киберугрозы.

2. $R = 0$ // исходный уровень риска.

3. Если $R = 0$, переходим к шагу 4, иначе – к шагу 5 // если риск для процесса еще не оценивался ($n = 0$).

4. $R = R\_proc = Cr$ // риск вычисляется на основе критичности процесса.

5. Если $n = 0$, $Pr = 0$, переходим к шагу 9.

6. Если $w_i > 0$, переходим к шагу 7, иначе – к шагу 10 // в текущей партии были аномалии.

7. $w = \frac{\sum_k Pr_k}{n}$ // $\sum_k Pr_k$ – сумма вероятностей аномалий для аномальных партий, из партии берем max Pr; w – доля партий с аномалиями, $0 \le w \le 1$.

8. Если $Pr + \log_n(1 +) \le 1$, $Pr = Pr + \log_n(1 + w)$, иначе $Pr = 1$ // вероятность реализации киберугрозы растет.

9. $R = R + Cr \cdot Pr$.

10. Если $norm > N\_norm$, переходим к шагу 11 // $norm$ – количество последовательных партий без аномалий, $N\_norm = 10$.

11. $coeff = (n – an)/n$ // an – общее количество партий с аномалиями.

12. Если $coeff > R\_coeff$, переходим к шагу 13, иначе – к шагу 14 // R_coeff – коэффициент, определяющий максимальное возможное снижение риска, $R\_coeff = 0{,}3$, определен экспериментальным путем.

13. $coeff = R\_coeff$.

14. $Pr = Pr – coeff$.

15. $R = R – Cr \cdot Pr$.

Выходными данными методики являются оценки уровня рисков киберустойчивости для отдельных процессов, сенсоров/актуаторов и технологического процесса очистки сточных вод в целом.

Обсуждение результатов. Выявление аномалий является одним из важнейших этапов обнаружения и оценки рисков киберустойчивости в технологическом процессе очистки сточных вод. В качестве обучающей выборки использован открытый набор данных, для формирования которого применялась методика, наиболее близкая к предложенному авторами формированию набора данных. Secure Water Treatment (SWaT) [37] создан с помощью программно-аппаратного стенда, который представляет собой уменьшенную копию реальной системы по очистке и дезинфекции воды. Моделируемый процесс состоит из шести последовательных подпроцессов:

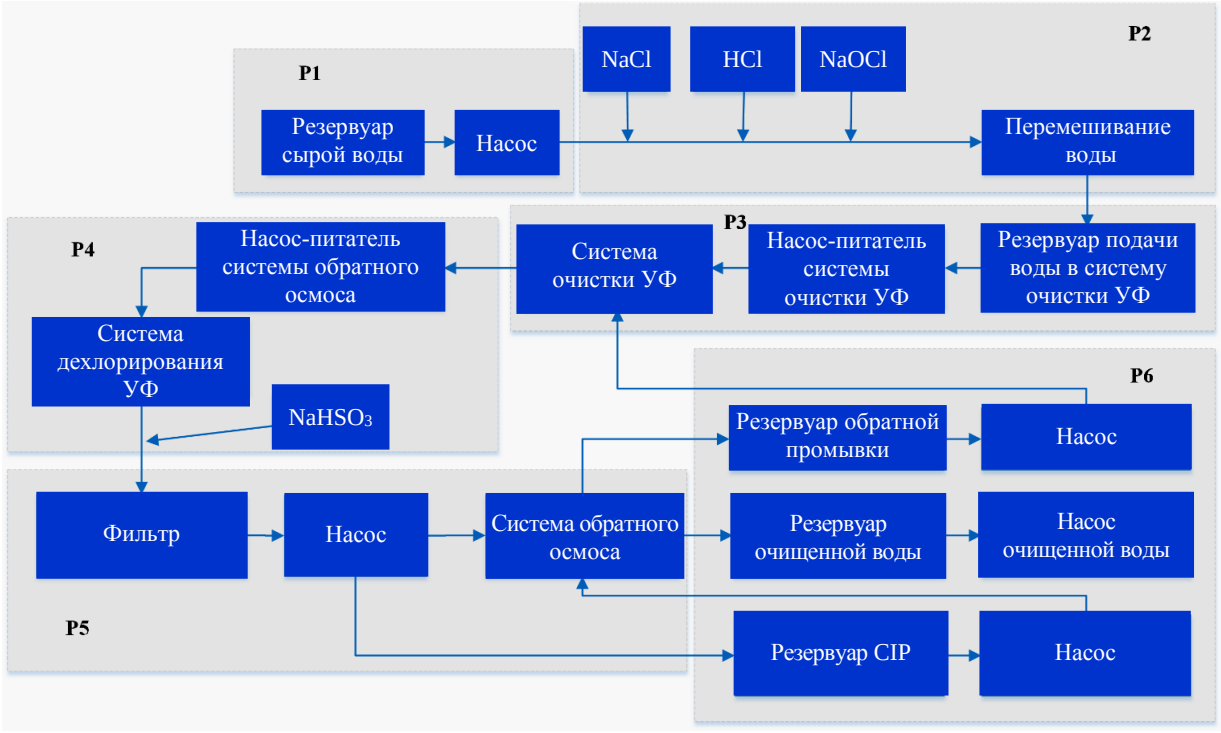

Рис.5. Схема технологического процесса, моделируемого стендом SWaT [37]

Записки Горного института. 2024. Т. 267. С. 488-500

забора грязной воды, добавления в нее необходимых химических веществ, фильтрации, дехлорирования с помощью ультрафиолетовых (УФ) ламп, подачи в систему обратного осмоса, отвода чистой воды и шлама. На рис.5 представлена схема моделируемого технологического процесса, подпроцессы обозначены буквами P1-P6, соответственно.

Коммуникационная часть тестового стенда SWaT выражена многоуровневой коммуникационной сетью, программируемыми логическими контроллерами, сервером и рабочей станцией диспетчерского контроля и сбора данных (SCADA), а также хранилищем исторических данных. Архитектура тестового стенда позволяет оперативному персоналу удаленно подключаться к инфраструктуре объекта. Набор данных имеет несколько версий, различающихся по типам собираемых данных, проводимым деструктивным воздействиям и общей продолжительности функционирования тестового стенда.

В выполненном эксперименте использовался вариант набора данных, в котором содержится 36 различных сценариев атак для подмены передаваемых данных. Деструктивные воздействия оказывались на разные физические устройства, относящиеся к разным технологическим подпроцессам. В наборе представлено 19 атак с целью модификации значений одного датчика, шесть атак для подмены значений двух и трех датчиков; семь атак на датчики, принадлежащие разным технологическим подпроцессам (табл.1).

Таблица 1

Характеристика аномалий в наборе данных SWaT

Тип записи в наборе	Число процессов, ставших целью деструктивного воздействия	Процессы	Число записей
Норма	0	0	399157
Аномалия	1	P1	4053
	1	P2	1809
	1	P3	37860
	1	P4	1700
	1	P5	1044
	2	P3, P4	1691
	2	P1, P3	1445
	2	P3, P6	697
	2	P4, P5	463

Для оценки эффективности обнаружения аномалий использовались показатели, которые характеризуют точность (precision) и полноту (recall) выявления аномалий. Показатель точности определяет долю записей, которые классифицированы как аномальные и действительно являются аномальными. Показатель полноты (recall) отражает долю аномальных записей, которые выявлены алгоритмом. Чем выше значения данных показателей, тем лучше эффективность выявления аномалий. Поскольку исследуемый набор данных несбалансированный, т.е. число нормальных записей значительно превышает число аномальных записей, был также использован показатель *F*1-мера, который представляет собой среднее гармоничное между точностью и полнотой. Кривые обучения и функции потерь представлены на рис.6. Выполнен анализ времени логического вывода и числа параметров модели. Последние показатели позволяют оценить необходимые вычислительные ресурсы и взвешенные средние этих показателей. Сделан сравнительный анализ результатов с методами, представленными в литературе, а именно решениями на основе автоэнкодеров MTS-DVGAN [12] и MTAD-CAN [13], а также нейронной сети долгой краткосрочной памяти (LSTM) и метода опорных векторов с автоэнкодером (deepSVDD), которые были обучены на том же наборе данных (табл.2). Для моделей MTS-DVGAN и MTAD-CAN информации по времени логического вывода и времени принятия решений в соответствующих научных статьях нет, поэтому в графах таблицы указан прочерк.

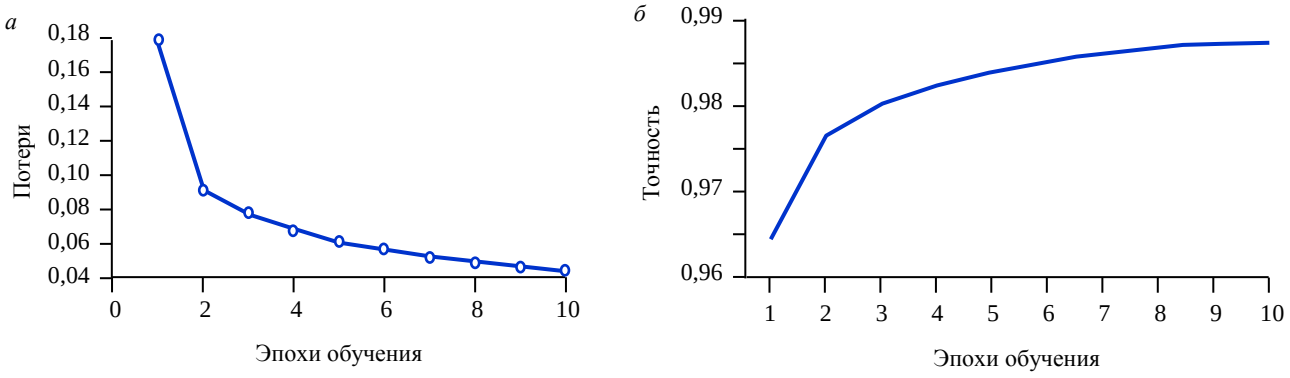

Рис.6. Функции потерь (*а*) и обучения (*б*) сверточной модели выявления аномалий

Эксперименты показали, что предложенная модель выявления аномалий показывает самую высокую полноту, а точность лишь немного уступает модели MTS-DVGAN. При этом предложенное решение отличается малым числом параметров, а время логического вывода для одной записи составляет $2{,}07 \cdot 10^{-5}$. Несмотря на то, что эти данные для модели MTS-DVGAN отсутствуют, с учетом архитектуры используемых в ней нейронных сетей – вариационный автокодировщик с двумя сетями короткой долгой памяти – можно предположить, что число параметров не меньше параметров модели LSTM, которая также основана на сети короткой долгой памяти. Следовательно, предложенная модель обладает высокой эффективностью и низкой вычислительной сложностью и может быть использована в системах с ограниченными вычислительными ресурсами.

Отметим, что обученные в режиме проектирования модели в дальнейшем используются для выявления аномалий, вычисления вероятностей успешной реализации угроз нарушения киберустойчивости и оценки рисков киберустойчивости технологического процесса очистки сточных вод.

Таблица 2

Результаты экспериментов по выявлению аномалий в наборе данных SWaT

Подход к выявлению аномалий	Точность	Полнота	*F*1-мера	Число параметров модели	Время принятия решения моделью, с
Предложенный подход	0,98	0,96	0,97	10 561	$2{,}07 \cdot 10^{-5}$
DeepSVDD	0,95	0,68	0,82	11 648	$2{,}43 \cdot 10^{-5}$
LSTM	0,98	0,71	0,82	66 035	$4{,}37 \cdot 10^{-5}$
MTS-DVGAN [10]	0,99	66,93	0,79	–	–
MTAD-CAN [11]	0,91	0,94	0,92	–	–

Заключение. Представлена методика выявления и оценки рисков киберустойчивости технологического процесса очистки сточных вод. Описаны этапы работы методики, включая формирование обучающего набора данных, обнаружение аномалий и оценку рисков. Каждый этап сопровождается отдельной методикой. Этап выявления аномалий в процессе является ключевым, поскольку от его точности зависят последующие оценки рисков. Предложен новый подход к предобработке входных данных для последующего обнаружения аномалий. Эксперименты показали, что при таком подходе используются достаточно простые нейронные сети, имеющие низкое время задержки на принятие решение, что позволяет обрабатывать интенсивные потоки данных. Благодаря небольшому числу параметров модели ее можно использовать в устройствах, обладающих ограниченными вычислительными ресурсами, например промышленных контроллерах, микрокомпьютерах и т.д. Таким образом, цель исследования, состоявшая в повышении точности обнаружения аномалий в технологическом процессе очистки сточных

вод для оценки рисков киберустойчивости с учетом ограничений по используемым ресурсам, достигнута.

Дальнейшие направления работ по этой задаче связаны с оптимизацией процедуры генерации изображений путем определения достаточного объема выборки данных, исследованием подходов на данных от технологических процессов и разработкой методик выявления аномалий в самих процессах. Планируется проведение экспериментов с вычислением рисков киберустойчивости, а также на собственном сгенерированном наборе данных.

ЛИТЕРАТУРА

1. *Balaram V., Copia L., Saravana Kumar U. et al.* Pollution of water resources and application of ICP-MS techniques for monitoring and management – A comprehensive review // Geosystems and Geoenvironment. 2023. Vol. 2. Iss. 4. № 100210. DOI: 10.1016/j.geogeo.2023.100210

2. *Chukaeva M., Petrov D.* Assessment and analysis of metal bioaccumulation in freshwater gastropods of urban river habitats, Saint Petersburg (Russia) // Environmental Science and Pollution Research. 2023. Vol. 30. Iss. 3. P. 7162-7172. DOI: 10.1007/s11356-022-21955-8

3. *Сафиуллин Р.Н., Афанасьев А.С., Резниченко В.В.* Концепция развития систем мониторинга и управления интеллектуальных технических комплексов // Записки Горного института. 2019. Т. 237. С. 322-330. DOI: 10.31897/PMI.2019.3.322

4. *Patokin D., Danilov A., Isakov A.* Environmental monitoring of natural waters in the zone of impact of an enterprise producing explosives // IOP Conference Series: Earth and Environmental Science. 2020. Vol. 578. № 012038. DOI: 10.1088/1755-1315/578/1/012038

5. *Смирнов Ю.Д., Матвеева В.А., Яковлев Н.М., Сахабутдинова Э.Р.* Анализ и оценка современных технологий очистки сточных вод на гальваническом производстве // Горный журнал. 2023. № 9. С. 55-60. DOI: 10.17580/gzh.2023.09.08

6. *Ромашева Н.В., Бабенко М. А., Николайчук Л.А.* Устойчивое развитие Арктического региона России: экологические проблемы и пути их решения // Горный информационно-аналитический бюллетень. 2022. № 10-2. С. 78-87 (in English). DOI: 10.25018/0236_1493_2022_102_0_78

7. *Чукаева М.А., Матвеева В.А., Сверчков И.П.* Комплексная переработка высокоуглеродистых золошлаковых отходов // Записки Горного института. 2022. Т. 253. С. 97-104. DOI: 10.31897/PMI.2022.5

8. *Русскевич Е.А.* Нарушение правил централизованного управления техническими средствами противодействия угрозам информационной безопасности // Journal of Digital Technologies and Law. 2023. Т. 1. № 3. С. 650-672. DOI: 10.21202/jdtl.2023.28

9. *Landauer M., Onder S., Skopik F., Wurzenberger M.* Deep learning for anomaly detection in log data: A survey // Machine Learning with Applications. 2023. Vol. 12. № 100470. DOI: 10.1016/j.mlwa.2023.100470

10. *Yuan Luo, Ya Xiao, Long Cheng et al.* Deep Learning-based Anomaly Detection in Cyber-physical Systems: Progress and Opportunities // ACM Computing Surveys. 2022. Vol. 54. Iss. 5. № 106. DOI: 10.1145/3453155

11. *Haili Sun, Yan Huang, Lansheng Han et al.* MTS-DVGAN: Anomaly detection in cyber-physical systems using a dual variational generative adversarial network // Computers & Security. 2024. Vol. 139. № 103570. DOI: 10.1016/j.cose.2023.103570

12. *Feng Xia, Xin Chen, Shuo Yu. et al.* Coupled Attention Networks for Multivariate Time Series Anomaly Detection // IEEE Transactions on Emerging Topics in Computing. Vol. 12. Iss. 1. P. 240-253. DOI: 10.1109/TETC.2023.3280577

13. *Zhukovskiy Y., Buldysko A., Revin I.* Induction Motor Bearing Fault Diagnosis Based on Singular Value Decomposition of the Stator Current // Energies. 2023. Vol. 16. Iss. 8. № 3303. DOI: 10.3390/en16083303

14. *Meleshko A., Shulepov A., Desnitsky V. et al.* Visualization Assisted Approach to Anomaly and Attack Detection in Water Treatment Systems // Water. 2022. Vol. 14. Iss. 15. № 2342. DOI: 10.3390/w14152342

15. *Pliatsios D., Sarigiannidis P., Lagkas T., Sarigiannidis A.* A Survey on SCADA Systems: Secure Protocols, Incidents, Threats and Tactics // IEEE Communications Surveys & Tutorials. 2020. Vol. 22. Iss. 3. P. 1942-1976. DOI: 10.1109/COMST.2020.2987688

16. *Cherdantseva Y., Burnap P., Blyth A. et al.* A review of cyber security risk assessment methods for SCADA systems // Computers & Security. 2016. Vol. 56. P. 1-27. DOI: 10.1016/j.cose.2015.09.009

17. *Mehmood A., Epiphaniou G., Maple C. et al.* A Hybrid Methodology to Assess Cyber Resilience of IoT in Energy Management and Connected Sites // Sensors. 2023. Vol. 23. Iss. 21. № 8720. DOI: 10.3390/s23218720

18. *Xirong Ning, Jin Jiang.* Design, Analysis and Implementation of a Security Assessment/Enhancement Platform for Cyber-Physical Systems // IEEE Transactions on Industrial Informatics. 2022. Vol. 18. Iss. 2. P. 1154-1164. DOI: 10.1109/TII.2021.3085543

19. *Teixeira A., Kin Cheong Sou, Sandberg H., Johansson K.H.* Secure Control Systems: A Quantitative Risk Management Approach // IEEE Control Systems Magazine. 2015. Vol. 35. Iss. 1. P. 24-45. DOI: 10.1109/MCS.2014.2364709

20. *Федорченко Е.В., Новикова Е.С., Котенко И.В. и др.* Система измерения защищенности информации и персональных данных для устройств интернета вещей // Вопросы кибербезопасности. 2022. № 5 (51). С. 28-46. DOI: 10.681/2311-3456-2022-5-28-46

21. *Дойникова Е.В., Котенко И.В.* Оценивание защищенности и выбор контрмер для управления кибербезопасностью. М.: Российская академия наук, 2021. 184 с.

22. *Jbair M., Ahmad B., Maple C., Harrison R.* Threat modelling for industrial cyber physical systems in the era of smart manufacturing // Computers in Industry. 2022. Vol. 137. № 103611. DOI: 10.1016/j.compind.2022.103611

23. *Palleti V.R., Adepu S., Mishra V.K., Mathur A.* Cascading effects of cyber-attacks on interconnected critical infrastructure // Cybersecurity. 2021. Vol. 4. № 8. DOI: 10.1186/s42400-021-00071-z

24. *Khalil S.M., Bahsi H., Korõtko T.* Threat modeling of industrial control systems: A systematic literature review // Computers & Security. 2024. Vol. 136. № 103543. DOI: 10.1016/j.cose.2023.103543

25. *Kaixing Huang, Chunjie Zhou, Yu-Chu Tian et al.* Assessing the Physical Impact of Cyberattacks on Industrial Cyber-Physical Systems // IEEE Transactions on Industrial Electronics. 2018. Vol. 65. Iss. 10. P. 8153-8162. DOI: 10.1109/TIE.2018.2798605

26. *Tushkanova O., Levshun D., Branitskiy A. et al.* Detection of Cyberattacks and Anomalies in Cyber-Physical Systems: Approaches, Data Sources, Evaluation // Algorithms. 2023. Vol. 16. Iss. 2. № 85. DOI: 10.3390/a16020085

27. *Conti M., Donadel D., Turrin F.* A Survey on Industrial Control System Testbeds and Datasets for Security Research // IEEE Communications Surveys & Tutorials. 2021. Vol. 23. Iss. 4. P. 2248-2294. DOI: 10.1109/COMST.2021.3094360

28. *Renjie Wu, Keogh E.J.* Current Time Series Anomaly Detection Benchmarks are Flawed and are Creating the Illusion of Progress // IEEE Transactions on Knowledge and Data Engineering. 2023. Vol. 35. Iss. 3. P. 2421-2429. DOI: 10.1109/TKDE.2021.3112126

29. *Котенко И.В., Федорченко Е.В., Новикова Е.С. и др.* Методология сбора данных для анализа безопасности промышленных киберфизических систем // Вопросы кибербезопасности. 2023. № 5 (57). С. 69-79. DOI: 10.21681/2311-3456-2023-5-69-79

30. *Антонова Е.С.* Моделирование процесса очистки сточных вод во флотационной установке с эжекционной системой аэрации с диспергатором // Безопасность в техносфере. 2017. Т. 6. № 1. С. 43-50. DOI: 10.12737/article_590199b9952dc2.23575176

31. *Ivanov A., Strizhenok A., Borowski G.* Treatment of methanol-containing wastewater at gas condensate production // Journal of Water and Land Development. 2022. Vol. 54. P. 84-93. DOI: 10.24425/jwld.2022.141558

32. *Bazgir O., Zhang R., Dhruba S.R. et al.* Representation of features as images with neighborhood dependencies for compatibility with convolutional neural networks // Nature Communications. 2020. Vol. 11. № 4391. DOI: 10.1038/s41467-020-18197-y

33. *Sharma A., Vans E., Shigemizu D. et al.* DeepInsight: A methodology to transform a non-image data to an image for convolution neural network architecture // Scientific Reports. 2019. Vol. 9. № 11399. DOI: 10.1038/s41598-019-47765-6

34. *Zhu Y., Brettin T., Fangfang Xia et al.* Converting tabular data into images for deep learning with convolutional neural networks // Scientific Reports. 2021. Vol. 11. № 11325. DOI: 10.1038/s41598-021-90923-y

35. *Jong-Ik Park, Sihoon Seong, JunKyu Lee, Cheol-Ho Hong.* Vortex Feature Positioning: Bridging Tabular IIoT Data and Image-Based Deep Learning: ArXiv. 2024. 23 p. (препринт)

36. Yuansheng Zhou, Sharpee T.O. Using Global t-SNE to Preserve Intercluster Data Structure // Neural Computation. 2022. Vol. 34. Iss. 8. P. 1637-1651. DOI: 10.1162/neco_a_01504

37. *Goh J., Adepu S., Junejo K.N., Mathur A.* A Dataset to Support Research in the Design of Secure Water Treatment Systems. Critical Information Infrastructures Security. Cham: Springer, 2017. P. 88-99. DOI: 10.1007/978-3-319-71368-7_8

Авторы: **Е.С.Новикова,** *канд. техн. наук, старший научный сотрудник, https://orcid.org/0000-0003-2923-4954* (*Санкт-Петербургский Федеральный исследовательский центр РАН, Санкт-Петербург, Россия*), **Е.В.Федорченко**, *канд. техн. наук, старший научный сотрудник, doynikova@comsec.spb.ru, https://orcid.org/0000-0001-6707-9153* (*Санкт-Петербургский Федеральный исследовательский центр РАН, Санкт-Петербург, Россия*), **М.А.Бухтияров,** *Full-stack разработчик, https://orcid.org/0009-0004-3964-2796* (*ООО «Webim», Москва, Россия*), **И.Б.Саенко,** *д-р техн. наук, главный научный сотрудник, https://orcid.org/0000-0002-9051-5272* (*Санкт-Петербургский Федеральный исследовательский центр РАН, Санкт-Петербург, Россия*).

Авторы заявляют об отсутствии конфликта интересов.