УДК 004.056

Обеспечение информационной безопасности киберфизических объектов на основе прогнозирования и обнаружения аномалий их состояния

Васильев В. И., Вульфин А. М., Гвоздев В. Е.,
Картак В. М., Атарская Е. А.

Постановка задачи: *обеспечение устойчивого функционирования киберфизических систем за счет совершенствования методов предиктивного анализа, направленных на выявление сбоев функционирования, вызванных действиями злоумышленника и ведущих к деградации киберфизических объектов (КФО), на основе выявления аномалий в технологических временных рядах параметров состояния КФО в рамках концепции расширенного обнаружения и устранения угроз кибербезопасности.* ***Целью работы*** *является повышение эффективности обнаружения аномалий наблюдаемых параметров киберфизических систем за счет совершенствования алгоритмов выявления аномалий в технологических временных рядах накапливаемых параметрах состояния КФО на основе интеллектуального анализа.* ***Используемые методы****: методы интеллектуального анализа многомерных технологических временных рядов с применением гетерогенного ансамбля детекторов для обнаружения аномалий в накапливаемых параметрах состояния киберфизического объекта. Модель обнаружения аномалий включает группу детекторов для одномерных временных рядов и детектор для многомерного временного ряда на основе нейросетевых автоенкодеров, модели изолирующего леса и оценки фактора локального выброса.* ***Новизна:*** *модель обнаружения аномалий на основе гетерогенного ансамбля детекторов. Отличие заключается в использовании нейросетевых автоенкодеров на основе долгой краткосрочной памяти для моделирования нормального поведения системы. При появлении новых типов аномалий или изменении характера текущих аномалий детектор на основе оценки ошибки восстановления образа сохраняет свою работоспособность.* ***Результат****: структурная схема системы обнаружения аномалий технологического процесса, основанная на применении методов предиктивного анализа собираемых данных телеметрии КФО и позволяющая выявить воздействия злоумышленника, получившего доступ в промышленную сеть управления технологическим процессом; алгоритм анализа технологических временных рядов и гетерогенная модель детекторов обнаружения аномалий, вызванных воздействием злоумышленника, пытающегося перехватить управление или навязать алгоритм управления киберфизическим объектом.* ***Практическая значимость:*** *предлагаемый подход направлен на совершенствование механизмов предиктивного анализа в составе систем обнаружения и устранения аномалий производственных и технологических процессов автоматизированных систем управления технологическими процессами. Применение системы возможно в составе комплекса средств защиты промышленной сети, выступающих в качестве источников событий безопасности для системы сбора и корреляции событий кибербезопасности.*

Ключевые слова: *киберфизический объект; временной ряд; детекторы аномалий; нейросетевой автоенкодер с долгой краткосрочной памятью.*

Библиографическая ссылка на статью:
Васильев В. И., Вульфин А. М., Гвоздев В. Е., Картак В. М., Атарская Е. А. Обеспечение информационной безопасности киберфизических объектов на основе прогнозирования и обнаружения аномалий их состояния // Системы управления, связи и безопасности. 2021. № 6. С. 90-119. DOI: 10.24412/2410-9916-2021-6-90-119.

Reference for citation:
Vasilyev V. I., Vulfin A. M., Gvozdev V. E., Kartak V. M., Atarskaya E. A. Ensuring information security of cyber-physical objects based on predicting and detecting anomalies in their state. *Systems of Control, Communication and Security*, 2021, no. 6, pp. 90-119 (in Russian). DOI: 10.24412/2410-9916-2021-6-90-119.

Введение

На современном этапе цифрой трансформации индустрии актуальными являются вопросы поддержания работоспособности киберфизических систем (КФС), т.е. обеспечения устойчивости протекающих в них физических процессов и непрерывности управления в условиях возможных внутренних и внешних целенаправленных деструктивных воздействий. Основным направлением развития систем защиты информации для обеспечения киберустойчивости КФС является реализация опережающей стратегии защиты (проактивная защита), основанной на предсказании угрозы (предиктивный анализ) и раннем обнаружении атак с целью адаптации системы к предполагаемому деструктивному воздействию.

Одним из современных подходов к построению систем защиты является концепция расширенного обнаружения и устранения угроз (XDR [1], Extended Detection and Response) – рис. 1, где *X* – это любой источник данных (информационно-телекоммуникационная инфраструктура, конечные системы, пользователи, киберфизические объекты (КФО)), *D* и *R* – обнаружение и реагирование. Подобные системы обеспечивают видимость и контекст на этапе анализа сложных угроз с возможностью приоритезации мер по их устранению на основе агрегации и анализа данных из множества источников.

Источниками данных для XDR являются:

1) системы анализа сетевого трафика (NTA, Network Traffic Analysis) информационно-телекоммуникационной среды КФС;
2) системы управления безопасностью и автоматизации реагирования (SOAR, Security Orchestration and Automated Response), объединяющие анализ контекста и контента в виде структурированных и неструктурированных данных в системах управления информацией и событиями безопасности (SIEM, Security Information and Event Management), реализуемый с помощью:
   - системы анализа безопасности поведения пользователей и сущностей (UEBA, User and Entity behavior Analytics);
   - системы обнаружения и реагирования на угрозы для конечных точек (EDR, Endpoint Threat Detection and Response);
   - системы обмена данными об угрозах (TI, Threat hunting);
3) системы обнаружения и устранения аномалий (ADM, Anomaly Detection and Mitigation) производственных и технологических процессов КФО.

В концепции расширенного обнаружения и устранения угроз XDR важная роль отводится предиктивному анализу, выступающему в качестве одного из методов обеспечения кибербезопасности КФС. Методы предиктивного анализа направлены на выявление предпосылок неполадок и сбоев функционирования, ведущих к деградации КФО в составе КФС, на основе анализа накапливаемых параметров их состояния. Основным инструментом предиктивного анализа является выявление аномалий в технологических временных рядах (ТВР) накапливаемых параметров состояния КФО. Под аномалией при этом понима-

ется отклонение в функционировании КФО или отклонения, связанные с нарушением взаимодействия устройств при обмене данными в составе КФО [2, 3].

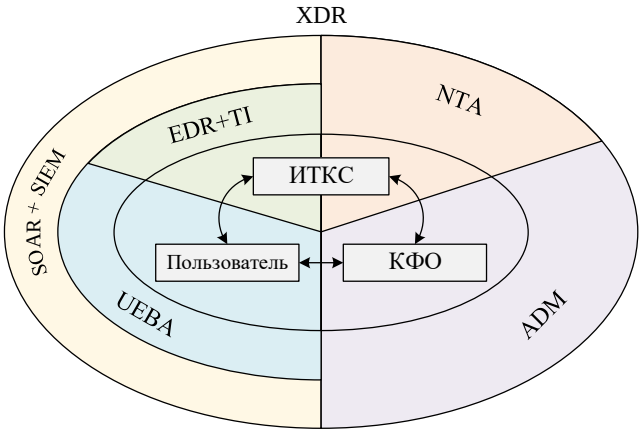


Рис. 1. Концепция расширенного обнаружения и устранения угроз (ИТКС – информационно-телекоммуникационная система)

Анализ систем обнаружения аномалий в рамках концепции расширенного обнаружения и устранения угроз кибербезопасности

Одна из возможных [4-6] классификаций методов обнаружения аномалий и направлений их использования в задачах обеспечения кибербезопасности КФС представлена на рис. 2.

Рис. 2. Методы обнаружения аномалий в рамках концепции расширенного обнаружения и устранения угроз

При построении подобных систем возникает необходимость сбора и обработки значительных объемов структурированных и слабоструктурированных данных со всех уровней КФС для формирования набора параметров, пригодных для оперативного анализа и выявления аномалий, возникающих в результате возможных действий злоумышленника. Ведущую роль при решении этой задачи играют методы интеллектуального анализа данных (ИАД) временных рядов параметров, характеризующих состояние КФО, и методы машинного обучения.

Применение подобных решений при обнаружении аномалий функционирования ИТКС, аномалий в поведении пользователей и аномалий параметрах КФО нашло свое отражение в ряде публикаций (таблица 1).

Таблица 1 – Применение методов ИАД и машинного обучения в задачах обнаружения аномалий в рамках концепции расширенного обнаружения и устранения угроз кибербезопасности

Тип системы	Пример
Системы анализа аномалий сетевого трафика на основе методов ИАД и машинного обучения	– модель обнаружения гибридных аномалий в высоконагруженных сетях связи на основе методов ИАД [7]; – платформа обнаружения аномалий для выявления кибератак на облачные вычислительные среды [8]; – комплексная система контроля и обеспечения безопасности сбора данных, реализующая мониторинг трафика в реальном времени, обнаружение аномалий, анализ воздействия, стратегии смягчения последствий [9, 10]; – система обнаружения аномалий на основе алгоритмов машинного обучения для устранения угроз кибербезопасности сетей Интернета вещей в умном городе [11]; – подход на основе кластерного анализа сетевого трафика для обнаружения кибератак, вызывающих аномалии в сетях критической информационной инфраструктуры газокомпрессорных станций [12]; – распределенная система обнаружения вторжений для систем диспетчерского управления и сбора данных [13]; – алгоритм обнаружения аномалий и система обнаружения вторжений с фильтрацией ложных срабатываний и возможностью подтверждения атаки [14]; – система обнаружения аномалий для обнаружения утечек конфиденциальной информации в сетевом трафике энергосистем [15]; – методология создания надежных наборов данных для обнаружения аномалий в АСУ ТП [16]
Анализ состояния конечных систем и поведения пользователя в задаче	– программная платформа для обнаружения аномалий работы конечной системы в режиме реального времени на основе анализа и оценки значимых выбросов наблюдаемых параметров [17];

Тип системы	Пример
обнаружения аномалий	– детектор аномалий для обнаружения атак на конечные системы и подсистема объяснения решения [18]; – фильтр событий для последующего анализа на основе метода глубокого обучения для обнаружения аномальной сетевой активности из журналов конечных систем в режиме реального времени [19]; – модели сетевых вторжений, основанных на учете человеческого фактора при реализации сложных сетевых атак [20]
Обнаружение аномалий работы КФО	– алгоритм, основанный на гауссовском процессе (метод непараметрического машинного обучения) для мониторинга состояния установок ветрогенераторов и выявления эксплуатационных аномалий [21]; – двухэтапная методология обнаружения аномалий в промышленных процессах [22]; – стратегия выбора датчика и обнаружение аномалий данных с помощью методов теории информации [23]; – компоненты онлайн-системы для диагностики и обнаружения аномалий трансформаторов силовой подстанции [24]; – анализ отклонения прогноза нейронной сети от параметров реального объекта для выявления аномалий на водоочистной станции [25]; – обнаружение аномального поведения с помощью метода контролируемой классификации на основе частично определенной логической функции, позволяющий извлекать закономерности из исторических измерений датчиков [26]

Основной задачей обнаружения аномалий работы КФО является разработка механизма, который не только позволяет выявлять аномалии состояния КФО, но и способен отличать имеющий место фактический отказ от проводимой кибератаки [25-27].

Рассмотрим возможную классификацию методов и алгоритмов обнаружения аномалий на основе интеллектуального анализа временных рядов параметров состояния КФО. Группа методов M_1 (рис. 2) основана на построении прогностической модели одномерных и многомерных временных рядов и дальнейшем пороговом сравнении прогноза модели и реальных данных, характеризующих состояние КФО:

– модели авторегрессии (ARIMA, Auto Regressive Integrated Moving Average) и нейросетевой регрессии NARX (Nonlinear autoregressive exogenous model);

– нелинейные предикторы на основе рекуррентных нейронных сетей (Recurrent Neural Network, RNN) и сетей с долгой краткосрочной памятью (Long Short-Term Memory, LSTM).

Группа методов M_2 (рис. 2) основана на применении моделей, обучаемых без учителя:

– метод опорных векторов с одним классом (One-Class Support Vector machine, SVM) – модель обучается на данных, не содержащих аномалий. Для задания порога отделения нормальных и аномальных данных необходимо иметь оценку их соотношения;
– метод изолирующего леса (Isolation Forest): ансамбль случайных деревьев решений на первых уровнях построения модели выделяет наиболее значимые аномальные данные;
– метрические методы (k-ближайших соседей, LOF (Local Outlier Factor)) основаны на оценке относительного взаимного положения данных в пространстве признаков;
– методы, основанные на кластерном анализе, оценивают удаленность точек в пространстве признаков от выделенных центров кластеров;
– методы, использующие нейросетевые автоенкодеры, строят модели, обучаемые на нормальных данных.

Группа методов M_3 (рис. 2) основана на построении моделей, обучаемых с учителем:

– классификация отдельных примеров наблюдений с помощью моделей, обучаемых с учителем, требует наличия размеченных исторических данных;
– методы предсказания дефектов и сбоев на основе специфических предвестников (классификация);
– прогнозирование оставшегося времени безотказной работы системы (задача регрессии).

Группа методов M_4 (рис. 2) понижает размерность признакового пространства описания состояния системы:

– метод главных компонент;
– вероятностный метод главных компонент.

Анализ киберфизического объекта и наблюдаемых параметров

Предложенный исследователями из Южной Кореи (Institute of ETRI, Daejeon, South Korea) [28, 29] набор данных собран в ходе эксплуатации стендовой АСУ ТП и дополнен результатами программно-аппаратного моделирования (HIL, Hardware-in-the-Loop) генерации энергии паровой турбиной и процесса гидроаккумулирования. Целью работы является исследование методов и алгоритмов обнаружения аномалий в таких киберфизических системах, как паровые турбины, водоочистные сооружения и электростанции. Первоначально были запущены три испытательных стенда: стенд турбины General Electronics, стенд паровых котлов Emerson и стенд MPS FESTO для водоочистки. Затем была построена система, которая объединила эти три стенда с программно-

аппаратным симулятором, имитирующим выработку тепловой и гидроаккумулирующей энергии. Первая версия набора данных содержит нормальные и аномальные ситуации, соответствующие 34 сценариям атак [28-31].

Технологические процессы на испытательном стенде показаны на рис. 3:

– процесс котла (P1);
– процесс турбины (P2);
– процесс водоподготовки (P3);
– HIL-моделирование (P4) сценариев выработки тепловой энергии и генерации гидроаккумулирования энергии.

Процессы котла и турбины используются для моделирования тепловой электростанции, а процесс очистки воды используется для моделирования гидроаккумулирующей электростанции.

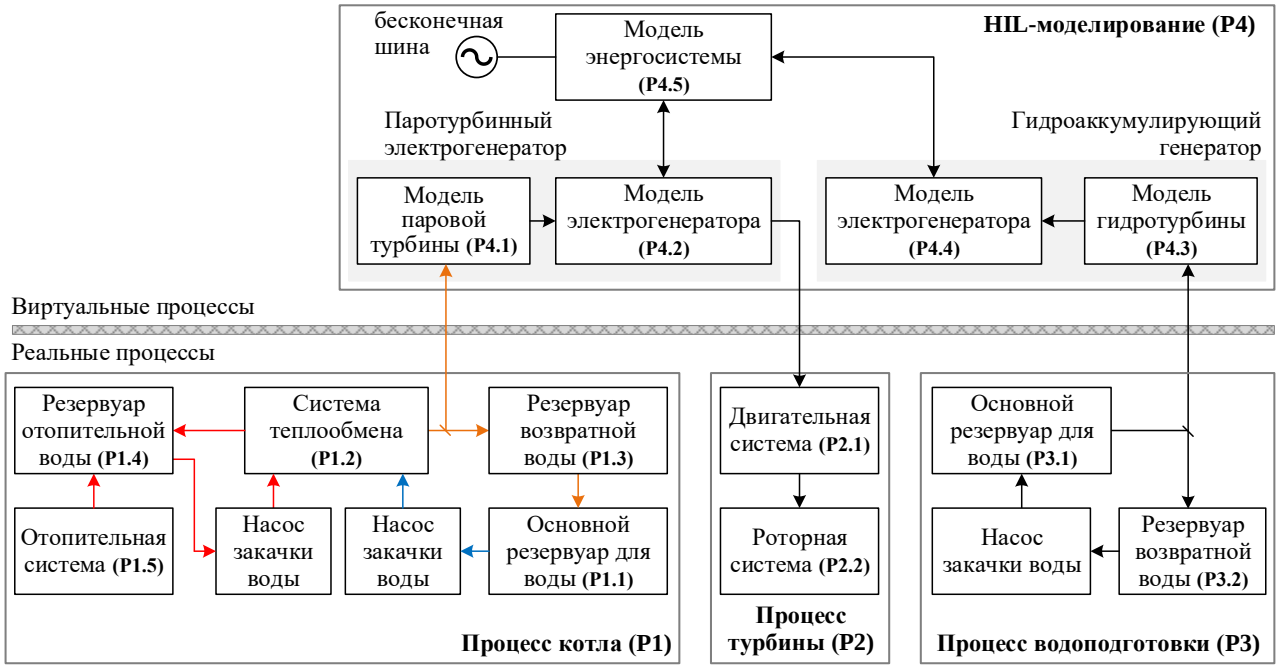


Рис. 3. Технологическая схема анализируемого стенда

Управление нагревателем в котле контролируется распределенной системой управления (РСУ) Emerson Ovation. Для управления скоростью вращения и мониторинга вибрации турбины используется контроллер РСУ Mark VIe компании General Electric. Процесс водоподготовки контролируется с помощью программируемого логического контроллера (ПЛК) Siemens S7-300, который управляет уровнем воды и работой насоса. В испытательном стенде HAI моделирование HIL проводилось с использованием системы dSPACE SCALEXIO, сопряжённой с со стендом с помощью ПЛК S7-1500, ПЛК (Siemens) и с устройствами удаленного ввода-вывода ET200 (рис. 4).

Сценарии работы системы задаются с помощью четырех переменных замкнутого контура управления, а именно: уставок (SPs), параметров процесса (PVs), управляющие воздействия (CVs) и параметров управления (CPs) (рис. 5).

При нормальной работе системы предполагается, что оператор управляет объектом в обычном режиме через HMI (Human-machine interface) и что пере-

№6. 2021
ISSN 2410-9916

менные симулятора, связанные с выработкой электроэнергии в HIL-симуляторе, изменяются. Оператор отслеживает значения PV, фиксируемые датчиками и отображаемые с помощью HMI, и задает SP для контролируемых устройств.

Планировщик задач HMI используется для периодической установки SPs и переменных HIL-симулятора на случайные или предопределенные значения в пределах нормального диапазона для имитации вариантов штатного сценария. Нормальные диапазоны значений SP определены путем экспериментального изменения значения каждого SP. Аномальное поведение возникало, когда некоторые параметры выходили за пределы нормального диапазона или находились в неожиданном состоянии вследствие атак, сбоев или отказов.

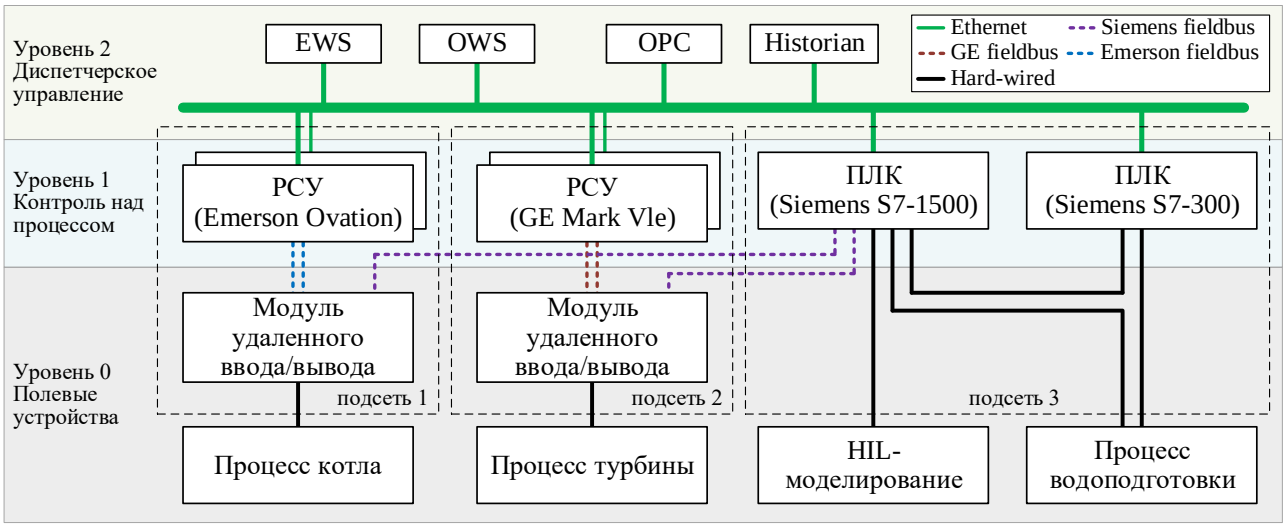


Рис. 4. Тестируемые компоненты и поток данных управления по уровням (EWS – Engineering Work Station; OWS – Operator Work Station; OPC – Open Platform Communications)

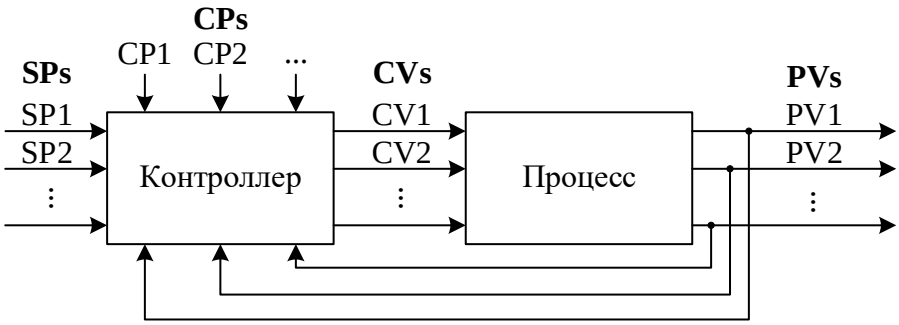


Рис. 5. Модель контура управления

Было проведено 50 атак, включая 25 примитивов атак и 25 комбинированных атак при одновременном выполнении сразу двух примитивов атаки. Сценарии атак реализуются с учетом цели атаки, времени атаки и метода для каждого контура управления с обратной связью. Примеры примитивов атак:

– закрыть клапан контроля давления, а затем вернуться в нормальное состояние и пытаться поддержать предыдущее значение датчика;

DOI: 10.24412/2410-9916-2021-6-90-119
URL: https://sccs.intelgr.com/archive/2021-06/06-Vasilyev.pdf

– уменьшить значение расхода воды в баке отработанной воды (P1_B3005), а затем восстановить его (скрывая изменения в HMI);
– открыть клапан контроля уровня, а затем восстановить его в виде трапециевидного профиля и пытаться поддержать предыдущее значение датчика;
– кратковременная атака, при которой на несколько секунд открывается клапан контроля уровня и восстанавливается нормальное состояние. Повторяется несколько раз.

Разработка системы обнаружения аномалий наблюдаемых параметров состояния киберфизического объекта

Структурная схема предложенной системы обнаружения аномалий технологического процесса, основанная на применении методов анализа собираемых данных телеметрии, позволяющая выявить действия злоумышленника, получившего доступ в промышленную сеть управления технологическим процессом, представлена на рис. 6. Технологические временные ряды представляют собой последовательность измерений, собранных с датчиков промышленных объектов. Аномалии представляют собой отрезки временного ряда. На этапе предварительной обработки входные данные подвергаются нормализации и фильтрации. С помощью метода скользящего окна из временных рядов набора данных формируются обучающие и тестовые выборки.

Для создания детектора аномалий используются данные о нормальном состоянии для построения модели нормального поведения. Обучающая выборка содержит только данные о нормальном поведении системы, тестовая – содержит данные как нормального класса, так и класса аномалий (одиночные атаки и их комбинации).

В данном исследовании применяется несколько алгоритмов для построения моделей машинного обучения (ML) с целью обнаружения аномалий. Рассматриваются следующие модели:

– детектор на основе нейросетевого (НС) автоенкодера LSTM для одномерного и многомерного ТВР;
– детектор выбросов с автоподстройкой порога (LOF-детектор);
– детектор аномалий на основе изолирующего леса (IFO-детектор);
– детектор аномалий на основе машины опорных векторов (One-class SVM).

На рис. 6 обозначены:

– X_R – нормализация каждого из рядов многомерного ТВР (1);
– X_R^P – сглаженные многомерные ТВР (2);
– X_R^W – скользящее окно длины W с шагом S, формирует набор отсчетов для анализа по каждому из ТВР (рис. 6, б) (3);
– подготовленные данные для построения, тестирования и использования ансамбля детекторов (4);

- ансамбль автоенкодеров на основе нейронной сети LSTM (5): детектор выбросов с автоподстройкой порога (LOF детектор); детектор аномалий на основе модели изолирующего леса (IFO, Isolation Forest); детектор аномалий на основе машины опорных векторов (One class SVM);
- многомерный детектор НС LSTM (рис. 6, в);
- суммирование оценок детекторов в каждом окне *W* одномерных ТВР (7);
- блок принятия решений (БПР) о наличии аномалий в окне анализа *W* одномерных ТВР (8);
- специалист по интеллектуальному анализу данных (9);
- оператор системы обнаружения аномалий (10).

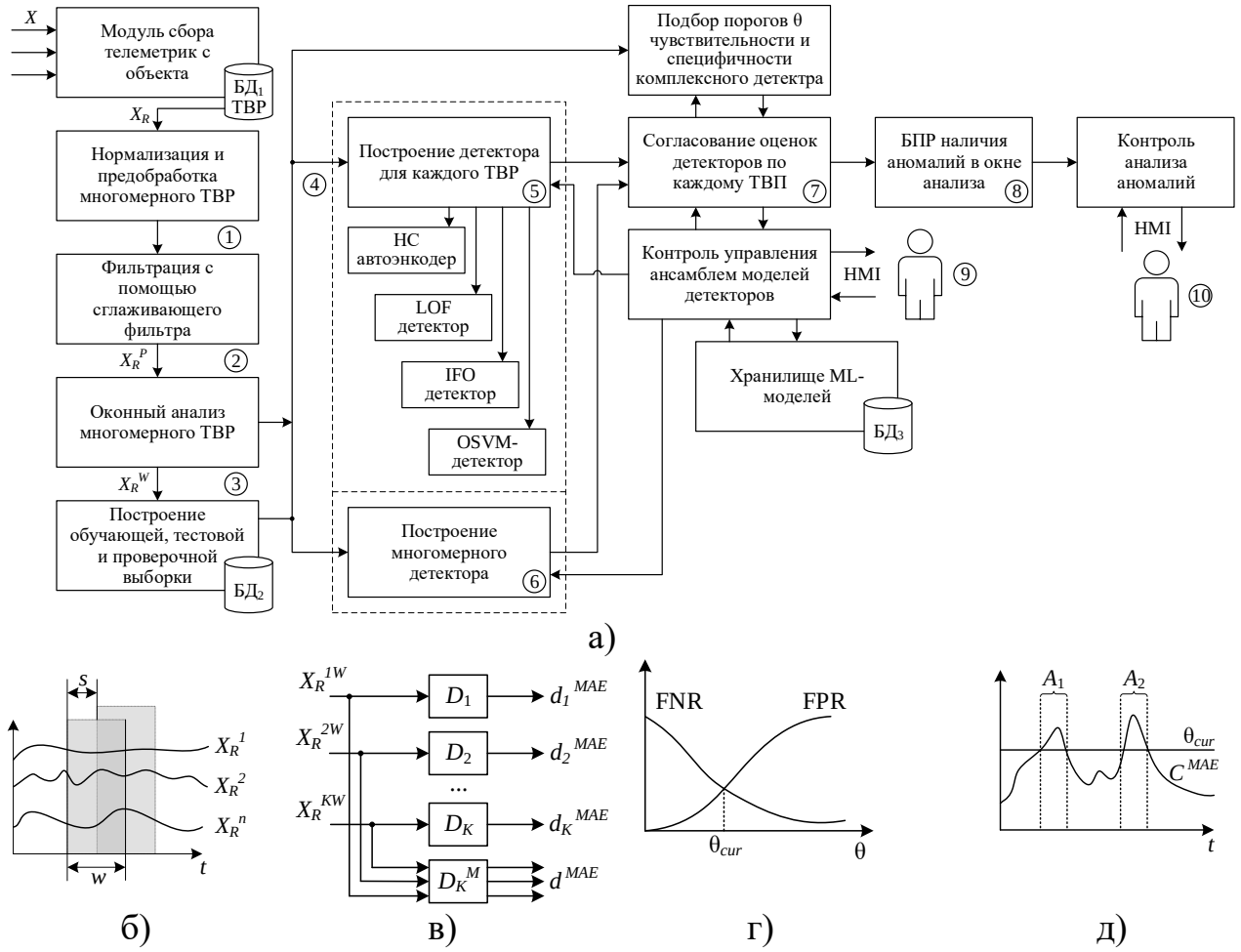


Рис. 6. Структурная схема системы обнаружения аномалий (а); процесс формирования с помощью скользящего окна фрагментов многомерного ТВР (б); многомерный нейросетевой детектор (в); подбор порогового значения для определения окна анализа на основе оценок относительного количества ложно-положительных (FPR) и ложно-отрицательных (FNR) срабатываний (г); визуализация разметки аномальных фрагментов ТВР на основе порогового сравнения ошибки восстановления образа с помощью детектора (д)

Нейросетевой автоенкодер (НА) в задаче обнаружения аномалий предназначен для восстановления (реконструкции) фрагмента ТВР. Обнаружение

аномалий осуществляется на основе порогового сравнения среднеквадратической (или абсолютной) ошибки между фактическими данными и восстановленным образом. Применяемая архитектура НА на основе долгой краткосрочной памяти (LSTM) является разновидностью архитектуры рекуррентных нейронных сетей глубокого обучения, способной к анализу долговременных зависимостей.

Автоенкодер состоит из двух частей (рис. 7):

- кодер – отображает входные данные $x \in R^{d_x}$ на внутреннее представление $z \in R^{d_z}$, $z = f\left(x, W_x\right)$, $W \in R^{d_z * d_x}$ – матрица весов;
- декодер – выполняет обратное отображение из внутреннего представления во входное пространство (реконструкция): $x = g\left(z\right) = g\left(f\left(x, W_x\right)\right)$, $W' \in R^{d_z * d_x}$ – матрица весов.

Процедура обучения автоенкодеров состоит в нахождении набора параметров: $\theta = \left(W\right)$, который минимизируют функцию потерь; $L\left(x, g\left(f\left(x\right)\right)\right)$, определяющую качество реконструкций образа – выходная реконструкция образа x должна быть как можно ближе к исходному входному вектору x. Типичный выбор для функции потерь – это среднеквадратичная ошибка:

$$L\left(x, g\left(f\left(x\right)\right)\right) = \left\| x - x \right\|_2^2 \cdot$$

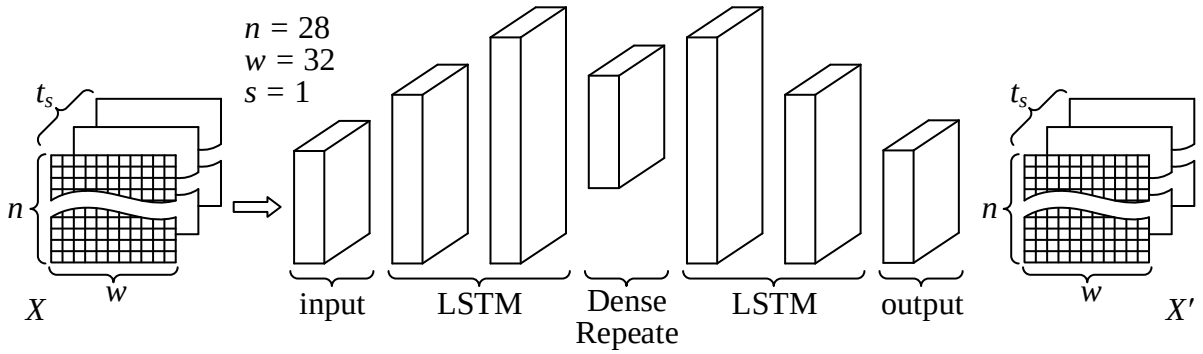


Рис. 7. Схема применяемого многомерного автоенкодера на основе LSTM, n – количество анализируемых параметров (количество ТВР), w – длина скользящего окна анализа, s – шаг скользящего окна анализа, t_S – глубина погружения в ТВР

Основным преимуществом применения моделей автоенкодеров в задаче обнаружения аномалий является возможность построения модели нормального поведения системы. При появлении новых типов аномалий или изменении характера текущих аномалий детектор на основе автоенкодера по-прежнему, оценивая ошибку реконструкции образа, способен выявлять подобные образы [32].

Метод изолированного леса (IFO) основан на предположении, что аномалии немногочисленны и отличаются от нормальных экземпляров данных. Для определения количества разбиений исходного набора данных, необходимых для выделения выбранного образца, используется рекурсивная процедура. Метод изолированного леса не строит модель нормального поведения системы,

непосредственно анализируя аномалии. Метод построения модели имеет линейную временную сложность и высокую производительность. При построении модели не учитываются временные зависимости в последовательности анализируемых отсчетов ТВР.

Модель на основе машины опорных векторов (One-class SVM) позволяет построить оптимальную гиперплоскость разделения объектов на два класса – нормальные и аномальные образцы, при этом допускается некоторое количество неправильных классификаций для поиска оптимального решения в неразделимых случаях.

Метод на основе оценки фактора локального выброса. Обеспечивает вычисление плотности расположения примеров в пространстве признаков относительно плотности расположения их соседей. Каждому экземпляру данных присваивается показатель аномальности – фактор локального выброса (LOF). Для любого экземпляра данных показатель LOF равен отношению оценки средней локальной плотности k ближайших соседей экземпляра к оценке локальной плотности для самого экземпляра.

Итоговая модель ансамбля детекторов для обнаружения аномалий в многомерном технологическом временном ряду, характеризующем ход технологического процесса, включает группу детекторов для одномерных ТВР и детектор для многомерного ТВР на основе нейросетевых автоенкодеров, LOF и IFO моделей (рис. 8).

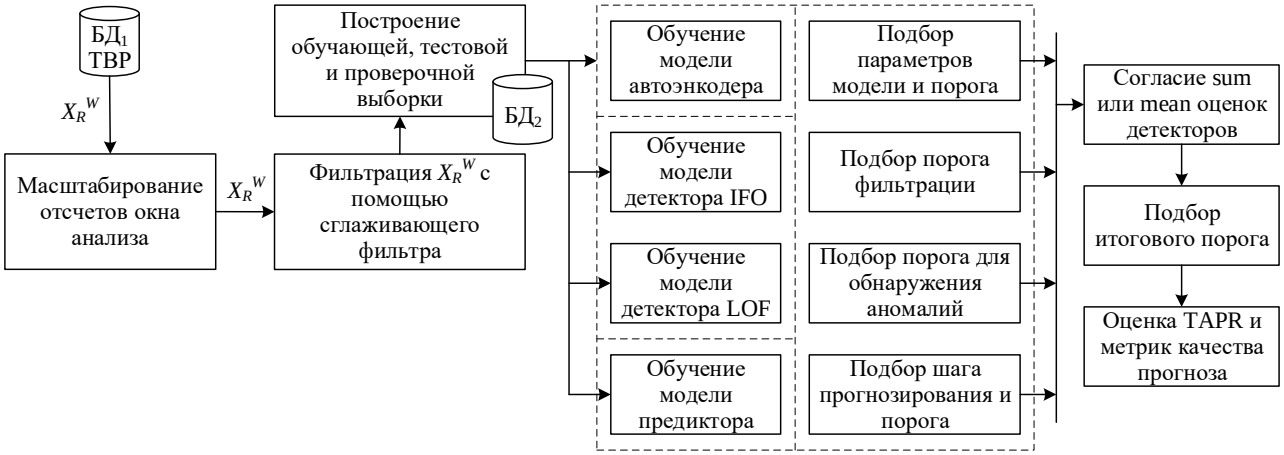


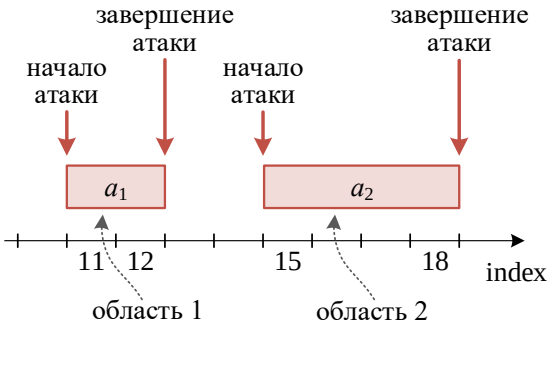
Рис. 8. Модель обнаружения аномалий на основе ансамбля детекторов

Оценка качества методов обнаружения аномалий в временных рядах параметров производится с помощью традиционных метрик качества классификации и TaPR – метрики оценки обнаружения аномалии и корректности границ аномалии во временных рядах. Как показано на рис. 9, конечная цель заключается в определении области действия атаки в совокупности отсчетов ТВР, например, область действия для двух атак a_1 и a_2.

Интегральный показатель TaPR (количество и правильность обнаружения аномалий во временном окне анализа) включает следующие частные показатели: оценка обнаружения аномалии, показатель TaR (сколько аномалий обнару-

жено), и оценка корректности границ обнаруженной аномалии, показатель TaP (насколько точно обнаруживается каждая аномалия) (рис. 10) [33].

Оценка качества в задачах распознавания аномалий наглядно иллюстрируется с помощью матрицы ошибок (таблица 2).



Index	Label	
10	0	
11	1	область 1
12	1	
13	0	
14	0	
15	1	
16	1	область 2
17	1	
18	1	
19	0	

Рис. 9. Модель обнаружения аномалий на основе ансамбля детекторов

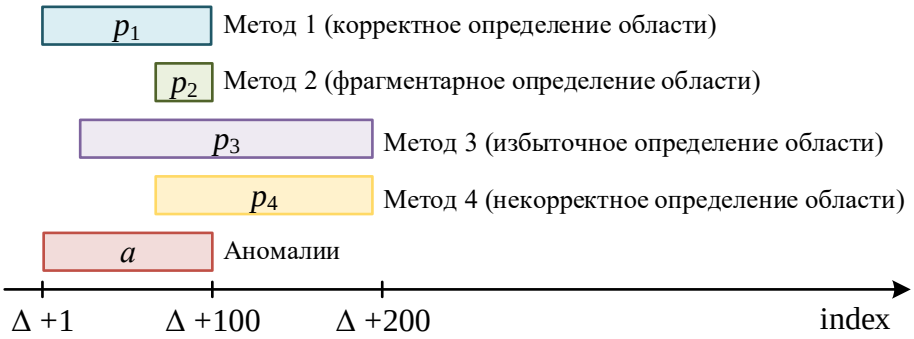


Рис. 10. Показатель TaPR

Таблица 2 – Матрица ошибок

	Принадлежит классу (P)	Не принадлежит классу (N)
Предсказана принадлежность определенному классу	*TP* – True Positive, классификатор верно отнес объект к рассматриваемому классу	*FP* – False Positive, классификатор неверно отнес объект к рассматриваемому классу
Предсказано отсутствие принадлежности к классу	*FN* – False Negative, классификатор неверно утверждает, что объект не принадлежит к рассматриваемому классу	*TN* – True Negative, классификатор верно утверждает, что объект не принадлежит к рассматриваемому классу

Показатели эффективности модели детектора:

– коэффициент точности, характеризующий долю верно классифицированных положительных объектов (precision):

$$PPV = \frac{TP}{TP + FP};$$

– коэффициент точности обнаружения, или полнота (recall):

$$TPR = \frac{TP}{TP + FN};$$

– доля истинно отрицательных классификаций, представляющая собой пропорцию отрицательных образцов, которые были корректно классифицированы как отрицательные (специфичность (specificity)):

$$TNR = \frac{TN}{TN + FP}.$$

Для общей оценки качества классификатора часто используют F_1 меру, учитывающую значения *precision* и *recall*:

$$F_1 = \left(\frac{Prec^{-1} + Recall^{-1}}{2} \right)^{-1}.$$

Показатель *F-beta* – это средневзвешенное гармоническое значение показателей *precision* и *recall*, достигающее оптимального значения при 1 и наихудшего при 0.

Эксперимент на натурных данных

Обобщенная схема алгоритма анализа ТВР в задаче обнаружения аномалий, вызванных воздействием злоумышленника, пытающегося перехватить управление или навязать алгоритм управления КФО, представлена на рис. 11.

Рис. 11. Обобщенная схема алгоритма анализа аномалий

Шаг 1. Предобработка и фильтрация данных. Обучающая выборка содержит 921603 примера без аномалий (атаки не проводились, нормальный режим работы системы) и 309604 примеров тестовых данных с тремя типа одиночных и комбинированных атак. Применяется нормализация количественных признаков – приведение к нулевому среднему и единичному стандартному отклонению, и выполняется преобразование категориальных переменных в количественные. Далее основной задачей является анализ взаимосвязи признаков.

Шаг 2. Отбор признаков. С помощью попарной корреляции Пирсона для переменных обучающей выборки построена тепловая карта, позволяющая оценить параметры с сильной линейной зависимостью. Удаление зависимых переменных позволит существенно ускорить обучение моделей обнаружения аномалий.

Выполняется поиск и удаление константных (22 признака), квазиконстантных (с порогом вариабельности (дисперсия) за период анализа (обучающая выборка) 0,005 удаляется 6 признаков) и коррелирующих признаков (с порогом 0,9 на основе матрицы попарной корреляции Пирсона – удаляются 23 признака) позволяет последовательно сократить количество анализируемых признаков до 28.

Шаг 3. Генерация оконных признаков. Строятся оконные признаки для группы детекторов. Скользящее окно длиной 90 отсчетов (экспертная оценка, являющаяся компромиссом для обнаружения длительных и кратковременных аномалий) с шагом 1 перемещается по каждому из ТВР 28 признаков. Для отсчетов, попавших в текущее скользящее окно анализа, рассчитываются следующие признаки:

- max – минимальное значение отсчета в окне;
- min – максимальное значение отсчета в окне;
- mean – среднее значение отсчетов в окне;
- std – среднеквадратичное отклонение отсчетов в окне.

Удаляются строки, содержащие неверные форматы данных или пропуски. Остается 1231118 записей, содержащих оконные признаки, определенные по описанной выше схеме, и элементы исходных ВР признаков – 147 признаков на каждую запись. В обучающей выборке окончательно содержится 921514 примеров, в тестовой – 309604 примеров.

Выполняется генерация оконных признаков для автоенкодера на основе LSTM. Для 28 ТВР с глубиной погружения в 32 отсчета строится с помощью скользящего окна с шагом 1 множество примеров обучающей и тестовой выборок.

Шаг 4-5. Построение детекторов аномалий и композиции детекторов. Исходный обучающий набор не содержит аномалий, вызванных действиями злоумышленника. Тестовый набор включает 6770 примеров, связанных с действиями злоумышленника (аномалии).

Объединенный набор данных содержит 1224315 примеров нормальной работы (отсчеты исходных 28 ВР и рассчитанные для каждого ВР признаки в скользящих окнах – по 5 признаков).

Параметры детекторов для каждого из ТВР приведены в (таблице 3 и 4). Параметры детектора на основе нейросетевого автоенкодера приведены в таблице 5. Процесс обучения нейронной сети показан на рис. 12.

Таблица 3 – Параметры модели детектора аномалий на основе изолирующего леса (Isolation Forest, IFO)

Параметр	Значение
Количество базовых оценок в ансамбле	128
Доля выбросов в наборе данных (используется при подгонке для определения порога оценки образцов) (contamination)	auto

Таблица 4 – Параметры детектор аномалий на основе модели оценки локального уровня выброса (Local Outlier Factor, LOF)

Параметр	Значение
Количество соседей, используемых по умолчанию для запросов	16

Таблица 5 – Архитектура многомерного нейросетевого автоенкодера LSTM

Слой	Функция активации	Размеры слоя	Регуляризация
Input	–	32, 28	–
LSTM (L1)	relu	32, 128	L2
LSTM (L2)	relu	64	–
RepeatVector (L3)	–	32, 64	–
LSTM (L4)	relu	32, 64	–
LSTM (L5)	relu	32, 128	–
Output	–	32, 28	–

Начальное обучение сети выполняется за 25 эпох, обучающая выборка делится на пакеты (батчи) размером 512 примеров.

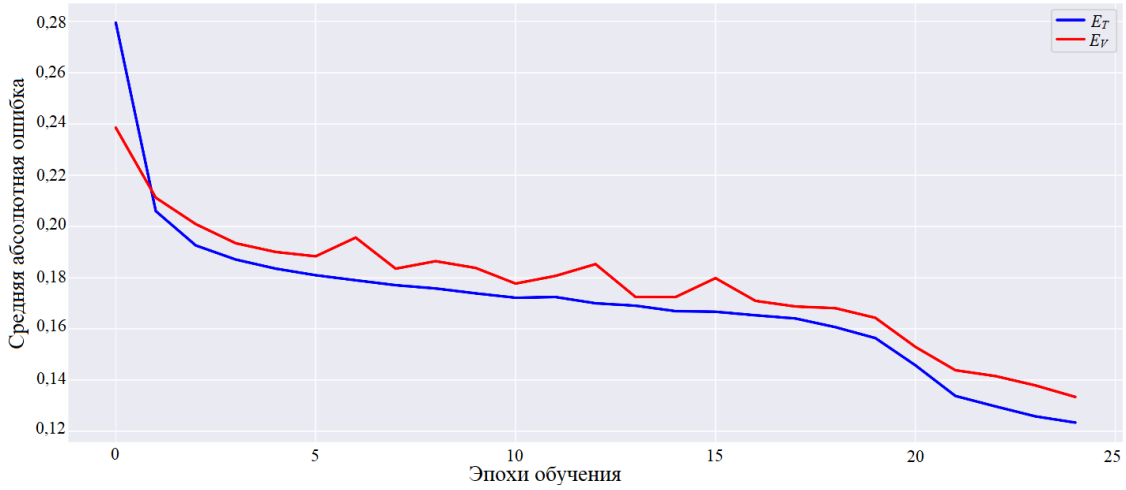


Рис. 12. Процесс начального обучения автоенкодера LSTM, E_T – суммарная квадратичная ошибка на обучающей выборке, E_V – суммарная квадратичная ошибка на проверочной выборке

№6. 2021
ISSN 2410-9916

Абсолютная ошибка (*MAE*) раскрывает ошибку восстановления образа с помощью автоенкодера по отношению к среднему значению расстояния между прогнозируемым моделью значением $f(x)$ и истинным значением y:

$$MAE = \frac{\sum_{i=1}^{n} \left| f(x_i) - y_i \right|}{n}.$$

Ошибка восстановления образа (Loss_mae) к концу начального обучения продолжает уменьшаться. Дообучение происходит 25 эпох, набор данных делится на пакеты (батчи) размером 2048. Итоговая гистограмма распределения ошибок восстановления образов с помощью автоенкодера приведена на рис. 13.

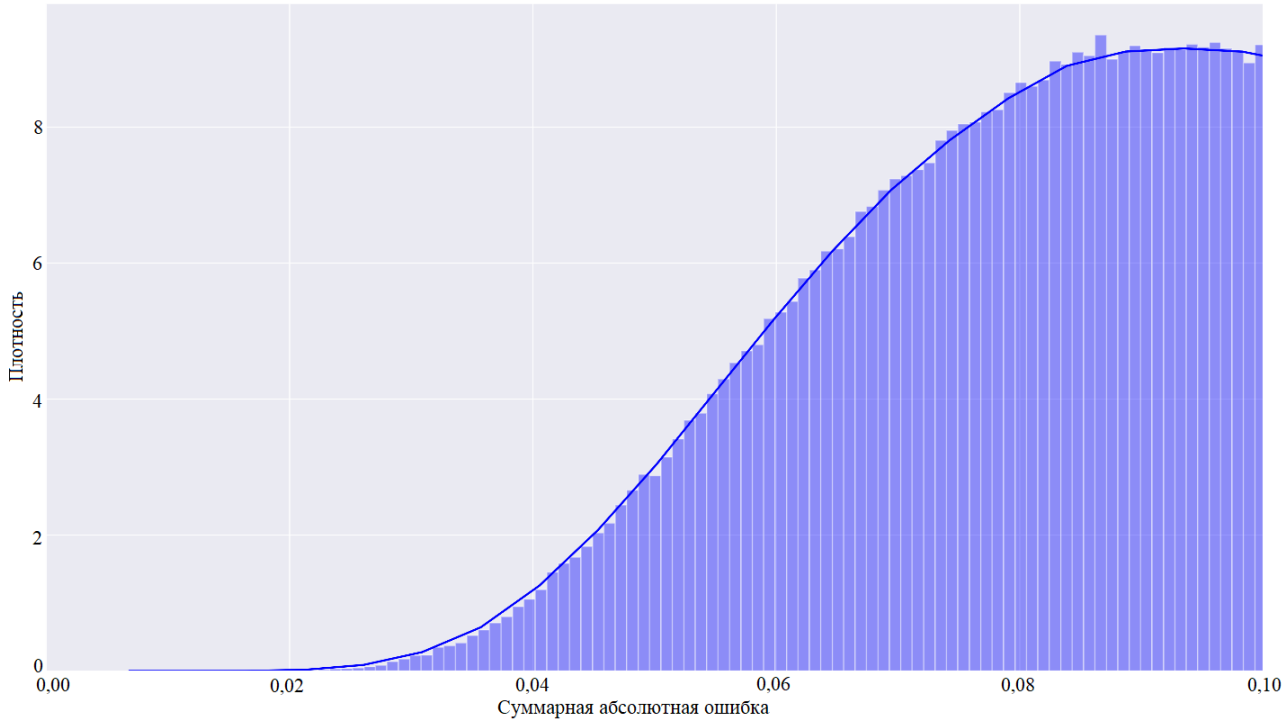


Рис. 13. Гистограмма распределения ошибок восстановления образа с помощью автоенкодера

Анализ результатов и оценка эффективности предложенного решения

Результаты разметки отсчетов ВР, попадающих в скользящие окна анализа для полного набора данных, включающего обучающую и тестовую выборки, приведены на рис. 14. Указана доля отсчетов, отнесенных к аномалиям, по отношению к общему числу отсчетов, определенная каждым детектором по каждой переменной.

Из рис. 14 видно, что детекторы на основе изолирующего леса значительно чаще относят примеры данных к аномальным, напротив, избирательность детекторов на основе LOF существенно лучше. Распределение суммарных оценок детекторов на основе изолирующего леса приведено на рис. 15.

DOI: 10.24412/2410-9916-2021-6-90-119
URL: https://sccs.intelgr.com/archive/2021-06/06-Vasilyev.pdf

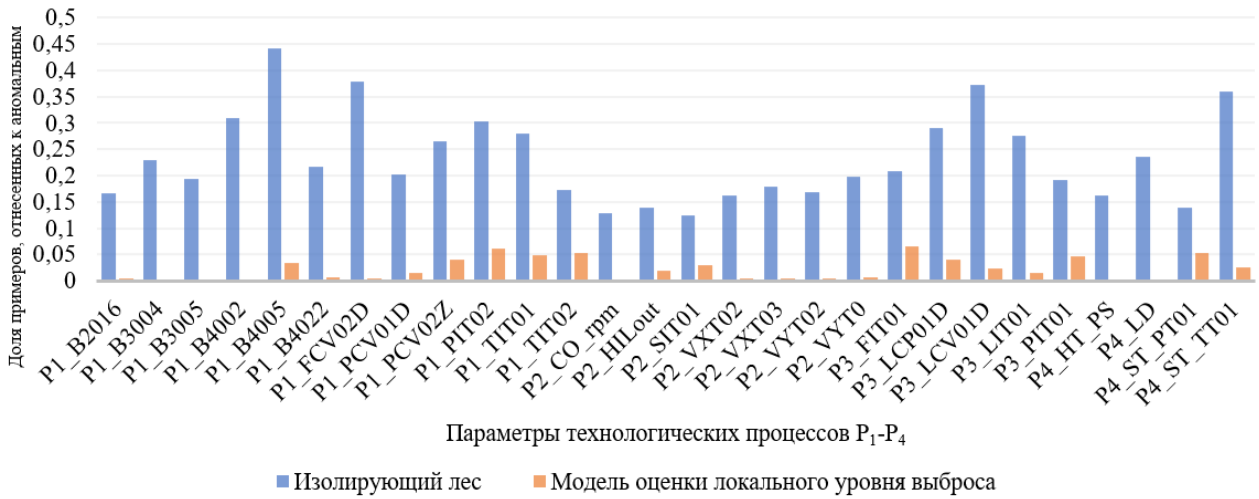


Рис. 14. Результаты разметки отсчетов детекторами на основе изолирующего леса и модели оценки локального уровня выброса: доля примеров (ось ординат), отнесенных к аномальным, по каждому параметру технологических процессов P_1-P_4 в обозначения набора данных (ось абсцисс)

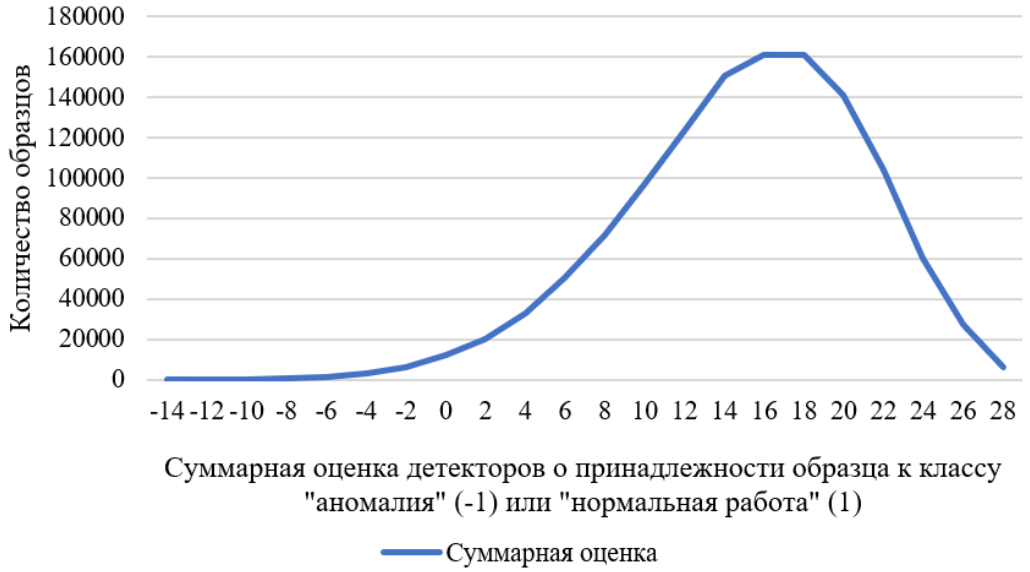


Рис. 15. Распределение суммарных оценок детекторов на основе изолирующего леса при оценке одномерных ТВР (ось абсцисс – суммарная оценка детекторов о принадлежности образца к классу аномалий «-1» или классу нормальной работы «1»; ось ординат – количество образцов)

Ключевым этапом является подбор порога чувствительности ансамбля детекторов. Определим порог для оценки принадлежности окна анализа к аномальному. Подбор оптимального значения порога для минимизации пропусков аномалий реализуем как поиск порога чувствительности и специфичности модели в диапазоне от $\Theta \in [10;29]$ с шагом 0,1. Исходный диапазон поиска определяется из рисунка 15 как отсечка, показывающая степень согласованности

детекторов по каждому ВР при отнесении текущего образца к нормальному или аномальному режиму работы.

Анализ чувствительности и F1-меры, взвешенных количеством примеров в каждом из классов оценок, позволяет выбрать приемлемое значение порога фильтрации аномалий и уменьшить количество ложных срабатываний.

Далее проанализируем гистограмму распределения ошибок восстановления образа с помощью автоенкодера (рис. 13). Из рисунка видно, что значение порога обнаружения аномалий может быть задано как значение в интервале (0,1-0,3). Подбор порога для оптимизации количества выявленных аномалий с помощью автоенкодера реализуем перебором по сетке: стартовое значение составляет 0,1, конечное – 0,35, шаг подбора – 0,05. Подобранное значение порога составляет 0,21. С установленным порогом качественно оценим распознавание аномалий каждого типа.

Стоит отметить, достаточно большое количество ложно положительных срабатываний ансамбля детекторов, для снижения которого необходимы дополнительные фильтры срабатываний на основе оценки временных аспектов реализации атак: минимальной длительности, потенциальной периодичности, минимального интервала времени между реализацией разных типов атак.

Сводная оценка качества работы детектора по обнаружению аномалий приведена в таблице 6. Итоговая оценка интегрального показателя TaPR для композиции детекторов: TaR = 0,993, TaP = 0,915.

Таблица 6 – Сводная таблица оценки качества работы детектора по обнаружению аномалий без постфильтрации

	Все атаки		Первая атака		Вторая атака		Третья атака	
	F1-мера	количество	F1-мера	количество	F1-мера	количество	F1-мера	количество
Нормальная работа	0,97	1224315	0,97	1225819	0,97	1229746	0,97	1230487
Аномалия, вызванная атакой	0,13	6770	0,10	5266	0,03	1339	0,02	598
Доля правильных классификаций (Accuracy)	0,95	1231085	0,95	1231085	0,95	1231085	0,95	1231085
Accuracy weighted avg (взвешенная количеством примеров доля правильных классификаций)	0,97	1231085	0,97	1231085	0,97	1231085	0,97	1231085
FP	58365		59395		61721		62288	
FN	2364		1890		289		115	
F_beta (0,5)	0,9835		0,9847		0,9882		0,9888	

В работе [34] рассмотрен подход к обнаружению аномалий на основе решения задачи классификации состояний объекта с помощью методов машинного обучения. Этапы предобработки, анализа значимости и отбора признаков за-

вершаются построением классификаторов: k ближайших соседей, комитета деревьев решений, и решающего дерева. Особенностью является объединение обучающей (примеры нормального функционирования объекта, не содержащие атак) и тестовой выборок (смесь нормальной работы и примеров, характеризующих реализованные атаки) с последующим исправлением дисбаланса количества примеров в классах нормальной и аномальной работы с помощью алгоритма увеличения числа примеров миноритарного класса (класс аномалий) SMOTE (Synthetic Minority Oversampling Technique). Далее аугментированная выборка разбивается на обучающую и тестовую в соотношении 70% и 30%. Итоговые характеристики качества классификации достигают оценки F1-меры на уровне 0,9976.

К недостаткам рассмотренного подхода стоит отнести следующее:

1) использование моделей, обучаемых с учителем, требует размеченного набора данных и первоначального объединения обучающей и тестовой выборок, предложенных разработчиками набора данных, – обучающая выборка в первоначальном варианте включает только данные, характеризующие нормальный режим работы объекта;
2) не в полной мере учитывается временная упорядоченность отсчетов данных о состоянии объекта – выполнена классификация изолированных временных отсчетов;
3) построенные модели позволяют с очень высокой точностью характеризовать отдельные последовательности отсчетов, без учета начала и завершения временного интервала, в течение которого реализуется атака злоумышленника, приводящая к изменениям параметров функционирования объекта. Т.е., невозможно оценить метрику TaP (насколько точно обнаруживается каждая аномалия).

В работе [35] также предложен подход, основанный на применении моделей, обучаемых с учителем:

- машина опорных векторов;
- комитет решающих деревьев;
- решающее дерево;
- классификатор k ближайших соседей;
- ансамбль стохастического градиентного бустинга (light gradient boosting machine, LightGBM).

На этапе конструирования новых признаков использован подход на основе вычисления статистических характеристик (среднее значение параметра, разброс, минимальное и максимальное значения) в скользящем окне анализа переменной длины. Существенным преимуществом предлагаемого решения при построении моделей классификации является принятие временной упорядоченности отсчетов параметров, характеризующих состояние объекта, и анализ влияния длины скользящего окна на итоговое качество классификации. Дальнейшая процедура перекрестной проверки также реализована для временных рядов параметров.

Характеристики качества классификации F1-меры находятся на уровне 0,9987. К сожалению, невозможно оценить метрику TaP (насколько точно обнаруживается каждая аномалия).

Однако, при построении моделей, обучаемых с учителем, необходима разметка примеров, поэтому авторы вновь объединяют исходные обучающую и тестовую выборки с последующим новым разбиением, что приводит к так называемой проблеме утечки данных [36].

В работе [33] предложено построение модели детектора аномалий на основе стекирования двунаправленных рекуррентных нейронных сетей (bidirectional Gated Recurrent Unit (GRU)), позволяющих учитывать временную природу анализируемых данных и частично решающих проблемы длительного и ресурсоемкого обучения нейронных сетей LSTM. Применение алгоритма автоматического подбора порога чувствительности на основе эвристик позволяет повысить качество итогового обнаружения аномалий. Полученная оценка взвешенной меры F1 составляет для тестовой выборки 0,977 (исходная выборка, предложенная создателями набора данных), рассчитаны значения метрик TaP = 0,968 и TaR = 0,805. Однако, для тестовой выборки сделаны предположения, что отдельные реализации одной атаки и вызванные ими аномалии в параметрах состояния объекта разделены временным интервалом не менее чем в 500 секунд. Атаки разных классов разделены интервалом более 2000 секунд. Предложенные допущения на порядок снижают количество ложно положительных срабатываний детектора.

Анализ таблицы 6 показывает, что количество ложноположительных срабатываний детектора существенно превышает общее количество примеров, связанных с реализацией атаки. Приняв во внимание дополнительные временные ограничения по возможности реализации злоумышленником атаки, предложенные в работе [33], добавим временной фильтр, позволяющий существенно снизить количество ложных срабатываний, лишь незначительно уменьшив количество корректно распознанных аномалий. Примем, что атаки разделены временным интервалом $t_1 = 350$ с, класс атаки меняется с интервалом не менее $t_2 = 1500$ с (таблица 7).

Количество ложноположительных срабатываний детектора удается снизить в 10-15 раз. Стоит отметить, что значительная доля атак каждого класса обнаруживается детектором достаточно уверенно, количество ложноотрицательных случаев (пропусков атак) для 2 и 3 типа атак находится на приемлемом уровне. Практическое применение подобной системы возможно в составе комплекса средств защиты промышленной сети, выступающих в качестве источников событий безопасности для системы сбора и корреляции событий информационной безопасности. Дальнейший анализ больших массивов гетерогенных данных о событиях безопасности и обнаружения инцидентов и угроз безопасности является основной, наиболее ресурсоемкой операцией, которая выявляет причинно-следственные связи между поступающими на обработку событиями. Операция корреляции позволяет выявлять вредоносную и аномальную активности, определять источник и цель атаки на основе анализа комплекса показателей, но не единичных инструментов и детекторов, и отсеивать значительное

количество ложноположительных срабатываний за счет сопоставления признаков атаки из разных источников.

Таблица 7 – Сводная таблица оценки качества работы детектора по обнаружению аномалий с применением постфильтрации ложноположительных срабатываний

Постфильтрация	Показатель		Все атаки		Первая атака		Вторая атака		Третья атака	
Нет	*TP*	*FP*	1165950	58365	1166424	59395	1168025	61721	1168199	62288
	FN	*TN*	2364	4406	1890	3376	289	1050	115	483
	F1-мера (по классам)		0,975	0,127	0,974	0,099	0,974	0,033	0,974	0,015
Есть	*TP*	*FP*	1220425	3890	1222329	3490	1226867	2879	1225941	4546
	FN	*TN*	2399	4371	2007	3259	358	981	128	470
	F1-мера (по классам)		0,997	0,582	0,998	0,542	0,998	0,377	0,998	0,167

Таким образом, особенностями предложенного авторами подхода к построению ансамбля детекторов аномалий по сравнению с аналогичными исследованиями является:

- использование моделей, не требующих разметки на классы нормального и аномального функционирования, что потенциально позволяет обнаруживать аномалии, вызванные новыми типами атак злоумышленника;
- использование исходных обучающей и тестовой выборок, предложенных авторами набора данных HAI 2.0, без их объединения и утечки данных о природе аномалий из тестовой в обучающую выборку;
- возможность работы с несбалансированной выборкой как в отношении бинарной классификации (нормальная работа – аномалия), так и в задаче обнаружения отдельных типов атак;
- отсутствие допущений о периодичности и длительности атак на этапе тестирования модели, поскольку допущение, подобные [33], позволили для тестовой выборки уменьшить количество ложноположительных срабатываний более чем в 10 раз и получить уровень оценки метрики F1 = 0,9781 по сравнению с F1 = 0,977 в [33];
- итоговая оценка качества обнаружения (интегральный показатель обнаружения аномалии и корректности границ аномалии во временных рядах) для композиции детекторов составляет TaR = 0,993, TaP = 0,915, что сравнимо (показатель TaR) и превосходит (показатель TaP) лучшие результаты исследователей [33] TaR = 0,968 и TaP = 0,805.

Корректность обнаружения аномалий первого типа с помощью гетерогенной модели детекторов составила 69%, второго типа – 78%, третьего типа – 80% с применением постфильтрации результатов на основе эвристик длительности и периодичности атак. Дальнейшая работа по подбору параметров глубины погружения в историю ВР параметров объекта и оптимизация архитектуры нейросетевого автоенкодера позволит добиться лучших результатов, снижая

количество ложных срабатываний. Перспективной является гетерогенная архитектура нейросетевого автоенкодера и модели LOF, а также добавление еще одного выходного фильтра, позволяющего учитывать временные особенности реализации атак и длительность вызванных изменений в наблюдаемых параметрах.

Заключение

Для выявления сложных атак злоумышленников, получивших доступ с сеть промышленного объекта, необходимо применение методов и инструментов расширенной аналитики данных, позволяющих выполнять оперативный анализ и выявление скрытых признаков злонамеренной активности на основе модели наблюдаемого КФО. Для построения модели обнаружения аномалий состояния КФО, вызванных действиями злоумышленника в промышленной сети в ходе реализации сложной сетевой атаки, нами был использован доступный набор данных, который был собран в ходе испытаний АСУ ТП промышленного объекта и дополнен результатами программно-аппаратного моделирования.

Предложена структурная схема системы обнаружения аномалий технологического процесса, основанная на применении методов предиктивного анализа собираемых данных телеметрии и позволяющая выявить воздействия злоумышленника, получившего доступ в промышленную сеть управления технологическим процессом.

Разработана гетерогенная модель ансамбля детекторов для обнаружения аномалий в многомерном технологическом временном ряду параметров, характеризующих ход технологического процесса. Модель включает группу детекторов для одномерных ТВР и детектор для многомерного ТВР на основе нейросетевых автоенкодеров, LOF и IFO моделей.

Разработан алгоритм интеллектуального анализа технологических временных рядов в задаче обнаружения аномалий наблюдаемых параметров состояния объектов АСУ ТП. Обнаружение аномалий по всем типам составило в среднем 65 % (первого типа – 69 %, второго типа – 78 %, третьего типа – 80 %) при принятии допущений о периодичности и длительности атак, что свидетельствует об эффективности решения поставленной задачи. Дальнейшее развитие фильтрации ложноположительных срабатываний на основе анализа временных характеристик потенциальных атак позволит повысить эффективность предлагаемого ансамбля детекторов.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-08-00668 и 19-07-00895.

Литература

1. Милославская Н. Г. Научные основы построения центров управления сетевой безопасностью в информационно-телекоммуникационных сетях. – М.: Горячая Линия-Телеком, 2021. – 432 с.

2. Lavrova D. S. An approach to developing the SIEM system for the Internet of Things // Automatic control and computer sciences. 2016. Vol. 50. № 8. P. 673-681.

3. Зегжда Д. П. Кибербезопасность цифровой индустрии. Теория и практика функциональной устойчивости к кибератакам. – М.: Горячая Линия-Телеком, 2020. – 560 с.

4. Pang G., Shen C., Cao L., Hengel A. V. D. Deep learning for anomaly detection: A review // ACM Computing Surveys (CSUR). 2021. Vol. 54. № 2. P. 1-38.

5. Шелухин О. И. Сетевые аномалии. Обнаружение, локализация, прогнозирование. – М.: Горячая Линия-Телеком, 2020. – 448 с.

6. Современные технологии аналитики в кибербезопасности // Газинформсервис [Электронный ресурс]. – URL: https://habr.com/ru/company/gaz-is/blog/480980/ (дата обращения 24.08.2021).

7. Monshizadeh M., Khatri V., Atli B. G., Kantola R., Yan Z. Performance evaluation of a combined anomaly detection platform // IEEE Access. 2019. Vol. 7. P. 100964-100978.

8. Moustafa N., Creech G., Sitnikova E., Keshk M. Collaborative anomaly detection framework for handling big data of cloud computing // 2017 military communications and information systems conference (MilCIS). IEEE. 2017. P. 1-6.

9. Ten C. W., Manimaran G., Liu C. C. Cybersecurity for critical infrastructures: Attack and defense modeling // IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans. 2010. Vol. 40. № 4. P. 853-865.

10. Ten C. W., Hong J., Liu C. C. Anomaly detection for cybersecurity of the substations // IEEE Transactions on Smart Grid. 2011. Vol. 2. № 4. P. 865-873.

11. Alrashdi I., Alqazzaz A., Aloufi E., Alharthi R., Zohdy M., Ming H. Ad-iot: Anomaly detection of iot cyberattacks in smart city using machine learning // 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC). IEEE. 2019. P. 305-310.

12. Kiss I., Genge B., Haller P., Sebestyén G. Data clustering-based anomaly detection in industrial control systems // 2014 IEEE 10th International Conference on Intelligent Computer Communication and Processing (ICCP). IEEE. 2014. P. 275-281.

13. Cruz T., Rosa L., Proença J., Maglaras L., Aubigny M., Lev L., Simoes P. A cybersecurity detection framework for supervisory control and data acquisition systems // IEEE Transactions on Industrial Informatics. 2016. Vol. 12. № 6. P. 2236-2246.

14. Tartakovsky A. G., Polunchenko A. S., Sokolov G. Efficient computer network anomaly detection by changepoint detection methods // IEEE Journal of Selected Topics in Signal Processing. 2012. Vol. 7. № 1. P. 4-11.

15. Keshk M., Sitnikova E., Moustafa N., Hu J., Khalil I. An integrated framework for privacy-preserving based anomaly detection for cyber-physical systems // IEEE Transactions on Sustainable Computing. 2019. Vol. 6. № 1. P. 66-79.

16. Gómez Á. L. P., Maimó L. F., Celdran A. H., Clemente F. J. G., Sarmiento C. C., Masa C. J. D. C., Nistal R. M. On the generation of anomaly detection datasets in industrial control systems // IEEE Access. 2019. Vol. 7. P. 177460-177473.

17. Hariharan A., Gupta A., Pal T. Camlpad: Cybersecurity autonomous machine learning platform for anomaly detection // Future of Information and Communication Conference. Springer, Cham. 2020. P. 705-720.

18. Siddiqui M. A., Stokes J. W., Seifert C., Argyle E., McCann R., Neil J., Carroll J. Detecting cyber attacks using anomaly detection with explanations and expert feedback // ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE. 2019. P. 2872-2876.

19. Tuor A., Kaplan S., Hutchinson B., Nichols N., Robinson S. Deep learning for unsupervised insider threat detection in structured cybersecurity data streams // Workshops at the Thirty-First AAAI Conference on Artificial Intelligence. 2017.

20. Chen S., Janeja V. P. Human perspective to anomaly detection for cybersecurity // Journal of Intelligent Information Systems. 2014. Vol. 42. № 1. P. 133-153.

21. Pandit R. K., Infield D. SCADA-based wind turbine anomaly detection using Gaussian process models for wind turbine condition monitoring purposes // IET Renewable Power Generation. 2018. Vol. 12. № 11. P. 1249-1255.

22. Quatrini E., Costantino F., Di Gravio G., Patriarca R. Machine learning for anomaly detection and process phase classification to improve safety and maintenance activities // Journal of Manufacturing Systems. 2020. Vol. 56. P. 117-132.

23. Liu L., Liu D., Zhang Y., Peng Y. Effective sensor selection and data anomaly detection for condition monitoring of aircraft engines // Sensors. 2016. Vol. 16. № 5. P. 623.

24. Catterson V. M., Rudd S. E., McArthur S. D., Moss G. On-line transformer condition monitoring through diagnostics and anomaly detection // 2009 15th International Conference on Intelligent System Applications to Power Systems. IEEE. 2009. P. 1-6.

25. Goh J., Adepu S., Tan M., Lee Z. S. Anomaly detection in cyber physical systems using recurrent neural networks // 2017 IEEE 18th International Symposium on High Assurance Systems Engineering (HASE). IEEE. 2017. P. 140-145.

26. Das T. K., Adepu S., Zhou J. Anomaly detection in industrial control systems using logical analysis of data // Computers & Security. 2020. Vol. 96. P. 101935.

27. Karimipour H., Geris S., Dehghantanha A., Leung H. Intelligent anomaly detection for large-scale smart grids // 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE). IEEE. 2019. P. 1-4.

28. Shin H. K., Lee W., Yun J. H., Kim H. HAI 1.0: HIL-based Augmented ICS Security Dataset // 13th USENIX Workshop on Cyber Security Experimentation and Test (CSET'20). Santa Clara, CA, 2020.

29. Choi S. HAI Security Dataset. HIL-based Augmented ICS (HAI) Security Dataset // Kaggle. ICS Security Dataset [Электронный ресурс]. – URL: https://www.kaggle.com/icsdataset/hai-security-dataset (дата обращения 24.08.2021).

30. Hwang W. S., Yun J. H., Kim J., Kim H. C. Time-series aware precision and recall for anomaly detection: considering variety of detection result and addressing ambiguous labeling // Proceedings of the 28th ACM International Conference on Information and Knowledge Management. 2019. P. 2241-2244.

31. Choi S., Yun J. H., Kim S. K. A comparison of ICS datasets for security research based on attack paths // International Conference on Critical Information Infrastructures Security. Springer, Cham. 2018. P. 154-166.

32. Pereira J. Unsupervised anomaly detection in time series data using deep learning // Master's thesis, Instituto Superior Técnico, University of Lisbon, 2018.

33. Bian X. Detecting Anomalies in Time-Series Data using Unsupervised Learning and Analysis on Infrequent Signatures // Journal of IKEEE. 2020. Vol. 24. № 4. P. 1011-1016.

34. Mokhtari S., Abbaspour A., Yen K. K., Sargolzaei A. A machine learning approach for anomaly detection in industrial control systems based on measurement data // Electronics. 2021. Vol. 10. № 4. P. 407.

35. Park S., Lee K. Improved Mitigation of Cyber Threats in IIoT for Smart Cities: A New-Era Approach and Scheme // Sensors. 2021. Vol 21. № 6. P. 1976.

36. Samala R. K., Chan H. P., Hadjiiski L., Koneru S. Hazards of data leakage in machine learning: a study on classification of breast cancer using deep neural networks // Medical Imaging 2020: Computer-Aided Diagnosis. International Society for Optics and Photonics. 2020. Vol. 11314. P. 1131416.

References

1. Miloslavskaya N. G. *Nauchnye osnovy postroeniya centrov upravleniya setevoj bezopasnost'yu v informacionno-telekommunikacionnyh setyah* [Scientific foundations of building network security control centers in information and telecommunication networks]. Moscow, Goryachaya liniya-Telekom, 2021. 432 p. (in Russian).

2. Lavrova D. S. An approach to developing the SIEM system for the Internet of Things. *Automatic control and computer sciences*, 2016, vol. 50, no. 8, pp. 673-681.

3. Zegzhda D. P. *Kiberbezopasnost' cifrovoj industrii. Teoriya i praktika funkcional'noj ustojchivosti k kiberatakam* [Cybersecurity of the digital industry. Theory and practice of functional resilience to cyber attacks]. Moscow, Goryachaya liniya-Telekom, 2020. 560 p. (in Russian).

4. Pang G., Shen C., Cao L., Hengel A. V. D. Deep learning for anomaly detection: A review. *ACM Computing Surveys (CSUR)*, 2021, vol. 54, no. 2, pp. 1-38.

5. Sheluhin O. I. *Setevye anomalii. Obnaruzhenie, lokalizaciya, prognozirovanie* [Network anomalies. Detection, localization, forecasting]. Moscow, Goryachaya liniya-Telekom, 2020. 448 p. (in Russian).

6. Sovremennye tekhnologii analitiki v kiberbezopasnosti [Modern technologies of analytics in cybersecurity]. *Gazinformservice.* Available at: https://habr.com/ru/company/gaz-is/blog/480980/ (accessed 24 August 2021) (in Russian).

7. Monshizadeh M., Khatri V., Atli B. G., Kantola R., Yan Z. Performance evaluation of a combined anomaly detection platform. *IEEE Access*, 2019, vol. 7, pp. 100964-100978.

8. Moustafa N., Creech G., Sitnikova E., Keshk M. Collaborative anomaly detection framework for handling big data of cloud computing. *2017 military communications and information systems conference (MilCIS). IEEE*, 2017, pp. 1-6.

9. Ten C. W., Manimaran G., Liu C. C. Cybersecurity for critical infrastructures: Attack and defense modeling. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 2010, vol. 40, no. 4, pp. 853-865.

10. Ten C. W., Hong J., Liu C. C. Anomaly detection for cybersecurity of the substations. *IEEE Transactions on Smart Grid*, 2011, vol. 2, no. 4, pp. 865-873.

11. Alrashdi I., Alqazzaz A., Aloufi E., Alharthi R., Zohdy M., Ming H. Ad-iot: Anomaly detection of iot cyberattacks in smart city using machine learning. *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC). IEEE*, 2019, pp. 305-310.

12. Kiss I., Genge B., Haller P., Sebestyén G. Data clustering-based anomaly detection in industrial control systems. *2014 IEEE 10th International Conference on Intelligent Computer Communication and Processing (ICCP). IEEE*, 2014, pp. 275-281.

13. Cruz T. et al. A cybersecurity detection framework for supervisory control and data acquisition systems. *IEEE Transactions on Industrial Informatics*, 2016, vol. 12, no. 6, pp. 2236-2246.

14. Tartakovsky A. G., Polunchenko A. S., Sokolov G. Efficient computer network anomaly detection by changepoint detection methods. *IEEE Journal of Selected Topics in Signal Processing*, 2012, vol. 7, no. 1, pp. 4-11.

15. Keshk M., Sitnikova E., Moustafa N., Hu J., Khalil I. An integrated framework for privacy-preserving based anomaly detection for cyber-physical systems. *IEEE Transactions on Sustainable Computing*, 2019, vol. 6, no. 1, pp. 66-79.

16. Gómez Á. L. P., Maimó L. F., Celdran A. H., Clemente F. J. G., Sarmiento C. C., Masa C. J. D. C., Nistal R. M. On the generation of anomaly detection datasets in industrial control systems. *IEEE Access*, 2019, vol. 7, pp. 177460-177473.

17. Hariharan A., Gupta A., Pal T. Camlpad: Cybersecurity autonomous machine learning platform for anomaly detection. *Future of Information and Communication Conference*, Springer, Cham, 2020, pp. 705-720.

18. Siddiqui M. A., Stokes J. W., Seifert C., Argyle E., McCann R., Neil J., Carroll J. Detecting cyber attacks using anomaly detection with explanations and expert feedback. *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE*, 2019, pp. 2872-2876.

19. Tuor A., Kaplan S., Hutchinson B., Nichols N., Robinson S. Deep learning for unsupervised insider threat detection in structured cybersecurity data streams. *Workshops at the Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

20. Chen S., Janeja V. P. Human perspective to anomaly detection for cybersecurity. *Journal of Intelligent Information Systems*, 2014, vol. 42, no. 1, pp. 133-153.

21. Pandit R. K., Infield D. SCADA-based wind turbine anomaly detection using Gaussian process models for wind turbine condition monitoring purposes. *IET Renewable Power Generation*, 2018, vol. 12, no. 11, pp. 1249-1255.

22. Quatrini E., Costantino F., Di Gravio G., Patriarca R. Machine learning for anomaly detection and process phase classification to improve safety and maintenance activities. *Journal of Manufacturing Systems*, 2020, vol. 56, pp. 117-132.

23. Liu L., Liu D., Zhang Y., Peng Y. Effective sensor selection and data anomaly detection for condition monitoring of aircraft engines. *Sensors*, 2016, vol. 16, no. 5, pp. 623.

24. Catterson V. M., Rudd S. E., McArthur S. D., Moss G. On-line transformer condition monitoring through diagnostics and anomaly detection. *2009 15th International Conference on Intelligent System Applications to Power Systems. IEEE*, 2009, pp. 1-6.

25. Goh J., Adepu S., Tan M., Lee Z. S. Anomaly detection in cyber physical systems using recurrent neural networks. *2017 IEEE 18th International Symposium on High Assurance Systems Engineering (HASE). IEEE*, 2017, pp. 140-145.

26. Das T. K., Adepu S., Zhou J. Anomaly detection in industrial control systems using logical analysis of data. *Computers & Security*, 2020, vol. 96, pp. 101935.

27. Karimipour H., Geris S., Dehghantanha A., Leung H. Intelligent anomaly detection for large-scale smart grids. *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE). IEEE*, 2019, pp. 1-4.

28. Shin H. K., Lee W., Yun J. H., Kim H. HAI 1.0: HIL-based Augmented ICS Security Dataset. *13th USENIX Workshop on Cyber Security Experimentation and Test (CSET'20)*. Santa Clara, CA, 2020.

29. Choi S. HAI Security Dataset. HIL-based Augmented ICS (HAI) Security Dataset. *Kaggle. ICS Security Dataset* Available at: https://www.kaggle.com/icsdataset/hai-security-dataset (accessed 24 August 2021).

30. Hwang W. S., Yun J. H., Kim J., Kim H. C. Time-series aware precision and recall for anomaly detection: considering variety of detection result and addressing ambiguous labeling. *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 2019, pp. 2241-2244.

31. Choi S., Yun J. H., Kim S. K. A comparison of ICS datasets for security research based on attack paths. *International Conference on Critical Information Infrastructures Security*. Springer, Cham, 2018, pp. 154-166.

32. Pereira J. *Unsupervised anomaly detection in time series data using deep learning.* Master's thesis, Instituto Superior Técnico, University of Lisbon, 2018.

33. Bian X. Detecting Anomalies in Time-Series Data using Unsupervised Learning and Analysis on Infrequent Signatures. *Journal of IKEEE*, 2020, vol. 24, no. 4, pp. 1011-1016.

34. Mokhtari S., Abbaspour A., Yen K. K., Sargolzaei A. A machine learning approach for anomaly detection in industrial control systems based on measurement data. *Electronics*, 2021, vol. 10, no. 4, pp. 407.

35. Park S., Lee K. Improved Mitigation of Cyber Threats in IIoT for Smart Cities: A New-Era Approach and Scheme. *Sensors*, 2021, vol 21, no. 6, pp. 1976.

36. Samala R. K., Chan H. P., Hadjiiski L., Koneru S. Hazards of data leakage in machine learning: a study on classification of breast cancer using deep neural networks. *Medical Imaging 2020: Computer-Aided Diagnosis. International Society for Optics and Photonics*, 2020, vol. 11314, pp. 1131416.

Статья поступила 24 августа 2021 г.

Информация об авторах

Васильев Владимир Иванович – доктор технических наук, профессор. Профессор кафедры вычислительной техники и защиты информации. Уфимский государственный авиационный технический университет. Область научных интересов: интеллектуальные системы управления; информационная безопасность. E–mail: vasilyev@ugatu.ac.ru

Вульфин Алексей Михайлович – кандидат технических наук. Доцент кафедры вычислительной техники и защиты информации. Уфимский государственный авиационный технический университет. Область научных интересов: интеллектуальный анализ данных; моделирование сложных технических систем. E–mail: vulfin.alexey@gmail.com

Гвоздев Владимир Ефимович – доктор технических наук, профессор. Профессор кафедры технической кибернетики. Уфимский государственный авиационный технический университет. Область научных интересов: интеллектуальный анализ данных, функциональная безопасность. E–mail: wega55@mail.ru

Картак Вадим Михайлович – доктор физико-математических наук, доцент. Заведующий кафедрой вычислительной техники и защиты информации. Уфимский государственный авиационный технический университет. Область научных интересов: большие данные, методы оптимизации, информационная безопасность. E–mail: kvmail@mail.ru

Атарская Елена Андреевна – магистрант кафедры вычислительной техники и защиты информации. Уфимский государственный авиационный технический университет. Область научных интересов: интеллектуальный анализ данных. E–mail: arskaya25@mail.ru

Адрес: 450008, Россия, г. Уфа, ул. К. Маркса, д. 12.

Ensuring information security of cyber-physical objects based on predicting and detecting anomalies in their state

V. I. Vasilyev, A. M. Vulfin, V. E. Gvozdev, V. M. Kartak, E. A. Atarskaya

Formulation of the problem. *Ensuring the stable functioning of cyber-physical systems by improving predictive analysis methods aimed at identifying operational failures caused by the actions of an attacker and leading to the degradation of cyber-physical objects (CPOs), based on the identification of anomalies in the technological time series of parameters of the state of CPOs within the framework of the concept of advanced detection and elimination of cyber security threats.* ***Purpose.*** *Increasing the efficiency of detecting anomalies in the observed parameters of cyber-physical systems by improving the algorithms for detecting anomalies in technological time series of the accumulated parameters of the state of CPOs based on intelligent analysis.* ***Methods*** *of intellectual analysis of multidimensional technological time series are used with the use of a heterogeneous ensemble of detectors to detect anomalies in the accumulated parameters of the state of a CPO. The anomaly detection model includes a group of detectors for a univariate time series and a detector for a multivariate time series based on neural network autoencoders, an isolation forest model, and an estimate of the local outlier factor.* ***Novelty:*** *anomaly detection model based on a heterogeneous ensemble of detectors. The difference lies in the use of neural network autoencoders based on long short-term memory to simulate the normal behavior of the system. When new types of anomalies appear or the nature of current*

anomalies changes, the detector, based on the evaluation of the image recovery error, retains its operability. ***Results:*** *Block diagram of a process anomaly detection system based on the use of predictive analysis methods for collected telemetry data of a CPO and allowing to identify the impact of an attacker who has gained access to an industrial process control network; an algorithm for analyzing technological time series and a heterogeneous model of detectors for detecting anomalies caused by an attacker trying to intercept control or impose a control algorithm on a CPO.* ***Practical relevance.*** *The proposed approach is aimed at improving the mechanisms of predictive analysis as part of systems for detecting and eliminating anomalies in production and technological processes of automated process control systems. The system can be used as part of a complex of industrial network protection tools that act as sources of security events for the system for collecting and correlating cybersecurity events.*

Key words: *cyber-physical object; time series; anomaly detectors; neural network autoencoder with long-short-term memory.*

Information about Authors

Vladimir Ivanovich Vasilyev – Dr. habil. of Engineering Sciences, Full Professor. Professor at the Department of Computer Engineering and Information Security. Ufa State Aviation Technical University. Field of research: intelligent control systems; information security. E–mail: vasilyev@ugatu.ac.ru

Alexey Mikhailovich Vulfin – Ph.D. of Engineering Sciences. Associate Professor at the Department of Computer Engineering and Information Security. Ufa State Aviation Technical University. Field of research: data mining; modeling of complex technical systems. E–mail: vulfin.alexey@gmail.com

Vladimir Efimovich Gvozdev – Dr. habil. of Engineering Sciences, Full Professor. Professor at the Department of Technical Cybernetics. Ufa State Aviation Technical University. Field of research: data mining, functional safety. E–mail: wega55@mail.ru

Vadim Mikhailovich Kartak – Dr. habil. of Physical and Mathematical Sciences, Docent. Head of the Department of Computer Engineering and Information Security. Ufa State Aviation Technical University. Field of research: big data, optimization methods, information security. E–mail: kvmail@mail.ru

Elena Andreevna Atarskaya – Master of Science at the Department of Computer Engineering and Information Security. Ufa State Aviation Technical University. Field of research: data mining. E–mail: arskaya25@mail.ru

Address: 450008, Russia, Ufa, Karl Marx Str., 12.