УДК 004

ПРИМЕНЕНИЕ САМООБУЧАЮЩИХСЯ МОДЕЛЕЙ ДЛЯ АВТОМАТИЧЕСКОГО ВЫЯВЛЕНИЯ АНОМАЛИЙ В ПОТОКАХ IOT-УСТРОЙСТВ

Пидгурская Е.С.[1], студент гр.АТПРб-22-1, 3 курс
[1]Иркутский национальный исследовательский университет, г.Иркутск

В эпоху экспоненциального роста Интернета Вещей (IoT) безопасность и надежность устройств стали критическими вызовами для инженеров и исследователей. С развитием умных городов, промышленного интернета вещей (IIoT) и автономных систем потоки данных от миллионов датчиков формируют сложную экосистему, уязвимую для кибератак и сбоев оборудования. Традиционные методы мониторинга, основанные на правилах и статических порогах, не справляются с динамичностью современных IoT-сред. Самообучающиеся модели машинного обучения открывают новые возможности для автоматического обнаружения аномалий, адаптируясь к изменяющимся условиям и минимизируя участие человека [1].

IoT-устройства - от промышленных датчиков до умных домашних систем - генерируют эксабайты данных, среди которых скрыты критически важные отклонения:

- Кибератаки: несанкционированный доступ, DDoS-атаки, эксфильтрация данных.
- Аппаратные сбои: перегрев двигателей, вибрации подшипников, аномальное энергопотребление.
- Контекстуальные аномалии: действия, нормальные в одних условиях, но опасные в других (например, доступ к системе вне рабочего времени).

По данным исследований, до 40% инцидентов в IoT остаются незамеченными при использовании классических методов из-за динамичности сред и камуфляжа атак под легитимную активность.

Аномалии в IoT делятся на три ключевых типа:

- Точечные: единичные события, резко отклоняющиеся от нормы (например, скачок температуры обмотки электродвигателя на 200%).
- Контекстуальные: отклонения, выявляемые только в конкретном контексте (например, передача данных ночью для датчика, активного только днем).
- Коллективные: последовательности событий, безвредных по отдельности, но опасных вместе (сканирование портов с нескольких устройств).

Главные сложности их выявления:

- Динамичность сред: в микросервисных архитектурах и CI/CD-цепочках «нормальное» поведение меняется ежечасно [2].

- Ресурсные ограничения: IoT-устройства имеют малую вычислительную мощность, что затрудняет локальную аналитику.

- Неочевидность угроз: APT-атаки маскируются под легитимные операции, обходя статические правила.

Эффективные системы объединяют краевые вычисления (Edge) и облачную аналитику, создавая гибридную архитектуру:

Уровень устройств: легковесные агенты собирают данные с датчиков (температура, вибрация, трафик) и применяют упрощенные модели (например, изолирующие леса) для первичной фильтрации.

Шлюзы (Edge): выполняют предобработку данных и локальное обучение моделей, сокращая объем передаваемой информации. Например, автоэнкодеры сжимают данные, выделяя признаки для облака.

Облачный уровень: централизованные ML-модели (LSTM, трансформеры) анализируют агрегированные данные, выявляя межустройные аномалии и переобучаясь на новых данных.

Пример: AWS IoT Device Defender использует 14-дневное окно для автоматического ретренинга моделей, ежедневно обновляя эталоны «нормы» для метрик вроде числа сбоев авторизации.

Самообучающиеся модели для IoT делятся на три категории:

Без учителя (Unsupervised):

Изолирующий лес (Isolation Forest): изолирует аномалии через случайные разбиения данных. Эффективен для точечных аномалий при малых вычислительных затратах [3].

Автоэнкодеры: восстанавливают входные данные, фиксируя аномалии по высокой ошибке реконструкции. Применяются в промышленности для мониторинга вибрации двигателей.

С учителем (Supervised):

Случайный лес (Random Forest): в тестах для умных домов показал точность до 98% в детектировании сетевых аномалий (сканирование портов, эксфильтрация).

Гибридные подходы: комбинация статистических методов и ML. Например, Z-оценка для первичного отбора + LSTM для анализа временных рядов.

Таблица 1 – Сравнение алгоритмов для IoT

Алгоритм	**Точность**	**Ресурсы**	**Лучший кейс**
Изолирующий лес	85–92%	Низкие	Предиктивный мониторинг двигателей
Случайный лес	92–98%	Средние	Обнаружение кибератак в умных домах
LSTM	89–95%	Высокие	Анализ сетевого трафика

Практические кейсы внедрения

Промышленность: Мониторинг электродвигателей

На заводе «Силовые машины» внедрена IoT-система на базе AWS IoT. Датчики вибрации и температуры передают данные на шлюзы, где автоэнкодеры сжимают информацию. В облаке LSTM-сети анализируют временные ряды, прогнозируя остаточный ресурс подшипников.

Результаты:

- Сокращение простоев на 40%;
- Увеличение срока службы оборудования на 25%.

Умные дома: Защита от сетевых атак

В экосистеме умного дома Яндекс.Умный дом для детектирования аномалий используется случайный лес. Модель обучается на метриках:

- Число TCP-соединений;
- Частота ошибок авторизации;
- Объем исходящего трафика.

При выявлении сканирования портов система блокирует устройство через интеграцию с AWS IoT Device Defender.

Интеграция обучения в реальном времени

Непрерывное обновление моделей - ключевое преимущество самообучающихся систем. Техники для его обеспечения: [4]

- Инкрементное обучение: модели (например, Hoeffding Trees) обновляются на лету, используя потоки данных без переобучения с нуля.
- Transfer Learning: предобученные модели для типовых устройств адаптируются под специфику новых датчиков.
- Контроль дрейфа данных: перекалибровка при отклонении распределений (например, методом ADWIN).

Важно: Для снижения ложных срабатываний после обновлений ПО IoT-устройств AWS IoT Device Defender вводит «режимы заморозки алертов» на 1–2 часа.

Перспективные направления

1. Swarm Intelligence: роевые алгоритмы для распределенного обучения без центрального сервера.

2. Нейроморфные вычисления: энергоэффективные чипы для обработки данных на устройстве.

3. Генеративные модели: синтез аномальных данных для обучения в условиях дисбаланса классов.

4. Объяснимый ИИ (XAI): интерпретация срабатываний для SOC-команд.

Самообучающиеся модели превращают IoT-безопасность и мониторинг из реактивных в проактивные системы. Гибридные архитектуры, сочетающие Edge-обработку и облачную аналитику, позволяют выявлять сложные аномалии в реальном времени, адаптируясь к изменениям среды. Внедрение таких решений уже дает измеримые результаты: снижение простоев оборудования на 30–50%, сокращение ложных срабатываний на 60%, предотвращение до 95% кибератак. Ключ к успеху - комбинация алгоритмов под задачи: изолирующие леса для ресурсоограниченных устройств, LSTM для анализа временных паттернов, случайный лес для классификации угроз. С развитием периферийных вычислений и swarm intelligence, самообучающиеся системы станут стандартом для IoT следующего десятилетия [5-7].

Список литературы:

1. RU2767713C1 - Способ создания и обновления профиля сети, содержащей IoT-устройства. Патент. 2025. URL: https://patents.google.com/patent/RU2767713C1/ru

2. Как выявлять аномалии в ИБ: стратегии, которые работают в реальных условиях. Securitymedia.org. 2025. URL: https://securitymedia.org/info/kak-vyyavlyat-anomalii-v-ib-strategii-kotorye-rabotayut-v-realnykh-usloviyakh.html

3. Золотов И. Оптимизация бизнес-процессов с помощью гибридных моделей искусственного интеллекта. Universum: технические науки. 2025. URL: https://7universum.com/ru/tech/archive/item/19955

4. Функции AWS IoT Device Defender. Amazon Web Services. 2025. URL: https://aws.amazon.com/ru/iot-device-defender/features/

5. Алгоритмы машинного обучения в контроле качества продукции. Synaptik. 2024. URL: https://synaptik.ru/blog/kak-it-resheniya-uluchshayut-kachestvo-produkczii/algoritmy-mashinnogo-obucheniya-v-kontrole-kachestva-produkczii-novyj-vzglyad-na-starye-problemy/

6. Обнаружение аномалий IoT сети в технологии умного дома. Научная статья. 2024. URL: https://bibl.vgltu.ru/ru/nauka/conference_article/12292

7. Удаленный мониторинг электродвигателей через IoT. Inner. 2025. URL: https://inner.su/articles/udalennyy-monitoring-i-upravlenie-elektrodvigatelyami-cherez-iot/