Перспективы использования ИИ и цифровых двойников для обеспечения безопасности IIoT

Пермяков Руслан Анатольевич

Советник директора по инновациям

III Форум «Технологии доверенного искусственного интеллекта», Москва 20 мая 2025

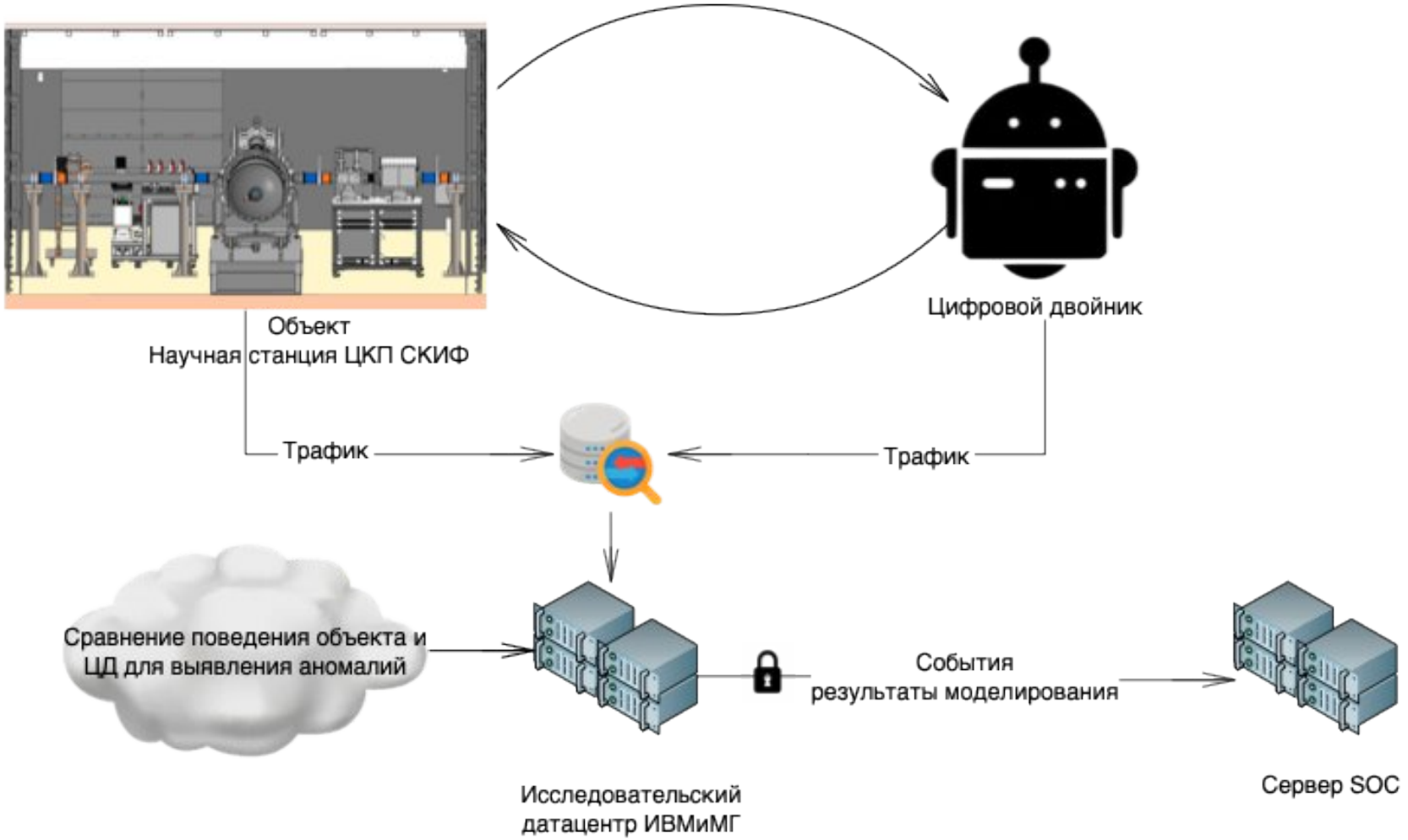
Цифровые двойники — определение

- **Цифровой двойник** — это виртуальная модель реального объекта, процесса или системы, которая точно воспроизводит его характеристики, поведение и взаимодействия в реальном времени или близком к реальному масштабе времени.
- **Основные характеристики цифрового двойника:**
  - Моделирование физического состояния и процессов.
  - Постоянная синхронизация с данными из реального объекта (IIoT-датчики, телеметрия).
  - Возможность проведения прогнозного анализа и виртуального тестирования.
- **В контексте безопасности IIoT:**
  - Цифровые двойники позволяют моделировать атаки, тестировать сценарии защиты, оценивать риски без воздействия на реальные системы.

Актуальность проблемы

- **Рост уязвимости IIoT-систем:**
  Расширение применения IIoT приводит к увеличению поверхностей атак на критически важные объекты промышленности, транспорта и энергетики.
- **Ограниченные возможности традиционных методов защиты:**
  Стандартные подходы к обеспечению кибербезопасности не учитывают сложную динамику и распределенность промышленных сетей.
- **Необходимость безопасного тестирования угроз и сценариев инцидентов:**
  Проведение испытаний на реальных объектах связано с высокими рисками сбоев и повреждений.
- **Появление новых требований к проактивной защите:**
  Современные системы безопасности должны обеспечивать не только реагирование на атаки, но и их предиктивное выявление и предотвращение.
- **Цифровые двойники как технологическая основа:**
  Использование цифровых двойников представляет собой перспективное направление для безопасного моделирования, анализа и оптимизации мер защиты IIoT-систем.

Вариант применения ЦД для решения вопросов безопасности



Актуальность применения ИИ в ЦД IIoT

- **Необходимость обработки больших потоков данных в реальном времени:**
  - IIoT-системы генерируют огромные объемы телеметрических данных, которые требуют интеллектуальной фильтрации, агрегации и анализа.
- **Сложность моделирования поведения объектов:**
  - Традиционные методы моделирования недостаточны для описания сложных нелинейных и стохастических процессов в промышленных системах.
- **Ранняя диагностика аномалий и предиктивное обслуживание:**
  - ИИ-алгоритмы повышают способность цифровых двойников обнаруживать скрытые отклонения в поведении объектов до наступления критических событий.
- **Адаптивность моделей к изменениям окружающей среды:**
  - Самообучающиеся ИИ-модели позволяют цифровым двойникам динамически подстраиваться под изменения в реальных объектах и внешних условиях.
- **Повышение эффективности тестирования сценариев безопасности:**
  - ИИ ускоряет процесс поиска оптимальных стратегий реагирования на потенциальные угрозы и оценки их последствий на виртуальных моделях.

Актуальность применения ИИ в ЦД IIoT

- **Умное моделирование процессов:**
  - ИИ дополняет цифровой двойник способностью прогнозировать сложные поведенческие сценарии на основе больших объемов данных.
- **Анализ и предсказание аномалий:**
  - ИИ-модели выявляют скрытые отклонения между данными реального объекта и виртуальной модели, обеспечивая раннее обнаружение угроз.
- **Автоматизация тестирования сценариев безопасности:**
  - Цифровой двойник с ИИ способен самостоятельно моделировать различные типы атак и оценивать их последствия без участия оператора.
- **Оптимизация стратегий реагирования:**
  - ИИ анализирует эффективность различных мер защиты и предлагает оптимальные стратегии предотвращения или минимизации ущерба.
- **Непрерывная адаптация моделей:**
  - Использование методов машинного обучения позволяет цифровому двойнику корректировать свои параметры в ответ на изменения в реальной среде.

Актуальность применения ИИ в ЦД IIoT

- **Умное моделирование процессов:**
  - ИИ дополняет цифровой двойник способностью прогнозировать сложные поведенческие сценарии на основе больших объемов данных.
- **Анализ и предсказание аномалий:**
  - ИИ-модели выявляют скрытые отклонения между данными реального объекта и виртуальной модели, обеспечивая раннее обнаружение угроз.
- **Автоматизация тестирования сценариев безопасности:**
  - Цифровой двойник с ИИ способен самостоятельно моделировать различные типы атак и оценивать их последствия без участия оператора.
- **Оптимизация стратегий реагирования:**
  - ИИ анализирует эффективность различных мер защиты и предлагает оптимальные стратегии предотвращения или минимизации ущерба.
- **Непрерывная адаптация моделей:**
  - Использование методов машинного обучения позволяет цифровому двойнику корректировать свои параметры в ответ на изменения в реальной среде.

Основные вызовы и ограничения

- **Неготовность объекта:**
  - Для целей ЦД объект должен предоставлять данные для анализа и эволюции экземпляра ЦД в достаточном количестве на всех этапах производственного и жизненного цикла.
- **Точность цифровых моделей:**
  - Недостаточная детализация или ошибки в построении цифрового двойника могут приводить к некорректной оценке угроз и искажению сценариев защиты.
- **Сложность синхронизации с реальными объектами:**
  - Поддержание актуальности данных требует высокой пропускной способности сетей, надежности сенсоров и точных механизмов обновления модели.
- **Риски атак на цифровые двойники:**
  - Цифровой двойник сам становится потенциальной целью кибератак, направленных на искажение прогнозов или маскировку реальных угроз.
- **Неопределенность поведения ИИ:**
  - Использование самообучающихся моделей может приводить к появлению непредсказуемых ошибок и отклонений в виртуальной симуляции.

Перспективные направления развития

- **Интеграция методов Privacy-Enhancing Technologies (PETs):**
  - Защита персональных и промышленных данных, обрабатываемых в цифровых двойниках, при обучении и функционировании ИИ.
- **Разработка механизмов постоянной верификации моделей:**
  - Внедрение процедур регулярной проверки корректности работы ИИ в цифровых двойниках на реальных и синтетических сценариях угроз.
- **Автоматизированное выявление новых классов угроз:**
  - Использование ИИ для обнаружения ранее неизвестных уязвимостей и аномалий в поведении сложных промышленных систем.
- **Сертификация цифровых двойников с ИИ:**
  - Разработка стандартов проверки надёжности, безопасности и доверенности цифровых двойников, интегрирующих технологии ИИ.
- **Адаптивная коэволюция цифрового двойника и реального объекта:**
  - Обеспечение динамического обновления виртуальной модели с учётом изменений в реальной системе и в ландшафте угроз.
- **Развитие объяснимого ИИ (Explainable AI):**
  - Создание моделей, способных предоставлять обоснования своих прогнозов и решений для повышения доверия к симулируемым сценариям безопасности.

Спасибо за внимание.

Пермяков Руслан Анатольевич
+7(913)916-21-56, , @pransk

Институт вычислительной математики и математической геофизики Сибирского отделения РАН

Почему PETs, а не только Zero Trust

ИВМиМГ

- Разные уровни защиты:
  Zero Trust контролирует доступ к данным, PETs защищают сами данные в процессе обработки.
- Угроза утечек через модели ИИ:
  PETs предотвращают извлечение персональных данных из обученных моделей, чего Zero Trust не обеспечивает.
- Противодействие внутренним угрозам:
  PETs защищают данные даже от легитимных участников системы (администраторов, моделей), в отличие от архитектур доступа.
- Необходимость глубинной защиты данных:
  В системах с ИИ защита периметра недостаточна — требуется контроль за приватностью на уровне обучения и вывода моделей.
- Комплементарность подходов:
  Zero Trust важен для защиты инфраструктуры, но без PETs невозможно обеспечить реальную защиту приватности данных в ГИС с ИИ.