УДК 004.056.53

МЕТОДИКА ПОСТРОЕНИЯ ПРОЗРАЧНЫХ И АДАПТИВНЫХ СИСТЕМ ОБНАРУЖЕНИЯ ВТОРЖЕНИЙ НА ОСНОВЕ УПРУГИХ КАРТ НЕЙРОННЫХ СЕТЕЙ

Зуев Д. П.
Научный руководитель – кандидат технических наук, доцент Штеренберг С. И.
СПБГУТ
zuev.dp@sut.ru

Введение

В условиях цифровизации критически важных секторов и госуправления системы обнаружения вторжений (IDS) являются ключевым элементом защиты информационных систем. Современные киберугрозы характеризуются высокой динамичностью, многоэтапностью и способностью к обходу сигнатурных методов обнаружения, что требует применения адаптивных решений на основе машинного обучения. При этом внедрение сложных моделей ИИ сталкивается с проблемой отсутствия интерпретируемости решений. Операторы центров мониторинга безопасности получают классификации без понимания причин принятого решения, что снижает доверие к системе, замедляет расследование инцидентов и увеличивает когнитивную нагрузку. Систематический анализ литературы выявил, что 88,36% исследований в области ИИ для систем безопасности не содержат элементов интерпретируемости. Проблема особенно актуальна для РФ, где защита критической информационной инфраструктуры требует полностью контролируемых, аудируемых и суверенных технологий. Анализ публикаций за последние 5 лет свидетельствует об отсутствии систематических исследований в РФ, сочетающих адаптивные нейросетевые архитектуры с механизмами интерпретируемости для задач обнаружения вторжений, что определяет научную новизну и практическую востребованность работы.

Основная часть

Классические самоорганизующиеся карты (SOM), несмотря на историческую популярность в задачах кластеризации сетевого трафика без учителя, обладают рядом фундаментальных ограничений [1]. Их жесткая предопределенная топология приводит к искажению истинной структуры многомерных данных, особенно при наличии сложных многообразий с переменной плотностью, что негативно сказывается на качестве обнаружения аномалий. Также визуальное расположение кластеров на карте не дает прямого объяснения причин классификации конкретного сетевого пакета, требуя дополнительного анализа весовых векторов [2]. Высокая чувствительность к гиперпараметрам и необходимость полного переобучения при изменении статистики трафика делают SOM непригодными для эксплуатации в динамичных сетевых средах. Существующие подходы к обеспечению интерпретируемости, такие как LIME и SHAP, применяемые к «черным ящикам», обладают высокой вычислительной сложностью и генерируют локальные объяснения, не обеспечивающие целостного понимания работы системы в реальном времени.

В рамках исследования разработана методика IDS на основе упругих карт, заменяющих жёсткую топологию SOM адаптивной структурой с синаптическими и латеральными связями, минимизирующими энергетическую функцию [3]. Карта динамически деформируется, отражая геометрию данных, включая сложные многообразия. Механизмы «растущего» и «разрывного» типа добавляют или удаляют узлы для непрерывной адаптации без переобучения. Оригинальность состоит в интерпретируемости на основе геометрии карты. Классификация поясняется

топологическим окружением узла, ошибкой квантования и визуальной подсветкой. Подход обеспечивает глобальную интерпретируемость, понятную оператору.

Теоретическая значимость исследования заключается в расширении теории самоорганизующихся структур за счет демонстрации преимуществ упругих карт в задаче анализа сетевого трафика, а также в создании нового класса методов интерпретируемого ИИ, основанных на топологической интерпретации геометрических представлений данных. Предложенная методика вносит вклад в развитие междисциплинарного подхода, для решения комплексных задач защиты информационных систем.

Практическая ценность разработки проявляется в нескольких ключевых сценариях применения на территории РФ. Для защиты критической информационной инфраструктуры методика обеспечивает полностью контролируемую и аудируемую архитектуру IDS, исключающую риски. В центрах мониторинга безопасности применение методики снижает количество ложных срабатываний за счет точного моделирования легитимного трафика и предоставляет операторам интуитивно понятные интерпретации в виде визуализированной топологии карты, что сокращает время расследования инцидентов. Техническая реализация методики возможна на базе инструментов анализа данных (например, платформы ViDaExpert для визуализации упругих карт) и может быть интегрирована как модуль в современные платформы SIEM/SOC [4]. Перспективы дальнейшего развития включают интеграцию с платформами киберразведки для сопоставления обнаруженных атак с отечественными базами угроз, адаптацию методики для edge-вычислений в распределенных сетях IoT.

Выводы

Проведён анализ ограничений классических самоорганизующихся карт в задачах обнаружения вторжений и разработана методика построения адаптивной архитектуры на основе упругих карт с обеспечением интерпретируемости решений через анализ топологической структуры. Предложенная методика устраняет фундаментальные недостатки топологии SOM, обеспечивает непрерывную адаптацию к изменяющемуся сетевому трафику без полного переобучения и генерирует глобально интерпретируемые объяснения решений. Результаты исследования создают основу для развития отечественных решений и формируют направление для дальнейших исследований.

Литература

1. Эйблс Дж., Кирби Т., Андерсон У., Миттал С., Рахими Ш., Баницеску И., Сил М. Создание объяснимой системы обнаружения вторжений с использованием самоорганизующихся карт [Электронный ресурс] // URL: https://arxiv.org/abs/2207.07465 (дата обращения: 01.02.2026).
2. Ландауэр М., Скопик Ф., Вурценбергер М., Хотвагнер В., Раубер А. Визуализация системных вызовов с использованием самоорганизующихся карт для обнаружения вторжений в систему // Труды 6-й Международной конференции по безопасности и приватности информационных систем (ICISSP 2020). – 2020. – С. 349–360.
3. Горбан А., Зиновьев А. Упругие главные графы и многообразия и их практическое применение // Computing. – 2005. – Т. 75. – С. 359–379.
4. Горбан А. Н., Питенко А., Зиновьев А. ViDaExpert: удобный инструмент для нелинейной визуализации и анализа многомерных векторных данных [Электронный ресурс] // URL: https://arxiv.org/abs/1406.5550 (дата обращения: 01.02.2026).