ISSN 2658-5014 (Print) СИИТ ISSN 2686-7044 (Online)
http://siit.ugatu.su 2026. Т. 8, № 4(28)
СИСТЕМНАЯ ИНЖЕНЕРИЯ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ
Научная статья

УДК 004.056 DOI: 10.54708/SIIT-2026-no4-p128 EDN: AVBPAI

Методика обнаружения, классификации и координированного реагирования на многовекторные атаки в децентрализованной киберфизической среде с замкнутой обратной связью по параметрам доверия

В. И. Петренко

Северо-Кавказский федеральный университет

Рассматривается задача защиты роботизированных агентов, функционирующих в децентрализованной киберфизической среде без доверенного центра, от многовекторных атак – координированных воздействий, сочетающих сетевые, поведенческие и физические каналы в рамках единого сценария. Предложена методика обнаружения и классификации, построенная как согласованная по входам и выходам композиция шести алгоритмических компонентов: группового зонирования с учётом доверия, гибридного обнаружения на основе генеративно-состязательной сети и обучения с подкреплением, федеративного обучения, многомерной реконструкции трафика, распределённой верификации на основе PoV-консенсуса и нейросетевой классификации вредоносных агентов. Отличительной особенностью является замкнутая двунаправленная обратная связь по параметрам доверия: подтверждённое обнаружение каскадно обновляет функцию доверия и инициирует перераспределение задач, а обновлённые уровни доверия перенастраивают веса федеративной агрегации подсистемы обнаружения. Имитационное моделирование в среде ns-3 (2 400 экспериментов, до 1 000 узлов) показало полноту обнаружения 0,97 при доле ложных срабатываний 2,6 %, сокращение времени обнаружения на 49,7 %, времени реагирования на 56,2 % и долю нейтрализованных атак 93,4 % при времени реакции не более 5 с; устойчивость эмпирически подтверждена при доле скомпрометированных агентов до 0,30, что согласуется с теоретической границей 1/3.

Роботизированные агенты; децентрализованные системы; глубокое обучение с подкреплением; доверенная оркестрация; механизм доверия; координация; интеллектуальные системы.

Введение

Развитие групповой робототехники, промышленного интернета вещей и распределённых беспилотных комплексов сформировало класс сред, в которых автономные агенты взаимодействуют без централизованного органа управления и единой точки доверия. Такие децентрализованные киберфизические среды (ДКФСр) характеризуются динамически меняющейся топологией, ресурсной ограниченностью и гетерогенностью участников, частичной наблюдаемостью состояния и, что принципиально, допустимостью присутствия скомпрометированных сущностей [Abs25, Пет23а, Zuo23]. Именно свойства среды, а не отдельных систем, определяют в этих условиях ландшафт угроз и возможности координированной защиты.

Актуальные сценарии воздействия на ДКФСр всё чаще реализуются как композиция элементарных атак, охватывающая несколько независимых каналов [Пет24]. Многовекторная атака (МВА) выделяется в самостоятельный объект исследования совокупностью признаков, не охватываемых классическими моделями одиночных воздействий: координированностью

Петренко В. И. Методика обнаружения, классификации и координированного реагирования на многовекторные атаки в децентрализованной киберфизической среде с замкнутой обратной связью по параметрам доверия // СИИТ. 2026. Т. 8, № 4(28). С. 129-144. DOI: 10.54708/SIIT-2026-no4-p128. EDN: AVBPAI.

Petrenko V. I. “Multi-vector attack detection, classification and coordinated response in a decentralized cyber-physical environment with closed-loop trust feedback” // SIIT. 2026. Vol. 8, no. 4(28), pp. 129-144. DOI: 10.54708/SIIT-2026-no4-p128. EDN: AVBPAI. (In Russian).

нескольких каналов в общем окне времени, нацеленностью на вычислительно ограниченных мобильных агентов и способностью сочетать сетевые, поведенческие и физические уязвимости в рамках единого сценария [Abs25, Пет24]. Обзоры методов обнаружения аномалий в киберфизических системах и федеративного обнаружения вторжений отмечают, что комбинирование векторов воздействия не покрывается ни одним из существующих классов решений в отдельности [Abs25, Zoh24, Anw25].

Дополнительный разрыв связан с архитектурой контура защиты. В преобладающих схемах «детекция → реагирование» результат обнаружения используется для формирования ответного действия, но не для адаптивной перенастройки самой подсистемы обнаружения [Ana25a, Bac21]. В условиях, когда атакующая сторона способна варьировать стратегию в течение целевой операции, разомкнутость контура ведёт к деградации качества детекции при смене распределения признаков. Отдельные подходы к управлению доверием на основе глубокого обучения [Ana25b] и распределённых реестров [Zuo23, Ana25a] сближают оценку доверия и обнаружение, однако не формируют строго определённого двунаправленного интерфейса между ними.

Вклад настоящей работы состоит в следующем. Во-первых, формализована модель многовекторной атаки и предложена методика её обнаружения и классификации, построенная как согласованная по входам и выходам композиция шести алгоритмических компонентов с единой шкалой показателей аномальности. Во-вторых, разработана система координированного реагирования с двухканальной схемой управляющих воздействий – через контур доверия и через контур оркестрации задач. В-третьих, формализована замкнутая двунаправленная обратная связь по параметрам доверия, для которой доказана каузальная корректность и обоснован перенос оценки устойчивости к скомпрометированным агентам (доля менее 1/3) на совокупный контур. В-четвёртых, характеристики методики и системы подтверждены имитационным моделированием в среде ns-3 объёмом 2 400 экспериментов. Отдельные алгоритмические компоненты опубликованы ранее и предметом новизны не являются; новыми результатами выступают их согласованная композиция, формализованный замкнутый контур доверия и экспериментальная верификация совокупного решения. Вклад работы носит интеграционно-архитектурный, а не алгоритмический характер: новизна сосредоточена, во-первых, в формализованном двунаправленном контуре доверия, не описанном ни в одной из предшествующих работ автора, и, во-вторых, в экспериментальном подтверждении того, что целевые показатели достигаются только композицией компонентов, но не каждым из них в отдельности. Настоящая статья и параллельно подготовленная работа автора о контекстно-зависимой модели управления доверием опираются на общий фундамент – модель ДКФСр и функцию доверия с асимметричной динамикой, введённые в [Пет23а], – однако решают разные задачи и не дублируют друг друга: в параллельной работе доверие используется как контекстно-зависимый мультипликативный коэффициент при оценке риска отдельных взаимодействий агентов, тогда как здесь оно выступает системообразующим сигналом замкнутого контура обнаружения многовекторных атак и координированного реагирования.

Обзор родственных работ и постановка задачи

К процедуре обнаружения многовекторных атак в ДКФСр одновременно предъявляются четыре требования: локальность оценок (телеметрия не покидает зону наблюдения), сохранение приватности данных агентов, распределённая верификация обнаружений (решение об атаке не принимается единственным узлом) и согласованность с функцией доверия, управляющей допуском агентов к операциям. Совместное выполнение этих требований очерчивает границы применимости трёх сложившихся классов решений.

Централизованные системы обнаружения вторжений [Zoh24] достигают высокого качества детекции ценой передачи сырой телеметрии в единый центр, что нарушает и локальность, и приватность, а сам центр образует единую точку отказа, несовместимую с природой ДКФСр. Системы на основе федеративного обучения (FL-IDS) [Anw25, Ami25] сохраняют приватность

за счёт обмена параметрами моделей вместо данных, однако решение об инциденте по-прежнему принимается без распределённой верификации, что делает их уязвимыми к отравлению локальных моделей. Гибридные решения на стыке машинного обучения и распределённого реестра [Пет23б] обеспечивают верифицируемость обнаружений, но функционируют изолированно от механизма доверия: факт обнаружения не изменяет права и роли агентов. Гибридные интеллектуальные системы обнаружения [Bac21] и методы поведенческой классификации трафика в контуре доверенного взаимодействия [Арх22] решают отдельные аспекты задачи, не образуя замкнутого контура. Сводное сопоставление классов приведено в табл. 1.

Таблица 1

Сопоставление классов существующих решений с предлагаемой методикой

Свойство	Централизованные IDS	FL-IDS	ML + распределённый реестр	Предлагаемая методика
Локальность оценок	нет	частично	нет	да
Сохранение приватности	нет	да	частично	да
Распределённая верификация	нет	нет	да	да
Согласованность с функцией доверия	нет	нет	нет	да
Применимость к комбинированным атакам	ограниченная	ограниченная	ограниченная	подтверждённая

Из табл. 1 видно, что ни один существующий класс решений не обеспечивает одновременного выполнения всех четырёх требований; предлагаемая методика закрывает их в совокупности.

От систем классов SIEM и SOAR предлагаемый подход отличается характером обратной связи. SIEM-системы фиксируют события безопасности ретроспективно и не изменяют поведение агентов в реальном времени; SOAR-системы автоматизируют реагирование, но по заранее заданным сценариям, не адаптирующимся к новым типам атак. В обоих случаях контур «обнаружение – доверие – реагирование» остаётся разомкнутым.

Средой функционирования выступает ДКФСр, формализуемая кортежем $\mathrm{DCPE} = \langle E, \mathrm{Comm}, \mathrm{Top}(t)\rangle$, где E – множество сущностей, Comm – множество взаимодействий, $\mathrm{Top}(t)$ – функция динамической топологии. Каждому агенту j сопоставлена функция доверия $\mathrm{Trust}(j,t) \in [0,1]$, вычисляемая как взвешенная свёртка компонентов надёжности, компетентности, честности и предсказуемости с асимметричной динамикой обновления (быстрое снижение, медленное восстановление) [Пет23а]. Многовекторная атака формализуется кортежем

$$MVA = \langle S,\ V,\ A_{\mathrm{act}},\ T_w,\ \rho\rangle\,, \tag{1}$$

где $S \subseteq A$ – подмножество агентов-целей; V – подмножество эксплуатируемых уязвимостей, $|V| \geq 2$; A_{act} – множество синхронизированных действий атакующих; T_w – окно координации, в пределах которого действия квалифицируются как составляющие одной атаки; ρ – доля скомпрометированных агентов.

Модель нарушителя соответствует византийской модели: скомпрометированный агент может произвольно отклоняться от протокола, посылать противоречивые сообщения и согласовывать действия с другими нарушителями [Lam82]. Рассматриваются три стратегии поведения нарушителей – случайная, оппозиционная и координированная. Допущением служит ограничение доли скомпрометированных агентов $\rho < 1/3$, при котором, согласно доказанной ранее теореме об устойчивости децентрализованной модели доверия [Lam82], сохраняется согласованность оценок доверия; полное доказательство приведено в [Пет23а] и здесь не воспроизводится.

Задача состоит в построении методики M обнаружения и классификации МВА и системы координированного реагирования, совместно удовлетворяющих четырём сформулированным

требованиям и замкнутых на функцию доверия $Trust(j,t)$ двунаправленной обратной связью. Построение методики и системы реагирования рассматривается в двух следующих разделах.

Методика обнаружения и классификации многовекторных атак

Методика строится как композиция шести алгоритмических компонентов, каждый из которых разработан и верифицирован в самостоятельных авторских работах [Теб26а, Пат24, Теб26б, Пет25а–Пет25в] и решает один из аспектов общей задачи. Все компоненты формируют показатели аномальности в единой шкале $[0,1]$, что обеспечивает их сцепляемость. Формально методика записывается композицией операторов

$$M = T_c \circ T_v \circ T_f \circ T_d \circ T_z, \tag{2}$$

где T_z – оператор группового зонирования; T_d – оператор локальной детекции (объединяет гибридное обнаружение и многомерную реконструкцию трафика); T_f – оператор федеративной агрегации; T_v – оператор распределённой верификации; T_c – оператор классификации. Запись читается справа налево: выход каждого оператора служит входом следующего, а итогом является метка класса атаки. Структурная схема композиции приведена на рис. 1.

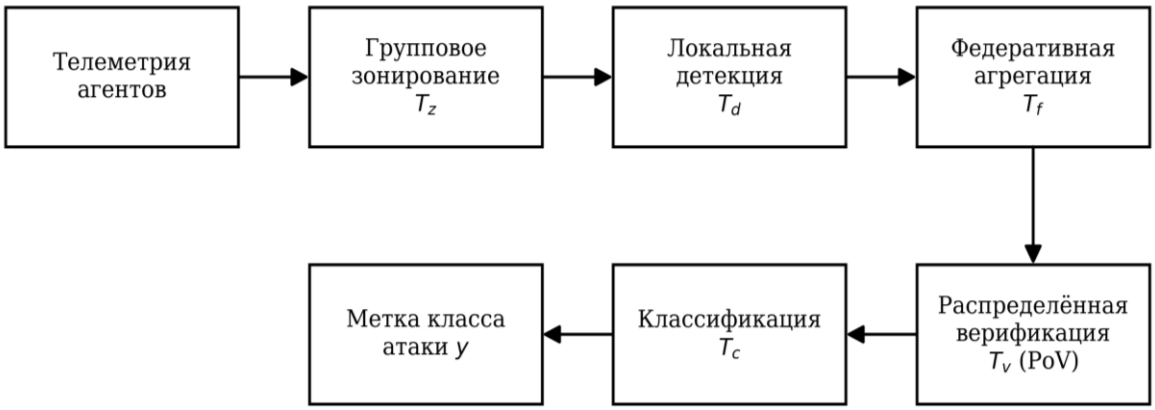


Рис. 1. Структурная схема методики: композиция операторов зонирования, локальной детекции, федеративной агрегации, распределённой верификации и классификации

Схема на рис. 1 отражает конвейерный характер обработки: телеметрия агентов последовательно проходит операторы группового зонирования T_z, локальной детекции T_d, федеративной агрегации T_f и распределённой верификации T_v, а завершается тракт оператором классификации T_c, выходом которого служит метка класса атаки y. Промежуточные показатели аномальности, передаваемые между операторами T_d, T_f и T_v, приведены к единой шкале $[0,1]$, что и обеспечивает их сцепляемость; порядок блоков на схеме соответствует композиции (2), читаемой справа налево. Компоненты рассматриваются в порядке следования блоков схемы.

Блок «Групповое зонирование» реализует оператор T_z, локализующий телеметрию: множество агентов A разбивается на непересекающиеся зоны $Z_1, \ldots, Z_L$, однородные по поведенческим характеристикам. За основу принят алгоритм группового зонирования MCMA-GZ [Теб26а], защищённый патентом [Пат24] и адаптированный здесь включением уровня доверия в признаковое описание агента:

$$x_j = \langle p_j, \mathrm{type}_j, \mathrm{Trust}_j, \beta_j \rangle, \tag{3}$$

где p_j – пространственные координаты; type_j – тип выполняемой задачи; Trust_j – текущий уровень доверия; β_j – доступная полоса пропускания. Попарное расстояние между агентами вычисляется взвешенной свёрткой трёх метрик:

$$D_{i,j} = \gamma_1 \|p_i - p_j\| + \gamma_2 [\tau_i \neq \tau_j] + \gamma_3 |\mathrm{Trust}_i - \mathrm{Trust}_j|, \quad (4)$$

где признаки p_j, τ_j и Trust_j определены в (3); $\gamma_1 + \gamma_2 + \gamma_3 = 1$ (рекомендованные значения 0,40; 0,25; 0,35), а квадратные скобки обозначают индикатор Айверсона. Слагаемое с разностью уровней доверия штрафует объединение в одной зоне агентов с различающимся поведенческим профилем.

Отнесение агента к зоне сочетает близость к центроиду с проверкой минимального порога доверия:

$$j \in Z_k \Leftrightarrow D(j, \mu_{Z_k}) < \theta_k \wedge \mathrm{Trust}_j \geq \mathrm{Trust}_{\min}, \quad (5)$$

где μ_{Z_k} – центроид зоны, θ_k – её радиус, $\mathrm{Trust}_{\min} = 0{,}40$, а расстояние $D(j, \mu_{Z_k})$ вычисляется по правилу (4). Условие (5) исключает попадание скомпрометированных агентов в общую обучающую выборку. Значение $\mathrm{Trust}_{\min} = 0{,}40$ установлено по результатам калибровки на валидационном наборе данных, содержащем размеченные образцы поведения добросовестных и скомпрометированных агентов. При указанном пороге полнота покрытия зон добросовестными агентами составляет не менее 0,98, тогда как доля скомпрометированных агентов, ошибочно включаемых в обучающую выборку, не превышает 0,02. Подробные результаты калибровки приведены в работе [Теб26а]. Оптимальное число зон, минимизирующее суммарные внутризональные и межзональные издержки, составляет $L^* = O(|A|^{2/3})$.

Блок «Локальная детекция», соответствующий оператору T_d, объединяет две ветви обработки. Первую из них образует гибридное обнаружение на основе генеративно-состязательной сети и обучения с подкреплением. Поведенческие классы атак представлены в реальной телеметрии редко и неравномерно: в наборах CIC IoT 2023 и N-BaIoT на каждый класс приходится в среднем менее 5% записей [Теб26б, Пет25а, Пет25б]. Дефицит аномальных данных компенсируется генеративно-состязательной сетью, расширяющей обучающую выборку, а адаптация к смене стратегии атакующего – агентом обучения с подкреплением, корректирующим классификацию поведенческих классов по мере поступления новых наблюдений [Теб26б]. Вторую ветвь того же блока образует многомерная реконструкция трафика. Локальным детектором каждой зоны служит гибридная модель, объединяющая автоэнкодер для безнадзорной детекции отклонений и каскад «разделяемая свёртка → LSTM → многоканальное внимание» для надзорной классификации [Пет25б]. Обе ветви сводятся в единый показатель аномальности зоны $S_{\mathrm{loc}}(j,t) \in [0,1]$, который и служит выходом оператора T_d. Итеративная обрезка нейронов сокращает число параметров модели с $1{,}5 \cdot 10^6$до $3 \cdot 10^5$и вычислительные затраты на 80% при сохранении точности классификации 99,1% на наборе CIC IoT 2023; время обработки одного образца составляет 12 мс на встраиваемой платформе с тактовой частотой 1,8 ГГц [Пет25б].

Блок «Федеративная агрегация» реализует оператор T_f, обучающий глобальную модель обнаружения алгоритмом федеративного обучения с импульсной агрегацией градиентов [Пет25а]: зоны обмениваются параметрами описанных выше локальных детекторов вместо сырой телеметрии, а вклад каждой зоны в агрегат взвешивается по среднему уровню доверия её участников. Такая схема обеспечивает сходимость глобальной модели за 30–40 раундов на выборке объёмом 10^6образцов при AUC = 0,99 и сокращает объём передаваемых данных на 50,2 % относительно централизованной схемы [Пет25а].

Блок «Распределённая верификация» реализует оператор T_v на основе консенсуса «доказательство голосования» (PoV). Однократное превышение порога аномальности в одной зоне не является достаточным основанием для квалификации события как МВА: при пороге, обеспечивающем приемлемую полноту, доля ложных срабатываний одиночного детектора неприемлемо высока для контура управления доверием. Зональный детектор при срабатывании формирует структурированное оповещение

$$m_k = \langle \mathrm{id}_{Z_k}, t, S_{\mathrm{loc}}, v, \mathrm{hash} \rangle, \quad (6)$$

содержащее идентификатор зоны, временную метку, значение показателя аномальности, нормированный вектор признаков v (не путать с множеством уязвимостей V из определения (1)) и хэш SHA-256 тела оповещения. Узлы-валидаторы голосуют за принятие блока по правилу PoV [Пет25в]:

$$\frac{1}{N}\sum_{i=1}^{N} V_i\,(k) > \theta_v, \tag{7}$$

где $V_i(k) \in \{0,1\}$ – голос валидатора iпо оповещению m_k; N – число валидаторов; $\theta_v = 0{,}5$ – порог голосования. Поверх правила (7) применяется условие, превращающее разрозненные локальные срабатывания в коллективный сигнал об атаке:

$$\mathrm{MVA}_{\mathrm{conf}} \Leftrightarrow \left|\{m_k : S_{\mathrm{loc}}(m_k) > \theta_S,\ t_k - t_0 \le \Delta T,\ V(k) = 1\}\right| \ge K_{\min}, \tag{8}$$

т. е. многовекторный характер атаки подтверждается только при согласованном срабатывании не менее $K_{\min}$ независимых зон в общем окне временной согласованности ΔT.

Одиночные срабатывания квалифицируются как локальные аномалии и не приводят к глобальной реакции – этим достигается основное снижение доли ложных срабатываний. За счёт обрезки модели валидации время верификации одного оповещения составляет 12 мс, объём памяти смарт-контракта – 180 МБ, что обеспечивает применимость на встраиваемых платформах [Пет25в].

Замыкает конвейер блок «Классификация», реализующий оператор T_c: подтверждённое по правилу (8) событие поступает на его вход, а результатом становится блок «Метка класса атаки y». Для комбинированных атак метка строится по правилу декартовой композиции базовых классов, что позволяет корректно помечать сценарии вида «Vampire + Sybil» или «DDoS + Stealth On-Off + сговор» без введения отдельного классификатора для каждой комбинации.

Вне последовательного тракта (2) работает шестой компонент методики – обнаружение вредоносных агентов при коллективном принятии решений. Отдельный класс воздействий не порождает аномалий трафика: скомпрометированный агент корректно выполняет служебные обмены, но систематически голосует против большинства или за заранее заданную альтернативу. Такие воздействия выявляются нейросетевым классификатором, анализирующим показатели уверенности агентов в выборе альтернатив [Пет22]. Этот компонент не анализирует трафик и потому не показан на рис. 1; его выход поступает непосредственно в контур управления доверием наравне с подтверждёнными обнаружениями, дополняя пять операторов схемы до шести алгоритмических компонентов методики.

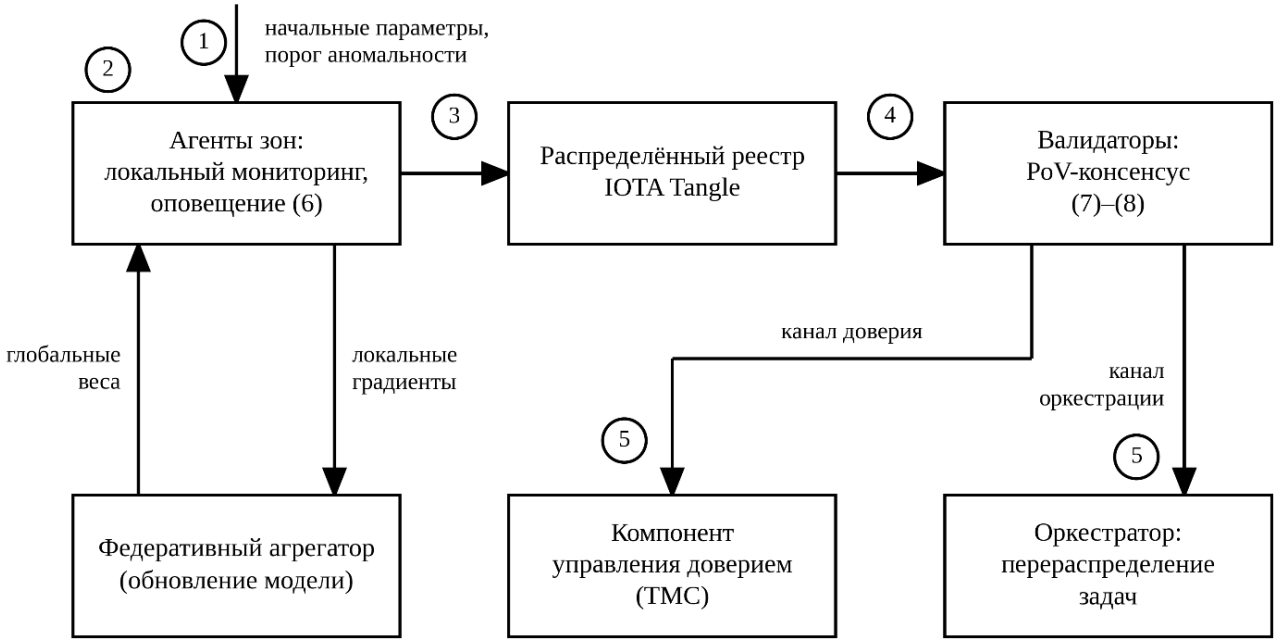
Выходом методики является показатель аномальности S_{loc} и метка класса y, которая передаётся в контур управления доверием.

Система координированного реагирования и замкнутый контур доверия

Система реагирования получает на вход подтверждённые по правилу (8) обнаружения и преобразует их в управляющие воздействия по двум независимым каналам: каналу доверия – через компонент управления доверием (ТМС), обновляющий функцию $Trust(j,t)$, и каналу оркестрации – через команду на пересмотр распределения подзадач для агентов с изменённым уровнем доверия. Архитектура системы показана на рис. 2.

На схеме рис. 2 выделены шесть функциональных блоков и два канала управляющих воздействий. Блок «Агенты зон: локальный мониторинг, оповещение (6)» формирует первичные обнаружения и передаёт оповещения в блок «Распределённый реестр IOTA Tangle», откуда они поступают в блок «Валидаторы: PoV-консенсус (7)-(8)». Подтверждённое обнаружение расходится от валидаторов по двум независимым направлениям: по каналу доверия – в блок «Компонент управления доверием (ТМС)», обновляющий функцию $\mathrm{Trust}(j,t)$, и по каналу оркестрации – в блок «Оркестратор: перераспределение задач». Обособленно от этого тракта работает блок «Федеративный агрегатор (обновление модели)», связанный с агентами зон

парой встречных потоков: локальные градиенты передаются от агентов к агрегатору, глобальные веса возвращаются обратно, чем и обеспечивается фоновое дообучение локальных детекторов.



Этапы работы системы: 1 — инициализация; 2 — локальный мониторинг; 3 — локальная реакция; 4 — распределённая верификация; 5 — координированное реагирование. Контур федеративного дообучения (слева) выполняется параллельно этапам 2–5.

Рис. 2. Архитектура системы координированного реагирования: локальный мониторинг, распределённый реестр, федеративный агрегатор и два канала управляющих воздействий

Работа системы организована в пять этапов, отмеченных на рис. 2 номерами. На этапе инициализации (*1*) на агенты загружаются начальные параметры локальной модели, порог аномальности устанавливается порог аномальности принимается равным 99-му процентилю эмпирического распределения ошибок реконструкции на калибровочной выборке нормального трафика: $\theta = Q_{0,99}(e)$, что соответствует ожидаемой доле ложных срабатываний локального детектора 1%. На этапе локального мониторинга (*2*) каждый агент непрерывно вычисляет ошибку реконструкции очередного наблюдения; превышение порога помечает событие как аномалию. На этапе локальной реакции (*3*) агент-детектор подписывает оповещение (*6*) приватным ключом и публикует транзакцию в распределённом реестре IOTA Tangle [Пет26а]. На этапе распределённой верификации (*4*) валидаторы асинхронно проверяют целостность и подпись сообщения и применяют правила (7) и (8); при выполнении условия (8) событие квалифицируется как многовекторная атака. На этапе координированного реагирования (*5*) подтверждённое обнаружение направляется по каналу доверия в компонент управления доверием, где пересчитываются уровни доверия участников атаки, и по каналу оркестрации – оркестратору, формирующему новое распределение подзадач; агенты зон при этом обновляют локальные правила фильтрации и изоляции. Федеративное дообучение глобальной модели, показанное на схеме левым контуром, выполняется параллельно этапам 2–5: агенты зон передают агрегатору локальные градиенты и получают от него глобальные веса.

Принципиальная особенность системы состоит в разнесении темпов двух контуров. Быстрый контур – публикация оповещения, его верификация и передача управляющих воздействий по каналам доверия и оркестрации – отрабатывает за единицы секунд, тогда как контур федеративного дообучения выполняется в фоновом режиме с интервалом порядка 15 мин. Локализация и ограничение атаки поэтому не ожидают завершения очередного раунда обучения:

замыкание контура обратной связью по доверию, отсутствующее в схемах «детекция → реагирование» [Ana25a, Бас21], достигается без увеличения времени реагирования. Количественная оценка этого эффекта приведена в разделе экспериментальной оценки.

Замыкание контура обеспечивается формализованной двунаправленной трансляцией сигналов между подсистемой обнаружения, системой реагирования и функцией доверия. Прямой канал начинается со структурного согласования двух семантически идентичных, но разнородных источников: зонного показателя S_{loc} и индивидуального показателя аномальности агента $S_{CAT\text{-}GNN}$, формируемого поведенческим детектором на основе графовой нейронной сети [Пет26б]. Эффективный показатель агента вычисляется линейной свёрткой

$$S_{\text{agent}}(a,t) = w_z S_{\text{loc}}(z(a),t) + w_g S_{\text{CAT-GNN}}(a,t), \tag{9}$$

где $z(a)$ – зона агента a; $w_z + w_g = 1$. По умолчанию принимается $w_z = w_g = 0{,}5$; при подтверждённом по (8) многовекторном характере атаки вес зонного источника увеличивается до $w_z = 0{,}7$, что отражает большую информативность коллективного сигнала. Свёртка выполняется на стороне ТМС без пересылки сырой телеметрии за пределы зоны, сохраняя требование локальности.

Для агентов из множества S_{conf}, входящих в состав подтверждённой атаки, применяется правило ужесточения, дифференцированное по классам атак:

$$S_{\text{eff}}(a,t) = \max\{S_{\text{agent}}(a,t),\ \kappa_y \cdot S^{\max}_{\text{agent}}\}, \quad a \in S_{\text{conf}}, \tag{10}$$

где $S^{\max}_{\text{agent}}$ – максимум эффективного показателя в множестве S_{conf}, а коэффициент ужесточения κ_y зависит от метки класса y: для классов «подрыв доверия», «нарушение кооперации» и классов нового типа $\kappa_y = 1{,}0$; для маршрутизирующих атак (sinkhole, wormhole, blackhole, Sybil) $\kappa_y = 0{,}9$; для утечки данных и манипуляции поведением $\kappa_y = 0{,}85$; для атак исчерпания ресурсов класса Vampire $\kappa_y = 0{,}75$. Дифференциация коэффициентов основана на оценке потенциального ущерба для целостности контура управления: атаки классов «подрыв доверия» и «нарушение кооперации» непосредственно дестабилизируют механизм доверия, поэтому требуют максимального штрафа; маршрутизирующие атаки и атаки исчерпания ресурсов воздействуют на периферийные каналы и допускают менее жёсткую реакцию без потери устойчивости системы (обоснование градаций приведено в модели угроз [Пет24]). Метка класса y в формуле (10) формируется блоком классификации T_c (оператор (2)) на основе данных, верифицированных распределённым консенсусом PoV (правила (7) и (8)), и не может быть подменена отдельным агентом: решение о классификации принимается коллегиально на основе агрегированных показателей аномальности, а хэш-сумма тела оповещения (6) гарантирует неизменность переданных признаков. Тем самым подмена метки со стороны скомпрометированного агента исключена на уровне архитектуры конвейера обработки.

Эффективный показатель преобразуется в фактор поведенческого штрафа сигмоидальной зависимостью

$$\text{PF}(a,t) = \frac{1}{1+\exp(-\lambda(S_{\text{eff}}(a,t)-\tau))}, \tag{11}$$

где $\lambda \in [3{,}10]$ – коэффициент крутизны; τ – порог чувствительности. Обновление доверия выполняется каскадно – синхронно для всего множества участников атаки:

$$T_{\text{upd}}(a) = T_{\text{prev}}(a)\big(1 - PF(a,t)\big), \quad a \in S_{\text{conf}}. \tag{12}$$

Каскадный характер отличает обработку МВА от индивидуальных инцидентов: вместо последовательного пересчёта для одного агента выполняется групповое обновление, согласованное с принципом синхронного изменения поведения агентов при выявлении угроз. Восстановление доверия после многовекторной атаки замедляется дополнительным штрафным коэффициентом:

$$T_{\text{recover}}(a, t+\Delta) = T_{\text{upd}}(a) + \beta_{MVA}\alpha^{+}\left(1 - T_{\text{upd}}(a)\right), \tag{13}$$

где α^{+} – асимметричный коэффициент роста функции доверия; $\beta_{\text{MVA}} = 0{,}5$. Асимметрия «быстрое падение – медленное восстановление» согласуется с практикой моделей репутации. Падение доверия ниже порога изоляции переключает конечный автомат состояний агента:

$$T_{\text{upd}}(a) < \theta_{\text{iso}} \Longrightarrow \sigma(a) := \sigma_i, \quad \theta_{\text{iso}} = 0{,}25, \tag{14}$$

где σ_i – изолированное состояние. Обновление доверия и потенциальная изоляция автоматически инициируют пересмотр политик доступа в смарт-контрактах генерации и распространения политик; команды передаются формализованным кортежем «агент – обновлённое доверие – класс атаки – временная метка – состояние автомата», а все этапы фиксируются в распределённом реестре, обеспечивая неизменяемость журнала инцидентов [Пет23б, Пет26а].

Обратный канал переносит обновлённые уровни доверия в подсистему обнаружения: веса федеративной агрегации формализуются как функция средних уровней доверия зон

$$\alpha_k = \frac{\text{Trust}_{\text{avg}}(Z_k)}{\sum_j \text{Trust}_{\text{avg}}(Z_j)}, \tag{15}$$

где $\text{Trust}_{\text{avg}}(Z_k)$– среднее значение $\text{Trust}(j,t)$ по агентам зоны Z_k. Зависимость (15) подавляет вклад зон с пониженным доверием в глобальную модель, чем реализуется адаптивная перенастройка детектора по итогам каждого цикла реагирования. Сводная схема двунаправленной трансляции сигналов приведена на рис. 3.

Схема на рис. 3 представляет замкнутый цикл управления. По прямому каналу выход блока «Подсистема обнаружения: S_{loc}, метка y» преобразуется соотношениями (9)–(11) в фактор поведенческого штрафа и поступает в блок «Функция доверия: каскадное обновление (12)–(14)», откуда управляющие команды передаются в блок «Смарт-контракты, оркестрация, изоляция агентов». По обратному каналу обновлённые уровни доверия поступают в блок «Веса федеративной агрегации α_k (15)» и возвращаются оттуда в подсистему обнаружения как перенастройка детектора. Тем самым цикл замыкается: результат обнаружения меняет доверие, а изменённое доверие меняет сам детектор, и система адаптируется к действиям нарушителя без участия оператора.

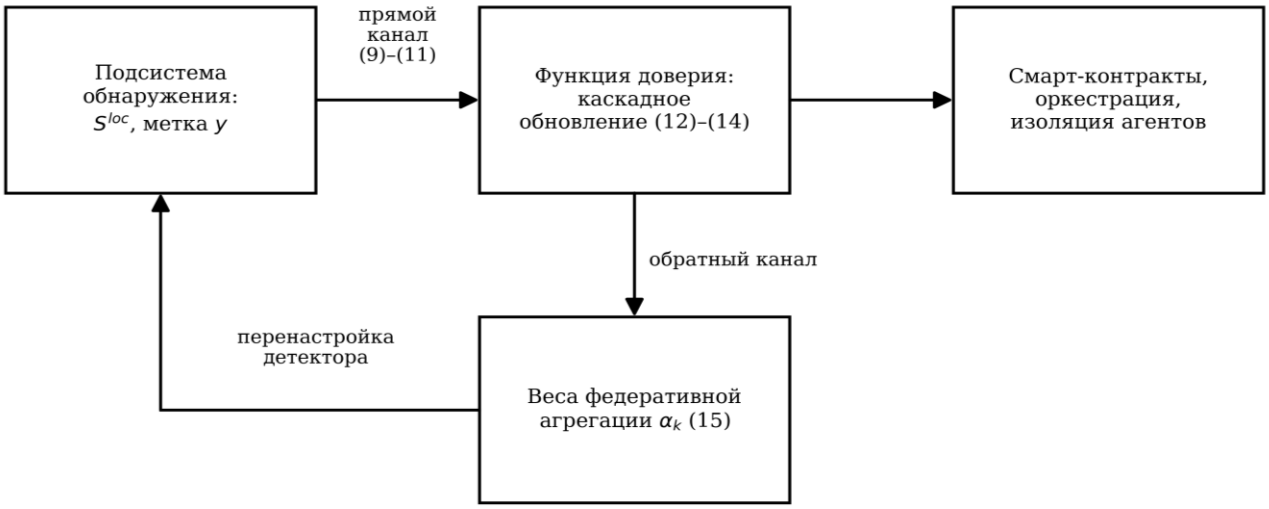


Рис. 3. Замкнутый контур «обнаружение – доверие – реагирование»: прямой канал (9)–(14) от показателей аномальности к обновлению доверия и командам смарт-контрактам; обратный канал (15) от доверия к весам федеративной агрегации

Каузальная корректность контура следует из ацикличности цепочки преобразований во временном измерении: значение T_{upd} в момент t зависит только от значения предыдущего цикла; фактор штрафа – только от результата обнаружения текущего момента; веса α_k очередного раунда агрегации – только от доверия, обновлённого по итогам предыдущего цикла. Каждое преобразование однотактово относительно своего канала, циклические зависимости отсутствуют. Отметим, что каскадное обновление (12) и веса (15) удовлетворяют тем же ограничениям, при которых доказана устойчивость децентрализованной функции доверия к византийским агентам при $f < |A|/3$ [Пет23а, Lam82]; это позволяет распространить оценку устойчивости на совокупный замкнутый контур. Различие темпов каналов – секунды для реагирования против минут для обучения – исключает взаимные колебания: быстрый канал локализует атаку, медленный плавно адаптирует модель.

От известных решений предлагаемую систему отличает двунаправленность обратной связи. В отличие от SIEM-систем факт обнаружения преобразуется в прямое управляющее воздействие на доверие и распределение задач без участия оператора; в отличие от SOAR-систем реагирование не ограничено заранее заданными сценариями – модель обнаружения непрерывно дообучается федеративным механизмом с весами (15).

Экспериментальная оценка

Верификация выполнена имитационным моделированием в среде ns-3 с интеграцией распределённого реестра IOTA Tangle. Число агентов-детекторов варьировалось от 5 до 20, число моделируемых узлов сети – от 100 до 1 000, длительность эксперимента – от 60 до 300 с, доля атак в трафике задавалась на уровнях 5, 15 и 30 %. Стратегии нарушителя (случайная, оппозиционная, координированная) моделировались по методике [Бас22]; генерация тестовых данных для сценариев атак на беспилотные аппараты опиралась на подходы [Бас22, Бас21]. План эксперимента образован полным перекрёстным комбинированием четырёх уровней числа узлов (100, 250, 500, 1 000), четырёх уровней числа детекторов (5, 10, 15, 20), трёх уровней длительности (60, 180, 300 с), трёх уровней доли атак (5, 15, 30%) и трёх стратегий нарушителя – всего 432 комбинации по пять повторений на каждую (2 160 прогонов); ещё 240 прогонов составили серию оценки масштабируемости (табл. 4). Всего проведено 2 400 экспериментов; доверительные интервалы уровня 95 % построены в соответствии с ГОСТ Р 8.736-2011[1]. Использованы наборы данных CIC IoT 2023 и N-BaIoT. Все базовые конфигурации сравнения (централизованная IDS, FL-IDS на стандартном FedAvg, изолированные локальные модели) реализованы и исполнены в той же среде ns-3, на тех же наборах данных, сценариях атак и плане эксперимента, что и предлагаемая методика; готовые значения показателей из цитируемых работ при построении таблиц 2 и 3 не использовались. Статистическая значимость различий подтверждена парным t-критерием Стьюдента с поправкой Холма-Бонферрони: сравнивались пары «интегрированная методика – каждая альтернатива таблицы 2» и «предлагаемая система – каждая базовая конфигурация таблицы 3», по 40 независимых прогонов на конфигурацию в каждом сравнении; для всех пар $|t| \geq 8{,}3$ при 78 степенях свободы ($p < 0{,}001$).

Помимо стандартных метрик качества классификации (Accuracy, Precision, Recall, F1, FPR) использованы показатель экономии трафика

$$\eta_{\text{traffic}} = 1 - \frac{V_{\text{FL}}}{V_{\text{raw}}}, \tag{16}$$

где V_{FL} и V_{raw} – объёмы данных, передаваемых при федеративной и централизованной схемах соответственно, а также суммарное время реакции

[1] ГОСТ Р 8.736–2011. Государственная система обеспечения единства измерений. Измерения прямые многократные. Методы обработки результатов измерений. Основные положения. М.: Стандартинформ, 2013. 20 с.

$$T_{\text{react}} = T_{\text{detect}} + T_{\text{alert}}, \tag{17}$$

складывающееся из времени обнаружения (от поступления аномального наблюдения до превышения порога) и времени доставки и подтверждения оповещения – далее времени реагирования (от публикации транзакции до выполнения условия (8)); результативность реагирования оценивалась долей успешно нейтрализованных атак η_{neut}.

Показатель (16) использован при сопоставлении методики с аналогами в табл. 2.

Таблица 2

Сопоставление интегрированной методики с аналогами и изолированными компонентами

Метод	Recall	FPR, %	η_{traffic}, %	T_{react}, с
Централизованная IDS	0,98	3,9	0	< 1
FL-IDS на стандартном FedAvg	0,93	5,2	41,7	4–6
Групповое зонирование MCMA-GZ [Теб26а]	0,96	2,9	–	3–5
GAN + RL [Теб26б]	0,92	3,4	–	3–5
Федеративное обучение [Пет25а]	0,96	3,1	50,2	3–5
CNN + LSTM + DPA [Пет25б]	0,99	2,8	–	3–5
PoV-верификация с обрезкой [Пет25в]	0,97	2,7	–	12 мс/блок
Классификатор вредоносных агентов при голосовании [Пет22]	0,94	2,1	–	2–4
Интегрированная методика	0,97	2,6	58,4	3–5

Интегрированная методика приближается к централизованной IDS по полноте обнаружения (0,97 против 0,98), не требуя передачи сырой телеметрии, и превосходит её по доле ложных срабатываний (2,6 % против 3,9 %) за счёт распределённой верификации (7)-(8), одновременно обеспечивая недоступную централизованным решениям экономию трафика 58,4 %. Относительно FL-IDS полнота выше на 0,04 при двукратном снижении FPR. При этом ни один из шести компонентов в изолированной форме не обеспечивает одновременно Recall ≥ 0,97 и FPR ≤ 2,6 %; эти показатели достигаются только их согласованной композицией (2), что подтверждает значимость синтеза как самостоятельного результата. Охват тестовых сценариев по строкам таблицы 2 следующий: пять изолированных компонентов, работающих на уровне трафика, оценивались на сценариях атак, порождающих сетевые аномалии; классификатор вредоносных агентов при коллективном голосовании [Пет22], по построению не анализирующий трафик, оценивался на сценариях манипуляции голосованием и приведён отдельной строкой; интегрированная методика и оба класса аналогов оценивались на полном наборе сценариев, включая манипуляции голосованием. Характеристики системы координированного реагирования в сопоставлении с референсными конфигурациями приведены в табл. 3.

Таблица 3

Сопоставление системы координированного реагирования с референсными конфигурациями

Конфигурация	Precision, %	Recall, %	F1, %	FPR, %	T_{react}, с	η_{neut}, %
Централизованная IDS (SIEM-класс)	95,8	95,1	95,5	1,2	≈ 9,9	61,2
FL-IDS без распределённого реестра	94,7	94,0	94,3	3,9	≈ 7,4	78,5
Изолированные локальные модели	90,5	88,2	89,3	5,4	≈ 11,4	52,7
Предлагаемая система	95,2	94,6	94,9	4,0	≤ 5,0	93,4

Суммарное время реакции (17) раскладывается по конфигурациям следующим образом (время обнаружения / время доставки и подтверждения оповещения, среднее ± 95 % ДИ): централизованная IDS – 6,16 ± 0,42 с / 3,70 ± 0,28 с (в сумме ≈ 9,9 с); FL-IDS – 4,86 ± 0,35 с /

2,52 ± 0,21 с (≈ 7,4 с); изолированные локальные модели – 7,92 ± 0,56 с / 3,45 ± 0,30 с (≈ 11,4 с); предлагаемая система – 3,10 ± 0,24 с / 1,62 ± 0,14 с (≤ 5,0 с). Тем самым предлагаемая система сокращает время обнаружения инцидентов на 49,7% (с 6,16 до 3,10 с) и время реагирования на 56,2 % (с 3,70 до 1,62 с) относительно централизованной конфигурации, обеспечивая долю нейтрализованных атак 93,4 % против 61,2 % при времени реакции не более 5с. Источник превосходства – асинхронное распространение оповещений через реестр и непрерывное дообучение модели. Ценой децентрализации является повышенный уровень ложных срабатываний: FPR системы реагирования (4,0%) превышает показатель SIEM-класса (1,2%). Этот компромисс смягчается двумя факторами: полуторакратным выигрышем в результативности нейтрализации и двукратным – во времени реакции, а также асимметричным правилом восстановления (13), ограничивающим последствия ложного штрафа для добросовестного агента.

Устойчивость к росту доли скомпрометированных агентов проверена в диапазоне $\rho \in [0{,}05;\ 0{,}30]$: методика сохраняет Recall ≥ 0,97 при FPR ≤ 3,5 %, а система реагирования – F1 ≥ 94 % при FPR = 4,0 %, что согласуется с теоретической границей $\rho < 1/3$. Превышение порога ведёт к деградации качества классификации, и это значение зафиксировано как граница применимости. Зависимость качества от ρ показана на рис. 4.

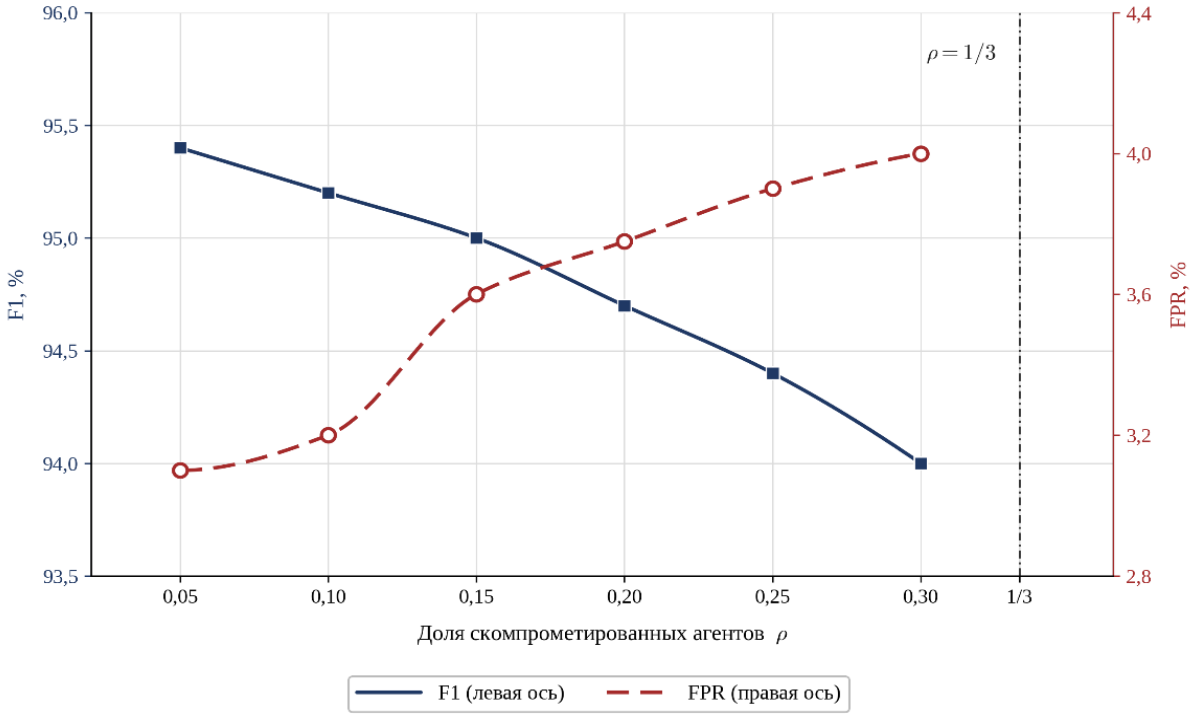


Рис. 4. Зависимость F1 и FPR от доли скомпрометированных агентов ρ в диапазоне 0,05-0,30; вертикальной линией отмечена теоретическая граница применимости $\rho = 1/3$

По результатам экспериментов снижение показателей в диапазоне до $\rho = 0{,}30$ остаётся плавным; резкая деградация качества наступает только за теоретической границей. Применимость к комбинированным атакам подтверждена тремя сценариями: «Vampire + Sybil» (сочетание маршрутизирующей и энергоистощающей составляющих), «DDoS + Stealth On-Off + сговор» и координированная стратегия скомпрометированных агентов в процессе коллективного принятия решений. Во всех случаях итоговая метка комбинированного класса сформирована корректно по правилу декартовой композиции базовых классов. Масштабируемость методики оценена при росте числа узлов сети; результаты приведены в табл. 4.

Таблица 4

Метрики масштабируемости интегрированной методики

Число узлов	Число детекторов	Accuracy	F1	T_{detect}, мс
100	5	0,996	0,997	142 ± 8
500	10	0,994	0,995	168 ± 11
1 000	20	0,990	0,991	196 ± 14

Табл. 4 характеризует поведение методики при десятикратном росте размерности задачи. Качество классификации при этом практически не меняется: Accuracy снижается с 0,996 до 0,990, F1 – с 0,997 до 0,991, то есть потеря по обоим показателям не превышает 0,006 и сопоставима с шириной доверительного интервала отдельного замера, а потому не является статистически значимой. Иначе ведёт себя время обнаружения: оно возрастает со 142 ± 8 до 196 ± 14 мс, однако прирост составляет 38 % при увеличении числа узлов на 900 %, т. е. зависимость выражено сублинейна. Такой характер роста объясняется устройством контура обнаружения. Число зональных детекторов увеличивается с 5 до 20 синхронно с размером сети, поэтому объём телеметрии, приходящейся на один детектор, растёт значительно медленнее общего объёма трафика; локальные фазы обнаружения при этом независимы и выполняются параллельно, так что увеличение их числа не удлиняет общий цикл. Оставшийся прирост задержки вносит фаза распределённой верификации, где сбор голосов по правилу (7) занимает тем больше времени, чем больше валидаторов участвует в голосовании; именно этот последовательный по своей природе этап и определяет наблюдаемые 38 %. Абсолютное значение времени обнаружения даже при максимальной конфигурации остаётся в пределах 200 мс – на порядок меньше интервала обновления глобальной модели, составляющего около 15 мин. Следовательно, методика применима к сетям масштаба тысячи узлов без пересмотра архитектуры.

Обсуждение и ограничения

Принципиальным ограничением подхода является граница $\rho = 1/3$, наследуемая от фундаментального результата теории византийского консенсуса [Lam82]: при большей доле скомпрометированных агентов согласованность децентрализованных оценок доверия не может быть гарантирована никакой схемой, и предлагаемая методика не составляет исключения. Верификация носит имитационный характер: среда ns-3 с интеграцией IOTA Tangle воспроизводит сетевую динамику и протокольные задержки, но не охватывает всей совокупности физических эффектов реальных робототехнических платформ. Натурные испытания на беспилотных летательных аппаратах и наземных платформах составляют предмет дальнейшей работы.

Разрыв темпов каналов обратной связи – секунды для оповещений против интервала федеративного обновления порядка 15 минут – означает, что атака нового, не представленного в модели типа локализуется быстрым каналом (изоляция, пересмотр задач), однако полная адаптация детектора к ней требует завершения очередного раунда обучения. Для операций с более жёсткими требованиями к адаптации интервал раундов подлежит сокращению с соответствующим ростом коммуникационных издержек. Кроме того, уровень ложных срабатываний фазы реагирования (4,0%) выше, чем у централизованных решений. Приемлемость этого уровня для контура доверия обеспечивается мультипликативной формой обновления (12) в сочетании с правилом восстановления (13): единичный ложный штраф снижает доверие добросовестного агента обратимо и не приводит к его изоляции. Перспективными направлениями развития являются натурная валидация, интеграция постквантовых криптографических примитивов в слой подписи оповещений и распространение методики на гибридные человеко-машинные коллективы.

Заключение

Предложена методика обнаружения, классификации и координированного реагирования на многовекторные атаки в децентрализованной киберфизической среде, отличающаяся от известных решений замкнутой двунаправленной обратной связью по параметрам доверия. Методика обнаружения построена как согласованная композиция шести алгоритмических компонентов с единой шкалой показателей аномальности и обеспечивает полноту обнаружения 0,97 при доле ложных срабатываний 2,6% и экономии трафика 58,4%. Система координированного реагирования преобразует подтверждённые обнаружения в управляющие воздействия по каналам доверия и оркестрации, сокращая время обнаружения на 49,7 %, время реагирования – на 56,2% и обеспечивая нейтрализацию 93,4% атак при времени реакции не более 5 с. Формализованная трансляция сигналов – прямой канал каскадного обновления доверия с дифференцированными по классам атак коэффициентами и обратный канал перенастройки весов федеративной агрегации – каузально корректна и сохраняет доказанную устойчивость к скомпрометированным агентам при их доле менее 1/3 (теоретическая граница), что эмпирически подтверждено в диапазоне до 0,30 при масштабировании до 1 000 узлов. Совокупность результатов образует замкнутый адаптивный контур «наблюдение – доверие – обнаружение – реагирование», в котором факт обнаружения атаки не только вызывает ответное действие, но и адаптивно перенастраивает саму подсистему обнаружения.

СПИСОК ЛИТЕРАТУРЫ | REFERENCES

[Abs25] Abshari D., Sridhar M. A Survey of anomaly detection in cyber-physical systems // arXiv preprint. 2025. DOI: 10.48550/arXiv.2502.13256.

Abshari D., Sridhar M. A Survey of anomaly detection in cyber-physical systems // arXiv preprint. 2025. DOI: 10.48550/arXiv.2502.13256.

[Ami25] Amine M. S., Nada F. A., Hosny K. M. Improved model for intrusion detection in the Internet of Things // Scientific Reports. 2025. Vol. 15, article 21547. DOI: 10.1038/s41598-025-92852-6. EDN: zzbifw.

Amine M. S., Nada F. A., Hosny K. M. Improved model for intrusion detection in the Internet of Things // Scientific Reports. 2025. Vol. 15, article 21547. DOI: 10.1038/s41598-025-92852-6. EDN: zzbifw.

[Ana25a] Anand R. V., Magesh G., Alagiri I., et al. Design of an improved model using federated learning and LSTM autoencoders for secure and transparent blockchain network transactions // Scientific Reports. 2025. Vol. 15, article 1615. DOI: 10.1038/s41598-024-83564-4.

Anand R. V., Magesh G., Alagiri I., et al. Design of an improved model using federated learning and LSTM autoencoders for secure and transparent blockchain network transactions // Scientific Reports. 2025. Vol. 15, article 1615. DOI: 10.1038/s41598-024-83564-4.

[Ana25b] Anand S., Narayanan A. V. Addressing rogue nodes and trust management: leveraging deep learning-enhanced hybrid trust to optimize wireless sensor networks management // J. of Robotics and Control. 2025. Vol. 6, no. 2. Pp. 846–861. DOI: 10.18196/jrc.v6i2.25600. EDN: opjcgf.

Anand S., Narayanan A. V. Addressing rogue nodes and trust management: leveraging deep learning-enhanced hybrid trust to optimize wireless sensor networks management // J. of Robotics and Control. 2025. Vol. 6, no. 2. Pp. 846–861. DOI: 10.18196/jrc.v6i2.25600. EDN: opjcgf.

[Anw25] Anwer R. W., Abrar M., Ullah M., et al. Advanced intrusion detection in the industrial Internet of Things using federated learning and LSTM models // Ad Hoc Networks. 2025. Vol. 178, art. 103991. DOI: 10.1016/j.adhoc.2025.103991. EDN: qargqw.

Anwer R. W., Abrar M., Ullah M., et al. Advanced intrusion detection in the industrial Internet of Things using federated learning and LSTM models // Ad Hoc Networks. 2025. Vol. 178, art. 103991. DOI: 10.1016/j.adhoc.2025.103991. EDN: qargqw.

[Lam82] Lamport L., Shostak R., Pease M. The Byzantine Generals Problem // ACM Transactions on Programming Languages and Systems. 1982. Vol. 4, no. 3. Pp. 382–401. DOI: 10.1145/357172.357176.

Lamport L., Shostak R., Pease M. The Byzantine Generals Problem // ACM Transactions on Programming Languages and Systems. 1982. Vol. 4, no. 3. Pp. 382–401. DOI: 10.1145/357172.357176.

[Zoh24] Zohourian A., Dadkhah S., Molyneaux H., et al. IoT-PRIDS: Leveraging packet representations for intrusion detection in IoT networks // Computers & Security. 2024. Vol. 146, article 104034. DOI: 10.1016/j.cose.2024.104034. EDN: iwfxnc.

Zohourian A., Dadkhah S., Molyneaux H., et al. IoT-PRIDS: Leveraging packet representations for intrusion detection in IoT networks // Computers & Security. 2024. Vol. 146, article 104034. DOI: 10.1016/j.cose.2024.104034. EDN: iwfxnc.

[Zuo23] Zuo J., Cao R., Qi J., et al. A Hierarchical Blockchain-Based Trust Measurement Method for Drone Cluster Nodes // Drones. 2023. Vol. 7, art. 627. DOI: 10.3390/drones7100627. EDN: kftljp.

Zuo J., Cao R., Qi J., et al. A Hierarchical Blockchain-Based Trust Measurement Method for Drone Cluster Nodes // Drones. 2023. Vol. 7, art. 627. DOI: 10.3390/drones7100627. EDN: kftljp.

[Арх22] Архипова А. Б., Медведев М. А. и др. Перспективы использования машинного обучения для классификации

Arkhipova A. B., Medvedev M. A., et al. Prospects of using machine learning for network traffic classification in

сетевого трафика в технологиях доверенного взаимодействия // Безопасность цифровых технологий. 2022. № 3 (106). С. 49–61. EDN: iyoemy.

[Бас21] Басан Е. С., Абрамов Е. С. и др. Метод обнаружения атак на систему навигации БПЛА // Информатика и автоматизация. 2021. Т. 20, № 6. С. 1368–1394. EDN: obobih.

[Бас22] Басан Е. С. и др. Генерация данных для моделирования атак на БПЛА с целью тестирования систем обнаружения вторжений // Информатика и автоматизация. 2022. Т. 21, № 6. С. 1290–1327. EDN: iqvybm.

[Вас21] Васильев В. И., Вульфин А. М., и др. Гибридная интеллектуальная система обнаружения атак на основе комбинации методов машинного обучения // МОИТ. 2021. Т. 9, № 3 (34). С. 1–11. EDN: yiubvm.

[Пат24] Пат. 2822581 РФ, МПК G16Y 30/10. Способ доверенного взаимодействия агентов в децентрализованной киберфизической среде на основе группового зонирования / В. И. Петренко, Ф. Б. Тебуева и др. № 2023131407; опубл. 09.07.2024, Бюл. № 19. EDN: dbskha.

[Пет22] Петренко В. И. и др. Метод обнаружения нарушений информационной безопасности в роевых робототехнических системах с использованием технологий машинного обучения // Вестник ВГУ. Серия: Сист. анализ и информ. технологии. 2022. № 1. С. 43–55. EDN: uoujgv.

[Пет23а] Петренко В. И., Тебуева Ф. Б. и др. Модель доверенного взаимодействия агентов в децентрализованной киберфизической среде // Вестник ДагГТУ. Технические науки. 2023. Т. 50, № 2. С. 134–141. EDN: qmazuq.

[Пет23б] Петренко В. И., Тебуева Ф. Б. и др. Метод доверенного взаимодействия агентов в децентрализованной киберфизической среде на основе технологии распределенного реестра // Прикасп. журн.: управление и высокие технологии. 2023. № 3 (63). С. 115–125. EDN: tyzdhc.

[Пет24] Петренко В. И., Копытов В. В. и др. Модель угроз нарушения информационной безопасности процесса доверенного взаимодействия устройств IoT-системы, формализующая сценарии многовекторных атак // Вестник СПГУТД. Серия 1: Естественные и технические науки. 2024. № 2. С. 126–133. EDN: nrobwe.

[Пет25а] Петренко В. И. и др. Масштабируемый метод обнаружения многовекторных атак на скомпрометированные устройства IoT-системы с использованием машинного обучения // Программная инженерия. 2025. Т. 16, № 4. С. 167–180. EDN: pnvcaw.

[Пет25б] Петренко В. И. и др. Методика обнаружения и противодействия многовекторным угрозам нарушения информационной безопасности децентрализованной IoT-системы // International Journal of Open Information Technologies. 2025. Т. 13, № 1. С. 14–24. EDN: dorket.

[Пет25в] Петренко В. И. и др. Имитационная модель масштабируемого метода выявления многовекторных атак с учетом ограничений вычислительных и информационных ресурсов IoT-устройств // Российский технологический журнал. 2025. Т. 13, № 5. С. 25–40. EDN: jkqmqm.

[Пет26а] Петренко В. И., Наджаджра М. Х. и др. Метод управления доверием в децентрализованных сетях БПЛА на основе контекстно-ориентированного консенсуса // Информационные технологии. 2026. Т. 32, № 5. С. 251–262. EDN: wlhdso.

[Пет26б] Петренко В. И., Наджаджра М. Х. и др. Метод распределенного контроля доступа при доверенном взаимодействии роботизированных агентов в децентрализованной киберфизической среде // Программная инженерия. 2026. Т. 17, № 6. С. 320–324. EDN: wjjhtw.

trusted interaction technologies // Digital Technology Security. 2022. No. 3 (106), pp. 49–61. (In Russian). EDN: iyoemy.

Basan E. S., et al. A method for detecting attacks on the UAV navigation system // Informatics and Automation. 2021. 20(6): 1368–1394. (In Russian). EDN: obobih.

Basan E. S., et al. Data generation for modeling attacks on UAVs for testing intrusion detection systems // Informatics and Automation. 2022. Vol. 21, no. 6, pp. 1290–1327. (In Russian). EDN: iqvybm.

Vasilyev V. I., Vulfin A. M., Gvozdev V. E., et al. Hybrid intelligent attack detection system based on a combination of machine learning methods // MOIT. 2021. Vol. 9, no. 3 (34), pp. 1–11. (In Russian). EDN: yiubvm.

Patent RU 2822581. Method of trusted interaction of agents in a decentralized cyber-physical environment based on group zoning / V. I. Petrenko, F. B. Tebueva, S. S. Ryabtsev, et al. Publ. 09.07.2024, Bull. no. 19. (In Russian). EDN: dbskha.

Petrenko V. I., et al. A method for detecting information security violations in swarm robotic systems using machine learning technologies // Proc. of Voronezh State University. Series: Systems Analysis and Information Technologies. 2022. No. 1, pp. 43–55. (In Russian). EDN: uoujgv.

Petrenko V. I., Tebueva F. B., Struchkov I. V., Ryabtsev S. S. Model of trusted interaction of agents in a decentralized cyber-physical environment // Vestnik DagSTU. Tech. Sci. 2023. 50(2): 134–141. (In Russian). EDN: qmazuq.

Petrenko V. I., Tebueva F. B., et al. A method of trusted interaction of agents in a decentralized cyber-physical environment based on distributed ledger technology // Caspian Journal: Control and High Technologies. 2023. No. 3 (63), pp. 115–125. (In Russian). EDN: tyzdhc.

Petrenko V. I., Kopytov V. V., et al. A threat model of information security violations of the trusted interaction process of IoT-system devices formalizing multi-vector attack scenarios // Vestnik of St. Petersburg State University of Technology and Design. Series 1. 2024. No. 2, pp. 126–133. (In Russian). EDN: nrobwe.

Petrenko V. I., et al. A scalable method for detecting multi-vector attacks on compromised IoT-system devices using machine learning // Software Engineering. 2025. Vol. 16, no. 4, pp. 167–180. (In Russian). EDN: pnvcaw.

Petrenko V. I., et al. A technique for detecting and countering multi-vector threats to information security of a decentralized IoT system // International Journal of Open Information Technologies. 2025. Vol. 13, no. 1, pp. 14–24. (In Russian). EDN: dorket.

Petrenko V. I., et al. A simulation model of a scalable method for detecting multi-vector attacks considering the computational and information resource constraints of IoT devices // Russian Technological Journal. 2025. Vol. 13, no. 5, pp. 25–40. (In Russian). EDN: jkqmqm.

Petrenko V. I., Najajra M. H., et al. A trust management method in decentralized UAV networks based on context-oriented consensus // Information Technologies. 2026. Vol. 32, no. 5, pp. 251–262. (In Russian). EDN: wlhdso.

Petrenko V. I., Najajra M. H., et al. A distributed access control method for trusted interaction of robotic agents in a decentralized cyber-physical environment // Software Engineering. 2026. Vol. 17, no. 6, pp. 320–324. (In Russian). EDN: wjjhtw.

[Теб26а] Тебуева Ф. Б. и др. Метод противодействия многовекторным атакам на агентов в децентрализованной среде Интернета вещей на основе группового зонирования // Кузнечно-штамповочное производство. Обработка материалов давлением. 2026. № 3. (В печати).

Tebueva F. B., et al. A method for countering multi-vector attacks on agents in a decentralized Internet of Things environment based on group zoning // Forging and Stamping Production. Material Working by Pressure. 2026. No. 3. (In press). (In Russian).

[Теб26б] Тебуева Ф. Б. и др. Метод обнаружения и классификации угроз нарушения информационной безопасности при реализации многовекторных атак на агентов в децентрализованной среде Интернета вещей // Информационные технологии. 2026. № 8. С. 437-448. EDN: eyiyjb.

Tebueva F. B., et al. A method for detecting and classifying information security threats in multi-vector attacks on agents in a decentralized Internet of Things environment // Information Technologies. 2026. No. 8. Pp. 437-448. (In Russian). EDN: eyiyjb.

ОБ АВТОРАХ | ABOUT THE AUTHORS

ПЕТРЕНКО Вячеслав Иванович
Северо-Кавказский федеральный университет, Россия.
vipetrenko@ncfu.ru ORCID: 0000-0003-4293-7013.
Зав. кафедрой, канд. техн. наук, доцент. Исс. в обл. кибербезопасности роботизированных агентов.

PETRENKO Vyacheslav Ivanovich
North-Caucasus Federal University, Russia.
vipetrenko@ncfu.ru ORCID: 0000-0003-4293-7013.
Head of Dept. PhD in Engineering, Assoc. Prof. Research in the field of cybersecurity of robotic agents.

МЕТАДАННЫЕ | METADATA

Заглавие: Методика обнаружения, классификации и координированного реагирования на многовекторные атаки в децентрализованной киберфизической среде с замкнутой обратной связью по параметрам доверия.

Авторы: Петренко В. И.

Аннотация: Рассматривается задача защиты роботизированных агентов, функционирующих в децентрализованной киберфизической среде без доверенного центра, от многовекторных атак – координированных воздействий, сочетающих сетевые, поведенческие и физические каналы в рамках единого сценария. Предложена методика обнаружения и классификации, построенная как согласованная по входам и выходам композиция шести алгоритмических компонентов: группового зонирования с учётом доверия, гибридного обнаружения на основе генеративно-состязательной сети и обучения с подкреплением, федеративного обучения, многомерной реконструкции трафика, распределённой верификации на основе PoV-консенсуса и нейросетевой классификации вредоносных агентов. Отличительной особенностью является замкнутая двунаправленная обратная связь по параметрам доверия: подтверждённое обнаружение каскадно обновляет функцию доверия и инициирует перераспределение задач, а обновлённые уровни доверия перенастраивают веса федеративной агрегации подсистемы обнаружения. Имитационное моделирование в среде ns-3 (2 400 экспериментов, до 1 000 узлов) показало полноту обнаружения 0,97 при доле ложных срабатываний 2,6 %, сокращение времени обнаружения на 49,7 %, времени реагирования на 56,2 % и долю нейтрализованных атак 93,4 % при времени реакции не более 5 с; устойчивость эмпирически подтверждена при доле скомпрометированных агентов до 0,30, что согласуется с теоретической границей 1/3.

Ключевые слова: Роботизированные агенты; децентрализованные системы; глубокое обучение с подкреплением; доверенная оркестрация; механизм доверия; координация; интеллектуальные системы.

Язык: Русский.

Статья поступила в редакцию 24 августа 2026 г.

Title: Multi-vector attack detection, classification and coordinated response in a decentralized cyber-physical environment with closed-loop trust feedback.

Authors: Petrenko V. I.

Abstract: The task of protecting robotic agents operating in a decentralized cyber-physical environment without a trusted center against multi-vector attacks—coordinated actions that combine network, behavioral, and physical channels within a single scenario—is considered. A detection and classification methodology is proposed, constructed as an input-output-consistent composition of six algorithmic components: trust-aware group zoning, hybrid detection based on a generative adversarial network and reinforcement learning, federated learning, multidimensional traffic reconstruction, distributed verification based on PoV consensus, and neural network classification of malicious agents. A distinctive feature is the closed bidirectional feedback loop over trust parameters: a confirmed detection cascades an update of the trust function and initiates task redistribution, while the updated trust levels retune the federated aggregation weights of the detection subsystem. Simulation modeling in the ns-3 environment (2,400 experiments, up to 1,000 nodes) demonstrated a detection recall of 0.97 with a false positive rate of 2.6%, a 49.7% reduction in detection time, a 56.2% reduction in response time, and a neutralization rate of 93.4% of attacks with a reaction time of no more than 5 s; robustness is empirically confirmed for a share of compromised agents of up to 0.30, consistent with the theoretical bound of one third.

Key words: Robotic agents; decentralized systems; deep reinforcement learning; trusted orchestration; trust mechanism; coordination; intelligent systems.

Language: Russian.

The article was received by the editors on 24 August 2026.