УДК 004.056:004.89:004.932

УГРОЗЫ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ В ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ ВИДЕОНАБЛЮДЕНИЯ И МЕТОДЫ ЗАЩИТЫ

Белкин Е.Н.

Начальник отдела научно-технического развития,

АО «АЛГОНТ»,

Калуга, Россия

Кондрашев А.С.

Первый заместитель генерального директора,

АО «АЛГОНТ»,

Калуга, Россия

Аннотация

В статье систематизированы угрозы информационной безопасности интеллектуальных систем оптико-электронного наблюдения, выполняющих детектирование, сопровождение, распознавание и классификацию объектов средствами компьютерного зрения и машинного обучения. Рассмотрены атаки на видеокамеры и каналы связи, подмена и повтор видеопотока, состязательные воздействия, отравление данных, внедрение закладок в модели, извлечение параметров, утечки биометрической информации, компрометация архивов и уязвимости вычислительной платформы. Отдельно обосновано приоритетное применение CPU для видеоаналитики при выполнении требований производительности: исключение проприетарного драйвера графического процессора (GPU) или нейронного процессора (NPU) и связанной среды исполнения уменьшает доверенную вычислительную базу и поверхность атаки. Предложена модель адаптивного доверенного контура видеоаналитики,

объединяющая оценку остаточного риска, выбор вычислительной платформы, независимую CPU-проверку результатов и безопасную деградацию.

Ключевые слова: интеллектуальное видеонаблюдение, система оптико-электронного наблюдения, компьютерное зрение, машинное обучение, информационная безопасность, состязательные атаки, видеопоток, GPU, CPU, NPU, доверенная вычислительная база.

INFORMATION SECURITY THREATS IN INTELLIGENT VIDEO SURVEILLANCE SYSTEMS AND PROTECTION METHODS

Belkin E.N.
Head of research and development department,
ALGONT JSC,
Kaluga, Russia

Kondrashev A.S.
First deputy general director,
ALGONT JSC,
Kaluga, Russia

Abstract

The article systematizes information security threats to intelligent electro-optical surveillance systems that perform object detection, tracking, recognition and classification using computer vision and machine learning. It considers attacks on cameras and communication channels, video stream substitution and replay, adversarial manipulation, data poisoning, model backdoors, model extraction, biometric data leakage, archive compromise and computing-platform vulnerabilities. Priority use of a central processing unit (CPU) for video analytics is justified when performance requirements are met: removing a proprietary graphics processing unit (GPU) or Neural Processing Unit (NPU) driver and its runtime from the architecture

reduces the trusted computing base and attack surface. An adaptive trusted video analytics perimeter model is proposed. It combines residual-risk assessment, computing-platform selection, independent CPU-based result verification and controlled fail-safe degradation.
Keywords: intelligent video surveillance, electro-optical surveillance system, computer vision, machine learning, information security, adversarial attacks, video stream, GPU, CPU, NPU, trusted computing base.

Введение

Интеллектуальная система оптико-электронного наблюдения (СОЭН) включает камеры, каналы передачи, вычислительные узлы, сервер управления видео, архив и модели компьютерного зрения. Она выполняет детектирование, сопровождение, распознавание и классификацию объектов, формирует метаданные и может передавать команды в системы контроля доступа, охранную сигнализацию и иные исполнительные контуры [10]. Поэтому ошибка видеоаналитики влияет не только на отображение оператору, но и на физическую безопасность и доказательственную ценность записи.

Поверхность атаки охватывает физическую сцену, камеру, сеть, видеопоток, модель, данные обучения, среду инференса, архив и интерфейс принятия решений. Для моделей актуальны уклонение, отравление, закладки и извлечение [2; 4–7; 9; 12; 13]. Для сетевой части — слабые учетные данные, уязвимые прошивки и отказ в обслуживании [3; 11; 14]. Платформенные уязвимости могут обходить меры на уровне модели.

Цель работы — систематизировать угрозы интеллектуальной СОЭН и предложить методы защиты с учетом выбора вычислительной базы. Использованы декомпозиция архитектуры, анализ публикаций и оценка остаточного риска. Научная новизна состоит в модели адаптивного доверенного контура видеоаналитики: вычислительная платформа выбирается одновременно

по риску и заданным показателям производительности (SLA), а критический результат ускорителя проходит независимую CPU-проверку.

Архитектура, активы и модель нарушителя

В СОЭН выделяются контуры сенсоров, встроенного ПО камер, сети, декодирования и видеоаналитики, управления видео и архива, подготовки моделей, рабочих мест и интеграций. Защищаются исходные кадры, метаданные, биометрические шаблоны и эмбеддинги, модели, наборы обучения, конфигурация правил, ключи, журналы и временные метки. Для критических применений самостоятельным активом является достоверность связи между событием, результатом модели и конкретным фрагментом записи.

Рассматриваются внешний сетевой нарушитель, физически присутствующий объект, легитимный пользователь, администратор отдельной подсистемы, поставщик модели и непривилегированный процесс на общем узле. Возможности включают предъявление подготовленного объекта, повтор потока, подмену прошивки или модели, модификацию обучающей выборки, эксплуатацию драйвера и удаление архива. Критична скрытная атака, при которой система сохраняет видимость штатной работы.

Классификация угроз интеллектуальной СОЭН

На сенсорном уровне применяются перекрытие поля зрения, засветка, расфокусировка, смещение камеры, загрязнение оптики, предъявление изображения экраном и физический состязательный объект. На уровне камеры используются заводские пароли, уязвимые сервисы, измененная прошивка и отключение записи. Анализ Default Credentials для веб-камер показывает возможность бокового перемещения [3], а исследование Mirai — масштабируемость компрометации IoT-устройств и DDoS [14].

В сети возможны перехват, подмена, задержка, удаление и повтор пакетов, перенаправление потока, изменение команд PTZ и времени. Шифрование без взаимной аутентификации и контроля свежести не исключает повтор ранее корректной записи. Архив атакуется путем удаления, выборочной модификации,

подмены временных меток, разрыва связи события с видео и несанкционированного экспорта.

Модельный уровень включает состязательное уклонение [2; 5; 7; 12; 13; 16], отравление данных [9], закладки [4; 17], подмену модели, извлечение параметров [22] и membership inference [20]. Последняя угрожает приватности биометрических наборов. Обычная точность на тестовой выборке не выявляет скрытый триггер и не характеризует устойчивость к физическим изменениям сцены. Если в состав СОЭН включен мультимодальный генеративный модуль, формирующий текстовые отчеты по видеопотоку, возникает угроза непрямой промпт-инъекции: текст в кадре может быть интерпретирован как инструкция и изменить поведение модели [8].

Платформенный уровень включает декодер, ОС, контейнер, библиотеку инференса, драйвер и ускоритель. Эксплуатация декодера возможна до анализа кадра. Компрометация GPU/NPU-драйвера затрагивает управление памятью, DMA и контекстами и действует ниже модельных фильтров. Основные угрозы и меры приведены в таблице 1.

Таблица 1 – Основные угрозы по контурам интеллектуальной СОЭН

Контур	**Основные угрозы**	**Последствия и приоритетные меры**
Сцена и сенсор	Перекрытие, засветка, смещение, экран, состязательный объект или патч.	Пропуск объекта или ложное событие. Контроль качества изображения, ориентиров, живости сцены, подтверждение другим сенсором.
Камера и прошивка	Заводские пароли, уязвимые сервисы, измененная прошивка, остановка записи.	Захват потока, ботнет, скрытая подмена. Уникальные ключи, secure boot, подписанная прошивка, сетевой запрет по умолчанию.
Сеть	Перехват, MITM, повтор, задержка, удаление кадров, подмена времени и команд PTZ.	Ложная картина обстановки. Взаимная аутентификация, шифрование, номера последовательности, контроль свежести и времени.
Модель и данные	Уклонение, отравление, закладка, подмена, извлечение, membership inference.	Ошибка распознавания, скрытый триггер, утечка модели или биометрии. Происхождение данных, подпись релиза, adversarial-тесты, квоты запросов.

Контур	Основные угрозы	Последствия и приоритетные меры
Вычислительный узел	Уязвимость декодера, ОС, контейнера, GPU/NPU-драйвера, runtime; межконтекстная утечка.	Выполнение кода, утечка кадров и модели, искажение результата, отказ. CPU-first, изоляция, минимизация компонентов, обновление и мониторинг.
Архив и управление	Удаление или изменение записи, подмена метаданных, подавление тревоги, злоупотребление оператором.	Утрата доказательств и неверное действие. WORM, хеш-цепочки, RBAC, усиленная аутентификация, независимый шлюз решения.

Известные факты реализации угроз

Физическая реализуемость состязательных воздействий подтверждена экспериментально. Устойчивые возмущения дорожных знаков сохраняли целевую ошибку при изменении условий съемки [15]. Патч перед телом снижал обнаружение человека камерой [21]. Специальная оправа обеспечивала уклонение и имперсонацию при распознавании лица [19]. Троянизированная нейросеть могла сохранять штатную работу без триггера и менять решение при его появлении [17].

Для GPU/NPU показаны практические побочные каналы через совместно используемые ресурсы [18] и извлечение характеристик DNN через переключение контекста [23; 1]. Для СОЭН это означает возможность утечки кадров, эмбеддингов и модели, искажения результата либо остановки группы потоков. Сводные факты приведены в таблице 2.

Таблица 2 – Результаты академических работ, значимые для СОЭН

Источник	Подтвержденный тип воздействия	Значение для СОЭН
[15]	Физически реализуемые возмущения сохраняли целевую ошибку классификации при разных условиях съемки.	Нельзя оценивать устойчивость только на цифровой тестовой выборке.
[21]	Патч снижал обнаружение человека автоматической камерой наблюдения.	Одежда или переносимый объект могут использоваться для уклонения от детектора.
[19]	Специальный аксессуар обеспечивал уклонение и имперсонацию при распознавании лица.	Биометрический допуск требует живости, качества и контекстной проверки.

Источник	Подтвержденный тип воздействия	Значение для СОЭН
[17]	Троянизированная нейросеть сохраняла нормальную работу без триггера и меняла результат при триггере.	Обычная оценка точности не выявляет закладку; нужен контроль происхождения и специальные тесты.
[14]	Массовая компрометация встроенных/IoT-устройств и их применение в DDoS.	Камера является сетевым узлом и источником атаки, а не только сенсором.
[18; 23]	Практические побочные каналы GPU и утечки признаков DNN через совместно используемые ресурсы.	Совместное использование ускорителя расширяет риск утечки модели и параметров нагрузки.
[1]	Систематизация атак через драйвер, память, DMA, планировщик, профилирование и изоляцию контекстов.	Проприетарный GPU/NPU-стек должен учитываться в доверенной базе; его исключение обосновывает CPU-first.

Обоснование приоритетного применения CPU

GPU/NPU добавляет в доверенную вычислительную базу драйвер ядра, runtime, пользовательские библиотеки, компилятор графов, прошивку, память устройства и средства профилирования. При проприетарной реализации независимый анализ ограничен. Уязвимость этого стека находится ниже контейнера и модели; совместное использование ускорителя создает дополнительные каналы утечки [1; 18; 23].

CPU также содержит уязвимости. Преимущество CPU-first состоит не в абсолютной безопасности, а в сокращении состава системы. Если число потоков, задержка и резерв обеспечиваются CPU, исключение GPU/NPU-драйвера и runtime устраняет специализированные интерфейсы, контексты, видеопамять и отдельный канал обновления. Поэтому для критической аналитики CPU является предпочтительным при подтвержденном SLA.

GPU/NPU допускается, когда стендовые испытания показывают невыполнение обязательного SLA на CPU. Ускоритель размещается на выделенном узле без недоверенной совместной нагрузки. Применяются IOMMU, белый список версий, подписи, минимальный runtime, запрет отладки и

профилирования, очистка памяти, квоты и watchdog. Результат GPU/NPU не должен напрямую вызывать критическое действие. Сравнение вариантов приведено в таблице 3.

Таблица 3 – Сравнение CPU- и GPU/NPU-вариантов с позиции безопасности

Критерий	**CPU без GPU/NPU-стека**	**GPU/NPU с проприетарным драйвером**	**Архитектурное решение**
Доверенная база	ОС, CPU, декодер, библиотека инференса.	Дополнительно драйвер ядра, runtime, библиотеки, прошивка, средства профилирования.	При равном SLA выбирать меньший состав.
Изоляция памяти	Механизмы ОС и процесса; остаются общесистемные аппаратные риски.	Отдельная видеопамять, контексты, DMA, кэши и специфичные интерфейсы.	Запрет недоверенной совместной нагрузки, IOMMU, очистка памяти.
Аудит и обновление	Обычно меньше специализированных компонентов и каналов обновления.	Зависимость от версии закрытого драйвера и согласованности нескольких компонентов.	Белый список версий, подписи, тестирование обновлений, быстрый откат.
Производительность	Достаточна не для всех моделей и числа потоков.	Высокая параллельная производительность.	Сначала стенд CPU; GPU/NPU только при нарушении обязательного SLA.
Критическое решение	Может выполнять анализ и независимую проверку.	Компрометация ускорителя не должна прямо вызывать действие.	CPU-шлюз, подтверждение несколькими признаками, режим отказа от решения.
Отказ	Перегрузка ведет к задержке или пропуску кадров.	Сбой драйвера может одновременно остановить много потоков.	Квоты, резерв, watchdog, перенос приоритетных задач на CPU.

Модель адаптивного доверенного контура видеоаналитики

Модель адаптивного доверенного контура видеоаналитики (МДКВ) содержит пять блоков: доверенный прием потока на CPU; диспетчер размещения; изолированное исполнение подписанной модели; независимый CPU-верификатор; шлюз решения. Приемник проверяет аутентичность камеры,

последовательность, время и криптографическую связь фрагментов. Шлюз разрешает действие, переводит систему в ограниченный режим либо передает событие оператору.

Для задачи j и платформы p, где p ∈ {CPU, GPU/NPU}, остаточный риск определяется как:

$$Rj(p) = \Sigma\, wi \cdot aij \cdot eip \cdot (1 - mip),$$

где wi — вес последствий угрозы i; aij — ее применимость; eip — экспозиция платформы; mip — эффективность мер. Для драйверных и межконтекстных угроз eip выше у GPU/NPU. Для тяжелой модели CPU может иметь больший риск недоступности. Платформа допустима при $Lj(p) \le Lmax$, $Bj(p) \ge Bmin$ и $Aj(p) \ge Amin$. Выбирается $p^* = \arg\min Rj(p)$ среди допустимых платформ; при близких значениях выбирается CPU.

CPU-верификатор выполняет независимые и более простые проверки: непрерывность траектории, допустимость скорости и размеров, стабильность класса, резкое изменение уверенности, живость сцены, согласование между камерами или другим сенсором. Он работает отдельным процессом и не использует GPU/NPU-драйвер. Критическое решение блокируется при расхождении либо требует подтверждения оператора.

При неподтвержденной версии драйвера, сбросах GPU/NPU, нарушении целостности модели, потере времени или аномалии результата приоритетные задачи переносятся на CPU. Для остальных снижается частота анализа или отключаются вторичные функции; запись исходного потока сохраняется. Новизна МДКВ состоит в совместном учете остаточного риска, измеренного SLA, независимой CPU-проверки и заранее определенной безопасной деградации.

Комплекс методов защиты

Камеры инвентаризируются, получают уникальные учетные данные и сертификаты. Неиспользуемые сервисы отключаются; прошивка проверяется; прямой доступ в Интернет запрещается. Изменения положения, фокуса, зон

приватности и кодирования журналируются. Физическое воздействие выявляется по резкости, яркости, статичности, сдвигу ориентиров и согласованности соседних камер.

Сегменты камер, управления, аналитики, архива и рабочих мест разделяются. Разрешается минимальный межсегментный обмен. Поток защищается взаимной аутентификацией и шифрованием; контролируются последовательность, временная метка, задержка, кодек, разрешение и частота кадров. Фрагменты связываются хеш-цепочкой или подписью. Резервирование и ограничение скорости снижают эффект DDoS.

Вычислительный узел использует secure boot, минимальную ОС, наименьшие привилегии, изоляцию декодера, read-only контейнеры, контроль целостности и централизованные обновления. Модель, предобработка, пороги и конфигурация подписываются как единый релиз. Состав компонентов и версии фиксируются для воспроизводимости расследования.

В жизненном цикле модели обеспечиваются происхождение данных, разделение ролей, версионирование, поиск выбросов и проверка предобученных весов. Перед внедрением проводятся тесты на цифровые и физические состязательные воздействия, триггеры, изменение освещения, масштаба, угла, сжатие и перекрытие. Состязательное обучение повышает устойчивость, но не дает универсальной гарантии [2; 12]. Для мультимодального генеративного модуля текст, извлеченный из кадра, рассматривается как недоверенный ввод; прямое выполнение содержащихся в нем инструкций запрещается, а потенциально опасные действия подтверждаются оператором [8]. Нужны мониторинг дрейфа и отказ от автоматического решения при низком качестве.

Архив хранится в неизменяемом или WORM-режиме с раздельными ключами, хеш-цепочками и изолированными резервными копиями. Метаданные содержат идентификаторы камеры, модели, конфигурации и платформы. Доступ к видео, биометрии и экспорту разделяется. Квоты и анализ запросов ограничивают извлечение модели [22]. Мониторинг объединяет потери кадров,

задержки, дрейф классов, сбросы GPU/NPU, ошибки памяти и смену версий. План реагирования предусматривает отзыв сертификата камеры, перенос на CPU, сохранение потока и проверку архива.

Заключение

Угрозы интеллектуальной СОЭН охватывают физическую сцену, камеру, сеть, модель, драйвер, архив и контур решений. Академические работы подтверждают практичность физических состязательных примеров, закладок, извлечения моделей, компрометации IoT-устройств и побочных каналов GPU/NPU. Защита одного уровня недостаточна.

При выполнении SLA CPU следует применять приоритетно, поскольку исключение проприетарного GPU/NPU-стека уменьшает доверенную базу. При необходимости GPU/NPU используется изолированно и не является единственным источником критического решения. МДКВ формализует этот выбор, независимую проверку и безопасную деградацию. Ограничения подхода — возможная недостаточность CPU и вероятностный характер проверок; параметры риска должны калиброваться на целевом объекте.

Библиографический список

1. Белкин Е.Н. Анализ кибератак на системы искусственного интеллекта через уязвимости в драйверах ускорителей // Дневник науки. – 2026. – № 6. – URL: https://dnevniknauki.ru/images/publications/2026/6/technics/Belkin.pdf (дата обращения: 24.06.2026).

2. Есипов Д.А., Бучаев А.Я., Керимбай А., Пузикова Я.В., Сайдумаров С.К., Сулименко Н.С., Попов И.Ю., Кармановский Н.С. Атаки на основе вредоносных возмущений на системы обработки изображений и методы защиты от них // Научно-технический вестник информационных технологий, механики и оптики. – 2023. – Т. 23. – № 4. – С. 720–733. – DOI: 10.17586/2226-1494-2023-23-4-720-733.

3. Жукова М.Н. Анализ уязвимости целевой информационной системы к MITRE ATT&CK Default Credentials технике Lateral Movement // Прикаспийский

журнал: управление и высокие технологии. – 2024. – № 4 (68). – URL: https://cyberleninka.ru/article/n/analiz-uyazvimosti-tselevoy-informatsionnoy-sistemy-k-mitre-att-ck-default-credentials-tehnike-lateral-movement (дата обращения: 24.06.2026).

4. Костогрызов А.И., Нистратов А.А. Анализ угроз злоумышленной модификации модели машинного обучения для систем с искусственным интеллектом // Вопросы кибербезопасности. – 2023. – № 5 (57). – С. 9–24. – DOI: 10.21681/2311-3456-2023-5-9-24.

5. Костюмов В.В. Обзор и систематизация атак уклонением на модели компьютерного зрения // International Journal of Open Information Technologies. – 2022. – Т. 10. – № 10. – С. 11–20.

6. Котенко И.В., Саенко И.Б., Лаута О.С., Васильев Н.А., Садовников В.Е. Атаки и методы защиты в системах машинного обучения: анализ современных исследований // Вопросы кибербезопасности. – 2024. – № 1 (59). – С. 24–37. – DOI: 10.21681/2311-3456-2024-1-24-37.

7. Лапина М.А., Ржевская Н.В., Котляров Д.В., Дюдюн Г.Д. Особенности организации атак на нейронные сети для распознавания образов // Auditorium. – 2023. – № 2 (38). – С. 97–103.

8. Мазин А.В., Белкин Е.Н. Анализ кибератак на алгоритмы искусственного интеллекта в системах видеонаблюдения и методы противодействия // Известия Института инженерной физики. – 2026. – № 1 (79). – С. 7–10.

9. Намиот Д.Е. Введение в атаки отравлением на модели машинного обучения // International Journal of Open Information Technologies. – 2023. – Т. 11. – № 3. – С. 58–68.

10. Рыжова В.А., Ярышев С.Н., Коротаев В.В. Интеллектуальные системы видеонаблюдения: учебное пособие. – СПб.: Университет ИТМО, 2021. – 107 с.

11. Стародубцев Ю.И., Закалкин П.В., Карасев С.В. Кибербезопасность систем видеонаблюдения в условиях осуществления информационно-технических

воздействий // Вопросы кибербезопасности. – 2025. – № 1 (65). – С. 136–146. – DOI: 10.21681/2311-3456-2025-1-136-146.

12. Фомичева С.Г., Беззатеев С.В. Механизмы защиты моделей машинного обучения от состязательных атак // T-Comm: Телекоммуникации и транспорт. – 2023. – Т. 17. – № 10. – С. 28–42. – DOI: 10.36724/2072-8735-2023-17-10-28-42.

13. Чехонина Е.А., Костюмов В.В. Обзор состязательных атак и методов защиты для детекторов объектов // International Journal of Open Information Technologies. – 2023. – Т. 11. – № 7. – С. 11–20.

14. Antonakakis M. et al. Understanding the Mirai Botnet // 26th USENIX Security Symposium. – 2017. – P. 1093–1110.

15. Eykholt K., Evtimov I., Fernandes E., Li B., Rahmati A., Xiao C., Prakash A., Kohno T., Song D. Robust Physical-World Attacks on Deep Learning Visual Classification // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2018. – P. 1625–1634.

16. Goodfellow I.J., Shlens J., Szegedy C. Explaining and Harnessing Adversarial Examples // 3rd International Conference on Learning Representations. – 2015.

17. Liu Y., Ma S., Aafer Y., Lee W.-C., Zhai J., Wang W., Zhang X. Trojaning Attack on Neural Networks // Network and Distributed System Security Symposium. – 2018. – DOI: 10.14722/ndss.2018.23291.

18. Naghibijouybari H., Neupane A., Qian Z., Abu-Ghazaleh N.B. Rendered Insecure: GPU Side Channel Attacks are Practical // Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. – 2018. – P. 2139–2153. – DOI: 10.1145/3243734.3243831.

19. Sharif M., Bhagavatula S., Bauer L., Reiter M.K. Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition // Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. – 2016. – P. 1528–1540. – DOI: 10.1145/2976749.2978392.

20. Shokri R., Stronati M., Song C., Shmatikov V. Membership Inference Attacks Against Machine Learning Models // 2017 IEEE Symposium on Security and Privacy. – 2017. – P. 3–18.
21. Thys S., Van Ranst W., Goedemé T. Fooling Automated Surveillance Cameras: Adversarial Patches to Attack Person Detection // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. – 2019. – P. 49–55.
22. Tramèr F., Zhang F., Juels A., Reiter M.K., Ristenpart T. Stealing Machine Learning Models via Prediction APIs // 25th USENIX Security Symposium. – 2016. – P. 601–618.
23. Wei J. et al. Leaky DNN: Stealing Deep-Learning Model Secret with GPU Context-Switching Side-Channel // 50th Annual IEEE/IFIP International Conference on Dependable Systems and Networks. – 2020. – P. 125–137. – DOI: 10.1109/DSN48063.2020.00031.