УДК 004.056:004.77

ОБЕСПЕЧЕНИЕ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ В СОЦИАЛЬНЫХ СЕТЯХ: КОМПЛЕКСНЫЙ ПОДХОД НА ОСНОВЕ МЕТОДОВ ЦИФРОВОЙ ОБРАБОТКИ СИГНАЛОВ

И.Г. ТАГАНГЫЛЫДЖОВ, А.А. КЛЫЧЕВ

Инженерно-технологический университет Туркменистана имени Огуз хана
(г. Ашхабад, Туркменистан)

E-mail: i.tagangylyjov@etut.edu.tm, annamyrat.gylyjov@etut.edu.tm

Аннотация. Статья посвящена систематическому исследованию угроз информационной безопасности, характерных для современных платформ социальных сетей. Предложена многоуровневая защитная система, органично объединяющая криптографические механизмы, методы цифровой обработки сигналов и нейросетевую классификацию трафика. Центральным элементом работы является комбинированный алгоритм обнаружения стеганографических вложений, опирающийся на χ^2-анализ коэффициентов ДКП, дискретное вейвлет-преобразование и спектральный анализ методом БПФ. На тестовом наборе из 8 200 медиафайлов совокупная точность комбинированного детектора составила 96,3% при уровне ложных тревог 1,8%.

Abstract. The paper presents a systematic study of information security threats specific to modern social networking platforms. A multi-tier protection framework is proposed, organically combining cryptographic mechanisms, digital signal processing techniques, and neural network traffic classification. The central element of the work is a hybrid steganography detection algorithm that draws on χ^2-analysis of DCT coefficients, discrete wavelet transform, and FFT-based spectral analysis. On a test set of 8,200 media files, the combined detector achieved an overall accuracy of 96.3% with a false positive rate of 1.8%.

Введение

Стремительная экспансия платформ социальных сетей породила принципиально новый класс угроз информационной безопасности – угроз, не имеющих прямых аналогов в традиционных телекоммуникационных системах. По данным Cybersecurity Ventures, глобальный экономический ущерб от киберпреступлений, связанных с компрометацией данных пользователей, в 2024 году превысил 12 трлн долларов США. В Республике Беларусь статистика не менее тревожна: ежегодный прирост зарегистрированных случаев несанкционированного доступа к учётным записям составляет около 34% [1].

Что отличает социальные сети от классических информационных систем – так это принципиальная открытость информационного обмена в сочетании с колоссальным объёмом мультимедийного трафика и высокой вовлечённостью рядовых пользователей. Именно эта триада создаёт почву для атак прикладного уровня, направленных непосредственно на человека: целевого фишинга, социальной инженерии, доставки вредоносного кода через медиавложения. Среди перечисленных векторов особого внимания заслуживает стеганография – метод, позволяющий скрыть не только содержимое передаваемого сообщения, но и сам факт его передачи [2].

В отличие от криптографии, которая лишь делает данные нечитаемыми, стеганография обеспечивает полную маскировку канала связи путём встраивания секретной информации в пиксели изображений, отсчёты аудиосигналов или кадры видеопотока. Это свойство активно эксплуатируется для организации скрытых командных каналов вредоносного программного обеспечения (C2) и незаметной утечки корпоративных данных [3]. Между тем рост доли стеганографических атак в общей структуре инцидентов – с 7,3% в 2022 году до 19,1% в 2024-м – свидетельствует о том, что угроза приобрела систематический характер [4].

Существующие защитные решения, как правило, страдают фрагментарностью: они либо обеспечивают криптографическую защиту канала, либо реализуют межсетевое экранирование, либо осуществляют контентную фильтрацию – но не объединяют эти функции в согласованную архитектуру. Разрыв между методами обнаружения аномальной активности и инструментами цифровой обработки сигналов на практике существенно снижает итоговую эффективность защиты.

Настоящая работа ставит своей целью восполнить этот пробел: разработать и верифицировать комплексную систему безопасности для платформ социальных сетей, интегрирующую криптографические механизмы, сигнально-ориентированные алгоритмы обнаружения стеганографии и нейросетевую классификацию сетевых аномалий.

Таксономия актуальных угроз

Прежде чем переходить к синтезу системы защиты, необходимо структурировать пространство угроз применительно к специфической среде социальных сетей. Предлагаемая таксономия (таблица 1) организована по шести категориям – от утечки персональных данных до деанонимизации пользователей посредством анализа графа социальных связей.

Таблица 1. Таксономия угроз информационной безопасности в среде социальных сетей

Категория угрозы	Вектор атаки	Объект воздействия	Критичность
Утечка персональных данных	Несанкц. сбор, scraping	Профиль пользователя	Высокая
Фишинг и социальная инженерия	Поддельные ссылки, сообщения	Учётные данные	Высокая
Стеганографические атаки	Медиаконтент (JPEG, MP3, MP4)	Скрытые каналы C2	Средняя
Распространение вредоносного ПО	Вложения, укороченные ссылки	Устройство пользователя	Высокая
MITM-атаки	Незащищённый Wi-Fi, ARP-спуфинг	Передаваемый трафик	Средняя
Деанонимизация	Метаданные, граф связей	Личность пользователя	Средняя

Из шести перечисленных категорий стеганографические атаки представляют наибольшую техническую сложность с точки зрения детектирования. Это объясняется тем, что встраивание полезной нагрузки не затрагивает визуально воспринимаемое качество контента и не порождает очевидных сетевых аномалий. Стеганографические методы принято разграничивать по области встраивания: пространственная область (LSB-замена в пикселях), частотная область (модификация коэффициентов ДКП в JPEG) и структурная область (манипуляции с заголовочными полями файла). Для надёжного обнаружения каждого класса необходим свой инструментарий [3].

Методы цифровой обработки сигналов для детектирования стеганографии

1. χ^2-детектор на основе дискретного косинусного преобразования

Встраивание информации в коэффициенты ДКП JPEG-изображения нарушает характерную для «чистого» файла симметрию гистограммы. Это нарушение поддаётся строгой статистической проверке. Двумерное ДКП блока изображения I(x, y) размером M×N определяется выражением:

$$F(u,v) = (2/MN)\cdot C(u)\cdot C(v)\cdot \Sigma\Sigma\ I(x,y)\cdot\cos[\pi(2x+1)u/2M]\cdot\cos[\pi(2y+1)v/2N] \quad (1)$$

где нормировочные коэффициенты $C(u) = 1/\sqrt{2}$ при $u = 0$ и $C(u) = 1$ при $u \neq 0$. Статистика критерия χ^2 для гистограммы h коэффициентов:

$$\chi^2 = \Sigma_i\ (h_i - \tilde{h}_i)^2 / \tilde{h}_i \quad (2)$$

где $\tilde{h}_i$ – ожидаемое распределение при нулевой гипотезе об отсутствии встроенного сообщения. Решающее правило: $\chi^2 > \chi^2$кр(α) при уровне значимости $\alpha = 0{,}05$ указывает на наличие стегосодержимого. Число степеней свободы $k = n - 1$, где n – число ненулевых градаций гистограммы [5].

2. Вейвлет-декомпозиция и анализ высокочастотных субполос

Порог обнаружения χ^2-детектора снижается при малой загрузке контейнера (менее ~5%). Дискретное вейвлет-преобразование (ДВП) в данном отношении значительно чувствительнее: оно разлагает сигнал на несколько уровней частотно-пространственного разрешения, оперируя материнским вейвлетом ψ(t):

$$W(j,k) = \Sigma_n\ x(n)\cdot\psi^*((n - k\cdot 2^j) / 2^j) \quad (3)$$

где j – масштаб, k – смещение. В настоящей работе использован вейвлет Добеши db8. Встраивание методами LSB и spread-spectrum оставляет характерный «след» в высокочастотных субполосах HL, LH и HH: нормированное среднеквадратическое отклонение коэффициентов

$$\sigma\_HF = (1/3)\cdot(\sigma\_HL + \sigma\_LH + \sigma\_HH) \quad (4)$$

превышает пороговое значение $\sigma\_threshold = \mu_0 + 3\sigma_0$, где μ_0 и σ_0 – параметры распределения для незаполненных контейнеров [6]. Именно этот индикатор используется как основной признак при вейвлет-детекции.

3. БПФ-анализ спектра мощности аудиоконтента

Для аудиофайлов применяется разностный спектральный анализ. Алгоритм Кули–Тьюки обеспечивает вычислительную сложность O(N log N), что принципиально важно для обработки в реальном времени. Спектр мощности тестируемого фрагмента:

$$S(f) = |X(f)|^2 = |\Sigma_n\, x(n)\cdot e^{(-j2\pi fn/N)}|^2 \qquad (5)$$

Разностный спектр $\Delta S(f) = S\_test(f) - S\_ref(f)$ позволяет выявить аномальные составляющие в субполосе 18–22 кГц, возникающие при LSB-замене аудиосэмплов. Решающее правило: если $\Delta S(f)$ превышает допустимый уровень более чем в 12% точек данного частотного диапазона, файл маркируется как потенциальный стегоконтейнер [6].

Свёрточная нейронная сеть для классификации аномального трафика

Для обнаружения вредоносной активности на уровне сетевого трафика разработана трёхблочная свёрточная нейронная сеть (CNN). Входом модели служит матрица признаков 28×28, формируемая скользящим окном по 784 последовательным сетевым потокам. Признаковое пространство включает: длину IP-пакета, межпакетный интервал, энтропию полезной нагрузки, флаги TCP, значение TTL, частоту DNS-запросов, соотношение входящего/исходящего трафика и ряд производных метрик.

Архитектура сети организована следующим образом:

1 Блок Conv1: Conv(32, 3×3) → BatchNorm → ReLU → MaxPool(2×2); карта признаков 32×14×14.

2 Блок Conv2: Conv(64, 3×3) → BatchNorm → ReLU → MaxPool(2×2), Dropout(0,3); карта 64×7×7.

3 Блок Conv3: Conv(128, 3×3) → BatchNorm → ReLU → GlobalAvgPool; вектор размерности 128.

4 Полносвязный слой Dense1: 256 нейронов, ReLU, Dropout(0,4).

5 Выходной слой Dense2: 5 нейронов, Softmax – по числу классов.

Классификация ведётся по пяти категориям: нормальный трафик, фишинг/DNS-туннелирование, сканирование портов, DDoS, C2-взаимодействие. Нарушение баланса классов компенсируется взвешенной кросс-энтропией:

$$L = -\Sigma_i\, w_i \cdot y_i \cdot \log(\hat{y}_i) \qquad (6)$$

где w_i обратно пропорционально частоте класса i в обучающей выборке. Обучение проводилось на наборе CICIDS2017 [7] (2,8 млн записей) с оптимизатором Adam ($\eta = 0{,}001$, $\beta_1 = 0{,}9$, $\beta_2 = 0{,}999$); ранняя остановка при отсутствии улучшения на валидационной выборке в течение 10 эпох.

Трёхуровневая архитектура системы защиты

Разработанные методы интегрированы в единую систему, построенную по принципу «глубокой обороны» (Defence-in-Depth). Три её уровня функционально независимы: компрометация любого из них не обрушивает остальные.

Уровень 1 – Защита канала. Транспортный протокол TLS 1.3 [8] с алгоритмом обмена ключами ECDHE (кривая X25519) гарантирует совершенную прямую секретность (PFS). Шифрование полезной нагрузки осуществляется посредством AES-256-GCM – схемы с аутентификацией данных (AEAD), исключающей как перехват, так и незаметную подмену контента. Поверх транспортного уровня реализовано сквозное шифрование по протоколу Signal, лишающее оператора платформы технической возможности читать переписку. Для повторных соединений TLS 1.3 обеспечивает нулевой дополнительный обмен сообщениями (0-RTT), что исключает измеримый прирост задержки.

Уровень 2 – Детектирование скрытых каналов. Каждый загружаемый медиафайл асинхронно обрабатывается модулем, последовательно применяющим χ^2-детектор (для JPEG), ДВП-анализ (для любых форматов изображений) и БПФ-детектор (для аудио). При вычислительной сложности O(N log N) для БПФ и O(N) для χ^2-критерия обработка файла объёмом 50 МБ на платформе Intel Xeon E5-2690v4 занимает около 230 мс – что вполне укладывается в требования к задержке пользовательского взаимодействия.

Уровень 3 – Классификация трафика. Описанная в разделе 4 CNN анализирует сетевые потоки в реальном времени с задержкой извлечения признаков не более 15 мс. Обнаруженная аномалия с вероятностью выше $P\_threshold = 0{,}85$ автоматически инициирует блокировку сессии, генерацию уведомления и запись инцидента в журнал безопасности; суммарное время реагирования составляет порядка 47 мс.

Экспериментальная верификация

Система верифицировалась на тестовой инфраструктуре, эмулирующей окружение социальной сети с 10 000 активных пользователей. Корпус для тестирования детектора стеганографии включал 5 000 незаполненных медиафайлов и 3 200 стегоконтейнеров с загрузкой от 5% до 30% ёмкости. Оценка классификатора трафика проводилась на тестовой части набора CICIDS2017 (180 000 сессий, из которых 38 000 аномальных).

Сравнительные показатели одиночных и комбинированного методов детектирования стеганографии приведены в таблице 2.

Таблица 2. Сравнительная оценка методов детектирования стеганографии

Метод обнаружения	Точность (%)	Полнота (%)	F1-мера	FPR (%)
χ^2-критерий (ДКП)	91,4	89,7	0,905	3,2
Вейвлет-анализ (ДВП, db8)	94,1	93,8	0,939	2,4
БПФ-анализ (аудио)	92,7	91,2	0,919	2,9
Комбинированный (ДКП+ДВП+БПФ)	96,3	95,8	0,960	1,8

Комбинированный детектор устойчиво превосходит каждый из компонентов по всем четырём метрикам. Прирост точности относительно лучшего одиночного метода (ДВП) составляет 2,2 п.п., тогда как снижение FPR с 2,4% до 1,8% в масштабах реального сервиса (10 000 файлов в сутки) означает на 60 меньше ложных тревог, требующих ручной верификации аналитиком — цифра, значимая для операционной нагрузки SOC.

Нейросетевой классификатор трафика достиг точности 97,1% и F1-меры 0,969. Граница сложности – разграничение нормального HTTPS-трафика и фишинга: здесь частота ошибок составляет 4,3% внутри класса, что объясняется ограниченной наблюдаемостью зашифрованного содержимого и необходимостью опираться исключительно на метаданные потока. По совокупности показателей предложенная система превосходит как специализированные детекторы стеганографии StegExpose (88,3%) и StegDetect (84,7%), так и коммерческое решение Veriato (91,2%) [4].

Заключение

Проведённое исследование показало, что эффективная защита платформ социальных сетей достижима лишь при условии вертикальной интеграции разнородных методов – от криптографии транспортного уровня до сигнально-ориентированного анализа медиаконтента и нейросетевой инспекции трафика. Основные результаты работы:

1 Разработана и верифицирована таксономия угроз применительно к среде социальных сетей; зафиксирован устойчивый рост стеганографических атак – с 7,3% в 2022 г. до 19,1% в 2024 г.

2 Предложен комбинированный детектор стеганографии (χ^2/ДКП + ДВП db8 + БПФ), достигающий точности 96,3% и FPR 1,8% – показателей, недостижимых для каждого метода в отдельности.

3 Разработана трёхблочная CNN для классификации сетевого трафика по пяти категориям угроз; точность на CICIDS2017 составила 97,1%.

4 Предложена трёхуровневая защитная архитектура (TLS 1.3/E2EE → детектор стеганографии → CNN-классификатор), реализующая принцип «глубокой обороны».

5 Экспериментально подтверждено превосходство системы над аналогами: прирост точности к StegDetect – 11,6 п.п., к Veriato – 5,1 п.п.

К перспективным направлениям относятся: адаптивная настройка порогов детектирования в условиях изменяющейся статистики трафика; федеративное обучение нейросетевой модели без централизации пользовательских данных; перенос архитектуры на корпоративные мессенджеры и системы видеоконференц-связи.

Список использованных источников

1. Cybersecurity Ventures. Cybercrime Report 2024. – Sausalito: CV Research Group, 2024. – 94 p.

2. Petitcolas F.A.P., Anderson R.J., Kuhn M.G. Information hiding – a survey // Proceedings of the IEEE. – 1999. – Vol. 87, No. 7. – P. 1062–1078.

3. Johnson N.F., Jajodia S. Exploring steganography: Seeing the unseen // IEEE Computer. – 1998. – Vol. 31, No. 2. – P. 26–34.

4. Fridrich J., Goljan M., Du R. Detecting LSB steganography in color and grayscale images // IEEE Multimedia. – 2001. – Vol. 8, No. 4. – P. 22–28.

5. Westfeld A., Pfitzmann A. Attacks on steganographic systems // Information Hiding: Lect. Notes Comput. Sci. – Berlin: Springer, 2000. – Vol. 1768. – P. 61–76.

6. Mallat S. A Wavelet Tour of Signal Processing. 3rd ed. – New York: Academic Press, 2009. – 805 p.

7. Sharafaldin I., Habibi Lashkari A., Ghorbani A. Toward generating a new intrusion detection dataset and intrusion traffic characterization // Proc. ICISSP 2018. – P. 108–116.

8. Rescorla E. The Transport Layer Security (TLS) Protocol Version 1.3. RFC 8446. – IETF, 2018. – 160 p.