УДК 004.056

ПОДХОД К ПОВЕДЕНЧЕСКОЙ КЛАССИФИКАЦИИ ЗАШИФРОВАННОГО TLS-ТРАФИКА БЕЗ РАСШИФРОВАНИЯ В ЗАДАЧАХ МОНИТОРИНГА БЕЗОПАСНОСТИ

Кормаков В.Р.[1]
Научный руководитель – Руссу В.Ю.[1]
[1]Военно-космическая академия им.А.Ф.Можайского
vka@mil.ru

Введение

Сегодня более 90 % интернет-трафика передаётся в зашифрованном виде с использованием протокола TLS. Причём в версии TLS 1.3 значительная часть служебной информации также скрыта шифрованием. В TLS 1.3 после ServerHello почти все сообщения handshake зашифрованы. Это серьёзно ограничивает применение традиционных методов Deep Packet Inspection (DPI), которые опираются на анализ содержимого пакетов.

Одновременно с этим действующие нормативные требования Российской Федерации в сфере информационной безопасности обязывают осуществлять мониторинг сетевого взаимодействия и фиксировать события безопасности [1, 2]. Возникает противоречие — контроль необходим, но содержимое трафика анализировать нельзя без применения MITM-механизмов, что в ряде случаев может не соответствовать действующим нормативным требованиям.

В таких условиях особенно актуальной становится разработка методов, позволяющих классифицировать зашифрованный трафик без его расшифрования — исключительно на основе метаданных и поведенческих характеристик соединений.

Основная часть

Анализ существующих сигнатурных DPI-подходов показал их зависимость от неизменности параметров соединения. В научных исследованиях также активно рассматриваются методы классификации зашифрованного трафика на основе машинного обучения [3]. Малейшие модификации Server Name Indication (SNI) или других идентификаторов заметно снижают эффективность таких методов. В условиях TLS 1.3, где полезная нагрузка полностью недоступна для анализа, их результативность падает ещё сильнее.

В рамках работы предложен поведенческий подход классификации TLS-трафика. Его основная идея проста: различные приложения формируют характерные, устойчивые профили сетевого поведения. Подобный подход ранее применялся для идентификации HTTPS-сервисов на основе анализа характеристик соединений [4].

Из каждого TLS-соединения извлекались признаки, не требующие расшифрования:
— версия TLS;
— набор «cipher suites»;
— JA3-отпечаток клиента;
— наличие и длина SNI;
— размеры первых десяти пакетов;
— направление передачи данных;
— продолжительность соединения;
— общий объём переданной информации;
— статистические параметры межпакетных интервалов.

На основе этих параметров формировался вектор признаков соединения. В качестве алгоритма классификации использовался Random Forest из 100 деревьев

решений. Для сравнения применялся сигнатурный метод, основанный на анализе SNI и JA3.

Эксперимент проводился на тестовой выборке из 5 200 TLS-соединений. В неё вошли TLS-соединения, относящиеся к трафику веб-сервисов, облачных хранилищ, потокового видео, мессенджеров и фоновой сетевой активности. Данные были разделены на обучающую (70 %) и тестовую (30 %) части.

Сигнатурный подход обеспечил точность 74,8 %, тогда как предложенная поведенческая модель достигла 92,4 % при значении F1-score 0,92. Более того, при изменении SNI точность сигнатурного метода снизилась на 18 %, в то время как поведенческая модель потеряла не более 4 %. Это демонстрирует её робастность к вариациям параметров соединения.

Таким образом, предложенный метод позволяет эффективно классифицировать зашифрованный TLS-трафик без анализа его содержимого.

Выводы

В работе предложен и экспериментально подтверждён поведенческий метод классификации TLS-трафика без расшифрования полезной нагрузки. Достигнутая точность 92,4 % существенно превышает показатели сигнатурного подхода.

Результаты показывают, что поведенческий анализ может применяться в системах мониторинга сети для предотвращения угроз безопасности в условиях повсеместного шифрования. Метод может применяться в системах обнаружения аномалий и регистрации событий безопасности без нарушения конфиденциальности передаваемых данных

Дальнейшие исследования целесообразно направить на расширение набора признаков, а также на адаптацию метода для анализа протокола QUIC (Quick UDP Internet Connections).

Литература

1. Приказ ФСТЭК России от 18.02.2013 № 21 «Об утверждении состава и содержания организационных и технических мер по обеспечению безопасности персональных данных при их обработке в информационных системах персональных данных».
2. Приказ ФСТЭК России от 25.12.2017 № 239 «Об утверждении требований по обеспечению безопасности значимых объектов критической информационной инфраструктуры Российской Федерации».
3. Anderson B., McGrew D. Machine Learning for Encrypted Traffic Classification: An Overview // IEEE Communications Surveys & Tutorials. 2017. Vol. 19, no. 4. P. 246–267.
4. Shbair W., Cholez T., Francois J., State R. A multi-level framework to identify HTTPS services // IEEE International Conference on Communications. 2016. P. 1–6.