МЕТОД АВТОМАТИЧЕСКОЙ КЛАССИФИКАЦИИ ЦИФРОВЫХ ОТПЕЧАТКОВ TLS-ПРОТОКОЛА

Ишкуватов С. М.[1], Бегаев А. Н.[2], Комаров И. И.[3]

DOI: 10.21681/2311-3456-2024-1-67-74

Цель исследования: *разработка метода классификации цифровых отпечатков TLS-протокола, обеспечивающего их автоматическое соотнесение с одной из известных групп или принятие решения об обнаружении новой реализации протокола.*

Методы исследования *базируются на положениях теории топологии, теории автоматов, теории множеств; использовании методов автоматической кластеризации, проведения натурного эксперимента и обработки экспериментальных данных.*

Результат: *мониторинг трафика на границе контролируемой зоны телекоммуникационной сети является основной составляющей обеспечения кибербезопасности. Одним из традиционных подходов для решения этой задачи является использование цифровых отпечатков как устройств, так и программной реализации телекоммуникационных протоколов.*

Несмотря на богатую историю развития методов автоматического определения реализации протокола на основе анализа цифровых отпечатков, эта задача в полной мере ещё не решена ввиду изменчивости как самих протоколов, так и телекоммуникационной инфраструктуры, определяющих вариативность конечных формы и значения соответствующего цифрового отпечатка.

В работе предлагается метод автоматической классификации цифровых отпечатков TLS-протокола, базирующийся на формальной оценке близости вариативных форм представления цифровых отпечатков и устойчивый к модификации их значений; приводятся данные по степени влияния значения порога близости на ошибки первого и второго рода в процессе кластеризации.

Полученные результаты в первую очередь ориентированы на применение в системах мониторинга трафика, но могут быть использованы и для решения других задач кибербезопасности.

Научная новизна *результатов определяется совокупностью авторских решений, связанных с обоснованием, введением и применением метрики для оценки близости цифровых отпечатков телекоммуникационных протоколов, устойчивой к модификации цифровых отпечатков клиентской реализации TLS-протокола, а также доказательным подтверждением реализуемости и получением значений показателей качества функционирования метода автоматической классификации цифровых отпечатков реализаций протокола TLS, применённого к известным базам данных цифровых отпечатков.*

Вклад авторов: *Ишкуватов С. М. – разработка метода автоматической классификации цифровых отпечатков протоколов, подготовка исходных данных, проведение эксперимента и визуализация результатов; Бегаев А. Н. – анализ опыта и перспективных сценариев применения периметровых систем мониторинга трафика, определение требований и ограничений исследования; Комаров И. И. – определение научно-методического аппарата и подходов к оценке близости цифровых отпечатков, разработка плана исследования.*

Ключевые слова: *кибербезопасность, мониторинг трафика, коммуникационный протокол, инцидент информационной безопасности, модель пассивного наблюдателя, вектор информативных признаков, мера близости, кластеризация, показатели качества.*

1 Ишкуватов Сергей Маратович, аспирант факультета безопасности информационных технологий, Университет ИТМО, Санкт-Петербург, Россия. E-mail: sysroot0@gmail.com

2 Бегаев Алексей Николаевич, кандидат технических наук, генеральный директор АО «Эшелон – Северо-Запад», Санкт-Петербург, Россия. E-mail: begaev@mail.ru

3 Комаров Игорь Иванович, кандидат физико-математических наук, доцент, доцент факультета безопасности информационных технологий, Университет ИТМО, Санкт-Петербург, Россия. E-mail: i_krov@mail.ru

THE AUTOMATIC METHOD OF TLS PROTOCOL DIGITAL FINGERPRINTS CLASSIFICATION

Ishkuvatov S. M.[4], Begaev A. N.[5], Komarov I. I.[6]

The purpose of the study *is to develop a method for classifying digital fingerprints of the TLS protocol, ensuring their automatic correlation with one of the known groups or making a decision on the discovery of a new protocol implementation.*

The research methods *are based on the principles of topology theory, automata theory, set theory, the use of automatic clustering methods, full-scale experiments, and experimental data processing.*

Results: *traffic monitoring of the telecommunications network-controlled zone border is a key component of ensuring cyber security. One of the traditional approaches to solving this problem is using of digital fingerprints (DF) of both devices and software implementation of telecommunication protocols. Despite the rich history of development automatically determining the protocol implementation methods based on the analysis of DF, this task has not yet been fully solved due to the variability of both the protocols themselves and the telecommunications infrastructure, which determine the variability of the corresponding DF final shape and value.*

The paper proposes an automatic the TLS protocol's DF classification method, based on a formal proximity assessment of the variable forms of DF and resistant to their values modification; data on the influence degree of the proximity threshold value on first and second kind errors in the clustering process are presented.

The results obtained are primarily focused on application in traffic monitoring systems but can also be used to solve other cybersecurity tasks.

Scientific novelty *is determined by a set of author's solutions related to the justification, introduction and application the telecommunication protocols DF proximity assessment metrics that are resistant to the TLS protocol's client implementations modifications, as well as evidence-based confirmation of the feasibility and obtaining the quality indicators values of the automatic classification DF TLS protocol's implementations method functioning applied to known DF databases.*

Keywords: *cybersecurity, informative features vector, digital fingerprint, proximity measure, clustering, digital fingerprint database.*

Введение

Мониторинг и глубокий анализ сетевого трафика на границах сетей является важной составляющей обеспечения кибербезопасности, позволяющий выявлять факты использования небезопасных протоколов, сетевые атаки и обнаруживать иные проблемы информационной безопасности (ИБ).

Одним из традиционных механизмов, используемых системами обнаружения вторжений, является фильтрование трафика по шаблонам – заранее подготовленному списку правил. Однако этот подход опирается на ретроспективный анализ и не позволяет выявлять не описанные ранее угрозы.

Сложность задачи *аналитического* анализа сетевого трафика определяется его имманентной изменчивостью и трудностью формальной интерпретации корректности, связанной как с естественным изменением конфигурации программно-аппаратных средств в процессе развития информационной системы, так и с целенаправленным противодействиям сетевым угрозам [1, 2].

Более того влиятельные транснациональные игроки телекоммуникационной отрасли[7] предпринимают специальные усилия по «запутыванию» протоколов для противодействия национальным цензурам или нежелательным для них способам использования ресурсов.

Промежуточным направлением между «шаблонным» и аналитическим анализом протоколов является подход, основанный на анализе цифровых отпечатков (ЦО), под которыми понимается набор параметров, характеризующий тот или иной протокол, а также позволяющий строить гипотезы относительно реализаций конкретного протокола. Некоторые подходы, связанные с использованием ЦО не только

4 Sergei M. Ishkuvatov, Ph.D. student, Faculty of Information Technology Security, ITMO University, St. Petersburg, Russia. E-mail: hieule250715@gmail.com
5 Alexey N. Begaev, Ph.D., CEO of JSC North-West Echelon, St. Petersburg, Russia. E-mail: begaev@mail.ru
6 Igor I. Komarov, Ph.D., (in Maht.), Associate Professor, Faculty of Information Technology Security, ITMO University, St. Petersburg, Russia. E-mail: i_krov@mail.ru
7 Больше протоколов для шифрования DNSзапросов. — URL: https:// vasexperts.ru/blog/tehnologii/bolsheprotokolovdlyashifrovaniyadnszaprosov/ (дата обращения: 10.10.2023).

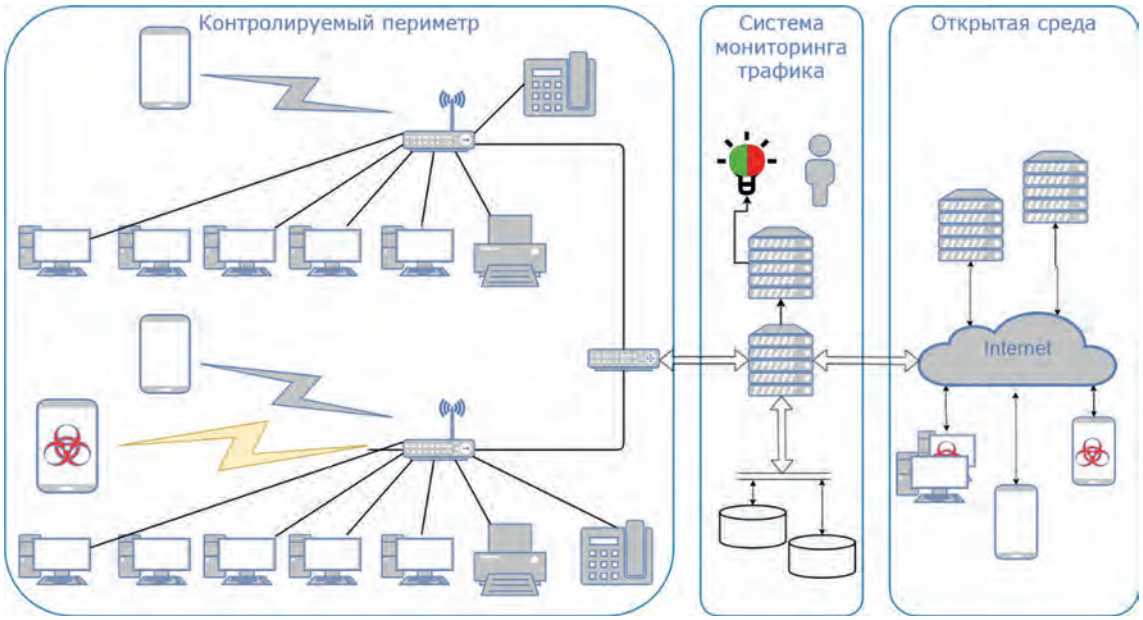


Рис. 1. Расположение систем поведенческого анализа сетевого трафика

программных реализаций, но устройств в целом, доведены[8] до корпоративных стандартов.

В работе предлагается метод, обеспечивающий автоматическую классификацию ЦО программной реализации TLS-протокола с использованием модели пассивного наблюдателя.

Постановка задачи и ограничения

Исследование предполагает (рис. 1) наличие корпоративной защищаемой инфраструктуры, которая находится внутри контролируемого периметра, но вынуждена взаимодействовать с внешней неконтролируемой средой. Взаимодействие осуществляется через систему мониторинга трафика, к которой применена модель «пассивного наблюдателя». Задача системы мониторинга – информирование администратора сети об обнаруженных инцидентах таких как появление новых абонентов/видов сетевых активностей или сокрытия фактов использования запрещённых организацией протоколов. Решение этой задачи осложняется:

- Распространением технологий сокрытия DNS запросов, таких как DNS over HTTPS, DNS over TLS, DNS over QUIC, DNS over CoAP, SecureDNS.
- Проблемами типа DNS leaks[9] – когда запрос локального узла к локальному ресурсу ошибочно перенаправляется глобальному DNS, раскрывая внутреннюю архитектуру сети на всем пути следования пакета [3].
- Расширением применения шифрованного TLS-рукопожатия Encrypted Client Hello (ECH), делающим невозможным[10] определения конечной точку TLS-сессии по значению поля ServerName из запроса клиента и проверки цепочки сертификатов, предоставляемых сервером.

В результате система мониторинга трафика сталкивается с рядом сложностей, в том числе, нарушающих функционирование прикладной системы:

- TLS-сессии станут практически неотличимы друг от друга традиционными методами;
- станет невозможным Sinkholing[11] – получение информации о заражении перенаправлением вредоносного трафика на сервер исследователя;
- ограниченное использование баз данных (БД) ЦО, например на основании дефакто-стандарта JA3[12];
- системы контент фильтрации смогут блокировать ресурсы только в случае явного обращения к запрещённому ресурсу по IPv4 адресу;
- невозможность выборочной блокировки ресурса без блокировки всех ресурсов, использующих эту сеть доставки контента CDN[13].

Преодоление указанных сложностей системы анализа трафика может реализовываться в следующих направлениях:

- поиск новых *информативных признаков* и *подходов к описанию* информативных признаков

8 СТО БР БФБО-1.7-2023 Стандарт Банка России «Безопасность финансовых (банковских) операций. Обеспечение безопасности финансовых сервисов с использованием технологии цифровых отпечатков устройств (принят и введён в действие приказом Банка России от 01.03.2023 N ОД-335)

9 Imana Basileal, Korolova Aleksandra, Heidemann John. Enumerating privacy leaks in DNS data collected above the recursive // NDSS: DNS Privacy Workshop. — 2018.

10 Encrypted Client Hello (ECH): часто задаваемые вопросы. — URL: https://support.mozilla.org/ru/kb/faq-encrypted-client-hello (дата обращения: 20.05. 2023).

11 Sinkholing — URL: https://encyclopedia.kaspersky.ru/glossary/sinkholing/ (дата обращения: 10.10.2023).

12 JA3 — A method for profiling SSL/TLS Clients [Электронный ресурс]. URL: https://github.com/salesforce/ja3 (дата обращения: 19.07.2020).13

13 Peng Gang. CDN: Content distribution network // arXiv preprint cs/0411069. — 2004.

реализаций протоколов для учёта их вариативности и постоянной мимикрии угроз;
- совместный анализ информативных признаков *разных протоколов* и на разных уровнях модели OSI [4];
- поиск *статистических закономерностей* вредоносного трафика, которые могут быть выявлены пассивным наблюдателем даже в случае использования сторонами шифрования.

Предпосылки исследования

Исторически сложилось два основных направления получения данных об устройствах и протоколах сети: активный и пассивный [5]. Несмотря на более широкие возможности активных методов, их применение не всегда возможно.

Исследования, посвящённые выделению и описанию информативных признаков, процедуре *пассивного* получения ЦО[14], а также его использования, в том числе в контексте HTTPS трафика[15], получали развитие по мере развития телекоммуникационных систем.

Результаты, полученные в работах [6 – 7] определили возможность применения ЦО для выявления угроз ИБ и легли в основу механизмов идентификации реализаций TLS-протокола, развиваемых проектами JA3 и JA3S, а также Cisco Mercury. В настоящее время в открытой БД ЦО проекта JA3 используется две формы: исходная полная запись признаков и *md5*-хеш этой полной формы.

Отдельную группу составляют работы, посвящённые проблеме классификации трафика, передаваемого в шифрованной сети [8–13]. Решения демонстрируют хорошие результаты по определению типа трафика на основании анализа нормализованных по времени и размерам распределений длин пакетов, как для отдельных TLS-сессий, так и всего канала VPN.

В работе[16] предлагаются обзор перспективных подходов, в том числе не ограниченных признаками TLS-рукопожатия, таких как цепи Маркова, описывающие сетевое взаимодействие сторон.

Анализ рынка систем анализа трафика позволяет выделить несколько ключевых проектов, характеризующих достигнутый практический уровень.

Характерным представителями систем, базирующихся на проекте JA3/JA3S является продукты Wireshark и Suricata[17] , использующие, в том числе, модели [14, 15].

Известна отечественная система анализа трафика для выявления атак PT Network Attack Discovery[18]. В контексте исследования особый интерес представляет библиотека OsDetectLib[19], которая по описанию разработчиков занимается определением операционных систем TCP-сессий. В репозитории на Github разработчик публикует открытую часть правил детектирования различных видов атак в формате Suricata[20]. Формат правил Suricata также имеет функционал автоматического получения ЦО TLS реализаций в формате JA3/JA3S и возможность описания ЦО TCP/IP. Однако правила, содержащие ЦО различных уровней, являются строгими и не предполагают оценок возможной близости, кроме того, в части случаев ЦО задаётся MD5-хешем, что хоть и делает правила более компактными, препятствует любой проверки на соответствие, кроме строгой.

Способ решения задачи и анализ полученных результатов

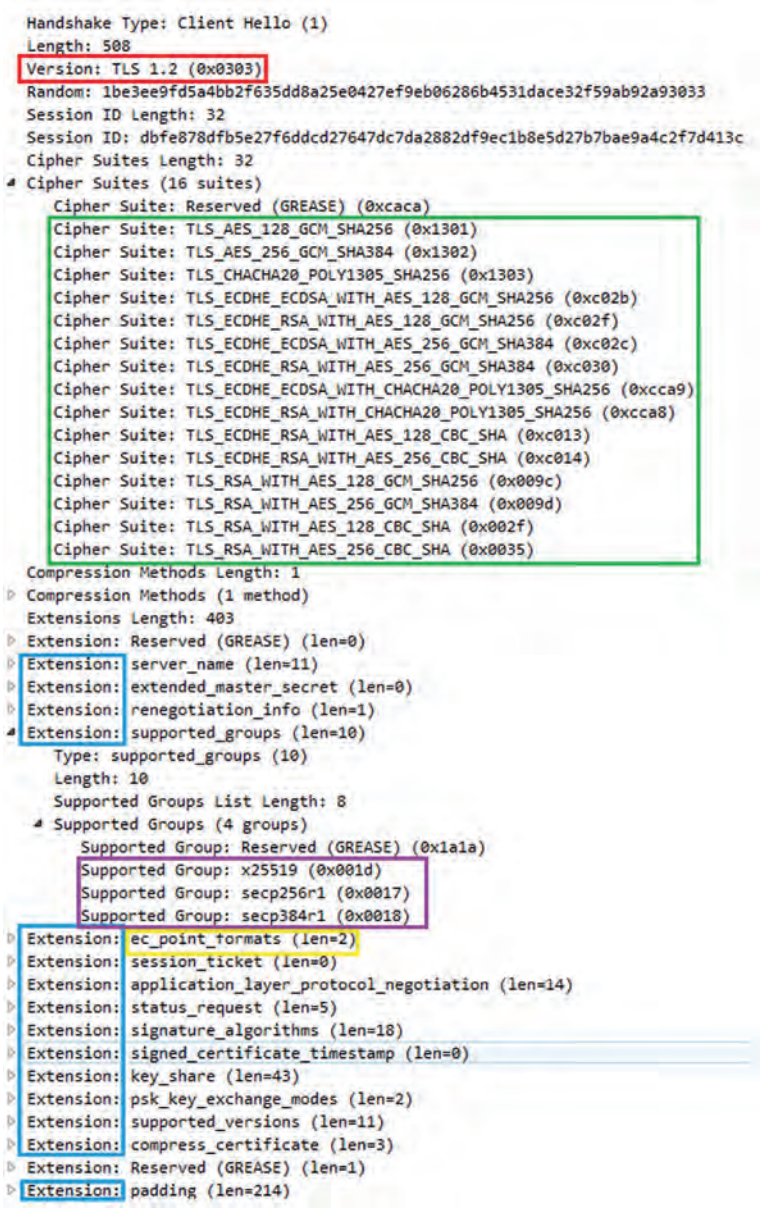
Для решения задачи автоматической классификации ЦО TLS-протоколов должны быть решены следующие частные задачи:
- определена номенклатура информативных признаков, доступных пассивному наблюдателю;
- выбрана единая (псевдоканоническая) форма записи наблюдаемых признаков;
- введена метрика близости ЦО, обеспечивающая формальную оценку расстояния ЦО в многомерном пространстве признаков;
- выполнена подготовка БД ЦО для использования в автоматическом режиме;
- предложен алгоритм кластеризации ЦО.

Под термином ЦО реализации TLS-протокола понимаются параметры, характеризующие именно эту *конкретную реализацию* протокола, именно *конкретную версию библиотеки*, реализующий этот протокол или *группу возможных версий*.

Для решения задач формирования ЦО клиентской реализации TLS-протокола наибольшее распространение в настоящее время получил алгоритм JA3 [16], интересны модели использования альтернативных баз представления признаков – проекта Mercury Cisco[21,22], [17, 18, 19] и LeeBrotherston[23] [20].

Критерием выбора информативных признаков TLS-протокола для использования в работе (рис. 2) являются: 1) возможность определения пассивным

14 Shu G., Lee D. Network protocol system fingerprinting a formal approach // Proceedings IEEE INFOCOM 2006. 25TH IEEE International Conference on Computer Communications. — IEEE, 2006. — Pp. 1–12.
15 HTTPS traffic analysis and client identification using passive SSL/TLS fingerprinting / Martin Husák, Milan Čermák, Tomáš Jirsík, Pavel Čeleda // EURASIP Journal on Information Security. 2016. Vol. 2016. Pp. 1–14.
16 Gancheva Z., Sattler P., Wüstrich L. TLS Fingerprinting Techniques // Network. — 2020. — URL: https://www.net.in.tum.de/fileadmin/TUM/NET/NET-2020-04-1/ NET-2020-04-1_04.pdf (online; accessed: 20.05.2023).
17 Suricata. Observe. Protect. Adapt. — URL: https://suricata.io/ (online; accessed: 10.10.2023).
18 PT Network Attack Discovery. — URL: https://ptsecurity.com/ru-ru/products/network-attack-discovery/ (дата обращения: 20.05.2023).
19 Результаты анализа трафика в 41 компании и новые возможности PT NAD. — URL: https://www.ptsecurity.com/upload/corporate/ru-ru/webinars/ics/PT_NAD_ 18_03.pdf (дата обращения: 20.05.2023).
20 Suricata PT Open Ruleset. — URL: https://github.com/ptresearch/AttackDetection (дата обращения: 20.05.2023).
21 Mercury: network fingerprinting and packet metadata capture - URL:https://github.com/cisco/mercury. (online; accessed: 21.10.2023).
22 Lee brotherston's work – URL: https://github.com/synackpse/tls-fingerprinting (online; accessed: 21.10.2023).
23 Brotherston Lee. Lee brotherston's work. — URL: https://github.com/synackpse/tls-fingerprinting (online; accessed: 20.05.2023).



а) Информативные признаки пакета TLS Client Hello, используемые для формирования ЦО

```
[JA3 Fullstring: 771,4865-4866-4867-49195-49199-49196-49200-52393-52392-49171-49172-156-157-47-53,0-23-65281-10-11-35-16-5-13-18-51-45-43-27-21,29-23-24,0]
[JA3: b32309a26951912be7dba376398abc3b]
```

б) Полный и хешированный ЦО в формате JA3, полученные из данных рис. 2.а)

Рис.2. – Формирования ЦО пакета TLS Client Hello

наблюдателем и 2) возможность их получения из распространённых БД ЦО.

Перечень информативных признаков включает:

- номер используемой версии протокола TLS – целое число (выделено красным);
- список поддерживаемых клиентской реализации алгоритмов шифрования Cipher Suites – последовательность 2-байтных символов (выделено зелёным);
- список опциональных параметров TLS Extensions последовательность 2-байтных символов (выделено голубым);
- хронология следования параметров TLS Extensions (формируется динамически);
- EC point formats (выделено жёлтым) – в случае, если этот тип поля присутствует только в одном ЦО, принимать расстояние равным 2;
- список Elliptic Curves – в случае, если этот тип поля присутствует только в одном ЦО, принимать расстояние равным 2.

Несмотря на возможность произвольной авторской модификации возможных форматов хранения ЦО, открывающих новые перспективы автоматиче-

ской обработки, в качестве *псевдоканонической формы записи* ЦО для выбран *полный формат JA3*. Выбор определяется его широким использованием в профессиональном сообществе, активным пополнением базы и поддержкой значительным числом программных продуктов, что упрощает вывод в практическое применение полученных результатов.

Введение метрики для оценки близости двух ЦО протокола предполагает определение метрического пространства [21] и способов обработки каждого из компонентов вектора признаков. Независимо от протокола все возможные признаки, описывающие ЦО можно разделить на следующие типы:

- флаги – булевские атрибуты, характеризующие наличие или отсутствие определённого признака у характеризуемой им реализации протокола;
- константное числовое значение;
- диапазон значений – применим для числовых значений, может определяться формулой, списком или границами интервалов значений;
- последовательность – обычно список мнемонических обозначений параметров с имеющей значение хронологией следования элементов друг за другом – для описания порядка следования и состав опциональных параметров.

В работе [21] предложен метод количественной оценки отличий $\Delta i(a,b)$ каждого из i значений компонента векторов признаков $A=<a_1, a_2...a_n>$ и $B=<b_1, b_2...bn>$. Она может определяться:

- если значения всех типов признаков a и b совпадают $\Delta i(a,b) = \emptyset$;
- в противном случае:
- для числовых констант – как абсолютное значение разности a и b: $\Delta i(a,b) = |(a - b)|$;
- для диапазонов – размер диапазона, образованного пересечением диапазонов $\Delta i(a,b) = M|a_i \cap b_i|$, (где M – мощность множества), возможно с модификацией: $\Delta i(a,b)=\frac{M|a_i \cap b_i|}{M|a_i \cup b_i|}$, обеспечивающей учёт относительной мощности сравниваемых множеств.
- для последовательностей – количественная оценка совпадения состава и порядка следования мнемоник, полученная как расстояние Левенштейна. Алгоритм определения расстояния Левенштейна lev(a,b) для последовательностей a и b с длинами $|a|$ и $|b|$ определяется (1), где $tail$ некоторой строки x – это строка всех символов x, кроме первого, а $x[n]$ – n-ный символ строки x, начиная $\emptyset$.

Доказана применимость настоящего подхода [21], основанная на том, что для подавляющего числа протоколов добавление или удаление параметра из списка, как правило, не приводит к нарушению общей хронологии следования параметров.

$$lev(a,b) = \begin{cases} |a| & \text{, если } |b| = 0 \\ |b| & \text{, если } |a| = 0 \\ lev(tail(a),tail(b)) & \text{, если } a|0| = b|0| \\ 1 + min \begin{cases} lev(tail(a),b) \\ lev(a,tail(b)) \\ lev(tail(a),tail(b)) \end{cases} & \text{, в остальных случаях} \end{cases} \quad (1)$$

Общим расстоянием между двумя отпечатками $LEV(A,B)$ следует считать сумму всех минимальных расстояний Левенштейна всех компонентов вектора признаков:

$$LEV(A,B) = \sum k \cdot \min(lev_i(a_i, b_i)), \quad (2)$$

где i – индекс вектора гиперпространства, в котором вычисляется (1) расстояние Левенштейна $lev_i(a_i, b_i)$; k – весовой коэффициент (коэффициент значимости) значения расстояния по вектору i.

Определение пространства информативных признаков и введение формальной метрики оценки близости ЦО предоставляет возможность автоматической классификации (рис. 3) всех известных наборов ЦО.

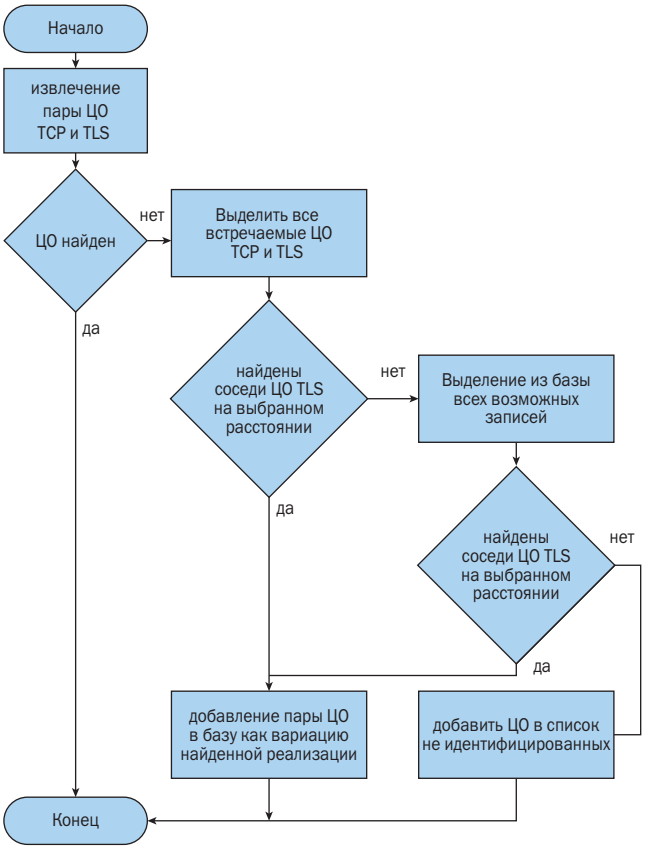


Рис.3. – Алгоритм автоматической классификации ЦО TLS-протокола

В ходе автоматической классификации ЦО реализаций протокола обнаружено явление *смещения центра кластеров родственных реализаций* по мере появления новых версий, как за счёт новых реализаций, так и за счёт использования новых библиотек. Этот факт должен учитываться при реализации подсистем ИБ, базирующихся на предлагаемом методе.

Таблица 1

Показатели качества функционирования метода автоматической классификации

Представление ЦО TLS: ***Источник*** (Полная запись ЦО) *Хеш-значение*	max расстояние близости	Корректно найденные соседи	Ошибки I рода (ложные соседи)	Ошибки II рода (элементы, ошибочно не включённые в класс)
Android Webkit (771,49195-49196-49199-49200-158-159-49161-49162-49171-49172-51-57-50-56-49159-49169-156-157-47-53-5-255,0-11-10-13,14-13-25-11-12-24-9-10-22-23-8-6-7--20-21-4-5-18-19-1-2-3-15-16-17,0-1-2) *f898478e132de326106e9eb8e861c1a2*	6	11	0	443
	20	16	1	438
	30	74	6	380
	50	85	325	369
Tor (769,49162-49172-136-135-57-56-49167-49157-132-53-49159-49161-49169-49171-69-68-51-50-49164-49166-49154-49156-150-65-4-5-47-49160-49170-22-19-49165-49155-65279-10-255,0-11-10,1-2-3-4-5-6-7-8-9-10-11-12-13-14-15-16-17-18-19-20-21-22-23-24-25,0-1-2) *581a3c7f54555512b8cd16e87dfe165b*	6	0	1	10
	20	1	1	9
	30	3	5	7
Kaspersky (771,4866-4867-4865-49200-49199-49192-49191-49196-49195-49188-49187-52392-52394-103-107-159-255,0-11-10-35-5-16-22-23-49-13-43-45-51-21,23-24-29,0-1-2) *aa63ca1ce311b0ff100de506d4d9b3ab*	6	20	0	19
	20	23	0	16
	30	24	135	15

Эксперимент и анализ полученных результатов

Эксперименты по автоматической классификация ЦО TLS-протокола проведены в два этапа.

Первый этап предполагает использование классических методов кластеризации для графического представления и визуальной интерпретации результатов. На рис. 4 представлен фрагмент дендрограммы,

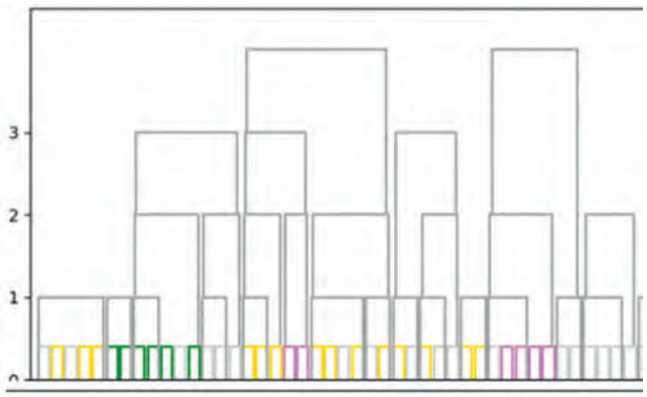


Рис.4. – Фрагмент дендрограммы кластеризации БД ЦО Cisco протокола TLS

где одинаковыми цветами обозначены родственные реализации протоколов, серым цветом обозначены все редко встречающиеся реализации. Такое представление позволяет визуально оценить корректность выбора порогового значения по размерам графов и количествам попаданий элементов разного цвета в один граф. Видно, что на представленном фрагменте наибольшее расстояние $LEV(A,B)$, объединяющее все известные реализации, равно 5.

Второй этап эксперимента нацелен на определение влияния порогового значения близости ЦО. Очевидно, что выбор высокого значения ведёт к увеличению ошибок первого рода (ложное принятие схожести ЦО), что в дальнейшем приведёт к слиянию разных групп ЦО и невозможности их разделения.

Снижение значения порога близости, хоть и ведёт к увеличению ошибок второго рода (ложное предположение о непохожести ЦО), однако вред таких ошибок менее значим, так как он ведёт к дроблению групп ЦО на более мелкие группы, которые при необходимости могут быть объединены на последующих этапах.

По результатам вычислительного эксперимента на текущей[24] БД ЦО Cisco, приведённой в формат JA3 и дополненной информацией из открытых баз[25] [24] получены результаты, представленные в таблице 1.

Полученные зависимости позволяют производить тонкую настройку прикладных систем с учётом степени важности ошибок первого и второго рода, например на основе рисковых моделей, например [22].

Выводы

В работе поставлена и решена задача разработки метода автоматической классификации цифровых отпечатков TLS-протокола. Предлагаемый метод базируется на совокупности ранее полученных результатов, связанных с обоснованием и выбором информативных признаков, введением метрики для оценки близости ЦО протоколов, обработкой открытых БД ЦО и использованием методов кластеризации данных.

Теоретические результаты подтверждены экспериментальным исследованием, в том числе, определяющим степень влияния порога близости ЦО на результат отнесения исследуемого ЦО к известным или новым кластерам.

Предлагаемые результаты ориентированы на применение в системах периметрового мониторинга трафика с использованием модели пассивного наблюдателя, однако могут найти применения и для ряда задач, требующих оценки аутентичности трафика, проходящего через канал.

24 База данных цифровых отпечатков Cisco – URL: https://github.com/cisco/mercury/blob/main/resources/fingerprint_db.json.gz (дата обращения: 20.09.2023).

25 Открытый формат представления цифровых отпечатков ja3 URL: https://github.com/trisulnsm/trisul-scripts/blob/master/lua/frontend_scripts/reassembly/ja3/prints/ja3fingerprint.json (дата обращения: 20.09.2023).

Литература

1. *Ворончихин И. С., Иванов И. И., Максимов Р. В., Соколовский С. П. Маскирование структуры распределённых информационных систем в киберпространстве // Вопросы кибербезопасности. 2019. № 6 (34). – С. 92–101. DOI: 10.21681/2311-3456-2019-6-92-101*
2. *Москвин А. А., Максимов Р. В., Горбачёв А. А. Модель, оптимизация и оценка эффективности применения многоадресных сетевых соединений в условиях сетевой разведки // Вопросы кибербезопасности. 2023. № 3 (55). – С. 13–22.*
3. *Tatang Dennis, Schneider Carl, Holz Thorsten. Largescale analysis of infrastructureleaking DNS servers // Detection of Intrusions and Malware, and Vulnerability Assessment: 16th International Conference, DIMVA 2019, Gothenburg, Sweden, June 19–20, 2019, Proceedings 16 / Springer. — 2019. —Pp. 353–373*
4. *Клименко Т. М., Акжигитов Р. Р. Обзор методов обнаружения распределённых атак типа» отказ в обслуживании» на основе машинного обучения и глубокого обучения //International Journal of Open Information Technologies. – 2023. – Т. 11. – №. 6. – С. 46–66.*
5. *Dangi A., Batra U. TLS Fingerprinting «A Passive Concept of Identification» //Artificial Intelligence and Machine Learning in Healthcare. – Singapore: Springer Nature Singapore, 2023. – С. 95-116.*
6. *Althouse J., Atkinson J., Atkins J. TLS fingerprinting with JA3 and JA3S //Salesforce. – 2019.*
7. *Rana S., Garg U., Gupta N. Intelligent Traffic Monitoring System Based on Internet of Things //2021 International Conference on Computational Performance Evaluation (ComPE). – IEEE, 2021. – С. 513–518.*
8. *Полянская М. С. Анализ подходов к обнаружению атак в зашифрованном трафике // Современные информационные технологии и ИТ-образование. 2021. Т. 17, No 4. С. 922–931. DOI: https://doi.org/10.25559/SITITO.17.202104.922-931*
9. *Ali Rasteh, Florian Delpech, Carlos AguilarMelchor et al. Encrypted internet traffic classification using a supervised spiking neural network // Neurocomputing. — 2022. — Vol. 503. — Pp. 272–282.*
10. *Gupta Neha, Jindal Vinita, Bedi Punam. Encrypted traffic classification using extreme gradient boosting algorithm // International Conference on Innova tive Computing and Communications: Proceedings of ICICC 2021, Volume 3 / Springer. — 2022. — Pp. 225–232.*
11. *Islam Faiz Ul, Liu Guangjie, Liu Weiwei. Identifying VoIP traffic in VPN tunnel via flow spatiotemporal features // Mathematical Biosciences and Engineering. — 2020. — Vol. 17, no. 5. — Pp. 4747–4772.*
12. *Islam F. U. et al. VoIP traffic detection in tunneled and anonymous networks using deep learning // IEEE Access. – 2021. – Т. 9. – С. 59783–59799.*
13. *Li K., Cui B. Malicious Encrypted Traffic Identification Based on Four-Tuple Feature and Deep Learning // Innovative Mobile and Internet Services in Ubiquitous Computing: Proceedings of the 15th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS-2021). – Springer International Publishing, 2022. – С. 199–208.*
14. *Sismis L., Korenek J. Analysis of TLS Prefiltering for IDS Acceleration //International Conference on Passive and Active Network Measurement. – Cham: Springer Nature Switzerland, 2023. – С. 85–109.*
15. *Deri L., Fusco F. Using Deep Packet Inspection in CyberTraffic Analysis //2021 IEEE International Conference on Cyber Security and Resilience (CSR). – IEEE, 2021. – С. 89–94.*
16. *Anderson Blake, McGrew David. Accurate TLS fingerprinting using destination context and knowledge bases // arXiv preprint arXiv:2009.01939. — 2020.*
17. *Anderson B., McGrew D. Tls beyond the browser: Combining end host and network data to understand application behavior //Proceedings of the Internet Measurement Conference. – 2019. – С. 379–392.*
18. *Varmarken J. et al. FingerprinTV: Fingerprinting Smart TV Apps //Proceedings on Privacy Enhancing Technologies (PoPETs). – 2022. – Т. 2022. – №. 3. – С. 606–629.*
19. *Kim H. et al. Revisiting TLS-Encrypted Traffic Fingerprinting Methods for Malware Family Classification //2022 13th International Conference on Information and Communication Technology Convergence (ICTC). – IEEE, 2022. – С. 1273–1278.*
20. *Heino J. et al. On usability of hash fingerprinting for endpoint application identification // 2022 IEEE International Conference on Cyber Security and Resilience (CSR). – IEEE, 2022. – С. 38–43.*
21. *Ишкуватов С. М., Швед В. Г., Филькова И. А. Метод оценки близости цифровых отпечатков реализаций протоколов // Информационно-методический журнал «Защита информации. Инсайд». — 2022. — № 2. — С. 29–33.*
22. *Беляев Е. А., Емельянова О. А., Лившиц И. И. Анализ методик оценки рисков информационной безопасности кредитно-финансовых организаций // Научно-технический вестник информационных технологий, механики и оптики. 2021. Т. 21, № 3. С. 437–441. DOI: 10.17586/2226-1494-2021-21-3-437-441*