Информационные процессы, Том 23, № 4, 2023, стр. 568–578.

ПЕРЕДАЧА ИНФОРМАЦИИ В КОМПЬЮТЕРНЫХ СЕТЯХ

О классификации трафика по метаданным протокола HTTP/3

А.А. Курапов[*,**], Д.Р. Шамсимухаметов[*], М.В. Любогощев[*], Е.М. Хоров[*]

[*] *Институт проблем передачи информации им. А.А. Харкевича Российской академии наук*

[**] *Московский физико-технический институт (национальный исследовательский университет)*

Поступила в редколлегию 1 ноября 2023 г. Принята 25 декабря 2023 г.

Аннотация—Новый протокол HTTP/3 использует протокол QUIC вместо стандартной связки протоколов TCP-TLS. Выбор QUIC обусловлен следующими его преимуществами. Во-первых, QUIC позволяет снизить задержку установления соединения за счет комбинирования транспортного рукопожатия и рукопожатия TLS. Это достигается за счет передачи, так называемых, транспортных параметров QUIC внутри сообщений TLS. Во-вторых, протокол QUIC предоставляет различные механизмы защиты от анализа трафика, такие как шифрование заголовков и содержимого первых служебных пакетов и дополнение пакетов нулевыми байтами до фиксированной длины. Данные механизмы применяются серверами для улучшения приватности, т.е. затруднения классификации трафика сторонним наблюдателем. В данной работе исследуется, как эти отличия протокола QUIC влияют на приватность трафика HTTP/3. Продемонстрировано, что, несмотря на одну из целей QUIC — повышение приватности, использование им разнообразных транспортных параметров, на практике, снижает приватность трафика HTTP/3 примерно на 5% в метрике F-score.

КЛЮЧЕВЫЕ СЛОВА: Классификация трафика, алгоритм «случайный лес», QUIC, TLS, Transport Parameters, метаданные HTTP/3

DOI: 10.53921/18195822_2023_23_4_568

1. ВВЕДЕНИЕ

В 2022 году Инженерный совет Интернета (англ.: Internet Engineering Task Force, IETF) стандартизовал новую версию протокола HTTP — HTTP/3 [1]. Согласно отчету интернет-ресурса CloudFlare [2], уже за год доля трафика HTTP/3 от всего мирового трафика HTTP выросла примерно до 40% для браузера Chrome и примерно до 35% для браузера Firefox. В отличие от предыдущих версий HTTP, работающих в связке с протоколом транспортного уровня Transmission Control Protocol (TCP) и протоколом защиты транспортного уровня Transport Layer Security (TLS), HTTP/3 работает в связке с протоколом QUIC [3]. QUIC сочетает в себе функции транспортного и прикладного уровней (см. рис. 1) и имеет ряд полезных для HTTP особенностей. Во-первых, QUIC использует протокол транспортного уровня UDP. Это позволяет приложениям более гибко контролировать доставку потоков данных. Во-вторых, объединение в себе функций протоколов транспортного и сессионного уровня позволяет протоколу QUIC существенно сократить задержку от старта потока до начала передачи полезных данных потока. При этом протокол HTTP/3 по умолчанию поддерживает все возможности связки протоколов TCP-TLS-HTTP (далее, для краткости, HTTP/2), такие как шифрование пользовательских данных, аутентификация сервера, последовательная и надежная доставка данных.

Трафик HTTP/3 шифруется с помощью встроенного в протокол QUIC протокола TLS 1.3 [4, 5]. Благодаря этому тип данных и их содержимое скрыты от стороннего наблюдателя. Другими

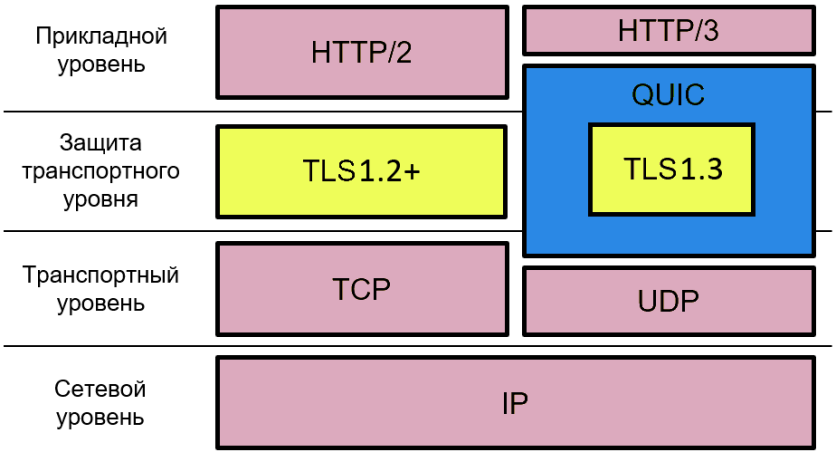

Рис. 1. Сравнение стеков протоколов для HTTP/2 и HTTP/3.

словами, информация о том, к какому типу трафика (видеотрафик, аудиотрафик, веб-трафик и т.д.) или сервису (YouTube, Instagram и т.д.) относится отдельный поток, не передается в открытом виде. Однако данная информация необходима операторам сотовой связи и администратором сетей Wi-Fi для обеспечения высокого качества обслуживания (англ.: Quality of Service, QoS) [6, 7] или предоставления бесплатного доступа к определенным ресурсам. Также эта информация используется для блокировки вредоносного или законодательно запрещенного трафика.

Задача определения типа трафика называется *классификацией трафика*. Объектом классификации трафика, как правило, является сетевой поток, идентифицируемый IP-адресами и портами клиента и сервера, а также используемым протоколом транспортного уровня.

Для классификации HTTP-трафика операторы связи и администраторы сетей Wi-Fi зачастую используют следующую уязвимость [8]. В современных версиях протокола TLS при установлении безопасного соединения в открытом виде передается параметр Server Name Indication (SNI). В данном параметре указывается доменное имя сервера, с которым клиент устанавливает соединение. В большинстве случаев это позволяет однозначно определить тип данных, передаваемых в данном потоке [9]. Для устранения данной уязвимости в настоящее время ведется разработка дополнения Encrypted ClientHello (ECH) протокола TLS [10]. Предложенное дополнение позволит скрыть SNI и ряд других уязвимых параметров, тем самым повысив конфиденциальность передачи данных с использованием протокола TLS и значительно усложнив задачу классификации трафика.

В протоколе QUIC также предусмотрены механизмы защиты от анализа трафика. Так, стандарт QUIC [3] специфицирует кадры QUIC PADDING, содержащие последовательности нулевых байтов. Эти кадры могут иметь произвольную длину и отправляются между служебными данными протокола TLS, т.е. потенциально затруднять классификацию трафика, изменяя размеры отправляемых пакетов. Кроме того, спецификацией [4] предусмотрено шифрование начальных пакетов QUIC и части их заголовков. Так, первые пакеты установления безопасного соединения в протоколе QUIC зашифрованы алгоритмом AEAD_AES_128_GCM [11] с помощью ключей, извлеченных из параметра Destination Connection ID клиента и используемой версии QUIC. Также QUIC шифрует часть данных заголовков этих пакетов, в частности, поле Packet Number.

Тем не менее в работе [12] продемонстрировано, что кадры QUIC PADDING не защищают от классификации трафика HTTP/3 по нешифрованным TLS-параметрам первых пакетов

полезной нагрузки. Работы [13,14] показывают, что потоки трафика HTTP/3 с высокой точностью можно классифицировать за счет анализа статистических признаков потока, таких как длины пакетов и временные интервалы их прихода.

Исследование [15] показывает, что даже в сценарии Encrypted ClientHello по оставшимся незащищенными параметрам можно с высокой точностью классифицировать трафик HTTP/2. В этой статье был разработан интерпретируемый алгоритм RB-RF, который классифицирует потоки по нешифрованному содержимому начальных сообщений ClientHello (CH) и ServerHello (SH) протокола TLS. Наконец, в работе [16] продемонстрировано, за счет каких TLS-параметров трафик HTTP/2 может быть классифицирован современными алгоритмами в сценарии ECH.

Таким образом, современные классификаторы трафика, основанные на анализе нешифрованных параметров оригинального протокола TLS 1.3 или статистических признаков потока, успешно справляются с классификацией HTTP/2 и HTTP/3 трафика даже в сценарии ECH. Более того, в литературе подробно изучен вопрос о важности каждого из этих признаков для классификации трафика [14,15]. Однако, протокол QUIC использует модифицированную версию протокола TLS [4]. Помимо полей и расширений обычного протокола TLS, начальный обмен сообщениями между клиентом и сервером может содержать специфичное для QUIC TLS-расширение Transport Parameters [3]. При помощи этого расширения клиент и сервер договариваются о транспортных параметрах, необходимых для обмена данными в течение соединения. В нем могут также содержаться так называемые новые транспортные параметры (англ.: New Transport Parameters) [3], определяющие функционал протокола QUIC, не описанный в оригинальной спецификации.

В зависимости от типа передаваемого трафика, клиент и сервер могут договориться использовать разные транспортные параметры. Однако существующие исследования пользовательской приватности при использовании протокола QUIC и вклада передающихся открытыми параметров протокола TLS в точность классификации трафика не рассматривали и не анализировали используемые различными потоками транспортные параметры. Знание же о вкладе данных параметров в точность классификации трафика HTTP/3 может продемонстрировать их практическую ценность для алгоритмов классификации, а также помочь сделать выводы о необходимости их шифрования в будущих версиях протокола QUIC.

В данной работе исследуется влияние транспортных параметров протокола QUIC на точность классификации трафика HTTP/3. Для получения результатов была собрана база данных из более чем 37000 потоков трафика HTTP/3. Также было разработано расширение алгоритма RB-RF для классификации трафика HTTP/3.

Дальнейшее изложение построено следующим образом. В разделе 2 описана структура протокола QUIC и его основные особенности, которые могут быть полезны для классификации трафика. Далее, в разделе 3 представлена собранная база данных. Раздел 4 содержит описание разработанного алгоритма. Наконец, в разделе 5 представлены численные результаты и дальнейшие планы исследования.

2. АНАЛИЗ РАБОТЫ ПРОТОКОЛА QUIC

Одним из важнейших преимуществ связки протоколов QUIC-HTTP над TCP-TLS-HTTP является снижение задержки установления соединения на одну задержку кругового пути (англ.: round-trip-time, RTT) за счет комбинированного криптографического и транспортного рукопожатия, схема которого приведена на рис. 2.

Перед отправкой первого HTTP/3-запроса клиент и сервер обмениваются QUIC-пакетами Initial и Handshake [3]. Основная полезная нагрузка данных пакетов — это QUIC-кадры CRYPTO. Данные кадры инкапсулируют сообщения TLS 1.3 и используются

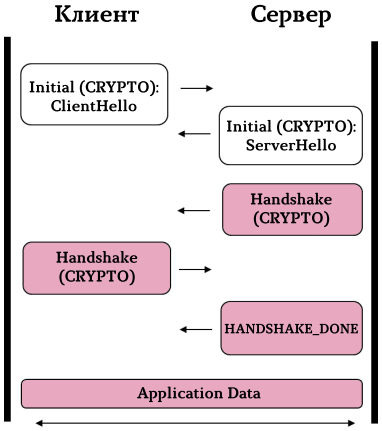

Рис. 2. Упрощенная схема рукопожатия QUIC

для установления безопасного соединения. Содержимое данных кадров шифруется. Однако ключи шифрования для кадров CRYPTO, содержащих первые два TLS-сообщения CH, SH, извлекаются из пакета Initial, отправляемого клиентом. Поэтому данные кадры могут быть расшифрованы любым промежуточным узлом в сети. Сообщение CH протокола QUIC содержит все те же поля и расширения, что и оригинальное TLS-сообщение CH (см. рис. 3). Кроме того, сообщение QUIC CH содержит расширение QUIC Transport Parameters (TP). Аналогично предусмотрено расширение TP сервера. Однако оно передается в зашифрованном виде в сообщении Encrypted Extensions (EE) TLS-«рукопожатия», а не в сообщении SH. Помимо кадров CRYPTO, пакеты Initial и Handshake могут содержать кадры PADDING, ACK и PING. Кадры PADDING дополняют пакеты Initial и Handshake последовательностью из нулевых байтов. Они применяются для того, чтобы дополнить длину пакета до минимального размера или затруднить анализ трафика по длинам пакетов. В свою очередь, кадры ACK используются, чтобы проинформировать отправителя пакетов об их успешной доставке. Наконец, кадры PING нужны для проверки достижимости клиента или сервера, а также для продления активности соединения при попытке протоколом уровня приложения закрыть его.

Record Type	Record Version	Record Len	Handshake Type	Message Len	Message Version	Random	Session ID Len
1 byte	2 bytes	2 bytes	1 byte	3 bytes	2 bytes	32 bytes	1 byte

Session ID	Cipher Suites Len	Cipher Suites	Compression Methods Len	Compression Methods	Extentions Len	Ext 1 Type
SID len bytes	2 bytes	CS len bytes	1 byte	CM len bytes	2 bytes	2 bytes

Ext 1 Len	Ext 1 Data	...	Ext i QUIC TP	Ext i QUIC TP Len	TP 1 Type	TP 1 Len
2 bytes	ext 1 len bytes	...	2 bytes	QUIC TP len bytes	1 | 2 | 4 | 8 bytes	1 byte

TP 1 Data	...	TP k Type	TP k Len	TP k Data	...	Ext n Type	Ext n Len	Ext n Data
TP 1 len	...	1 | 2 | 4 | 8 bytes	1 byte	TP k len	...	2 bytes	2 bytes	ext n len bytes

Рис. 3. Структура сообщения ClientHello в протоколе QUIC

В зависимости от версии QUIC, расширение TP имеет идентификатор TLS-расширения 57 или 65445 [3,17] и содержит транспортные параметры трех типов. Во-первых, в нем могут передаваться до 17 стандартных транспортных параметров, описанных в оригинальной спецификации QUIC. Во-вторых, расширение TP может содержать зарезервированные транспортные параметры. Идентификаторы Type таких параметров удовлетворяют условию Type = 31 * N + 27, где N — целое число [3]. Сервисы могут самостоятельно назначать идентификатор таких параметров, а также их содержимое. Поэтому учет зарезервированных транспортных параметров может повысить точность классификации некоторых сервисов. В-третьих, протокол QUIC оставляет возможность регистрировать новые транспортные параметры [3] согласно [18]. Так, например, [19] определяет транспортный параметр enable_multipath. Он используется для передачи данных по разным путям в рамках одного соединения (здесь путь задается парой IP-адрессов и парой портов клиента и сервера). Другой транспортный параметр, min_ack_delay [20], задает минимально допустимую задержку отправки кадров ACK в потоке QUIC. Также существуют транспортные параметры, предложенные компанией Google: google_connection_options, initial_rtt, user_agent, google_version, параметры Facebook:
facebook_partial_reliability, и другие. Эти параметры могут быть использованы для реализации новых функций протокола QUIC, не описанных в оригинальной спецификации. Длина их идентификатора кодируется первыми двумя битами и может принимать итоговое значение 1, 2, 4 или 8 байт (см. рис. 3). Таким образом, стандарт QUIC оставляет большую свободу для расширения функционала протокола. Учет параметров, информирующих об этом новом функционале, может позволить с большей точностью классифицировать использующие их сервисы.

3. БАЗА ДАННЫХ

Для исследования влияния транспортных параметров QUIC на точность классификации трафика база данных из работы [16] была дополнена трафиком популярных сервисов компаний Google и Meta, использующих протокол HTTP/3. Собранная база данных содержит буферизованный потоковый видеотрафик сервисов YouTube, Facebook. Также в базе представлены потоки видеотрафика сервисов коротких видеоклипов Instagram Reels и YouTube Shorts. Наконец, база содержит более 30000 потоков веб-трафика перечисленных и других популярных сервисов Google и Meta.

Таблица 1. Распределение исследуемых потоков трафика HTTP/3 по классам

Тип трафика	**Сервис**	**Число потоков**	**Паттерн SNI**
буферизованный потоковый видеотрафик	YouTube	605	r.*-.*googlevideo.*
	Facebook	3009	.*video.*fbcdn.net, scontent.*fbcdn.net
видеотрафик сервисов коротких видеоклипов	Instagram Reels	541	.*instagram.*fbcdn.*, scontent.*cdninstagram.*
	YouTubeShorts	1050	r.*-.*googlevideo.*
веб-трафик	Meta	4089	.*facebook.*, .*fbcdn.*, .*cdn.fbsbx.*', .*instagram.*
	Google	28073	.*googlevideo.*, .*.gvt.*, .*google.*, .*youtube.*, .*ytimg.*, .*ggpht.*, .*gstatic.*

В таблице 1 для каждого сервиса указан паттерн SNI, по которому производилась разметка. В случае совпадающих паттернов SNI для трафика разного типа, разметка осуществлялась по времени сбора трафика. Анализ собранной базы данных показал, что сервисы Google и Meta применяют новые транспортные параметры QUIC, описанные в документах [20–36].

4. РАСШИРЕНИЕ АЛГОРИТМА RB-RF ДЛЯ ТРАФИКА HTTP/3

Одним из лучших интерпретируемых алгоритмов классификации трафика по полезной нагрузке является RB-RF [15]. Его работа проходит в два этапа. Во-первых, RB-RF осуществляет перекомпоновку байтов полезной нагрузки сообщений CH и SH каждого потока в единый байт-вектор, согласно описанию в работе [15]. На втором этапе полученный байт-вектор подается на вход алгоритма случайный лес. При этом каждый байт полученного вектора используется как отдельный признак.

RB-RF изначально был разработан для трафика HTTP/2. С помощью данного алгоритма были исследованы все стандартные открытые параметры TLS на предмет их важности для классификации трафика. Однако трафик HTTP/3 содержит и другие открытые параметры, описанные в разделе 2, учет которых может повысить точность классификации.

original_destination_connection_id	max_idle_timeout	stateless_reset_token	max_udp_payload_size
4 bytes	4 bytes	2 bytes	4 bytes

initial_max_data	initial_max_stream_data_bidi_local	initial_max_stream_data_bidi_remote	initial_max_stream_data_uni
4 bytes	4 bytes	2 bytes	4 bytes

initial_max_streams_bidi	initial_max_streams_uni	ack_delay_exponent	max_ack_delay
4 bytes	4 bytes	2 bytes	2 bytes

disable_active_migration	preferred_address	active_connection_id_limit	initial_source_connection_id
2 bytes	2 bytes	4 bytes	4 bytes

New TP 1 Data	...	New TP 23 Data	1st Reserved TP Type	2nd Reserved TP Type	3rd Reserved TP Type
2 bytes	46 bytes	2 bytes	8 bytes	8 bytes	8 bytes

Рис. 4. Перекомпоновка QUIC Transport Parameters

Чтобы расширить алгоритм RB-RF для учета транспортных параметров протокола QUIC, вектор перекомпонованных TLS-параметров дополняется следующими полями. Во-первых, включаются стандартные транспортные параметры из исходной спецификации QUIC, как показано на рис. 4. Не включается только один из стандартных параметров, retry_source_connection_id, так как он не отправляется в сообщении CH. Для значений оставшихся 16 стандартных транспортных параметров в перекомпонованном байт-векторе выделены поля с размерами, указанными рис. 4. Если размер значения транспортного параметра меньше указанного на рисунке, то его содержимое дополняется нулями. Если же больше, то значение обрезается до этой длины. Всего на такие параметры выделено 52 байта. Затем, в байт-векторе представлены значения 23 новых транспортных параметров, которые использовались потоками из собранной базы данных. Для каждого нового транспортного параметра длина его значения в байт-векторе ограничена двумя байтами. Наконец, в вектор включены 24 байта идентификаторов первых трех (при их наличии) зарезервированных параметров. В случае отсутствия любого из транспортных параметров в потоке, его значение в байт-векторе заполняется нулями.

По аналогии с базовым алгоритмом RB-RF, после перекомпоновки CH, SH, TP полученные векторы объединяются в один байт-вектор, который затем подается на вход алгоритма случайный лес.

5. АНАЛИЗ РЕЗУЛЬТАТОВ

В данном разделе исследуется точность классификации трафика HTTP/3 в сценарии ECH с помощью разработанного расширения алгоритма RB-RF.

Протокол ECH на данный момент находится на стадии разработки и почти не используется в Интернете. Поэтому, для моделирования ECH-трафика, полезная нагрузка всех сообщений CH и SH из собранной базы данных была преобразована согласно актуальному описанию протокола ECH [10]. В каждом эксперименте база данных разбита на обучающую, проверочную и тестовую подвыборки в соотношении 7:1:2 соответственно. Усреднение результатов произведено по 150 запускам. В качестве метрики точности классификации использована стандартная метрика F-score [37].

Для сравнения точности классификации оригинального и разработанного алгоритмов на исследуемой базе потоков, описанной в разделе 3, рассматриваются два эксперимента. В первом, исследуется задача классификации видеопотоков сервисов Instagram Reels, YouTube Shorts, Facebook Video, YouTube Video. Для ее решения рассматривается подвыборка базы данных, содержащая буферизованный потоковый видеотрафик и видеотрафик сервисов коротких видеоклипов. Во втором, исследуется задача детектирования видеопотоков данных сервисов на фоне разнообразного веб-трафика этих и других сервисов Google и Meta. Для второго эксперимента используется вся собранная база данных.

5.1. Классификация видеотрафика

В эксперименте была оценена эффективность расширенного алгоритма RB-RF с использованием открытых данных рукопожатия TLS и транспортных параметров (признаки HTTP/3) по сравнению с его оригинальной версией, использующей исключительно TLS признаки (признаки HTTP/2), в задаче классификации видеотрафика.

Таблица 2. Сравнение эффективности оригинального и расширенного алгоритмов RB-RF при классификации видеотрафика Google и Meta.

№	класс	F-score алгоритма RB-RF	
		признаки HTTP/3	**признаки HTTP/2**
1	Instagram Reels	55,3%	52,6%
2	YouTube Shorts	90,9%	87,1%
3	Facebook Video	94,2%	91,0%
4	YouTube Video	80,3%	68,5%
-	среднее значение	**80,2%**	74,8%

Таблица 2 демонстрирует улучшение F-score при применении расширенного алгоритма для всех тестовых классов. Учет транспортных параметров QUIC в качестве дополнительных признаков к стандартным TLS-параметрам, позволяет повысить средний F-score на 5% в этом сценарии. Результаты подтверждают, что интеграция транспортных параметров QUIC повышает точность классификации, улучшая общую производительность алгоритма RB-RF.

5.2. Классификация веб- и видеотрафика

Во втором эксперименте в выборке дополнительно содержится веб-трафик сервисов Google и Meta. Из результатов, приведенных в таблице 3, можно видеть, что учет транспортных

параметров позволяет повысить точность на 6%. Однако даже учет транспортных параметров QUIC не позволяет добиться приемлемого качества классификации при таком разнообразии трафика. В частности, средний F-score не превышает 50%, а минимальный по классам — для потоков Instagram Reels — 22%.

Таблица 3. Сравнение эффективности оригинального и расширенного алгоритмов RB-RF на всей собранной базе данных.

№	класс	F-score алгоритма RB-RF	
		признаки HTTP/3	**признаки HTTP/2**
1	Instagram Reels	22,0%	9,1%
2	YouTube Shorts	28,7%	6,7%
3	Facebook Video	60,7%	67,5%
4	YouTube Video	30,2%	27,6%
5	Meta Web	58,8%	52,6%
6	Google Web	97,0%	95,8%
-	среднее значение	49,6%	43,2%

Итак, из представленных результатов можно сделать вывод, что, несмотря на неудовлетворительное в целом качество классификации, учет транспортных параметров протокола QUIC позволяет повысить качество классификации примерно на 5% в метрике F-score. Другими словами, передача транспортных параметров протокола QUIC в незашифрованном виде приводит к снижению приватности пользователей.

6. ЗАКЛЮЧЕНИЕ

В данной работе была исследована приватность трафика HTTP/3 в сценарии Encrypted ClientHello. Для этого была собрана база данных видеотрафика различных сервисов Google и Meta, а также веб-трафика этих компаний. В качестве инструмента для исследования приватности применялся известный из литературы интерпретируемый алгоритм классификации трафика RB-RF. При этом признаковое пространство данного алгоритма было расширено на транспортные параметры QUIC. Было выявлено, что механизмы повышения приватности QUIC не являются эффективными на практике. Более того, передача доступных для анализа стороннему наблюдателю разнообразных транспортных параметров QUIC снижает приватность трафика HTTP/3 примерно на 5% в метрике F-score. Тем не менее, при достаточно разнообразной базе данных, точность классификации трафика остается неудовлетворительной. Поэтому в дальнейшем планируется расширить признаковое пространство классификатора RB-RF на статистические признаки потоков, такие как длины и интервалы между приходами пакетов, для исследования других каналов утечки приватности.

СПИСОК ЛИТЕРАТУРЫ

1. Bishop Mike. HTTP/3. RFC 9114. 2022. — June. https://www.rfc-editor.org/info/rfc9114.

2. Examining HTTP/3 usage one year on. `https://blog.cloudflare.com/http3-usage-one-year-on/`.

3. Iyengar Jana, Thomson Martin. QUIC: A UDP-Based Multiplexed and Secure Transport. RFC 9000. 2021. — May. https://www.rfc-editor.org/info/rfc9000.

4. Thomson Martin, Turner Sean. Using TLS to Secure QUIC. RFC 9001. 2021. — May. https://www.rfc-editor.org/info/rfc9001.

5. Rescorla Eric. The Transport Layer Security (TLS) Protocol Version 1.3. RFC 8446. 2018. — August. https://www.rfc-editor.org/info/rfc8446.

6. Salman Ola, Elhajj Imad H, Kayssi Ayman, Chehab Ali. A review on machine learning–based approaches for Internet traffic classification // Annals of Telecommunications. 2020. Vol. 75. Pp. 673–710.

7. Kurapov Anton, Shamsimukhametov Danil, Liubogoshchev Mikhail, Khorov Evgeny. CloudETC: a Privacy-Preserving Encrypted Traffic Classification Platform for QoS in Wi-Fi // 2023 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom) / IEEE. 2023. Pp. 244–246.

8. Li Fangfan, Razaghpanah Abbas, Kakhki Arash Molavi et al. liberate,(n) a library for exposing (traffic-classification) rules and avoiding them efficiently // Proceedings of the 2017 Internet Measurement Conference. 2017. Pp. 128–141.

9. Shamsimukhametov D, Liubogoshchev M, Khorov E, Akyldiz IF. Are Neural Networks the Best Way for Encrypted Traffic Classification? // 2021 International Conference Engineering and Telecommunication (En&T) / IEEE. 2021. Pp. 1–5.

10. Eric Rescorla. TLS Encrypted Client Hello: Internet-Draft draft-ietf-tls-esni-17: IETF, 2023.

11. McGrew David, Igoe Kevin. AES-GCM Authenticated Encryption in the Secure Real-time Transport Protocol (SRTP). RFC 7714. 2015. — Dec. https://www.rfc-editor.org/info/rfc7714.

12. Barman Ludovic, Siby Sandra, Wood Christopher et al. This is not the padding you are looking for! On the ineffectiveness of QUIC PADDING against website fingerprinting // arXiv preprint arXiv:2203.07806. 2022.

13. Luxemburk Jan, Hynek Karel, Čejka Tomáš. Encrypted traffic classification: the QUIC case // 2023 7th Network Traffic Measurement and Analysis Conference (TMA) / IEEE. 2023. Pp. 1–10.

14. Almuhammadi Sultan, Alnajim Abdullatif, Ayub Mohammed. QUIC Network Traffic Classification Using Ensemble Machine Learning Techniques // Applied Sciences. 2023. Vol. 13, no. 8. P. 4725.

15. Shamsimukhametov Danil, Kurapov Anton, Liubogoshchev Mikhail, Khorov Evgeny. Is Encrypted ClientHello a Challenge for Traffic Classification? // IEEE Access. 2022. Vol. 10. Pp. 77883–77897.

16. Шамсимухаметов Д.Р., Курапов А.А., Любогощев М.В., Хоров Е.М.. Неразличимость трафика по открытым параметрам TLS при использовании Encrypted ClientHello // Информационные процессы. 2023. Vol. 23, no. 2. Pp. 231–240.

17. Thomson Martin, Turner Sean. Using Transport Layer Security (TLS) to Secure QUIC: Internet-Draft draft-ietf-quic-tls-13: Internet Engineering Task Force, 2018. — June. Work in Progress. https://datatracker.ietf.org/doc/draft-ietf-quic-tls/13/.

18. Cotton Michelle, Leiba Barry, Narten Dr. Thomas. Guidelines for Writing an IANA Considerations Section in RFCs. RFC 8126. 2017. — June. https://www.rfc-editor.org/info/rfc8126.

19. Liu Yanmei, Ma Yunfei, Coninck Quentin De et al. Multipath Extension for QUIC: Internet-Draft draft-ietf-quic-multipath-06: Internet Engineering Task Force, 2023. — Oct. Work in Progress. https://datatracker.ietf.org/doc/draft-ietf-quic-multipath/06/.

20. Iyengar Jana, Swett Ian, Kühlewind Mirja. QUIC Acknowledgement Frequency: Internet-Draft draft-ietf-quic-ack-frequency-07: Internet Engineering Task Force, 2023. — Oct. Work in Progress. https://datatracker.ietf.org/doc/draft-ietf-quic-ack-frequency/07/.

21. Schinazi David, Rescorla Eric. Compatible Version Negotiation for QUIC. RFC 9368. 2023. — May. https://www.rfc-editor.org/info/rfc9368.

22. Iyengar Jana, Swett Ian. QUIC Loss Detection and Congestion Control. RFC 9002. 2021. — May. https://www.rfc-editor.org/info/rfc9002.

23. Quantum Readiness test. https://github.com/quicwg/base-drafts/wiki/Quantum-Readiness-test.

24. Banks Nick. QUIC Connection ID Based Initial Routing: Internet-Draft draft-banks-quic-cibir-01: Internet Engineering Task Force, 2022. — Mar. Work in Progress. https://datatracker.ietf.org/doc/draft-banks-quic-cibir/01/.

25. Ferrieux Alexandre, Hamchaoui Isabelle, Lubashev Igor, Tikhonov Dmitri. Packet Loss Signaling for Encrypted Protocols: Internet-Draft draft-ferrieuxhamchaoui-quic-lossbits-03: Internet Engineering Task Force, 2020. — Jan. Work in Progress. https://datatracker.ietf.org/doc/draft-ferrieuxhamchaoui-quic-lossbits/03/.

26. Thomson Martin. Greasing the QUIC Bit. RFC 9287. 2022. — Aug. https://www.rfc-editor.org/info/rfc9287.

27. Google Transport Parameters. `https://github.com/google/quiche/tree/main/quiche/quic`.

28. Huitema Christian. Quic Timestamps For Measuring One-Way Delays: Internet-Draft draft-huitema-quic-ts-02: Internet Engineering Task Force. Work in Progress. https://datatracker.ietf.org/doc/draft-huitema-quic-ts/02/.

29. Huitema Christian. Quic Timestamps For Measuring One-Way Delays: Internet-Draft draft-huitema-quic-ts-03: Internet Engineering Task Force. Work in Progress. https://datatracker.ietf.org/doc/draft-huitema-quic-ts/03/.

30. Facebook Partial Reliablity. `https://github.com/facebook/mvfst/blob/main/quic/QuicConstants.h`.

31. Iyengar Jana, Swett Ian. Sender Control of Acknowledgement Delays in QUIC: Internet-Draft draft-iyengar-quic-delayed-ack-00: Internet Engineering Task Force. Work in Progress. https://datatracker.ietf.org/doc/draft-iyengar-quic-delayed-ack/00/.

32. Schinazi David, Rescorla Eric. Compatible Version Negotiation for QUIC: Internet-Draft draft-ietf-quic-version-negotiation-10: Internet Engineering Task Force, 2022. — Sep. Work in Progress. https://datatracker.ietf.org/doc/draft-ietf-quic-version-negotiation/10/.

33. Iyengar Jana, Swett Ian. QUIC Acknowledgement Frequency: Internet-Draft draft-ietf-quic-ack-frequency-01: Internet Engineering Task Force, 2021. — Oct. Work in Progress. https://datatracker.ietf.org/doc/draft-ietf-quic-ack-frequency/01/.

34. Iyengar Jana, Swett Ian, Kühlewind Mirja. QUIC Acknowledgement Frequency: Internet-Draft draft-ietf-quic-ack-frequency-06: Internet Engineering Task Force. Work in Progress. https://datatracker.ietf.org/doc/draft-ietf-quic-ack-frequency/06/.

35. Liu Yanmei, Ma Yunfei, Coninck Quentin De et al. Multipath Extension for QUIC: Internet-Draft draft-ietf-quic-multipath-04: Internet Engineering Task Force, 2023. — Mar. Work in Progress. https://datatracker.ietf.org/doc/draft-ietf-quic-multipath/04/.

36. Liu Yanmei, Ma Yunfei, Coninck Quentin De et al. Multipath Extension for QUIC: Internet-Draft draft-ietf-quic-multipath-05: Internet Engineering Task Force. Work in Progress. https://datatracker.ietf.org/doc/draft-ietf-quic-multipath/05/.

37. Powers David. Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation // Mach. Learn. Technol. 2008. — 01. Vol. 2.

38. Rokach Lior, Maimon Oded. Top-Down Induction of Decision Trees Classifiers—A Survey // Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on. 2005. — 12. Vol. 35. Pp. 476 – 487.

Traffic Classification Based on HTTP/3 Metadata Analysis

A.A. Kurapov, D.R. Shamsimukhametov, M.V. Liubogoshchev, E.M. Khorov

The new HTTP/3 protocol uses QUIC instead of the standard combination of protocols TCP-TLS because of the following advantages. First, QUIC allows reducing the connection establishment delay by combining the transport handshake and the TLS handshake. Second, QUIC provides various protection mechanisms against traffic analysis, such as encrypting the headers and content of the first service packets and padding packets with zero bytes to a fixed length. These mechanisms are used by servers to improve privacy, i.e.,

to make it more difficult for a third-party observer to classify traffic. This paper investigates how these QUIC protocol features affect the privacy of HTTP/3 traffic. It shows that despite one of QUIC's goals, i.e., improving privacy, the use of various transport parameters, in fact, reduces the privacy of real HTTP/3 traffic by about 5% in the F-score metric.

KEYWORDS: Traffic Classification, Random Forest, QUIC, TLS, Transport Parameters, HTTP/3 Metadata