Научная статья
УДК 004.7
DOI:10.31854/1813-324X-2022-8-4-109-118



# Способ и алгоритм определения типа трафика в шифрованном канале связи

**Сергей Маратович Ишкуватов**, sysroot0@gmail.com

Национальный исследовательский университет ИТМО,
Санкт-Петербург, 197101, Российская Федерация

**Аннотация:** *В статье предложен способ определения состава протоколов, применяемых в IPsec-канале связи, на основе закономерностей хронологии следования и* длин пакетов с шифрованной нагрузкой*. Рассмотрены характерные информативные признаки протоколов. Приведен алгоритм, позволяющий получить значения длин ESP-пакетов, содержащих произвольные пользовательские данные, для распространенных режимов работы IPsec-туннеля.*

**Ключевые слова:** *шифрование, пассивный наблюдатель, Virtual Private Network, IPsec-туннель*

**Ссылка для цитирования**: Ишкуватов С.М. Способ и алгоритм определения типа трафика в шифрованном канале связи // Труды учебных заведений связи. 2022. Т. 8. № 4. С. 109–118. DOI:10.31854/1813-324X-2022-8-4-109-118

# Method and Algorithm for Determining the Type of Traffic in an Encrypted Communication Channel

**Sergei Ishkuvatov**, sysroot0@gmail.com

ITMO University,
St. Petersburg, 197101, Russian Federation

**Abstract:** *The article proposes a method for determining the composition of protocols used in IPsec communication channel, based on the regularities of the chronology and the lengths of encrypted load packets. The characteristic informative features of the protocols are considered. An algorithm is given to obtain the length values of ESP packets containing arbitrary user data for common modes of IPsec tunnel operation.*

**Keywords:** *encryption, passive observer, Virtual Private Network, IPsec tunnel*

**For citation:** Ishkuvatov S. Method and Algorithm for Determining the Type of Traffic in an Encrypted Communication Channel. *Proc. of Telecom. Universities.* 2022;8(4):109–118. (in Russ.) DOI:10.31854/1813-324X-2022-8-4-109-118

## Введение

Распространение каналов Virtual Private Network (VPN), способных скрывать от пассивного наблюдателя передаваемые данные, и, как следствие – возможность обхода сетевых блокировок и политик корпоративных сетей, делает актуальным задачу определения *видов протоколов*, используемых внутри шифрованных каналов связи (ШКС). Ряд научных работ демонстрирует возможность определения передаваемых видов протоколов в VPN-канале с помощью нейронных сетей, анализирующих объемные и интервальные зависимости передаваемых сторонами порций данных [1–4]. Они рассматривают методы классификации исследуемых видов трафика для разных протоколов обеспечения защиты данных (OpenVPN, IPsec, TLS), но только в контексте выявления в них *одного вида трафика*. Нейронные сети, анализирующие трафик, в качестве входных признаков используют ограниченный набор доступных параметров: время регистрации

или длительность интервалов между пакетами, их длину и, иногда – направление передачи. При этом применяется корреляция между этими параметрами вложенного протокола и протокола обеспечения защиты данных.

В общем случае ШКС может содержать один вид трафика или их различные комбинации. В первом случае размеры и интервалы пакетов определяются исключительно свойствами этого трафика; во втором – общий трафик будет содержать признаки каждого из них, что увеличивает вероятность ложного распознавания или не распознавания используемых протоколов. Вместе с тем, следует отметить наличие устойчивых закономерностей между наблюдаемыми внешними характеристиками шифрованного трафика и предаваемыми внутри него данными. Особую ценность этим закономерностям придает то, что они могут быть выявлены и проанализированы с использованием низкозатратных вычислительных процедур, что открывает возможность *разработки методов и средств определения видов протоколов внутри ШКС*. Кроме того, при анализе передаваемых данных важно определять и в последствии учитывать параметры конфигурации канала, которые будут одинаково сказываться на характеристиках всех передаваемых в этом канале протоколов.

**Задача исследования**

Таким образом, задачей исследования является *разработка и проверка продуктивности алгоритма определения видов протоколов, используемых внутри ШКС* на основании закономерностей между наблюдаемыми внешними характеристиками шифрованного трафика и передаваемыми внутри него данными. Ее решение предполагает *выявление этих закономерностей* и *формализацию правил*, позволяющих определить используемые сетевые протоколы в IPsec-трафике без применения нейронных сетей.

Предлагаемый алгоритм должен обеспечивать работоспособность и при наличии нескольких видов трафика в ШКС. Алгоритм должен быть применим с произвольного момента времени и не должен иметь необходимости учитывать данные, передаваемые только в определенных фазах соединения (например, данных протокола Internet Key Exchange при установлении соединения).

**Описание условий и режимов функционирования исследуемого ШКС**

*Модель ШКС*

В работе используется типовая модель ШКС, которая применяется для передачи данных через сеть Интернет таким способом, чтобы передаваемая информация была защищена от чтения и модификации узлами сети по пути следования пакетов.

Стандартными этапами информационного обмена в ШКС являются:

1) установление соединения;

2) согласование применения алгоритмов шифрования, проверки целостности и аутентификации сторон;

3) упаковка передаваемых данных в пакеты, обеспечивающие шифрование и контроль целостности.

Так как ШКС всегда добавляет к передаваемым данным служебные заголовки, это уменьшает значение MTU (*аббр. от англ*. Maximum Transmission Unit) шифрованного канала, которое влечет необходимость фрагментировать крупные пакеты и уменьшать максимальные допустимые размеры сегментов (например, в протоколе TCP) для сокращения фрагментации.

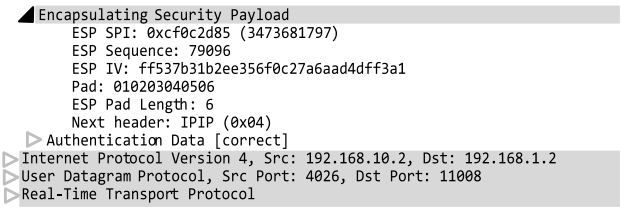
В контексте решаемой задачи IPsec [5] состоит из протоколов, обеспечивающих следующий функционал: Encrypted Secure Payload (ESP) [6] – шифрование передаваемой информации; Authentication Header (AH) – целостность и аутентификацию источника; Internet Security Association and Key Management Protocol (ISAKMP) – первичную настройку соединения, взаимную аутентификацию и обмен ключами.

Определены 2 режима функционирования:

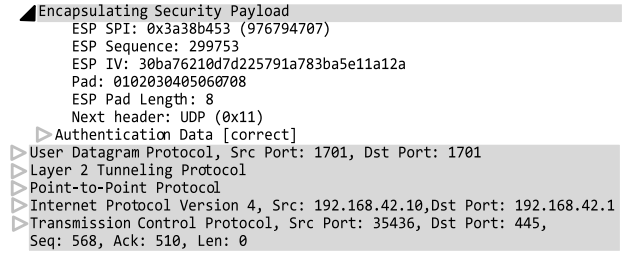
– транспортный (шифруются только данные IP-пакета, но сохраняется исходный IP-заголовок);

– тоннельный (весь исходный IP-пакет, включая информацию об отправителе и получателе, содержится в шифрованной нагрузке нового пакета; часто используется совместно с L2TP).

Тогда в зависимости от используемых конфигураций типовая структура данных, наблюдаемых в ШКС, представлена на рисунке 1.




a)




b)

**Рис. 1. Стеки протоколов внутри ESP-пакетов: только IP-заголовок (a), стек с L2TP (b)**

*Fig. 1. Protocol Stacks Inside ESP packets: IP Header Only (a), Stack with L2TP (b)*

Таким образом, пользовательская нагрузка в каждом ESP-пакете будет содержать составляющую постоянной длины, зависящую от режима функционирования тоннеля (рисунок 2; условные обозначения по таблице 1).

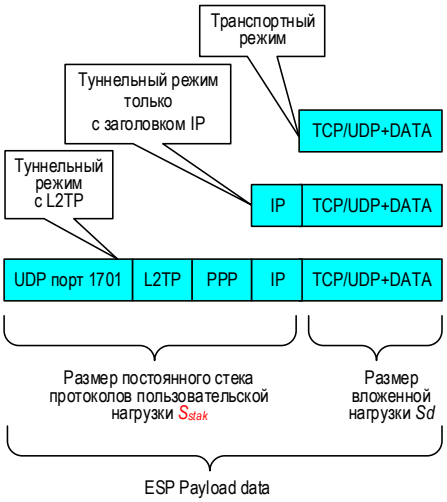



**Рис. 2. Структура ESP-пакета**

*Fig. 2. ESP Packet Structure*

Кроме того, на протяжении всего времени существования каждого конкретного тоннеля предполагается неизменность размеров полей Initialization Vector, Authentication Data и алгоритма шифрования для всех ESP-пакетов (рисунок 3; на белом фоне – не шифруемые данные).

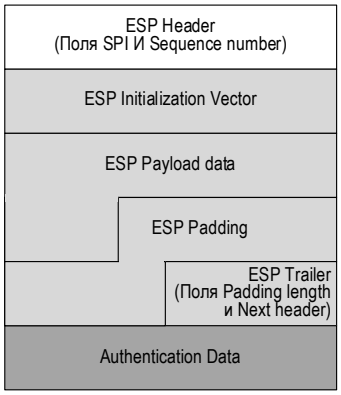



**Рис. 3. Схема расположения служебных полей в ESP-пакете**

*Fig. 3. Layout of Service Fields in the ESP Packet*

*Определение и характеристики анализируемых данных*

Пассивному наблюдателю ШКС, не имеющему ключей шифрования, для анализа доступны только следующие параметры: общие длины ESP-пакетов, значения 32-битового идентификатора тоннеля (SPI) и порядковые номера (Sequence number) для пакетов каждого из направлений (см. рисунок 3). Скрытые шифрованием поля, исключая пользовательскую нагрузку, имеют небольшой набор возможных длин, определяемых параметрами канала. Сумма длин всех полей и пользовательской нагрузки равна длине ESP-пакета.

В таблице 1 приводятся возможные размеры служебных заголовков для разных режимов работы IPsec. Если обозначить размер постоянного стека протоколов пользовательской нагрузки $S_{stak}$ (вложенной в ESP), то это позволит объединить записи формул для транспортного или туннельного режима, приведенные в работе [7].

Стандартными средствами мониторинга трафика обеспечивается получение только размера каждого ESP-пакета в сессии ($S_{ESP}$). На основании собранных длин пакетов становится возможным определить используемый в сессии размер блока алгоритма шифрования. Устанавливать конкретный использующийся алгоритм шифрования нет необходимости, т. к. в дальнейшем для расчетов используется только размер блока $Bl$. Результирующая длина ESP-пакета за вычетом константы всегда кратна $Bl$ (16 или 32, а для режима AEAD – 4 байтам). Несмотря на то, что в некоторых случаях размер поля $AuT_{ESP}$ может быть определен активным воздействием на ШКС (например, утилитами типа ike-scan [8]), прямому наблюдению средствами мониторинга он не доступен и может оцениваться только статистически.

## Определение закономерностей функционирования ШКС

Эмпирическим путем выявлен ряд зависимостей между типом трафика в канале связи и распределением длин ESP-пакетов, осуществляющих их передачу. На основании этого выдвигается гипотеза о наличии закономерности между длиной ESP-пакета и пользовательского вложения. Кроме того, апостериорно определены закономерности между *номенклатурой протоколов, одновременно присутствующих в ШКС*. Это позволяет делать априорные предположения о наличии «сопутствующих» протоколов при обнаружении одного из них. Например, при обнаружении сессии RTP-протокола следует предполагать наличие протоколов сигнализации и установления соединения SIP, MGCP или H.323, а также протокола управления передачей в реальном времени RTCP.

Примером описываемой закономерности является зависимость *размера шифрованной нагрузки* ESP-пакета *от размера, передаваемого в нем пакета нагрузки*, определяемая организацией структуры самого протокола. Размер накладных расходов служебных заголовков ESP-пакета (ESP Overhead) зависит от выбранного метода шифрования и размера поля Padding, который, в свою очередь, определяется остатком от деления размера вложенных данных на размер блока шифрования. В сети Интернет доступны калькуляторы длин IPsec-пакетов для произвольных вариаций стеков вложенных в IPsec протоколов [9, 10], которые позволяют вычислить размеры служебных заголовков в зависимости от используемого типа шифрования.

**ТАБЛИЦА 1. Обозначение, назначение и размеры полей протокола IPsec [7]**

*TABLE 1. Designation, Purpose and Sizes of IPsec Protocol Fields [7]*

| Обозначение | Описание | Длина |
|---|---|---|
| $H_{ESP}$ | Поле ESP Header | 8 байт |
| $Bl$ | Размер блока алгоритма шифрования | – 16 байт (aes128, blowfish128, cast128)<br>– 8 байт (des, 3des)<br>– 4 байта (AEAD) |
| $IV_{ESP}$ | Поле ESP Initialization Vector | Равен $Bl$, но в режиме работы AEAD 16 байт |
| $Tr_{ESP}$ | ESP Trailer | 2 байта |
| $AuT_{ESP}$ | Authentication Data | – 12 байт (96 bit HMAC включая: md5, sha1, aesxcbc, aescmac, sha256_96)<br>– 16 байт (128 bit HMAC включая: md5_128, aes128gmac, aes192gmac, aes256gmac)<br>– 20 байт (160 bit HMAC, sha1_160)<br>– 24 байта (192 bit HMAC)<br>– 32 байта (256 bit HMAC) |
| $Pad_{ESP}$ | ESP Padding | Определяется остатком от деления общего размера всего стека протоколов, нагрузки вложенного пакета и $Tr_{ESP}$ на $Bl$<br>Длина поля от 0 до ($Bl$ – 1) байт |
| $Sd$ | Размер вложенной нагрузки без постоянного стека протоколов | Длина пользовательской нагрузки, начиная с заголовков транспортных протоколов |
| $S_{stak}$ | Размер постоянного стека протоколов пользовательской нагрузки. | Суммарный размер всех заголовков пользовательской нагрузки до начала транспортного уровня:<br>– 0 байт при транспортном режиме;<br>– 20 байт для туннельного режима в случае вложения только заголовка IP Header.<br>Более 20 байт для туннельного режима в случае присутствия заголовков помимо заголовка IP Header |
| $S_{ESP}$ | Полный размер порции данных ESP, содержащей все служебные заголовки и полезную вложенную нагрузку, не включает вышестоящие протоколы | |

Другой полезной особенностью протокола IPsec в контексте определения вида передаваемого трафика может быть сохранение значения приоритета вложенного трафика в поле Type of Service (ToS). В большинстве случаев значение поля копируется из пакета вложения в заголовок ESP-пакета для сохранения возможности приоритизации чувствительного к задержкам трафика [11]. Таким образом, шифрованный трафик в большом количестве случаев будет иметь DSCP, (*аббр. от англ*. Differentiated Services Code Point), подобный исходному, что может являться дополнительным признаком при определении некоторых типов нагрузки.

Статистические закономерности вложенного трафика будут проявляться и в транспортирующем его трафике IPsec. Разбив сигнал на небольшие интервалы времени $\Delta t$, можно построить распределение длин всех пакетов, встретившихся в этом интервале времени. При визуализации серии пакетов одинаковой длины будут создавать в таком распределении линии для соседних интервалов или одиночные точки, повторяющиеся с определенной периодичностью.

В части случаев несколько линий появляются только совместно, например, при передаче данных равными порциями большого размера. В этом случае линию создаст максимальный размер пакета и остаток порции. Абсолютные значения длин, передаваемых по сети IPsec-пакетов, будут отличаться в зависимости от стека протоколов внутри канала, поэтому при исследовании следует рассматривать относительные значения длин и возможные диапазоны значений исходных не упакованных данных.

Длины ESP-пакетов для всех режимов с аутентификацией (ATH) можно вычислить как:

$$S_{ESP} = H_{ESP} + Tr_{ESP} + IV_{ESP} + AuT_{ESP} + Sd + S_{stak} + Pad_{ESP}, \quad (1)$$

где $Pad_{ESP}$ – размер поля ESP Padding, определяемый остатком от деления (mod) суммы размеров вложенной пользовательской нагрузки полностью ($Sd + S_{stak}$) и константной длины $Tr_{ESP}$ (см. выражение 2) на размер блока выбранного алгоритма шифрования ($Bl$).

Таким образом, сумма $Sd + S_{stak} + Tr_{ESP} + Pad_{ESP}$ всегда кратна $Bl$.

$$Pad_{ESP} = \begin{cases} 0, & \text{если } ((Sd + S_{stak} + Tr_{ESP}) \bmod Bl) = 0 \\ Bl - ((Sd + S_{stak} + Tr_{ESP}) \bmod Bl), & \text{в остальных случаях.} \end{cases} \quad (2)$$

Выражение ниже устанавливает зависимость между длиной пользовательских данных и параметрами ШКС; оно может быть использовано для вычисления границ размеров вложенного пакета пользовательской нагрузки, а также для получения длины исходных пользовательских данных при известных $S_{ESP}$, $Bl$, $AuT_{ESP}$ и $S_{stak}$:

$$Sd = S_{ESP} - H_{ESP} - Tr_{ESP} - IV_{ESP} - AuT_{ESP} - S_{stak} - Pad_{ESP}.$$

**Алгоритм определения видов протоколов, используемых внутри ШКС**

Обобщенный алгоритм определения видов протоколов, используемых внутри ШКС, предполагает выполнение следующих операций.

1) Подготовить список отличительных признаков протоколов, содержащих характерные распределения длин пакетов, начиная с заголовков транспортного уровня включительно (TCP, UDP), хронологические и интервальные закономерности (см. Обсуждение результатов).

2) Определить значение размера блока алгоритма шифрования $Bl$ (см. таблицу 1): вычисляется как минимальная разница между любыми двумя зарегистрированными значениями длин SESP.

3) Определить возможные значения $AuT_{ESP}$, величина должна безостаточно делиться на $Bl$, т. е. для любой встречаемой длины $S_{ESP}$ должно выполняться равенство $(S_{ESP} - H_{ESP} + IV_{ESP} + AuT_{ESP}) \bmod Bl = 0$.

4) Для каждого распределения длин из подготовленного списка признаков по формуле (1) вычислить поправку до размера ESP-пакета, для возможных значений $Bl$, $AuT_{ESP}$ и $S_{stak}$.

5) Искать совпадения признаков, полученных на предыдущем этапе в передаваемом трафике. При нахождении четкого совпадения сохранение параметров $Bl$, $AuT_{ESP}$ и $S_{stak}$ и в дальнейшем использование только их в этом конкретном IPsec-тоннеле.

6) При обнаружении стабильно передаваемых и легко прогнозируемых протоколов (таких как потоковое аудио или видео) следует вычесть из суммарного распределения значения, вносимые этими протоколами, для облегчения поиска других протоколов.

**Эксперимент**

Для проверки продуктивности предлагаемого алгоритма проведен эксперимент по моделированию работы IPsec-тоннеля на базе решения StrongSwan, в котором поочередно передавались протоколы:

– Samba и FTP в момент загрузки файлов;

– VoIP, содержащий пакеты установления соединения SIP и речевые пакеты RTP G.729, G.711, G723.1;

– RDP (*аббр. от англ.* Remote Desktop Protocol);

– BitTorrent в различных режимах.

Все пакеты шифрованного трафика записывались в PCAP-файл утилитой tcpdump. Протоколировались SPI ESP-пакетов каждого направления, время регистрации и длина пакета нагрузки. Для фрагментированных IP-пакетов записывалась итоговая длина и время регистрации первого. Это позволило находить хронологические закономерности, распределение длин ESP-пакетов. Командой «ip xfrm state» сохранялись сессионные ключи, которые использовались для проверки корректности работы алгоритма. Для записей с *большой дисперсией длин пакетов* параметр размера блока алгоритма шифрования $Bl$ определяется минимальной разницей между любыми двумя зарегистрированными значениями длин $S_{ESP}$. В некоторых случаях $Bl$ может быть определен как максимальный результат безостаточного деления каждой встречаемой длины пакета $S_{ESP}$ за вычетом $H_{ESP} + IV_{ESP} + AuT_{ESP}$ на 16,8 и 4.

Для автоматического обнаружения видов протоколов внутри ШКС реализованы программные модули-детекторы, базирующиеся на выделенных закономерностях распределения длин пакетов. На вход детекторов помимо сигнала последовательно подавались возможные значения параметров $Bl$, $AuT_{ESP}$ и $S_{stak}$.

Результаты, полученные обработчиком на подготовленном датасете для каждого типа детектируемого трафика, сведены в таблицу 2, где: $A$ – количество срабатываний детекторов с правильно подобранной конфигурацией; $B$ – с любой вариацией ошибочного определения конфигурацией или вложенного стека.

**ТАБЛИЦА 2. Результаты обработки файлов записей**

*Table 2. Results of Processing Record Files*

| Тип трафика | Количество файлов записей | Количество срабатываний детекторов | | | |
|---|---|---|---|---|---|
| | | если трафик присутствует в записи | | отсутствующих в записи типов трафика | |
| | | *A* | *B* | *A* | *B* |
| RTP G.711 | 100 | 100 | 0 | 0 | 0 |
| RTP G.723.1 | 100 | 100 | 0 | 0 | 0 |
| RTP G.729 | 100 | 100 | 0 | 0 | 0 |
| RDP | 100 | 66 | 476 | 0 | 0 |
| FTP | 100 | 85 | 106 | 0 | 0 |
| BitTorrent | 50 | 46 | 303 | 0 | 105 |

Количество корректных результатов детектирований протоколов с правильно подобранными параметрами $Bl$, $AuT_{ESP}$ и $S_{stak}$ содержатся в графе «с правильно подобранной конфигурацией». Отдельной графой приведено количество корректного детектирования этого протокола с любыми некорректными значениями из $Bl$, $AuT_{ESP}$ и $S_{stak}$. Такое сравнение позволяет выбрать наиболее подходящие протоколы обработчики для первоначального определения значений параметров $Bl$, $AuT_{ESP}$ и $S_{stak}$. Эти па-

раметры смогут использоваться другими детекторами как основные. Для оценки общего качества детектирования приведены поля с количествами ложных срабатываний также с разделением на корректные конфигурации и ошибочные. Такое разделение демонстрирует возможность уменьшения ложных срабатываний детектора четким заданием правильных значений параметров $Bl$, $AuT_{ESP}$ и $S_{stak}$.

В результате анализа установлено, что в ряде случаев детекторы срабатывали на близких, но неправильных значениях $AuT_{ESP}$ (например, 12 и 16 байт) по причине округления длин пользовательской нагрузки до их кратности размеру $Bl$. В части случаев неопределенность снимается за счет срабатывания других детекторов, результирующие значения длин которых попадают в разные диапазоны кратности $Bl$.

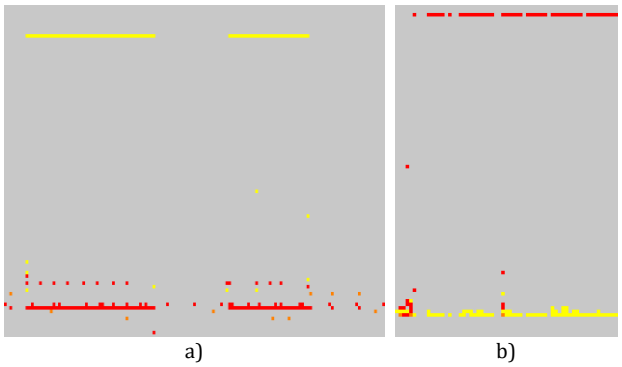
В ходе эксперимента подтверждена продуктивность «вычитания» данных идентифицированного протокола из общего трафика канала на примере дуплексной передачи RTP-пакетов. Такая процедура не затронула случайные наложения длин от других протоколов ввиду высокой регулярности данных RTP-сеанса.

## Обсуждение результатов

Для удобства анализа и интерпретации результатов распределения длин ESP-пакетов визуализировались в виде графиков для интервалов Δt = 0,5 секунд и длиной пакета, деленной на $Bl$.

*Samba и FTP*

На рисунке 4 представлено распределение длин пакетов протоколов Samba и FTP (желтый и красный цвета обозначают направления передачи, оранжевым цветом обозначено совпадение значений).



a) b)

**Рис. 4. Распределение длин ESP-пакетов, содержащих сессии Samba (a) и FTP (b) во время передачи файлов**

*Fig. 4. Distribution of ESP Packet Lengths Containing Samba (a) and FTP (b) Sessions During File Transfer*

При загрузке файла с сервера с использованием протоколов на базе TCP в трафике *загружающего* всегда будут преобладать пакеты подтверждений получения очередных порций данных (TCP Acknowledgment). В пакетах *от сервера* всегда будет преобладать максимальный возможный размер пакетов, определяемый максимальным размером передаваемого по сети сегмента Maximum Segment Size (MSS), который, в свою очередь, определяется минимальным значением MTU сети на пути следования пакетов. Таким образом, полученное значение в поле MSS TCP-пакетов от одного отправителя будет отличаться у разных адресатов в зависимости от их способа доступа в сеть [12]. Такой баланс свойственен всем протоколам с короткими запросами и объемными ответами: FTP, Samba, HTTP.

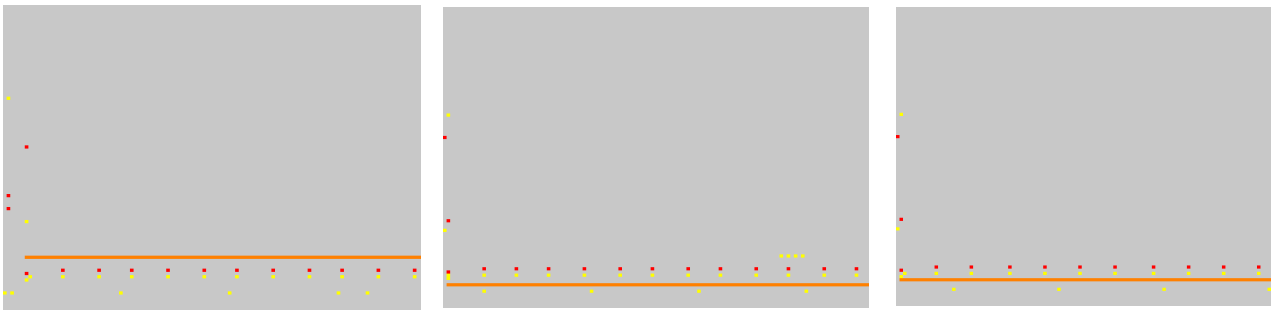
Однако для каждого из них могут выявляться характерные закономерности длин запросов и ответов на них. В Linux реализации протокола Samba версии 2 заметны закономерности передачи трафика: пакет Read AndX Response фиксированной длины всегда предваряет серию пакетов максимального размера объемной передачи данных.

*VoIP*

Для части протоколов возможно преобладание пакетов фиксируемого размера и частоты, определяемые ограничениями протоколов или особенностями их реализаций. При передаче VoIP, в обоих направлениях передачи будет наблюдаться стабильная по длине и интервалу времени составляющая длин RTP-пакетов (рисунок 5). Значение суммарной длины получившегося пакета будет определяться размером порции данных кодека, используемого в сеансе, и размером служебных заголовков, в том числе заголовка ESP.

Средний интервал повторения RTP-пакетов во время разговора также стабилен, и определяется реализацией (размером порций), что позволяет на низко загруженных каналах с малым количеством одновременных передач и различных протоколов выявить такие характерные признаки из общего спектра. При использовании RTP PCMA (G.711 с характеристиками: скорость – 64 кбит/с; пакетная скорость – 50 пакетов/с; размер полезной – 160 байт; полный размер RTP-пакета – 172 байта) размер полного пакета ESP ~ 248 байт, период ~ 20 миллисекунд. При использовании кодеков с большей степенью сжатия размер RTP-пакетов может сократиться, или уменьшится частота появления таких пакетов. Несколько VoIP-сеансов в один момент времени не изменят распределения длин пакетов, но кратно своему количеству уменьшат средний интервал между RTP-пакетами. Если представить изображение в виде тепловой с количествами длин в интервале времени, то можно установить моменты наложения нескольких сеансов.

Также, определив параметры передаваемой RTP-сессии, можно вычесть точки, создаваемые ею из общего распределения, тем самым облегчив распознавание оставшихся протоколов.



a) b) c)

**Рис. 5. Распределение длин ESP-пакетов, содержащих VoIP-сеансы с кодеками G.711(a), G.729(b) и G723.1(c)**

*Fig. 5. Distribution of ESP Packet Lengths Containing VoIP Sessions with Codecs G.711(a), G.729(b) and G 723.1(c)*

*RDP*

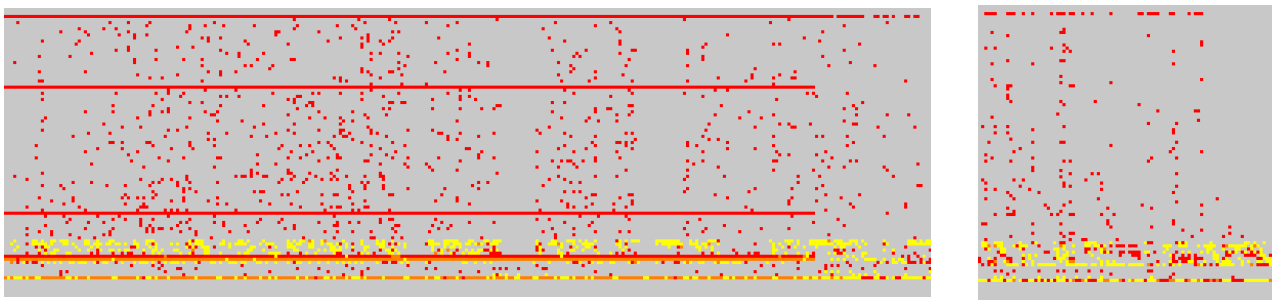
Трафик сеанса удаленного рабочего стола по протоколу RDP имеет широкий спектр встречающихся длин пакетов, однако рассмотрение его отдельно по направлениям даёт возможность выявить различия в наблюдаемых диапазонах значений длин и в объемах передаваемых данных (рисунок 6). На рисунке видно, как на большом количестве сеансов проявляются одновременно 3 линии с постоянными значениями длин. Также заметен четкий верхний предел размеров пакетов клиента желтого цвета.

В статье [13] рассматривается возможность определения типа пользовательской активности исследованием исходного RDP-трафика, анализируются статистические характеристики длин передаваемых исходных пакетов, интервалы и тип протоколов передачи. При передаче трафика RDP в IPsec-тоннеле в зависимости от конфигурации тоннеля (обуславливающей значения $Bl$, $AuT_{ESP}$ и $S_{stak}$) абсолютные значения длин увеличатся на вычисляемые размеры служебных заголовков, а характеристики, связанные только с интервалами активности не изменятся. Таким образом, можно сделать вывод о потенциальной возможности определения типа пользовательской активности в RDP-сеансе, в случае отсутствия в канале других видов трафика, существенно искажающих общую картину распределения в момент анализа.

*BitTorrent*

Клиент BitTorrent во время своей работы использует несколько различных протоколов и принимает входящие соединения от множества других клиентов. При отсутствии активных раздач или загрузок превалирует протокол распределенных хеш-таблиц (DTH, *аббр. от англ.* Distributed Hash Table) [14], использующийся для децентрализованного поиска участников сети. DTH-пакеты присутствуют в обоих направлениях и имеют узкий диапазон длин и, как следствие, создают линии на распределении, кроме того, часто появляются ICMP-пакеты Destination Unreachable от недошедших DHT-пакетов. Также выделяется протокол uTP, пакеты Acknowledgment которого создают четкую линию в распределении длин (рисунок 7). Кроме того, в пакетах BitTorrent в поле DSCP часто устанавливается значение Class Selector 3.

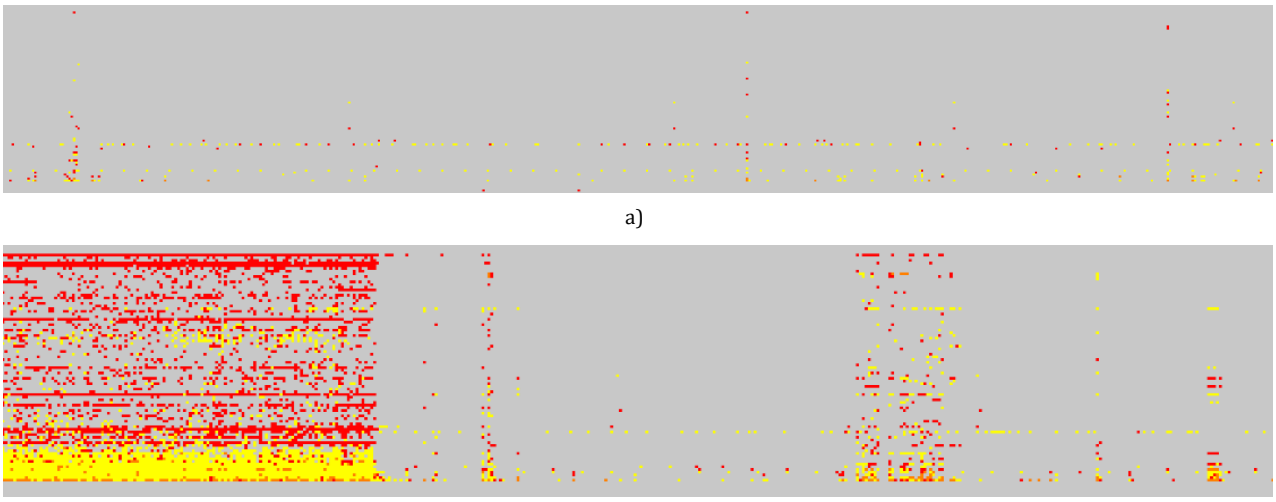
В отличие от остальных рассмотренных протоколов, при функционировании клиента BitTorrent происходит взаимодействие с большим количеством узлов, которые имеют разные типы доступа в Интернет. Следствием этого является широкий диапазон значений параметра Maximum Segment Size для TCP-сессий разных абонентов, что представлено на нижнем рисунке (см. рисунок 7b), где рядом с максимальным значением можно наблюдать несколько линий.



a) b)

**Рис. 6. Распределение длин ESP-пакетов во время сеансов удаленного рабочего стола Windows с горизонтальными линиями (a) и без них (b)**

*Fig. 6. Distribution of ESP Packet Lengths During Windows Remote Desktop Sessions with(a) and Without Horizontal Lines (b)*



a)

b)

**Рис. 7. Распределение длин ESP-пакетов во время функционирования клиента BitTorrent: a) без активных передач, с заметным преобладанием протокола DHT; b) во время загрузки файла**

*Fig. 7. Distribution of ESP Packet Lengths During the Operation of the BitTorrent Client: a) Without Active Transfers, with a Noticeable Predominance of the DHT Protocol; b) During File Upload*

### Оценка продуктивности алгоритма

Качество распознавания конкретного протокола в ШКС определяется номенклатурой и информативностью его признаков, а также выбранным интервалом, на котором эти признаки должны проявляться. В ходе эксперимента наблюдались ложные обнаружения RTP пакетов G.729 в RDP-трафике, что было обусловлено периодическим случайным попаданием длин, содержащих пакеты TCP Acknowledgment и RTP пакетов G.729 в одну градацию длин шифрованных пакетов. При увеличении длительности интервала, на котором делается вывод о наличии в трафике VoIP сессий минимум до 6 с, ложные срабатывания исчезали. Выбор малой длины интервала приведет к росту ложных распознаваний различных протоколов. При выборе слишком большой длины признака повышается вероятность несрабатывания этого признака на коротких сессиях.

Препятствиями применению предлагаемого способа могут быть использование искусственной фрагментации пакетов, рандомизация значений длин пакетов или одновременное использование большого количества разных протоколов, взаимно размывающих признаки друг друга. В статье [15] рассматривается противодействие статистическим атакам на IPsec посредством добавления в каждый транспортирующий пакет дополнительной служебной информации. Такое искусственное добавление данных позволяет рандомизировать объем передачи и обеспечивает возможность дополнительной фрагментации пакетов. Однако такое добавление неинформативных данных имеет ряд недостатков:

– необходимость поддержки такого функционала обеими сторонами коммуникации;

– рандомизация длин передаваемых шифрованных пакетов, без добавления неинформативных пакетов в произвольные моменты времени не скрывает график активности абонентов, сильные всплески передач трафика и приблизительный баланс приемо-передачи;

– использование сопряжено со значительным увеличением объема трафика за счет добавления в него неинформативных данных; увеличение объемов трафика не всегда технически возможно (особенно при использовании мобильных сетей связи) и часто связано с увеличением затрат на передачу этих данных.

Приведенные недостатки ограничивают массовое использование способов противодействия рандомизацией длин, что позволяет применять описываемые подходы в подавляющем большинстве используемых IPsec-каналов.

### Выводы

В работе предложен новый способ определения состава протоколов передающихся в IPsec-канале связи. Он базируется на низкозатратном анализе совокупностей информационных признаков, на основе закономерностей хронологии и длин пакетов.

Конкретизация предлагаемого способа выполнена в виде алгоритма, обеспечивающего обнаружения различных видов трафиков при одновременном нахождении их в ШКС. Продуктивность предложенного способа и реализующего его алгоритма подтверждается результатами эксперимента.

В ходе эксперимента была подтверждена возможность низкологикозатратного определения типа трафика, передаваемого в ШКС. В качестве признаков для описания типа трафика используются одновременно разделенные по направлениям характерные константы длин пакетов данных, их последовательности и периодичности.

Практическая значимость результатов заключается в возможности выявлять факты использования определенных протоколов в ШКС, что позволит, например, операторам связи блокировать, замедлять или отдельно тарифицировать соответствующие виды трафика. Кроме того, классификация типов трафика может реализовываться системами пассивного мониторинга для установления фактов использования на контролируемом участке сети запрещенных протоколов (например, BitTorrent, TOR) и информировании администраторов о таких фактах.

Диапазоны длин ESP-пакетов для одинаковых исходных данных смещаются в зависимости параметров используемого алгоритма шифрования, что может повлиять на точность распознавания различными нейронными сетями [1–4], если они не были корректно обучены на работу в некоторых редко используемых режимах. Для компенсации таких отклонений длин следовало бы обучать нейронные сети нешифрованными образцами трафика для каждой комбинации параметров шифрования в отдельности, вычисляя результирующие длины ESP-пакетов по формуле (1).

Рассмотренный в статье алгоритм может быть адаптирован к другим протоколам VPN, позволяющим точно вычислить объем накладных расходов на добавление служебных заголовков и шифрование пакетов вложенной пользовательской нагрузки. Наиболее актуальными примерами таких протоколов являются Wireguard и OpenVPN в режимах, не использующих сжатие трафика.

**Список источников**


1. Rasteh A., Delpech F., Aguilar-Melchor C., Zimmer R., Shouraki S.B., Masquelier T. Encrypted Internet Traffic Classification Using a Supervised Spiking Neural Network // arXiv preprint arXiv:2101.09818. 2022. URL: https://arxiv.org/pdf/2101.09818 (дата обращения 07.01.2022).
2. Gupta N., Jindal V., Bedi P. Encrypted Traffic Classification Using eXtreme Gradient Boosting Algorithm // Proceedings of the International Conference on Innovative Computing and Communications (ICICC 2021, Delhi, India, February 2021). Advances in Intelligent Systems and Computing (AISC). Vol. 1394. Singapore: Springer, 2022. PP. 225–232. DOI:10.1007/978-981-16-3071-2_20
3. Draper-Gil G., Lashkari A.H., Mamun M., Ghorbani A. Characterization of Encrypted and VPN Traffic Using Time-Related // Proceedings of the 2nd International Conference on Information Systems Security and Privacy (ICISSP, Rome, Italy, 19–21 February 2016). 2016. PP. 407–414. DOI:10.5220/0005740704070414
4. Islam F.U., Liu G., Liu W. Identifying VoIP traffic in VPN tunnel via Flow Spatio-Temporal Features // Mathematical Biosciences and Engineering. 2020. Vol. 17. Iss. 5. PP. 4747–4772. DOI:10.3934/mbe.2020260
5. Kent S., Seo K. Security Architecture for the Internet Protocol. No. rfc4301. 2005.
6. Atkinson R. IP Encapsulating Security Payload (ESP). No. rfc1827. 1995.
7. Xenakis C., Laoutaris N., Merakos L., Stavrakakis I. A generic characterization of the overheads imposed by IPsec and associated cryptographic algorithms // Computer Networks. 2006. Vol. 50. Iss. 17. PP. 3225–3241. DOI:10.1016/j.comnet.2005.12.005
8. Дмитренко А. Изучаем и выявляем уязвимости протокола IPsec // Хакер. 2015 URL: https://xakep.ru/2015/05/13/ipsec-security-flaws (дата обращения 06.04.2022)
9. Akhter A. IPSec Packet Size Calculator // Cisco Community. URL: https://community.cisco.com/legacyfs/online/legacy/4/8/7/27784-IPSec_Calculator_NAT_GRE-Key.htm (дата обращения 24.01.2022)
10. encapcalc // GitHub. URL: http://github.com/dmbaturin/encapcalc (дата обращения 09.06.2022)
11. Pérez J.A., Cabrera V.Z.C., Jenecek J. Quality of Service Analysis of site to site for IPSec VPNs for realtime multimedia traffic // Advanced Int'l Conference on Telecommunications and Int'l Conference on Internet and Web Applications and Services (AICT-ICIW'06, Guadelope, French Caribbean, 19–25 February 2006). 2006. URL: https://www.its.bldrdoc.gov/media/33388/per_j_slides1.pdf [Accessed 16th January 2022]
12. Ишкуватов С.М., Комаров И.И. Анализ аутентичности трафика на основании данных цифровых отпечатков реализаций сетевых протоколов // Научно-технический вестник информационных технологий, механики и оптики. 2020. Т. 20. № 5. С. 747–754. DOI:10.17586/2226-1494-2020-20-5-747-754
13. Lapczyk L., Skillicorn D.B. Activity Detection from Encrypted Remote Desktop Protocol Traffic // arXiv preprint arXiv:2008.02685. 2020. DOI:10.48550/arXiv.2008.02685
14. Urdaneta G., Pierre G., Steen M.V. A survey of DHT security techniques // ACM Computing Surveys. 2011. Vol. 43. Iss. 2. PP. 1–49. DOI:10.1145/1883612.1883615
15. Kiraly C., Teofili S., Bianchi G., Cigno R.L., Nardelli M., Delzeri E. Traffic Flow Confidentiality in IPsec: Protocol and Implementation // Proceedings of the 3rd IFIP WG 9.2, 9.6/11.6, 11.7/FIDIS International Summer School on the Future of Identity in the Information Society (Karlstad University, Sweden, 4–10 August 2007). The International Federation for Information Processing. Vol. 262. Boston: Springer, 2007. PP. 311–324. DOI:10.1007/978-0-387-79026-8_22

**References**

1. Rasteh A., Delpech F., Aguilar-Melchor C., Zimmer R., Shouraki S.B., Masquelier T. Encrypted Internet Traffic Classification Using a Supervised Spiking Neural Network. *arXiv preprint arXiv:2101.09818*. 2022. URL: https://arxiv.org/pdf/2101.09818 [Accessed 07.01.2022]
2. Gupta N., Jindal V., Bedi P. Encrypted Traffic Classification Using eXtreme Gradient Boosting Algorithm. *Proceedings of the International Conference on Innovative Computing and Communications, ICICC 2021, February 2021, Delhi, India. Advances in Intelligent Systems and Computing (AISC).* Singapore: Springer; 2022. vol.1394. p.225–232. DOI:10.1007/978-981-16-3071-2_20
3. Draper-Gil G., Lashkari A.H., Mamun M., Ghorbani A. Characterization of Encrypted and VPN Traffic Using Time-Related. *Proceedings of the 2nd International Conference on Information Systems Security and Privacy, ICISSP, 19–21 February 2016, Rome, Italy*. 2016. p.407–414. DOI:10.5220/0005740704070414
4. Islam F.U., Liu G., Liu W. Identifying VoIP traffic in VPN tunnel via Flow Spatio-Temporal Features. *Mathematical Biosciences and Engineering*. 2020;17(5):4747–4772. DOI:10.3934/mbe.2020260
5. Kent S., Seo K. *Security Architecture for the Internet Protocol*. No. rfc4301. 2005.
6. Atkinson R. IP *Encapsulating Security Payload (ESP)*. No. rfc1827. 1995.
7. Xenakis C., Laoutaris N., Merakos L., Stavrakakis I. A generic characterization of the overheads imposed by IPsec and associated cryptographic algorithms. *Computer Networks*. 2006;50(17):3225–3241. DOI:10.1016/j.comnet.2005.12.005
8. Dmitrenko A. Studying and identifying IPsec protocol vulnerabilities. *Khaker*. 2015. (in Russ.) URL: https://xakep.ru/2015/05/13/ipsec-security-flaws [Accessed 06th April 2022]
9. Akhter A. IPSec Packet Size Calculator. *Cisco Community*. URL: https://community.cisco.com/legacyfs/online/legacy/4/8/7/27784-IPSec_Calculator_NAT_GRE-Key.htm [Accessed 24th January 2022]
10. GitHub. *encapcalc*. URL: http://github.com/dmbaturin/encapcalc [Accessed 09th June 2022]
11. Pérez J.A., Cabrera V.Z.C., Jenecek J. Quality of Service Analysis of site to site for IPSec VPNs for realtime multimedia traffic. *Advanced Int'l Conference on Telecommunications and Int'l Conference on Internet and Web Applications and Services AICT-ICIW'06, 19–25 February 2006, Guadelope, French Caribbean*. 2006. URL: https://www.its.bldrdoc.gov/media/33388/per_j_slides1.pdf [Accessed 16th January 2022]
12. Ishkuvatov S.M., Komarov I.I. Traffic Authenticity Analysis Based on Digital fingerprint Data of Network Protocol Implementations. Scientific and Technical Journal of Information Technologies, Mechanics and Optics. 2020;20(5):747–754. DOI:10.17586/2226-1494-2020-20-5-747-754
13. Lapczyk L., Skillicorn D.B. Activity Detection from Encrypted Remote Desktop Protocol Traffic. *arXiv preprint arXiv:2008.02685*. 2020. DOI:10.48550/arXiv.2008.02685
14. Urdaneta G., Pierre G., Steen M.V. A survey of DHT security techniques. *ACM Computing Surveys*. 2011;43(2):1–49. DOI:10.1145/1883612.1883615
15. Kiraly C., Teofili S., Bianchi G., Cigno R.L., Nardelli M., Delzeri E. Traffic Flow Confidentiality in IPsec: Protocol and Implementation. *Proceedings of the 3rd IFIP WG 9.2, 9.6/11.6, 11.7/FIDIS International Summer School on the Future of Identity in the Information Society, 4–10 August 2007, Karlstad University, Sweden. The International Federation for Information Processing. vol.262*. Boston: Springer; 2007. p.311–324. DOI:10.1007/978-0-387-79026-8_22

Статья поступила в редакцию 11.09.2022; одобрена после рецензирования 25.11.2022; принята к публикации 28.11.2022.
The article was submitted 11.09.2022; approved after reviewing 25.11.2022; accepted for publication 28.11.2022.

## Информация об авторе:

**Ишкуватов Сергей Маратович**
аспирант факультета безопасности информационных технологий Национального исследовательского университета ИТМО
https://orcid.org/0000-0002-4006-3693