УДК 004.891.3 Вестник УрФО № 1(43) / 2022, с. 74–83

Старун И.Г., Югансон А.Н.

DOI: 10.14529/secur220109

РАЗРАБОТКА АЛГОРИТМА КЛАССИФИКАЦИИ ШИФРОВАННОГО ТРАФИКА НА ОСНОВЕ LIGHTGBM[1]

С ростом количества угроз в сети Интернет растет и популярность технологии шифрования. При этом часть полезной нагрузки в результате шифрования перестает быть видимой. Для эффективной реализации многих сценариев обеспечения ИБ требуется идентификация протокола шифрования и типа приложения, поэтому актуальной становится задача классификации шифрованного трафика. Лидирующее по популярности место среди способов классификации занимает машинное обучение. При этом наилучшие результаты достигаются с помощью глубокого обучения, но этот подход имеет и обратную сторону – высокую вычислительная сложность, требующую больших ресурсов для работы в режиме реального времени. Поэтому в данном исследовании внимание сфокусировано на классификации шифрованного трафика с помощью классических алгоритмов машинного обучения. Рассмотрена первая часть сценария классификации – разделение трафика на VPN и non-VPN. Предложен алгоритм поиска оптимальной модели с помощью AutoML. В результате получена модель на основе алгоритма LightGBM. Эксперименты проводились на основе известного набора данных ISCXVPN2016. Оценка качества на тестовой выборке показала следующие результаты: Accuracy = 94.08%, Precision = 92.85%, Recall = 96.07%, F1-measure = 94.43%. Эти оценки превосходят предыдущие решения по 3 из 4 ключевых метрик классификации.

Ключевые слова: *шифрованный трафик, классификация трафика, машинное обучение, VPN, AutoML, градиентный бустинг, LightGBM, информационная безопасность.*

Starun I.G., Iuganson A.N.

DEVELOPMENT OF THE ALGORITHM FOR CLASSIFICATION OF ENCRYPTED TRAFFIC BASED ON LIGHTGBM

As the number of threats on the Internet grows, so does the popularity of encryption technology. In this case, part of the payload because of encryption ceases to be visible. The effective implementation of many information security scenarios requires identification of the encryption protocol and application type, so the task of classifying encrypted traffic becomes relevant. Machine learning is one of the most popular classification methods. At the same time, the best results are achieved using deep learning, but this approach also has a downside - high computational complexity, which requires large resources to work in real time. Therefore, in this study, attention is focused on the classification of encrypted traffic using classical machine learning algorithms. The first part of the classification scenario is considered – the division of traffic into VPN and non-VPN. An algorithm for finding the optimal model using AutoML is proposed. As a result, a model based on the LightGBM algorithm was obtained. The experiments were carried out on the well-known ISCXVPN2016 dataset. The quality assessment on the test sample showed the following results: Accuracy = 94.08%, Precision = 92.85%, Recall = 96.07%, F1-measure = 94.43%. These scores outperform previous solutions on 3 out of 4 key classification metrics.

Keywords: *encrypted traffic, traffic classification, machine learning, VPN, AutoML, gradient boosting, LightGBM, information security.*

[1] Работа выполнена в Университете ИТМО при финансовой поддержке Министерства науки и высшего образования Российской Федерации в рамках проекта 2019-0898 «Многоуровневое управление сложными техническими системами».

Введение

Количество шифрованного трафика в Интернете растет от года к году, а доля использования HTTPS протокола приближается к 90% и продолжает расти. Такая популярность шифрования связана с возросшей потребностью в обеспечении информационной безопасности, а использование технологии VPN также дает пользователям возможность обходить местные блокировки ресурсов и анонимизировать свой цифровой след [1, 16–17]. В результате шифрования часть сведений перестает быть видимой, что усложняет задачу классификации трафика. Для эффективной реализации многих сценариев защиты от угроз ИБ в сетях при анализе сетевого трафика зачастую требуется идентификация протокола шифрования и типа приложения, к которому этот трафик относится. Такая классификация полезна не только для предотвращения атак и обнаружения аномалий, но и для анализа поведения пользователя, управления трафиком, контроля производительности приложений [4, 13].

Выделяют 3 подхода к классификации сетевого трафика – на основе анализа портов [5, 6], путем анализа полезной нагрузки [5, 7–8] и с помощью машинного обучения. Первый сопоставляет каждое приложение с соответствующим номером порта (например, порт 20 для FTP). Актуальность такого подхода заметно снизилась из-за внедрения динамического распределения портов. Второй способ слабо подходит для шифрованного трафика, так как выделить полезную нагрузку после шифрования практически невозможно. Поэтому наиболее популярным подходом к классификации шифрованного трафика в последние годы стало использование машинного обучения. Его можно разделить на 2 большие группы: классификация на основе классических алгоритмов и глубокое обучение (deep learning).

Широкое применение для решения задачи классификации шифрованного трафика приобрели методы глубокого обучения с помощью нейросетей. С их помощью достигаются наилучшие результаты. Однако у их использования есть и обратная сторона – высокая вычислительная сложность. Если крупные корпорации и компании могут себе позволить использование нейросетей для анализа трафика, то для малого и среднего бизнеса зачастую это становится непозволительной роскошью. Следовательно, при выборе метода анализа трафика необходимо достичь баланса между качеством классификации и вычислительной сложностью [2].

Помимо глубокого обучения, задачи классификации решаются с помощью классических алгоритмов машинного обучения, таких как логистическая регрессия, деревья решений, случайный лес, градиентный бустинг и других. Модели на основе этих алгоритмов значительно менее требовательны к вычислительным ресурсам, поэтому могут использоваться пользователями с более низким порогом входа.

Предметом настоящего исследования является классификация шифрованного трафика с помощью классических алгоритмов машинного обучения без использования deep learning. Предложен алгоритм поиска оптимальной модели с применением инструментов AutoML для тонкой настройки алгоритма градиентного бустинга LightGBM и выбора признаков на основе оценки их влияния на итоговую модель.

Дальнейшая часть статьи организована следующим образом. Раздел 2 представляет собой обзор предыдущих работ. В разделе 3 описывается набор данных, на основе которого проводились эксперименты. Четвертый раздел посвящен методологии исследования. В разделе 5 представлены результаты эксперимента. Шестая часть работы отведена под обсуждение полученных результатов и их сравнение с предыдущими решениями. В заключительной части подводятся итоги и обсуждаются дальнейшие перспективы.

Обзор предыдущих работ

Первая статья, в которой упоминается датасет ISCXVPN2016, была опубликована в 2016 году. Авторы сгенерировали большой объем шифрованного трафика, извлекли из него временные признаки и на их основе построили классификаторы. В качестве метрик были выбраны Precision и Recall. В задаче разделения трафика на VPN и non-VPN лучший результат был достигнут при использовании алгоритмов KNN и C4.5 – 0.89–0.9 в зависимости от вида трафика [9].

Большое количество работ посвящено классификации с помощью глубокого обучения [18-23]. Целью настоящего исследования является повышение качества классификации шифрованного трафика в условиях ограниченных вычислительных мощностей, поэтому было принято решение отказаться от использования нейронных сетей. К тому же существует мнение, что высокие показатели, демонстрируемые нейронными сетями, связаны с их способностью к адаптации к конкретному набору данных [15], что затрудняет масштабирование и перенос модели в новые условия, так как при этом результаты могут значительно упасть. Однако среди подобных работ отдельно стоит отметить статью [3], в которой авторы первыми предложили использовать AutoML для поиска оптимального решения данной задачи. Они применили его для получения наиболее эффективной архитектуры нейронной сети. В настоящем исследовании предложенный подход был адаптирован для тонкой настройки гиперпараметров модели.

В статье [10] авторы сосредоточились на второй части задачи классификации шифрованного трафика – идентификации приложения, сгенерировавшего шифрованный трафик, уже после разделения на VPN и non-VPN. В качестве основной метрики была ис-

пользована доля правильных ответов (Accuracy), а наилучшие результаты показали алгоритмы на основе градиентного бустинга – 89,03% для VPN и 93,19% для non-VPN.

В работе [11] авторы поднимают тему эффективного отбора признаков (Feature selection) и сокращения размерности для снижения вычислительной сложности итоговых моделей. С этой целью использован метод анализа основных компонентов (PCA) и метод опорных векторов (SVM) для выбора признаков из набора данных. Другой подход к сокращению сложности вычислений рассмотрен в [12]. Авторы предложили предварительно отбирать ключевые признаки методами дисперсионного анализа (ANOVA) и опорных векторов (SVM).

Переход от тяжелых нейронных сетей к усовершенствованной предобработке данных (временных признаков) предложен в статье [13]. Авторы предварительно используют метод DSSR для перемасштабирования временных диапазонов, чтобы затем использовать стандартные классификаторы. В комбинации с корреляционным анализом для выбора итогового набора признаков получены результаты, значительно превышающие предшественников. Основным недостатком предложенного метода является то, что он может преобразовывать признаки только поблочно, что вызывает задержку обнаружения, зависящую от длины окна. К тому же требует отдельного внимания вопрос утечки информации (Data leakage) о распределении в тестовой выборке в обучающий набор, так как в исследовании сначала реализовано преобразование данных, а уже затем разделение на Train и Test.

В одной из последних на момент проведения исследования работ [14] реализована комбинация нескольких методов машинного обучения для получения оптимальной модели. Авторы сначала нормализуют данные, затем выбирают 15 признаков, оказывающих наибольшее значение на результат. Затем данные балансируются, чтобы избежать проблемы несбалансированности классов, после чего подбираются оптимальные гиперпараметры модели. Лучший результат был получен с помощью алгоритма XGBoost – каждая из метрик precision, recall, accuracy и f1-measure немного превысила 93%.

Описание набора данных

Как было сказано выше, набор данных ISCXVPN2016 был представлен и описан в исследовании [9]. Чтобы создать репрезентативный набор данных, авторы зафиксировали реальный трафик, созданный участниками лаборатории. Они создали учетные записи для пользователей Алисы и Боба, чтобы они могли пользоваться такими сервисами, как Skype, Facebook и т. д. Полный список из 7 захваченных протоколов и приложений представлен в таблице 1. Для каждого из них трафик был сгенерирован двумя способами: путем обычного сеанса и сеанса через VPN. Таким образом, был получен набор из 14 категорий трафика общим объемом 28 ГБ. Для захвата использовались утилиты Wireshark и tcpdump.

Таблица 1

Перечень захваченных протоколов и приложений для ISCXVPN2016

Трафик	Содержимое
Web browsing	Firefox и Chrome
Email	SMPTS, POP3S и IMAPS
Chat	ICQ, AIM, Skype, Facebook и Hangouts
File Transfer	Skype, FTPS и SFTP с помощью Filezilla и внешней службы
Streaming	Vimeo и Youtube
VoIP	Facebook, Skype и голосовые звонки Hangouts (длительностью 1 час)
P2P	uTorrent и Transmission (Bittorrent)

Сгенерированный трафик далее рассматривался как двунаправленный поток, где под потоком следует понимать последовательность пакетов с одинаковыми значениями исходного IP-адреса, IP-адреса назначения, исходного порта, порта назначения и протокола (TCP или UDP). С помощью программы ISCXFlowMeter потоки разбивались на отрезки одинаковой временной продолжительности (timeout), по которым затем рассчитывались значения признаков. Полный список полученных функций и их описание представлено в таблице 2. Всего было использовано 4 значения timeout – 15, 30, 60 и 120 секунд.

Затем в оригинальной статье авторы выделили и протестировали 2 сценария классификации:

1. Сценарий А: сначала реализуется классификация трафика на VPN и non-VPN (сценарий А1), а затем проводится раздельная классификация этих двух видов трафика по типам приложений и протоколов (сценарий А2).

2. Сценарий Б: единый набор трафика сразу классифицируется по типам приложений и протоколов без предварительного разделения на VPN и non-VPN.

Настоящее исследование сфокусировано только на сценарии А1, а именно на задаче разделения шиф-

Таблица 2

Список временных признаков в ISCXVPN2016

Группа признаков	Признаки	Описание
Duration	duration	Длительность потока
Fiat (Forward Inter Arrival Time)	total_fiat, max_fiat, min_fiat, mean_fiat,	Время между двумя пакетами, отправляемыми в прямом направлении (всего, минимальное, максимальное, среднее)
Biat (Backward Inter Arrival Time)	total_biat, max_biat, min_biat, mean_biat	Время между двумя пакетами, отправляемыми в обратном направлении (всего, минимальное, максимальное, среднее)
Flowiat (Flow Inter Arrival Time)	mean_flowiat, max_flowiat, min_flowiat, std_flowiat	Время между двумя пакетами, отправленными в любом направлении (среднее, минимальное, максимальное, стандартное отклонение)
Active	mean_active, max_active, min_active, std_active	Время, в течение которого поток был активен до перехода в режим ожидания (среднее, минимальное, максимальное, стандартное отклонение)
Idle	mean_idle, max_idle, min_idle, std_idle	Время, в течение которого поток простаивал до того, как стал активным (среднее, минимальное, максимальное, стандартное отклонение)
Fb-psec	FlowBytesPerSecond	Количество байт потока в секунду
Fp-psec	FlowPktsPerSecond	Количество пакетов потока в секунду

рованного трафика на VPN и non-VPN. Этот этап классификации напрямую влияет на общий процесс, так как чем точнее предварительная классификация трафика, тем качественнее данные, поступающие на вход сценария A2.

Этапы проведения исследования (методология)

Для проведения исследования был использован язык программирования Python 3.8. Всего для получения итоговой модели классификации было выполнено 7 шагов:

1. Разделение исходного набора данных на обучающую и тестовую выборки в соотношении 80:20 с помощью функции train_test_split из библиотеки sklearn. При этом проводилась стратификация по целевому признаку, чтобы представленность каждого класса в выборках была сопоставимой. Тестовая выборка в дальнейшем использовалась только на этапе оценки итоговой модели.

2. Стандартизация данных с помощью StandartScaler по формуле (1):

$$x_{norm_i} = \frac{(x_i - x_{mean})}{SD}, \quad (1)$$

где x_i – исходное значение признака в выборке, x_{mean} – среднее значение признака в обучающем наборе, SD – стандартное отклонение признака в обучающем наборе.

3. Поиск оптимальной модели и ее гиперпараметров с помощью LightAutoML. Рассматривались такие модели, как CatboostClassifier, LGBMClassifier и линейные классификаторы. При этом комбинация нескольких моделей (бэггинг) не рассматривалась, чтобы не слишком усложнять модель.

4. Выбор признаков (Feature selection) с помощью LightAutoML. Использовался быстрый метод, который рассчитывает важность функций по встроенному методу LGBM.

5. Подбор оставшихся гиперпараметров. Этот шаг необходим ввиду того, что инструмент LightAutoML для экономии времени уделяет недостаточно внимания подбору некоторых ключевых параметров, таких как количество деревьев (num_estimators), скорость обучения (learning_rate) и максимальная глубина дерева (max_depth).

5. Компоновка итоговой модели и ее обучение на тренировочной выборке. Выбор оптимальной границы разделения классов (threshold).

6. Оценка качества итоговой модели на тестовой выборке. В качестве метрик использовались классические метрики для задачи классификации, которые уже упоминались в работе: Accuracy, Precision, Recall и F1-Measure.

Результаты исследования

Как и в большинстве предыдущих исследований, для сценария A1 наиболее качественные решения были получены при обработке потоков с таймаутом в 15 секунд.

Таблица 3

Оптимальные гиперпараметры модели LGBMClassifier

Гиперпараметр	Значение
feature_fraction	0.6872700594236812
num_leaves	244
bagging_fraction	0.8659969709057025
min_sum_hessian_in_leaf	0.24810409748678125
reg_alpha	2.5361081166471375e-07
reg_lambda	2.5348407664333426e-07
learning_rate	0.15
max_depth	50
num_estimators	2300

Наилучшие результаты были продемонстрированы моделью градиентного бустинга LGBMClassifier. Оптимальные гиперпараметры для модели представлены в таблице 3.

При этом оптимальная граница разделения классов на обучающем наборе составила 0.42. В дальнейшем она была использована при оценке итоговой модели.

Оценка важности признаков (feature importance) для модели представлена на рис. 1. Из исходных 25 признаков было принято решение оставить первые 17 по важности, так как в этом случае достигались лучшие метрики на обучающем наборе.

Итоговый перечень выбранных для модели признаков выглядит следующим образом: 'duration', 'total_fiat', 'total_biat', 'min_fiat', 'min_biat', 'max_fiat', 'max_biat', 'mean_fiat', 'mean_biat', 'flowPktsPerSecond', 'flowBytesPerSecond', 'min_flowiat', 'mean_active', 'mean_idle', 'max_flowiat', 'mean_flowiat' и 'std_flowiat'.

Для оценки качества модели использовались

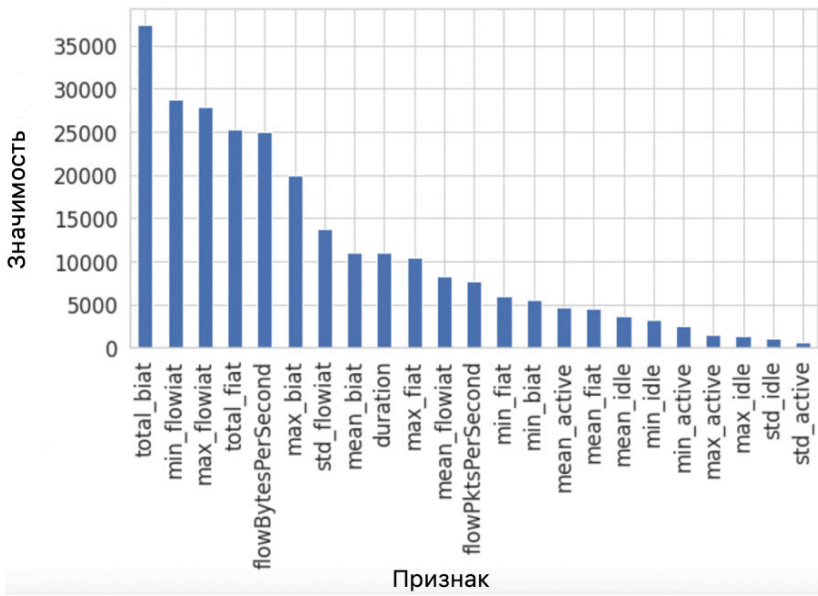

Рис. 1. Оценка важности признаков для модели встроенным методом LGBM

стандартные метрики классификации – Accuracy (2), Precision (3), Recall (4) и F1-measure (5).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (2)$$

$$Precision = \frac{TP}{TP + FP}, \quad (3)$$

$$Accuracy = \frac{TP}{TP + FN}, \quad (4)$$

$$F_1 - measure = \frac{Precision * Recall}{Precision + Recall}, \quad (5)$$

где TP – True Positive – количество истинно-положительных ответов, TN – True Negative – количество истинно-отрицательных ответов, FN – False Negative – количество ложноотрицательных ответов, FP – False Positive – количество ложноположительных ответов.

Оценка качества модели на тестовой выборке

показала следующие результаты: Accuracy = 94.08%, Precision = 92.85%, Recall = 96.07%, F1-measure = 94.43%.

ROC-кривая итоговой модели представлена на рисунке 2. Площадь под графиком, или AUC, составляет 0.98, что что эквивалентно доле пар объектов противоположных классов (VPN и non-VPN), которые модель верно упорядочила.

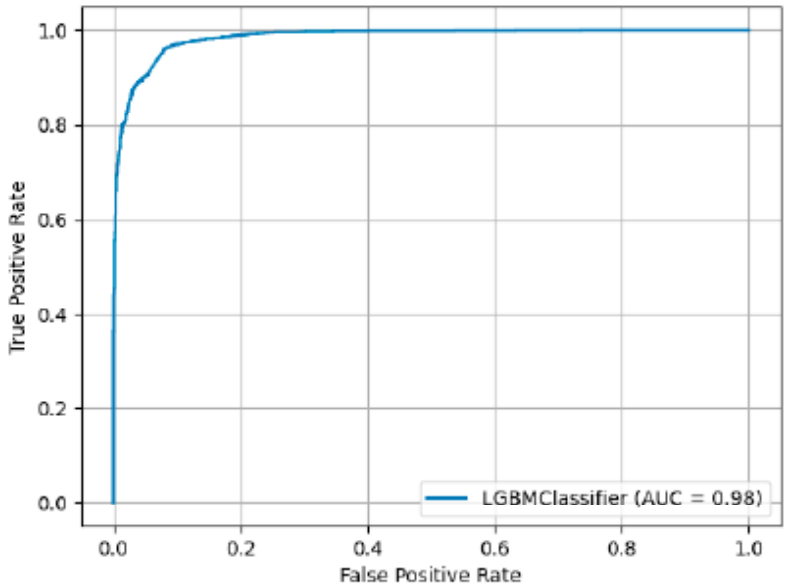

Рис. 2. ROC-AUC кривая полученной модели

Таблица 4

Сравнение качества классификации с другими решениями

Исследование	Год	Количество признаков	Accuracy	Precision	Recall	F1-measure
[9]	2016	25	-	90.6%	-	-
[3]	2021	25	-	85.21%	84.91%	85.57%
[10]	2021	8	88%	-	-	-
[14]	2021	15	93.02%	93.04%	93.02%	93.03%
Настоящее исследование	2022	17	94.09%	92.85%	96.07%	94.43%

Обсуждение результатов

Сравнение качества классификации предложенной модели с другими решениями без использования глубокого обучения представлено в таблице 4 и на рис. 3. Прочерки в таблице и пустые колонки на графике означают отсутствие данных.

Таким образом, предложенное решение превосходит предыдущие исследования по 3 из 4 ключевых метрик (Accuracy, Recall и F1-measure), незначительно уступая лишь по метрике Precision работе [14]. Полученная модель демонстрирует лучшую долю правильных ответов, а также лучшее гармоническое среднее между точностью и полнотой.

Важно отметить, что в зависимости от приоритетов классификации можно влиять на метрики Precision и Recall с помощью сдвига границы классификации. Предложенное в работе значение threshold, равное 0.42, можно считать оптимальным для полученной модели.

Еще одно интересное замечание связано с тем, как перекликается выбор признаков в текущем исследовании с работой [14]. Там авторы остановились на 15 лучших признаках, из которых 13 входят в полученный в настоящей работе итоговый список. При этом их эксперименты показали, что группы признаков Active и Idle не оказывают значительного влияния на качество классификации, в то время как в текущем исследовании было принято включить их в итоговую модель как значимые.

Заключение

Предложенный в работе алгоритм подбора оптимальной модели для классификации шифрованного трафика показал высокие результаты, опережая по большинству ключевых метрик предыдущие решения на основе классического машинного обучения. Это говорит о высокой эффективности AutoML подхода для поиска оптимальных параметров алгоритмов машин-

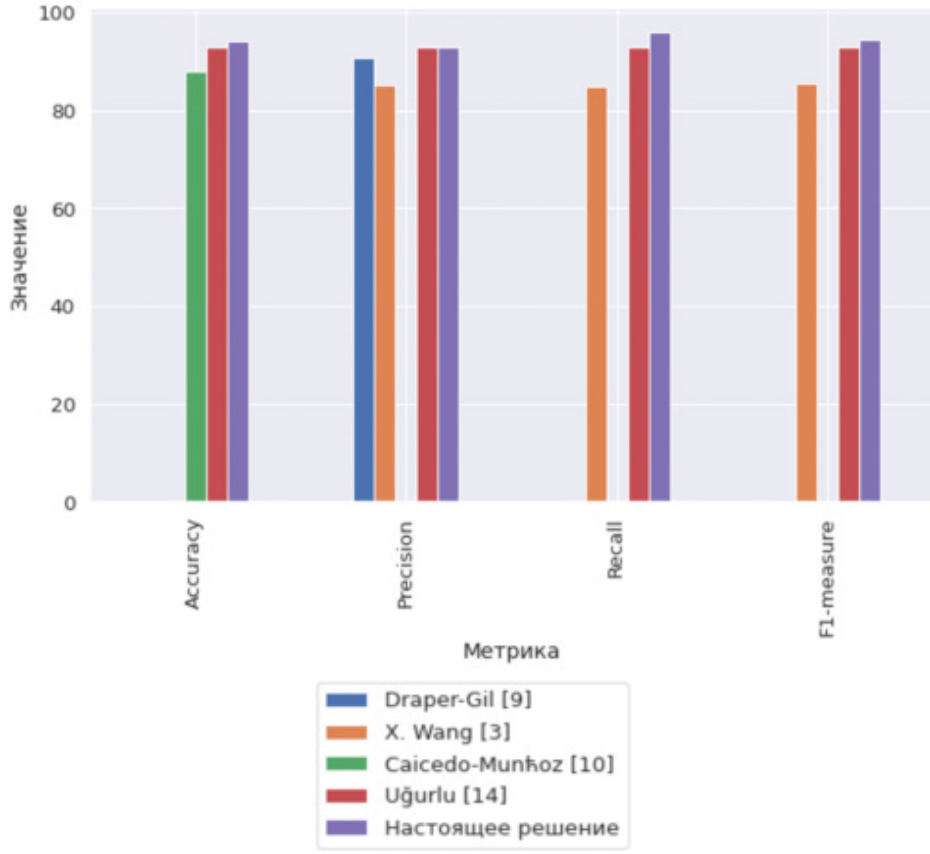

Рис. 3. Сравнение с существующими решениями

ного обучения и выбора признаков. Еще одной ключевой особенностью разработанного решения является выбор алгоритма LightGBM в качестве классификатора. В предыдущих работах также фигурировали алгоритмы на основе градиентного бустинга, но предпочтение отдавалось моделям CatBoost и XGBoost.

Дальнейшие исследования будут сосредоточены на второй части сценария A для идентификации конкретного приложения или протокола, сгенерировавшего шифрованный трафик.

Литература

1. Старун И.Г., Югансон А.Н., Гатчин Ю.А. Построение математической модели расчета комплексной оценки VPN: ст. - Вестник ТГТУ, том. 25, выпуск 4, 2019, с. 535-546.

2. Lu,B.; Luktarhan,N.; Ding, C; Zhang,W. ICLSTM: Encrypted Traffic Service Identification Based on Inception-LSTM Neural Network. Symmetry, 2021, 13, 1080. https:// doi.org/10.3390/sym13061080

3. X. Wang et al., "Evolutionary Algorithm-Based and Network Architecture Search-Enabled Multiobjective Traffic Classification," in IEEE Access, vol. 9, pp. 52310-52325, 2021

4. B. Yamansavascilar, M. A. Guvensan, A. G. Yavuz, and M. E. Karsligil, "Application identification via network traffic classification", In: Proc. of 2017 Int. Conf. Comput. Netw. Commun. ICNC 2017, pp. 843–848, 2017.

5. Megantara, A.A., Ahmad, T. ANOVA-SVM for Selecting Subset Features in Encrypted Internet Traffic Classification (2021) International Journal of Intelligent Engineering and Systems, 14 (2), pp. 536-546.

6. G. Cheng and S. Wang, "Traffic classification based on port connection pattern", In: Proc. of 2011 Int. Conf. Comput. Sci. Serv. Syst. CSSS 2011 - Proc., pp. 914–917, 2011.

7. H. K. Lim, J. B. Kim, K. Kim, Y. G. Hong, and Y. H. Han, "Payload-based traffic classification using multi-layer LSTM in software defined networks", Appl. Sci., Vol. 9, No. 12, 2019.

8. F. Dehghani, N. Movahhedinia, M. R. Khayyambashi, and S. Kianian, "Real-time traffic classification

based on statistical and payload content features", In: Proc. of - 2010 2nd Int. Work. Intell. Syst. Appl. ISA 2010, pp. 26–29, 2010.

9. G. Draper-Gil, A. H. Lashkari, M. Mamun, and A. A. Ghorbani,"Characterization of encrypted and VPN traffic using time-related features," in Proc. 2nd Int. Conf. Inf. Syst. Security Privacy, 2016,pp. 407–414

10. Caicedo-Munћoz JA, Espino AL, Corrales JC, RendoћЬn A.QoS-Classifier for VPN and Non-VPN tranћAic based on time-related features.Computer Networks 2018; 144: 271-279. doi: 10.1016/j.comnet.2018.08.008

11. A. Saber, B. Fergani, and M. Abbas, "Encrypted Traffic Classification: Combining Over-and Under-Sampling through a PCA-SVM", In: Proc. of - PAIS 2018 Int. Conf. Pattern Anal. Intell. Syst., pp. 1–5, 2018.

12. Achmad Akbar Megantara, Tohari Ahmad, ANOVA-SVM for Selecting Subset Features in Encrypted Internet Traffic Classification. International Journal of Intelligent Engineering and Systems, Vol.14, No.2, 2021

13. R. Nigmatullin, A. Ivchenko and S. Dorokhin, "Differentiation of Sliding Rescaled Ranges: New Approach to Encrypted and VPN Traffic Detection," 2020 International Conference Engineering and Telecommunication (En&T), 2020, pp. 1-5, doi: 10.1109/EnT50437.2020.9431285

14. Uğurlu, M., Doğru, İ.A., Arslan, R.S. A new classification method for encrypted internet traffic using machine learning (2021) Turkish Journal of Electrical Engineering and Computer Sciences, 25 (9), pp. 2450-2468.

15. Felipe Peter. Analysis of the ISCX VPN-nonVPN Dataset 2016 for Encrypted Network Traffic Classification, Tsinghua University, [Электронный ресурс] - pp. 1-5, 2018

16. A Glance through the VPN Looking Glass: IPv6 Leakage and DNS Hijacking in Commercial VPN Clients / V. C. Perta [et al.] // Proceedings of Conference: 15th Privacy Enhancing Technologies, 30 June – 02 July 2015, Philadelphia, USA. – Philadelphia, 2015. – P. 77 – 91.

17. Brissaud P, Francћlois J, Chrisment I, Cholez T, Bettan O. Transparent and Service-Agnostic Monitoring of Encrypted Web TranћAic. IEEE Transactions on Network and Service Management 2019; 16 (3): 842-856

18. Lu, B., Luktarhan, N., Ding, C., Zhang, W. ICLSTM: Encrypted traffic service identification based on inception-LSTM neural network (2021) Symmetry, 13 (6), ст. № 1080

19. Wang, W.; Zhu, M.; Wang, J.; Zeng, X.; Yang, Z. End-to-end encrypted traffic classification with one-dimensional convolution neural networks. In Proceedings of the 2017 IEEE International Conference on Intelligence and Security Informatics (ISI), Beijing, China, 22–24 July 2017; pp. 43–48.

20. Lotfollahi, M.; Siavoshani, M.J.; Zade, R.S.H.; Saberian, M. Deep packet: A novel approach for encrypted traffic classification using deep learning. Soft Comput. 2020, 24, 1999–2012.

21. Zou, Z.; Ge, J.; Zheng, H.; Wu, Y.; Han, C.; Yao, Z. Encrypted Traffic Classification with a Convolutional Long Short-Term Memory Neural Network. In Proceedings of the 20th IEEE International Conference on High Performance Computing and Communications; 16th IEEE International Conference on Smart City; 4th IEEE International Conference on Data Science and Systems, HPCC/SmartCity/DSS 2018, Exeter, UK, 28–30 June 2018; pp. 329–334.

22. Xu, L.; Dou, D.; Chao, H.J. ETCNet: Encrypted Traffic Classification Using Siamese Convolutional Networks. In Proceedings of the Workshop on Network Application Integration/CoDesign (NAI'20), Virtual Event, New York, NY, USA, 14 August 2020; ACM: New York, NY, USA, 2020; p. 3.

23. Song, M.; Ran, J.; Li, S. Encrypted Traffic Classification Based on Text Convolution Neural Networks. In Proceedings of the 2019, IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT), Dalian, China, 19–20 October 2019; pp. 432–436.

References

1. Starun I.G., Iuganson A.N., Gatchin Ju.A. Postroenie matematicheskoj modeli rascheta kompleksnoj ocenki VPN: st. - Vestnik TGTU, tom. 25, vypusk 4, 2019, s. 535-546.

2. Lu,B.; Luktarhan,N.; Ding, C; Zhang,W. ICLSTM: Encrypted Traffic Service Identification Based on Inception-LSTM Neural Network. Symmetry, 2021, 13, 1080. https:// doi.org/10.3390/sym13061080

3. X. Wang et al., "Evolutionary Algorithm-Based and Network Architecture Search-Enabled Multiobjective Traffic Classification," in IEEE Access, vol. 9, pp. 52310-52325, 2021

4. B. Yamansavascilar, M. A. Guvensan, A. G. Yavuz, and M. E. Karsligil, "Application identification via network traffic classification", In: Proc. of 2017 Int. Conf. Comput. Netw. Commun. ICNC 2017, pp. 843–848, 2017.

5. Megantara, A.A., Ahmad, T. ANOVA-SVM for Selecting Subset Features in Encrypted Internet Traffic Classification (2021) International Journal of Intelligent Engineering and Systems, 14 (2), pp. 536-546.

6. G. Cheng and S. Wang, "Traffic classification based on port connection pattern", In: Proc. of 2011 Int. Conf. Comput. Sci. Serv. Syst. CSSS 2011 - Proc., pp. 914–917, 2011.

7. H. K. Lim, J. B. Kim, K. Kim, Y. G. Hong, and Y. H. Han, "Payload-based traffic classification using multi-layer LSTM in software defined networks", Appl. Sci., Vol. 9, No. 12, 2019.

8. F. Dehghani, N. Movahhedinia, M. R. Khayyambashi, and S. Kianian, "Real-time traffic classification based on statistical and payload content features", In: Proc. of - 2010 2nd Int. Work. Intell. Syst. Appl. ISA 2010, pp. 26–29, 2010.

9. G. Draper-Gil, A. H. Lashkari, M. Mamun, and A. A. Ghorbani,"Characterization of encrypted and VPN traffic using time-related features," in Proc. 2nd Int. Conf. Inf. Syst. Security Privacy, 2016,pp. 407–414

10. Caicedo-Munђoz JA, Espino AL, Corrales JC, RendoђБn A.QoS-Classifier for VPN and Non-VPN traпђAic based on time-related features.Computer Networks 2018; 144: 271-279. doi: 10.1016/j.comnet.2018.08.008

11. A. Saber, B. Fergani, and M. Abbas, "Encrypted Traffic Classification: Combining Over-and Under-Sampling through a PCA-SVM", In: Proc. of - PAIS 2018 Int. Conf. Pattern Anal. Intell. Syst., pp. 1–5, 2018.

12. Achmad Akbar Megantara, Tohari Ahmad, ANOVA-SVM for Selecting Subset Features in Encrypted Internet Traffic Classification. International Journal of Intelligent Engineering and Systems, Vol.14, No.2, 2021

13. R. Nigmatullin, A. Ivchenko and S. Dorokhin, "Differentiation of Sliding Rescaled Ranges: New Approach to Encrypted and VPN Traffic Detection," 2020 International Conference Engineering and Telecommunication (En&T), 2020, pp. 1-5, doi: 10.1109/EnT50437.2020.9431285

14. Uğurlu, M., Doğru, İ.A., Arslan, R.S. A new classification method for encrypted internet traffic using machine learning (2021) Turkish Journal of Electrical Engineering and Computer Sciences, 25 (9), pp. 2450-2468.

15. Felipe Peter. Analysis of the ISCX VPN-nonVPN Dataset 2016 for Encrypted Network Traffic Classification, Tsinghua University, [Электронный ресурс] - pp. 1-5, 2018

16. A Glance through the VPN Looking Glass: IPv6 Leakage and DNS Hijacking in Commercial VPN Clients / V. C. Perta [et al.] // Proceedings of Conference: 15th Privacy Enhancing Technologies, 30 June – 02 July 2015, Philadelphia, USA. – Philadelphia, 2015. – P. 77 – 91.

17. Brissaud P, Francђlois J, Chrisment I, Cholez T, Bettan O. Transparent and Service-Agnostic Monitoring of Encrypted Web TraпђAic. IEEE Transactions on Network and Service Management 2019; 16 (3): 842-856

18. Lu, B., Luktarhan, N., Ding, C., Zhang, W. ICLSTM: Encrypted traffic service identification based on inception-LSTM neural network (2021) Symmetry, 13 (6), ст. № 1080

19. Wang, W.; Zhu, M.; Wang, J.; Zeng, X.; Yang, Z. End-to-end encrypted traffic classification with one-dimensional convolution neural networks. In Proceedings of the 2017 IEEE International Conference on Intelligence and Security Informatics (ISI), Beijing, China, 22–24 July 2017; pp. 43–48.

20. Lotfollahi, M.; Siavoshani, M.J.; Zade, R.S.H.; Saberian, M. Deep packet: A novel approach for encrypted traffic classification using deep learning. Soft Comput. 2020, 24, 1999–2012.

21. Zou, Z.; Ge, J.; Zheng, H.; Wu, Y.; Han, C.; Yao, Z. Encrypted Traffic Classification with a Convolutional Long Short-Term Memory Neural Network. In Proceedings of the 20th IEEE International Conference on High Performance Computing and Communications; 16th IEEE International Conference on Smart City; 4th IEEE International Conference on Data Science and Systems, HPCC/SmartCity/DSS 2018, Exeter, UK, 28–30 June 2018; pp. 329–334.

22. Xu, L.; Dou, D.; Chao, H.J. ETCNet: Encrypted Traffic Classification Using Siamese Convolutional Networks. In Proceedings of the Workshop on Network Application Integration/CoDesign (NAI'20), Virtual Event, New York, NY, USA, 14 August 2020; ACM: New York, NY, USA, 2020; p. 3.

23. Song, M.; Ran, J.; Li, S. Encrypted Traffic Classification Based on Text Convolution Neural Networks. In Proceedings of the 2019, IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT), Dalian, China, 19–20 October 2019; pp. 432–436.

СТАРУН Игорь Геннадьевич, магистрант факультета безопасности информационных технологий федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский университет ИТМО». Россия, 197101, Санкт-Петербург, Кронверкский пр., д. 49, лит. А. E-mail: starun.igor@yandex.ru.

STARUN Igor Gennadievich, master student of the Faculty of Secure Information Technologies of the Federal State Autonomous Educational Institution of Higher Education "National Research University ITMO", Russia, 197101, St. Petersburg, Kronverksky pr., 49, lit. A. E-mail: starun.igor@yandex.ru.

ЮГАНСОН Андрей Николаевич, кандидат технических наук, доцент факультета безопасности информационных технологий федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский университет ИТМО». Россия, 197101, Санкт-Петербург, Кронверкский пр., д. 49, лит. А. E-mail: a_yougunson@corp.itmo.ru.

IUGANSON Andrey Nikolaevich, Ph.D., Associate Professor of the Faculty of Secure Information Technologies of the Federal State Autonomous Educational Institution of Higher Education "National Research University ITMO", Russia, 197101, St. Petersburg, Kronverksky pr., 49, lit. A. E-mail: a_yougunson@corp.itmo.ru.