КЛАССИФИКАЦИЯ ЗАШИФРОВАННОГО ТРАФИКА МОБИЛЬНЫХ ПРИЛОЖЕНИЙ МЕТОДОМ МАШИННОГО ОБУЧЕНИЯ

Шелухин О.И.[1], Барков В.В.[2], Полковников М.В.[3]

Цель работы: *сравнительный анализ эффективности алгоритмов классификации Naive Bayes, C4.5, AdaBoost, SVM, Random Forest в случаях, когда объектом классификации является трафик мобильных приложений, Instagram, Почта Mail.ru, Pikabu, Сбербанк Онлайн, Hearthstone, Skype, использующих шифрование.*

Метод исследования: *экспериментальный метод для реализации сбора и классификации трафика мобильных устройств. Измерительный метод для определения числовых значений атрибутов собранного трафика. Сравнительный метод для сравнения эффективности алгоритмов классификации. Метод анализа для выявления наилучших характеристик обучающего набора данных и классифицируемого потока сетевого трафика.*

Результаты: *разработан программный комплекс «Система анализа трафика» для сбора и классификации сетевого трафика. Собрано более двух миллионов сетевых пакетов от шести приложений, передающих зашифрованный трафик. Использование алгоритма InfoGain показало, что для обеспечения высокого качества классификации трафика приложений, использующих шифрование, достаточно ограничиться тринадцатью атрибутами. Классификатор Random Forest является самым медленным, однако имеет наилучшие показатели оценки качества классификации, среди исследуемых алгоритмов. Размер обучающей выборки алгоритма Random Forest для достижения достаточно высокого качества классификации мобильных приложений может не превышать 300 потоков. Для обеспечения высокого качества классификации потоков – достаточно анализировать от 16 до 58 пакетов в потоке в зависимости от приложения. Дальнейшее увеличение количества пакетов в потоке не приводит к заметному улучшению качества классификации.*

Ключевые слова: *интеллектуальный анализ мобильного трафика, объём обучающей выборки, алгоритмы, сетевой трафик, пакет, поток, протокол, эффективность, random forest, svm, c4.5, adaboost, naive bayes.*

DOI: *10.21681/2311-3456-2018-4-21-28*

Постановка задачи

Проблема определения сотовым оператором, какими приложениями воспользовался тот или иной пользователь сети нужны для составления статистики наиболее часто используемых приложений. Подобная статистика приложений помогает не только отслеживать состояние сети, выявлять сбои, но и при необходимости ограничивать доступ к сетевым ресурсам, которые с точки зрения информационной безопасности могут нанести вред пользователю.

Внедрение методов машинного обучения позволяет производить автоматическую классификацию [1, 2], анализ и фильтрацию вредоносных и нежелательных мобильных приложений сетевого трафика [3, 4, 5].

Вредоносные мобильные приложения [6] могут представлять собой угрозу целостности или доступности данных, а нежелательные – угрозу конфиденциальности. Классификация трафика мобильных приложений, использующих шифрование, не подразумевает его дешифрацию. Данные, содержащиеся внутри пакетов, остаются конфиденциальными и доступны только лишь пользователю и удалённому узлу.

Мобильные приложения, использующие шифрование трафика можно разделить на три группы [7, 8]. К первой группе относятся приложения, использующие протокол шифрования транспортного уровня SSL/TLS [9] совместно с протоколом прикладного уровня HTTPS [10]. Примерами таких приложений являются Google, Facebook, Сбербанк и пр. Ко второй группе относятся приложения, использующие протокол P2P [11] с шифрованием (BitTorrent, MuTorrent, Vuze и пр.). К третьей группе относятся приложения, использующие помимо протоколов шифрования транспортного уровня собственные протоколы шифрования. Примерами таких приложения являются Skype [12], WhatsApp, Telegram и пр.

В ситуации сложного определения типа шифрования трафика, с целью классификации трафика мобильных приложений, использующих шифрование, целесообразно использовать методы машинного обучения [13].

Технология сбора трафика мобильных приложений

Для формирования базы данных трафика мобильных приложений был разработан программный комплекс «Система анализа трафика», включающий

1 Шелухин Олег Иванович, доктор технических наук, профессор, МТУСИ, заведующий кафедрой «Информационная безопасность», Москва, Россия. E-mail: sheluhin@mail.ru

2 Барков Вячеслав Валерьевич, МТУСИ, старший преподаватель кафедры «Информационная безопасность», Москва, Россия. E-mail: viacheslav.barkov@gmail.com

3 Полковников Михаил Вадимович, МТУСИ, магистрант кафедры «Информационная безопасность», Москва, Россия. E-mail: mnxamoto@mail.ru

Рис. 1. Схема сбора мобильного трафика

в себя сервер баз данных, сервер приложений, Web-приложение и клиентское ПО для мобильных устройств под управлением операционной системы Android (мобильный клиент).

Процесс сбора трафика с использованием программного комплекса «Система анализа трафика», а также взаимодействие компонентов программного комплекса между собой и с внешними мобильными приложениями представлен на рисунке 1.

На смартфон или планшет под управлением операционной системы Android установлен мобильный клиент программного комплекса2 «Система анализа трафика». Данный клиент перехватывает пакеты сетевого трафика заданных приложений, которые также установлены на данном устройстве.

Перехваченные пакеты сетевого трафика отправляются на сервер приложений программного комплекса «Система анализа трафика», установленном на серверной ЭВМ, управляемой операционной системой Windows Server 2016.

Сервер приложений программного комплекса «Система анализа трафика» группирует пакеты сетевого трафика в потоки и с помощью сервера баз данных сохраняет данные в базу данных.

Обмен данными между компонентами программного комплекса «Система анализа трафика» осуществляется через глобальную сеть Интернет с использованием протокола HTTP в формате JSON. Сервер приложения включает в себя Web-службу, которая предоставляет клиентам REST API, с помощью которого можно получить доступ к функциям сбора пакетов сетевого трафика, управления наборами данных, создания и обучения классификаторов, классификации и другим функциям.

С использованием программного комплекса был собран трафик мобильных приложений трёх категорий:

Таблица 1.

Характеристики используемого набора данных по типу приложений при анализе сетевых пакетов и потоков

Приложение	Обучающая выборка		Тестовая выборка	
	пакеты	потоки	пакеты	потоки
Почта Mail.ru	162517	3356	79612	1644
Сбербанк	156648	3303	80482	1697
Skype	146315	3329	73443	1671
Пикабу	167182	3325	84220	1675
Инстраграмм	1286600	3357	629695	1643
Hearthstone	151298	3330	75876	1670
Всего	2070562	20000	1023328	10000

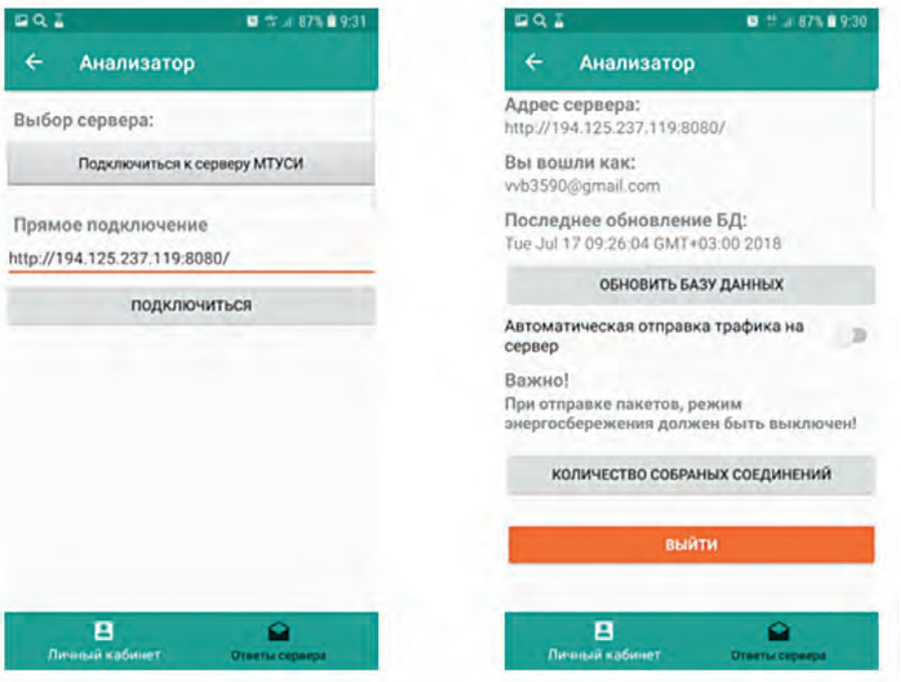


Рис. 2. Графический пользовательский интерфейс мобильного клиента программного комплекса «Система анализа трафика»: подключение к серверу и информация о сервере.

«С шифрованием трафика», «Без шифрования трафика», «С частичным шифрованием трафика».

В ходе сбора трафика мобильных приложений, использующих шифрование, были собраны потоки сетевого трафик 6 приложений: Instagram, Почта Mail.ru, Pikabu, Сбербанк-Онлайн, Hearthstone, Skype. В таблице 1 приведены числовые характеристики собранных сетевых пакетов и потоков для обучающей и тестовой выборок.

Для проведения эксперимента и формирования исходных данных на мобильные устройства под управлением операционной системы Android 5.0

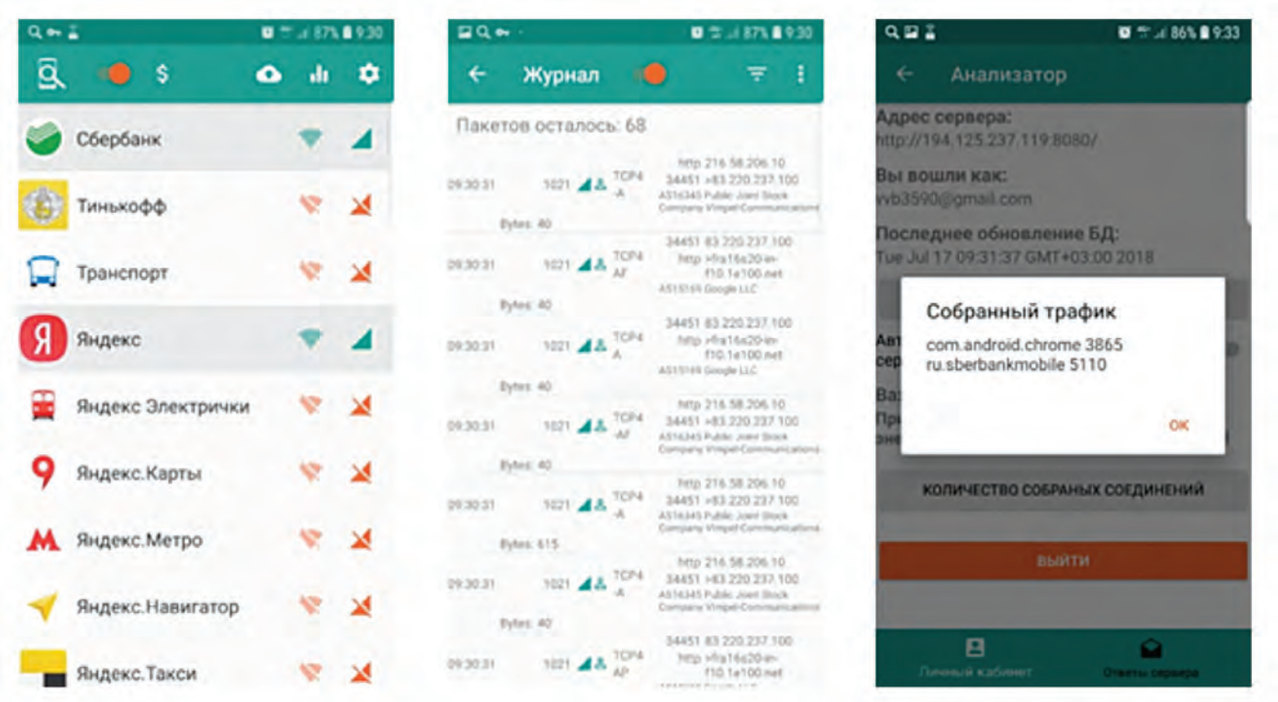


Рис. 3. Графический интерфейс пользователя мобильного клиента программного комплекса «Система анализа трафика»: Выбор приложений, по которым будут перехватываться пакеты; Журнал перехваченных пакетов; Просмотр количества собранных потоков

[17] и выше был разработан и установлен мобильный клиент программного комплекса «Система анализа трафика». На рисунках 2а и 2б представлен процесс подключения мобильного клиента к серверу приложений программного комплекса «Система анализа трафика».

На рисунке 3 представлен графический пользовательский интерфейс мобильного клиента программного комплекса «Система анализа трафика» в процессе настройки для перехвата трафика указанных приложений и отправки его на сервер (рисунок 3а); в процессе перехвата трафика (рисунок 3б); в процессе просмотра количества собранных потоков (рисунок 3в).

Результаты классификации трафика мобильных приложений, использующих шифрование

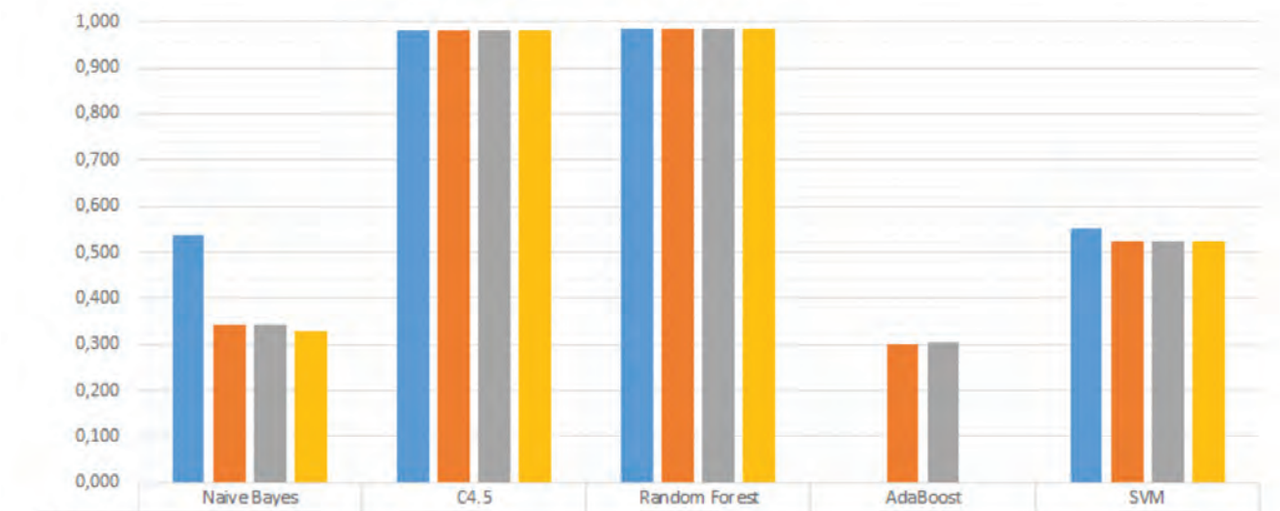
Воспользовавшись алгоритмом выбора атрибутов InfoGain [http://www.cs.waikato.ac.nz/ml/weka] из 23 исходных атрибутов было выделено 13 [18]:

1. Средний размер порции данных со стороны сервера (AverageSizeDataOnTransportLayerFromServer);
2. Средний размер пакета со стороны сервера (AverageSizeOnTransportLayerFromServer);
3. Кпд сервера – количество переданной нагрузки прикладного уровня, деленное на общее количество переданной нагрузки прикладного и транспортного уровня (EfficiencyOfServer);
4. Адрес клиента (FirstIP);
5. Размер полезной нагрузки сетевого уровня со стороны сервера (NetworkLayerPayloadSizeFromServer);
6. Соотношение полезной нагрузки - во сколько раз клиент передал больше байт информации, чем сервер (RatioOfData);
7. Адрес сервера (SecondIP);
8. Стандартное отклонение размера порции данных со стороны клиента (StandardDeviationOfDataOnTransportLayerFromClient);
9. Стандартное отклонение размера порции данных со стороны сервера (StandardDeviationOfDataOnTransportLayerFromServer);
10. Стандартное отклонение размера пакета со стороны клиента (StandardDeviationOfPacketSizeFromClient);
11. Стандартное отклонение размера пакета со стороны сервера (StandardDeviationOfPacketSizeFromServer);
12. Размер полезной нагрузки транспортного уровня со стороны клиента (TransportLayerPayloadSizeFromClient);
13. Размер полезной нагрузки транспортного уровня со стороны сервера (TransportLayerPayloadSizeFromServer);

Для оценки эффективности алгоритмов классификации Naive Bayes, C4.5, AdaBoost, SVM, Random Forest [14, 15, 16] использовались следующие метрики информационного поиска [19, 20, 21]: Precision (Точность), Recall (Полнота), F-Measure (F-мера), ROC-кривые (Receiver Operating Characteristic Curve), AUC (Area Under Curve) - площадь под ROC-кривой. В результате обработки экспериментальных данных получены количественные результаты, представленные в виде усредненных гистограмм на рисунке 4.

Анализ приведенных результатов показывает, что наилучшие результаты классификации показывают алгоритмы C4.5 и Random Forest. На рисунке 5 представлены временные интервалы в миллисекундах, которые потребовались исследуемым классификаторам на обучение и тестирование. Как видно, самыми «быстрыми» на этапе обучения оказались Naive Bayes, C4.5 и AdaBoost, а на этапе тестирования – C4.5, Random Forest, AdaBoost и SVM. Самыми «быстрыми» классификаторами на обоих фазах являются: C4.5 и AdaBoost.

Хотя классификатор AdaBoost и является самым «быстрым», он имеет самые худшие результаты оцен-



	Naive Bayes	C4.5	Random Forest	AdaBoost	SVM
Precision	0,535	0,980	0,985	0,000	0,552
Recall	0,344	0,980	0,985	0,302	0,524
Accuracy	0,344	0,980	0,985	0,303	0,524
F-мера	0,327	0,980	0,985	0,000	0,521

Рис.4. Усреднённые по приложениям значения метрик оценки эффективности исследуемых классификаторов

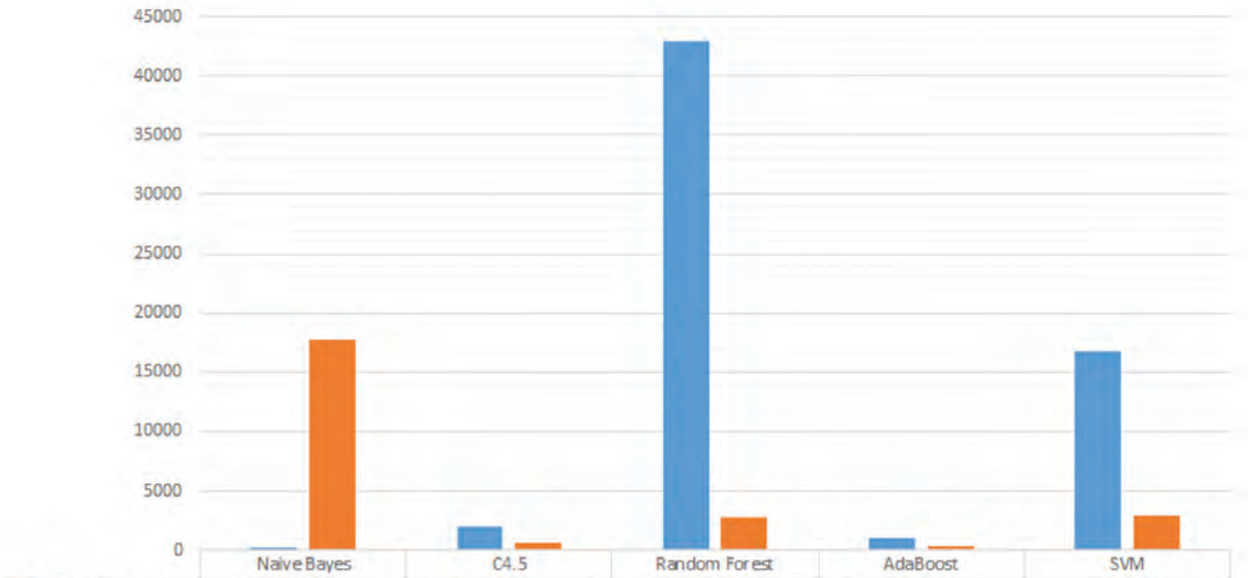


Рис.5. Временные диапазоны обучения и тестирования классификаторов

ки качества классификации. Классификатор Random Forest является самым «медленным», однако имеет наилучшие показатели оценки классификации.

На рисунках 6 представлены зависимости метрик оценки эффективности классификатора Random Forest от объёма обучающей выборки для исследуемых классов. На основе представленных зависимостей можно сделать следующие выводы: для достаточно высокого качества классификации (достоверность более 90%) достаточно обучить алгоритм Random Forest на 300 потоках.

В таблице 2 приведены значения AUC для алгоритма Random Forest, показывающие высокую достоверность классификации рассмотренных приложений.

Рассмотрим влияние количества пакетов в классифицируемом потоке на качество классификации зашифрованного трафика. На рисунках 7 отражены зависимости метрик оценки эффективности классификатора Random Forest от количества пакетов в классифицируемом потоке.

Представленные зависимости позволяют сделать вывод о том, что для обеспечения высокого качества клас-

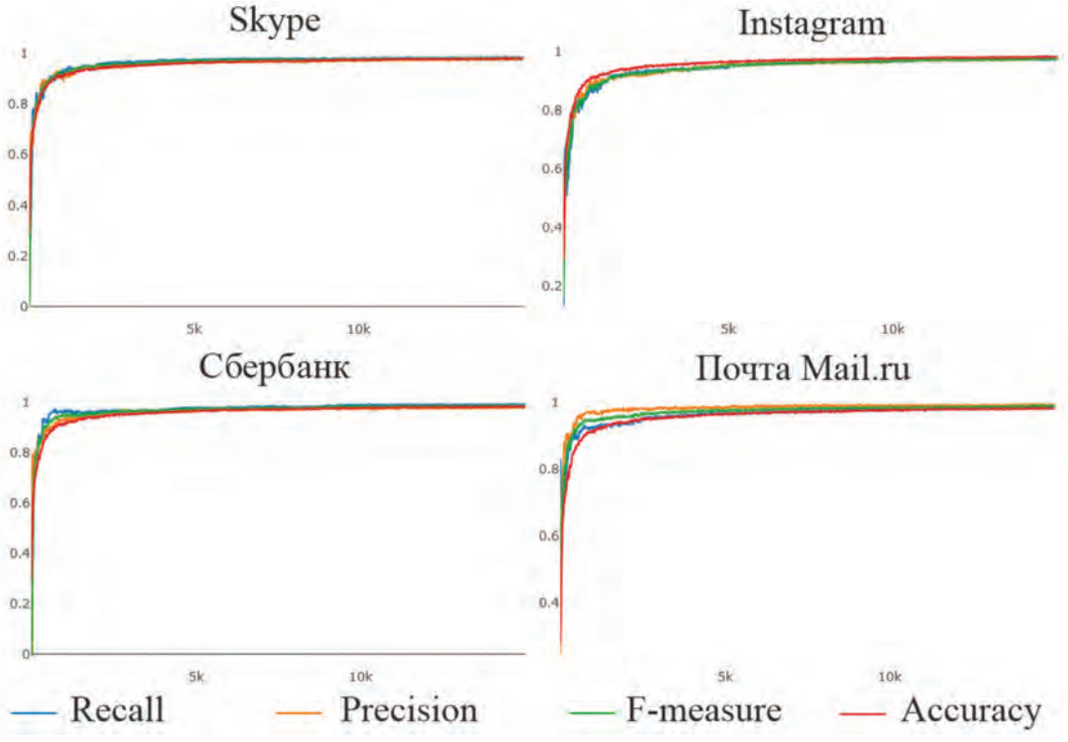


Рис.6. Зависимости метрик оценки эффективности классификатора Random Forest от объёма обучающей выборки

Таблица 2.

Значения AUC для алгоритма Random Forest

Класс	Instagram	Почта Mail.ru	Skype	Сбербанк	Hearthstone	Пикабу
ROC-AUC	0,9904	0,9885	0,9809	0,9920	0,9888	0,9797

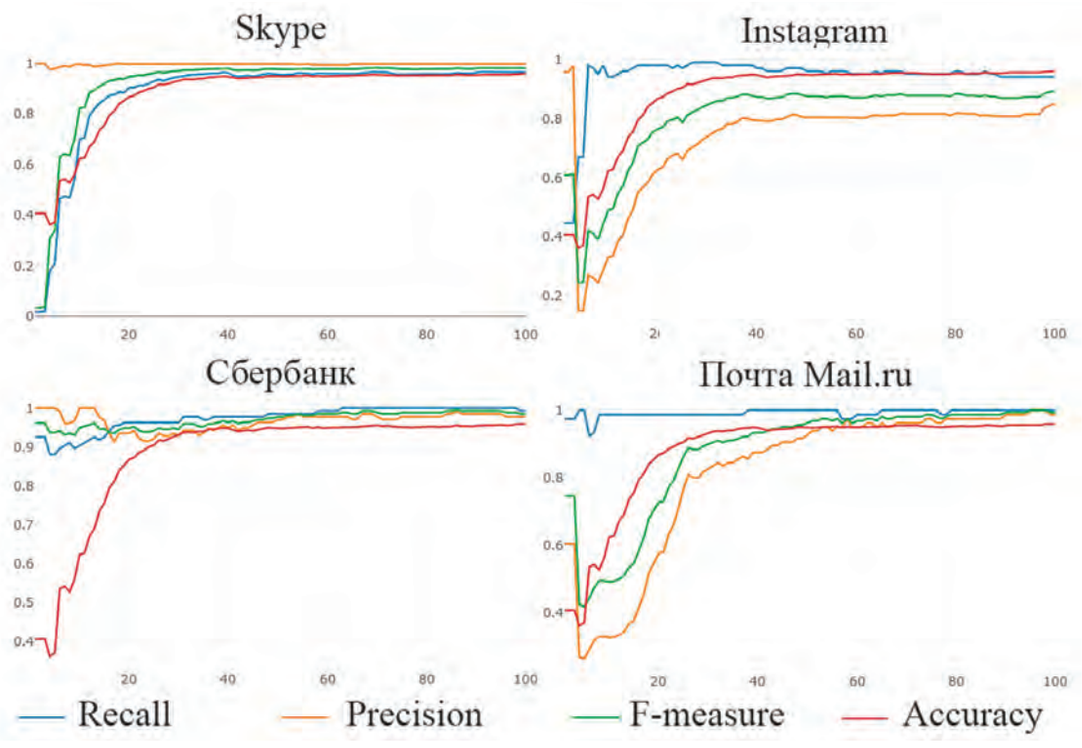


Рис.7. Зависимости метрик оценки эффективности классификатора Random Forest от количества пакетов в классифицируемом потоке для классов

сификации потока достаточно анализировать от 16 до 58 пакетов в потоке в зависимости от приложения. Дальнейшее увеличение количества пакетов в потоке не приводит к заметному улучшению качества классификации.

Выводы

На основании использования алгоритма InfoGain показано, что для обеспечения высокого качества классификации рассмотренных приложений, использующих шифрование при передаче данных, достаточно ограничиться тринадцатью атрибутами. Классификатор Random Forest является самым медленным, однако имеет наилучшие показатели оценки качества классификации.

Размер обучающей выборки алгоритма Random Forest для достаточно высокого качества классификации (достоверность более 90%) может не превышать 300 потоков. Для обеспечения высокого качества классификации потоков достаточно анализировать от 16 до 58 пакетов в потоке в зависимости от приложения. Дальнейшее увеличение количества пакетов в потоке не приводит к заметному улучшению качества классификации.

Рецензент: *Басараб Михаил Алексеевич, доктор физико-математических наук, профессор, МГТУ им. Н.Э. Баумана, Москва, Россия.* *E-mail: bmic@mail.ru*

Литература:

1. Ali Abdalla B.M., Jamil H.A., Hamdan M., Bassi J.S., Ismail I., Marsono M.N. Multi-stage Feature Selection for On-Line Flow Peer-to-Peer Traffic Identification. In: Mohamed Ali M., Wahid H., Mohd Subha N., Sahlan S., Md. Yunus M., Wahap A. (eds) Modeling, Design and Simulation of Systems. AsiaSim 2017. Communications in Computer and Information Science. Vol 752. Springer, Singapore. 26 August 2017. P. 509-523 DOI: 10.1007/978-981-10-6502-6_44
2. Santiago Egea Gómez, Belén Carro Martínez, Antonio J. Sánchez-Esguevillas, Luis Hernández Callejo. Ensemble network traffic classification: Algorithm comparison and novel ensemble scheme proposal. Computer Networks. Vol. 127. 2017. P. 68-80. DOI: 10.1016/j.comnet.2017.07.018
3. Шелухин О.И., Ерохин С.Д., Ванюшина А.В. Классификация IP-трафика методами машинного обучения. Горячая линия – телеком, 2018, 276 с.
4. Jamuna A., Vinoth Ewards S.E. Survey of Traffic Classification using Machine Learning. International Journal of Advanced Research in Computer Science. Vol. 4. No.4. March-April 2013. P. 65-70.
5. Jun Zhang, Yang Xiang, Yu Wang, Wanlei Zhou, Yong Xiang, Yong Guan Network Traffic Classification Using Correlation Information. IEEE Transactions on Parallel and Distributed systems. 2013. Vol. 24. P. 104 – 117. DOI: 10.1109/TPDS.2012.98

6. Шелухин О.И., Смычек М.А., Симонян А.Г. Фильтрация нежелательных приложений трафика подвижной радиосвязи для обнаружения угроз информационной безопасности. Радиотехнические и телекоммуникационные системы. 2018. № 1. с. 87-98.
7. Riyad Alshammari, A. Nur Zincir-Heywood, «Identification of VoIP encrypted traffic using a machine learning approach», Journal of King Saud University - Computer and Information Sciences, 2015, Vol. 27. P. 77. DOI: 10.1016/j.jksuci.2014.03.013
8. Daniel J. Arndt, A. Nur Zincir-Heywood, «A Comparison of three machine learning techniques for encrypted network traffic analysis», Computational Intelligence for Security and Defense Applications (CISDA) 2011 IEEE Symposium on. 2011. P. 107-114. DOI: 10.1109/CISDA.2011.5945941
9. Sung-Min Kim, Young-Hoon Goo, Myung-Sup Kim, Soo-Gil Choi, Mi-Jung Choi. A method for service identification of SSL/TLS encrypted traffic with the relation of session ID and Server IP – IEEE Security Privacy. 19-21 Aug. 2015. P. 4. DOI: 10.1109/APNOMS.2015.7275373
10. M. Prandini, M. Ramilli, W. Cerroni, F. Callegati. Splitting the HTTPS Stream to Attack Secure Web Connections. IEEE Security Privacy. 03 December 2010. P. 6. DOI: 10.1109/MSP.2010.190
11. X. Cheng, G. Dang. The P2P communication technology research based on Internet of things. 08 December 2014. P. 2. DOI: 10.1109/WARTIA.2014.6976225
12. Yi-Hui Lin, Shan-Hsiang Shen, Ming-Hong Yang, De-Nian Yang, Wen-Tsuen Chen, «Privacy-preserving deep packet filtering over encrypted traffic in software-defined networks», Communications (ICC) 2016 IEEE International Conference on. 2016. P. 1-7. DOI: 10.1109/ICC.2016.7510993
13. T.T.Nguyen, G. Armitage. A survey of techniques for internet traffic classification using machine learning. IEEE Communications Surveys and Tutorials. 2008. Vol.10. No.4. P. 56-76. DOI: 10.12691/jcsa-4-1-4
14. Костин Д.В., Шелухин О.И. Сравнительный анализ алгоритмов машинного обучения для проведения классификации сетевого зашифрованного трафика // T- Comm : Телекоммуникации и транспорт. 2016. № 9. стр. 46-52.
15. Soysal, Murat,Schmidt, Ece Guran. “Machine learning algorithms for accurate flow-based network traffic classification: Evaluation and comparison”. Performance Evaluation. 2010. Vol.67. No.6. P. 451-467. DOI: 10.1016/j.peva.2010.01.001
16. S. Bagui, X. Fang, K. Ezhil, S. C. Bagui and J. Sheehan, “Comparison of machine-learning algorithms for classification of VPN network using time-related features,” Journal of Cyber Security Technology. 2017. Vol. 1. No.2, P. 108-126. DOI: 10.1080/23742917.2017.1321891
17. Филлипс Б., Стюарт К., Марсикано К. Android. Программирование для профессионалов. Питер. 2017. 688 с. ISBN: 978-5-4461-0413-0
18. T. Nguyen and G. Armitage. “A Survey of Techniques for Internet Traffic Classification using Machine Learning”. IEEE Communications Surveys and Tutorials. 2008. Vol.11. No.3. P.37-52. DOI:10.1109/SURV.2008.080406
19. Шелухин О.И., Ванюшина А.В., Габисова М.Е. Фильтрация нежелательных приложений Интернет-трафика с использованием алгоритма классификации Random Forest // Вопросы кибербезопасности. 2018. №2 (26). С.44-51
20. Ehsan Mahdavi, Ali Fanian, Homa Hassannejad. Ehsan Mahdavi, Ali Fanian, Homa Hassannejad. Encrypted Traffic Classification Using Statistical Features. The ISC Int›l Journal of Information Security. January 2018. Vol. 10. No.1. P. 29–43. DOI: 10.22042/isecure.2018.95316.390
21. Yuichi Kumano, Shingo Ata, Nobuyuki Nakamura, Yoshihiro Nakahira, Ikuo Oka, «Towards real-time processing for application identification of encrypted traffic», Computing Networking and Communications (ICNC) 2014 International Conference on. 2014. P. 136-140. DOI: 10.1109/ICCNC.2014.6785319

CLASSIFICATION OF ENCRYPTED MOBILE APP TRAFFIC USING THE MACHINE LEARNING METHOD

Sheluhin O.[1], Barkov V.[2], Polkovnikov M.[3]

Purpose: Comparative efficiency analysis of the classification algorithms Naive Bayes, C4.5, AdaBoost, SVM and Random Forest where the classification object is the traffic of mobile applications Instagram, Mail.ru Mail, Pikabu, Sberbank Online, Hearthstone and Skype using encryption.

Research methods: Experimental method to collect and classify mobile device traffic. Measuring method to determine numerical values of collected traffic attributes. Comparative method to compare the efficiency of classification algorithms. Analysis method to identify the best characteristics of the training dataset and the network traffic flow being classified.

1 Oleg Sheluhin, Doctor of Technical Sciences, Professor. Moscow Technical University of Communication and Informatics, Head of Information Security Department, Moscow, Russia. E-mail: sheluhin@mail.ru

2 Vyacheslav Barkov, Moscow Technical University of Communication and Informatics, Senior Lecturer for Information Security Department, Moscow, Russia. E-mail: viacheslav.barkov@gmail.com

3 Mikhail Polkovnikov, Moscow Technical University of Communication and Informatics, Master’s Degree Student at Information Security Department, Moscow, Russia. E-mail: mnxamoto@mail.ru

Results: The Traffic Analysis System software package was developed for network traffic collection and classification. More than two million network packets were collected from six applications transferring encrypted traffic. As was shown by the use of InfoGain algorithm, a maximum of thirteen attributes is enough to ensure a high quality classification of encryption-using app traffic. Although the slowest, the Random Forest classifier has the best classification quality assessment indicators among the algorithms under study. To achieve a high enough quality classification of mobile applications, the size of the RF algorithm learning sample may be limited to 300 flows. High quality of flow classification can be achieved by just analyzing from 16 to 58 packets per flow, depending on the application. A further increase in the number of packets per flow will hardly bring about any noticeable improvement in the classification quality.

Keywords: intelligent analysis of mobile traffic, training sample size, algorithms, network traffic, packet, flow, protocol, efficiency, random forest, svm, c4.5, adaboost, naive bayes.

References

1. Ali Abdalla B.M., Jamil H.A., Hamdan M., Bassi J.S., Ismail I., Marsono M.N. Multi-stage Feature Selection for On-Line Flow Peer-to-Peer Traffic Identification. In: Mohamed Ali M., Wahid H., Mohd Subha N., Sahlan S., Md. Yunus M., Wahap A. (eds) Modeling, Design and Simulation of Systems. AsiaSim 2017. Communications in Computer and Information Science. Vol 752. Springer, Singapore. 26 August 2017. P. 509-523 DOI: 10.1007/978-981-10-6502-6_44
2. Santiago Egea Gómez, Belén Carro Martínez, Antonio J. Sánchez-Esguevillas, Luis Hernández Callejo. Ensemble network traffic classification: Algorithm comparison and novel ensemble scheme proposal. Computer Networks. Vol. 127. 2017. P. 68-80. DOI: 10.1016/j.comnet.2017.07.018
3. Sheluhin O.I., Erokhin S.D., Vanyushina A.V. Classification of IP traffic using machine learning methods. Hot line – telecom. 2018. P. 276.
4. Jamuna A., Vinoth Ewards S.E. Survey of Traffic Classification using Machine Learning. International Journal of Advanced Research in Computer Science, March-April. 2013. Vol 4. No.4. P. 65-70.
5. Jun Zhang, Yang Xiang, Yu Wang, Wanlei Zhou, Yong Xiang, Yong Guan Network Traffic Classification Using Correlation Information. IEEE Transactions on Parallel and Distributed systems. 2013. Vol. 24. P. 104 – 117, DOI: 10.1109/TPDS.2012.98
6. Sheluhin O.I., Smychek M.A., Simonyan A.G. filtering of unwanted mobile radio traffic applications to detect information security threats. Radio engineering and telecommunication systems. 2018. No.1. P. 87-98.
7. Riyad Alshammari, A. Nur Zincir-Heywood, «Identification of VoIP encrypted traffic using a machine learning approach», Journal of King Saud University - Computer and Information Sciences, 2015, Vol. 27. P. 77. DOI: 10.1016/j.jksuci.2014.03.013
8. Daniel J. Arndt, A. Nur Zincir-Heywood, «A Comparison of three machine learning techniques for encrypted network traffic analysis», Computational Intelligence for Security and Defense Applications (CISDA) 2011 IEEE Symposium on. 2011. P.107-114. DOI: 10.1109/CISDA.2011.5945941
9. Sung-Min Kim, Young-Hoon Goo, Myung-Sup Kim, Soo-Gil Choi, Mi-Jung Choi. A method for service identification of SSL/TLS encrypted traffic with the relation of session ID and Server IP - IEEE Security Privacy. 19-21 Aug. 2015. P. 4. DOI: 10.1109/APNOMS.2015.7275373
10. M. Prandini, M. Ramilli, W. Cerroni, F. Callegati. Splitting the HTTPS Stream to Attack Secure Web Connections - IEEE Security Privacy. 03 December 2010. P. 6. DOI: 10.1109/MSP.2010.190
11. X. Cheng, G. Dang. The P2P communication technology research based on Internet of things. 08 December 2014. P. 2. DOI: 10.1109/WARTIA.2014.6976225
12. Yi-Hui Lin, Shan-Hsiang Shen, Ming-Hong Yang, De-Nian Yang, Wen-Tsuen Chen, «Privacy-preserving deep packet filtering over encrypted traffic in software-defined networks», Communications (ICC) 2016 IEEE International Conference on. 2016. P. 1-7. DOI: 10.1109/ICC.2016.7510993
13. T.T.Nguyen, G. Armitage. A survey of techniques for internet traffic classification using machine learning. IEEE Communications Surveys and Tutorials. 2008. Vol.10. No.4. P. 56-76. DOI: 10.12691/jcsa-4-1-4
14. D. Kostin, O. Sheluhin. Comparative Analysis of Machine Learning Algorithms for Classifying Networked Encrypted Traffic // T-Comm: Telecommunications and Transport. 2016. No.9. P. 46-52.
15. Soysal, Murat,Schmidt, Ece Guran. "Machine learning algorithms for accurate flow-based network traffic classification: Evaluation and comparison". Performance Evaluation. 2010. Vol. 67. No.6. P. 451-467. DOI: 10.1016/j.peva.2010.01.001
16. S. Bagui, X. Fang, K. Ezhil, S. C. Bagui and J. Sheehan, "Comparison of machine-learning algorithms for classification of VPN network using time-related features," Journal of Cyber Security Technology. 2017. Vol. 1. No.2. P. 108-126. DOI: 10.1080/23742917.2017.1321891
17. Phillips B., Stuart K., Marsicano K. Android. Programming for professionals. Peter. 2017. 688 c. ISBN: 978-5-4461-0413-0
18. T. Nguyen and G. Armitage. "A Survey of Techniques for Internet Traffic Classification using Machine Learning". IEEE Communications Surveys and Tutorials. 2008. Vol. 11. No.3. P. 37-52. DOI:10.1109/SURV.2008.080406
19. Sheluhin OI, Vanyushina AV, Gabisova ME Filtering unwanted applications of Internet traffic using the Random Forest classification algorithm // Cybersecurity issues. 2018. No.2 (26). P.44-51.
20. Ehsan Mahdavi, Ali Fanian, Homa Hassannejad. Ehsan Mahdavi, Ali Fanian, Homa Hassannejad. Encrypted Traffic Classification Using Statistical Features. The ISC Int'l Journal of Information Security. January 2018. Vol. 10. No.1. P. 29–43. DOI: 10.22042/isecure.2018.95316.390
21. Yuichi Kumano, Shingo Ata, Nobuyuki Nakamura, Yoshihiro Nakahira, Ikuo Oka, «Towards real-time processing for application identification of encrypted traffic», Computing Networking and Communications (ICNC) 2014 International Conference on. 2014. P. 136-140. DOI: 10.1109/ICCNC.2014.6785319