УДК 004.852

РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ, ИСПОЛЬЗУЮЩЕГО МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ, ДЛЯ ОПРЕДЕЛЕНИЯ ЗАШИФРОВАННОГО ТРАФИКА СЕТИ TOR

Д.И. Деревцов
(Университет ИТМО, г. Санкт-Петербург)
Научный руководитель - к.т.н., доцент А.Ю. Кузнецов
(Университет ИТМО, г. Санкт-Петербург)

Введение. В современном мире в целях защиты собственной информационной инфраструктуры многие организации используют межсетевые экраны, позволяющие осуществлять контроль и фильтрацию проходящего через него сетевого трафика в соответствии с заданными правилами. Применение пользователями различных программ, позволяющих получить доступ к заблокированным ресурсам, может крайне негативно отразиться на безопасности такой инфраструктуры.

Одним из наиболее популярных анонимных способов коммуникации является распределенная сеть TOR, представляющая собой систему прокси-серверов и позволяющая устанавливать анонимное сетевое соединение, защищённое от прослушивания и предоставляющее передачу данных в зашифрованном виде. Его особенностью является то, что сообщения неоднократно шифруются и потом отсылаются через несколько сетевых узлов, а сам трафик очень сильно схож с обыкновенным HTTPS трафиком [1].

Цель. Целью данной работы является исследование существующих методов машинного обучения, а также разработка модели, позволяющей с достаточно высокой точностью определить наличие трафика распределенной сети TOR.

Базовые положения исследования. Гипотеза, лежащая в основе данной работы, состоит в следующем. Пользователь локальной сети запускает приложение, генерирующее трафик сети TOR, при этом на его компьютере или компьютерах других пользователей локальной сети также присутствуют приложения, генерирующие различные типы трафика, например, FTP, HTTP, HTTPS. Сеть TOR использует протокол шифрования TLS между клиентом, маршрутизаторами сети и конечным сервером, из этого следует, что трафик сети TOR должен иметь сходные характеристики с любым другим трафиком, использующим TLS, например с HTTPS. Отсюда следует, что если в характеристиках трафика удастся выявить различия, то трафик сети TOR можно будет однозначно определить среди любого другого трафика, использующего TLS [2].

Для того, чтобы определить такие различия в характеристиках TOR и HTTPS трафика, следует применить методы машинного обучения [3]. Их использование для определения закономерностей информации, содержащейся в пакетах, необходимо для прогнозирования трафика, содержащегося в потоке TLS.

Промежуточный результат. Разработаны алгоритмические и программные средства, способные определить наличие зашифрованного трафика распределенной сети TOR.

Практический результат. Разработано программное обеспечение, использующее методы машинного обучения и позволяющее определить наличие в локальной сети зашифрованного трафика распределенной сети TOR с достаточно высокой точностью.

Список литературы.

1. Inc Tor Project. (2019, February) torproject [Online]. https://torproject.org/docs/faq.html
2. Рашка С. Python и машинное обучение / пер. с англ. А. В. Логунова. – М.: ДМК Пресс, 2017. – 418 с.: ил.
3. S., Nguyen, T., & Armitage, G. Zander, "Automated traffic classification and application identification using machine learning," in In Local Computer Networks, 2005. 30th Anniversary, 2005, pp. 250-257.