Далее к показателям скрытого слоя подбираются весовые коэффициенты, предназначенные для функции активации. На выходном слое программа получает результат в виде авторизованного и неавторизованного пользователя.

Программный комплекс состоит из программных модулей распознавания личности по голосу, изображению и видео. Созданная модель нейронной сети включает в себя базу данных из более 1000 биометрических образов распознавания личности по голосу. Авторами самостоятельно подобрана конфигурация весовых коэффициентов для получения наиболее точных результатов обучения, предназначенных для определения авторизованного и неавторизованного пользователя информационной системы.

Исследование проводится при финансовой поддержке Московского технического университета связи и информатики в рамках научного проекта № 40469-20/2022-к.

Список использованных источников

1. Ismagilova A.S. and Lushnikov N.D. Learning Neural Network for Multifactor Authentication Using Biometric Technologies. 4th International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA), Lipetsk, Russian Federation, 2022, p. 416–420. DOI: http://dx.doi.org/10.1109/SUMMA57301.2022.9973920.

Исмагилова Альбина Сабирьяновна, д.ф-м.н., профессор, заведующий каф. управления информационной безопасностью, ismagilovaas@yandex.ru.
Лушников Никита Дмитриевич, ассистент каф. управления информационной безопасностью, luschnikovnikita@yandex.ru.

УДК 004.056

ИДЕНТИФИКАЦИЯ ПОЛЬЗОВАТЕЛЕЙ В СЕТЕВОМ ТРАФИКЕ

Ю.В. Мурашко
Московский технический университет связи и информатики,
г. Москва

Задача идентификации субъектов информационной системы является одной из основных задач в современной теории информационной безопасности [1]. В частности, идентификация пользователей компьютерной сети на текущий момент является задачей крайне актуальной, поскольку активное развитие технических средств анонимизации позволяет потенциальному злоумышленнику совершить противоправные действия, скрыв при этом информацию о себе.

Целью работы является оценка возможности использования применения различных наборов поведенческих характеристик для

повышения эффективности идентификации пользователей компьютерных сетей.

Идентификации пользователей по поведенческим характеристикам

Каждый конкретный пользователь генерирует уникальный сетевой трафик, который определяется его поведенческими привычками и характеристиками сетевых сессий. Следовательно, существует довольно сильная корреляция между уже собранным сетевым трафиком и новыми данными, собранными за определенный период времени.

Методология

В работе рассматривается метод идентификации пользователя на стороне оператора сети, состоящий из двух этапов. На первом этапе фиксируются статические параметры, больше характеризующие устройство пользователя, а не его самого. На втором этапе динамически отслеживается активность пользователя в сети (рис 1).

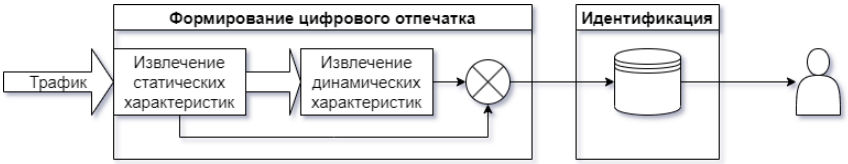

Рисунок 1 – Процесс идентификации

Набор характеристик

Извлекаемые на первом этапе предлагаемого подхода выделяются данные из заголовка IP-пакета, которые позволят провести первичную идентификацию [2]:

- Страна, из которой было инициировано подключение;
- Время начала соединения;
- Время конца соединения;
- IP-адрес источника;
- Имя хоста источника;
- IP-адрес назначения;
- Имя хоста назначения;
- Порт источника TCP;
- Порт назначения TCP.

После первичной идентификации осуществляется извлечение динамических характеристик и уточнение идентификации, за счет уникальности генерации потока трафика во время пользовательской активности в сети. Отмечается [3] [4] [5] [6], что группировка и анализ

потоков трафика обеспечивает более детальное понимание сети и может оказаться полезным, при извлечении следующих параметров потока:

- Количество записей пользователей в журналах трафика;
- Средняя продолжительность сеанса;
- Средняя разница во времени между временем начала сеанса;
- Среднее количество отправленных байт;
- Среднее количество полученных байт;
- Среднее количество отправленных пакетов;
- Среднее количество полученных пакетов;
- Количество уникальных IP-адресов источника;
- Количество уникальных исходных портов;
- Количество уникальных IP-адресов назначения;
- Количество уникальных портов назначения;
- Количество использованных приложений.

Эксперимент

Для проведения вычислительного эксперимента сгенерированный сетевой трафик был собран у нескольких пользователей с помощью анализатора трафика WireShark. Полученные данные были сохранены в формате JSON. После анализа полученных данных было решено преобразовать данные, при этом каждый объект принадлежал к классу с последующей целью использования данных для обучения алгоритмов машинного обучения с учителем. Эти объекты были скомпилированы в таблицу CSV с целью обучения алгоритмов машинного обучения для решения задачи идентификации пользователя по сетевому трафику, которая сводится к решению задачи классификации.

В таблице 1 приведены результаты эксперимента по классификации пользовательского трафика с использованием наивного байесовского классификатора.

Таблица 1 – Результаты классификации

Класс	Precision	Recall	F1-мера
Пользователь_1	0.81	0.99	0.89
Пользователь_2	0.98	0.63	0.77
Пользователь_3	1.00	0.29	0.46
Пользователь_4	0.94	0.62	0.75
Точность	0.8434		

Заключение

В работе выполнена задача по классификации данных, отражающие поведения пользователя в сети. На основе результатов, полученных в ходе вычислительного эксперимента, было получено, что точность идентификации пользователя составила более 84%. Было доказано, что на основе характеристик данных, передаваемых по сети, можно

идентифицировать конкретного пользователя с достаточно высокой точностью.

Список использованных источников

1. Зенков А. Основы информационной безопасности. 2022.
2. Gai V. et al. Model and Algorithms for User Identification by Network Traffic // Графикон-конференции по компьютерной графике и зрению, Vol. 31, 2021. pp. 1017-1027.
3. Verde N. V. et al. No NAT'd user left behind: Fingerprinting users behind NAT from NetFlow records alone // 2014 IEEE 34th International Conference on Distributed Computing Systems, 2014. pp. 218-227.
4. Alotibi G. et al. Behavioral-based feature abstraction from network traffic // Iccws 2015-The Proceedings of the 10th International Conference on Cyber Warfare and Security, 2015. pp. 1-9.
5. Alotibi G. et al. User profiling from network traffic via novel application-level interactions // 2016 11th International conference for internet technology and secured transactions (ICITST), 2016. pp. 279-285.
6. Clarke N., Li F., Furnell S. A novel privacy preserving user identification approach for network traffic // computers & security, Vol. 70, 2017. pp. 335-350.

Мурашко Юрий Викторович, аспирант каф. информационной безопасности, МТУСИ, yu.v.mur@gmail.com

УДК 004.056.53

СРЕДНЕКВАДРАТИЧЕСКАЯ ОШИБКА ФИЛЬТРАЦИИ В КАЧЕСТВЕ МЕТРИКИ РАССТОЯНИЯ ДЛЯ КОРРЕЛЯЦИОННЫХ МАТРИЦ

А.И. Плаван
Поволжский государственный университет телекоммуникаций и информатики, г. Самара

Ключевые слова: обнаружение аномалий, корреляционная матрица, метрика расстояния, среднеквадратическая ошибка фильтрации.

Поиск и обнаружение аномалий в сетевом трафике является актуальной задачей, так как количество информации, передаваемой по сети, год от года увеличивается, а сами аномалии могут являться признаком проведения хакерской атаки. При этом особый интерес представляют методы, позволяющие обнаруживать ранее неизвестные атаки, например методы, основанные на методах статистического анализа, машинного обучения.