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.

Если разделить сетевую активность некоторой корпоративной сети по часам, то для каждого интервала времени можно вычислить статистические характеристики, описывающие типичный сетевой трафик на этом интервале. Набор значений этих характеристик представляет профиль нормального поведения системы. Затем производится сравнение значений характеристик для текущего трафика с таким профилем для соответствующего интервала времени. Для определения как наличия отклонения, так и величины отклонения от профиля нормального поведения можно использовать метрики расстояния.

Одной из наиболее важных статистических характеристик является корреляционная функция или корреляционная матрица. В работе Herdin M. и др. [1] была предложена метрика расстояния для сравнения корреляционных матриц (CMD, correlation matrix distance) и получены значимые результаты при проверке этой метрики для мобильного радиоканала. В работе [2] автором был представлен способ обнаружения аномалий, основанный на вычислении минимальной среднеквадратической ошибки фильтрации. Можно показать, что представленный способ также можно использовать для определения расстояния между корреляционными матрицами.

Для проверки были проанализированы записи трафика на транзитном канале к вышестоящему интернет-провайдеру, полученные MAWI (Measurement and Analysis on the WIDE Internet) Working Group. 10 записей длительностью 15 минут были собраны с 15 по 24 ноября 2021 года. Для каждой записи была вычислена автокорреляционная матрица R, которая затем сравнивалась с некоторой известной матрицей R_0. За R_0 была принята корреляционная матрица, полученная для записи трафика 15 ноября. Таким образом значение расстояния между ней и первой записью должно быть равным нулю, так как эти матрицы полностью совпадают. Наибольшее значение должно показать, в какой день статистические свойства трафика наиболее сильно отличались от первого дня наблюдений.

Для каждой записи было вычислено два значения расстояния: значение метрики CMD [1] и значение минимальной среднеквадратической ошибки фильтрации [2]. Полученные результаты представлены на рисунке 1.

По графику можно видеть, что в целом, модуль расстояния для обоих способов изменяется синхронно. В случае совпадения корреляционных матриц 15 ноября значение расстояния, как и ожидается, равняется нулю, а максимум наблюдается 23 ноября. Этот день является праздничным в Японии, где собирался трафик MAWI, что возможно и привело к изменению структуры трафика и послужило причиной заметного отклонения значений.

Предложенный способ вычисления расстояния на основе минимальной среднеквадратической ошибки фильтрации можно применять для определения величины отклонения корреляционных матриц друг от друга, что может быть использовано при обнаружении аномалий в сетевом

трафике. Предложенный автором алгоритм предположительно имеет меньшую вычислительную сложность, чем алгоритм CMD.

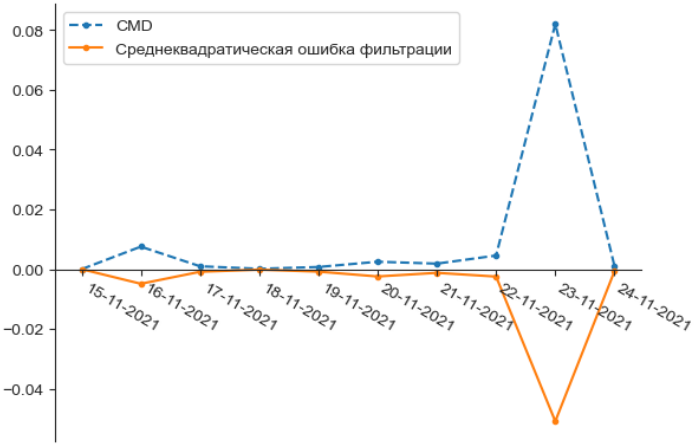

Рисунок 1 – Значение расстояния между корреляционными матрицами R и R_0 в разные дни

Для того, чтобы считаться настоящей метрикой, функция должна удовлетворять ряду условий. Необходимо определить все условия, которым удовлетворяет функция, используемая в данном способе, чтобы её можно было отнести к метрикам, псевдометрикам, квазиметрикам или другому типу, определить временную сложность алгоритма и сравнить ее с другими известными метриками для определения расстояния между корреляционными матрицами.

Список использованных источников

1. Herdin M. Correlation Matrix Distance, a Meaningful Measure for Evaluation of Non-Stationary MIMO Channels / M. Herdin [и др.] // 2005 IEEE 61st Vehicular Technology Conference. – Stockholm, Sweden: IEEE, 2005. – Т. 1. – С. 136-140.

2. Плаван А.И. Обнаружение аномалий сетевого трафика на основе фильтрации линейного преобразования трафика по критерию минимума среднеквадратической ошибки / А.И. Плаван // II Всероссийская научная школа-семинар «Современные тенденции развития методов и технологий защиты информации». – М.: Московский технический университет связи и информатики, 2022. – С. 189-198.

Плаван Алексей Игоревич, аспирант каф. информационной безопасности, aleksej-plavan@ya.ru