

ПОВЫШЕНИЕ КАЧЕСТВА ОБНАРУЖЕНИЯ АНОМАЛИЙ НА ОСНОВЕ УПРАВЛЯЕМОЙ СЕГМЕНТАЦИИ ИНФОРМАЦИОННЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ТЕЛЕКОММУНИКАЦИОННОГО ТРАФИКА

Макеенко А.А.¹

Научный руководитель - д.т.н., проф. Лебедев И.С.¹

¹СПб ФИЦ РАН

Введение

Обнаружение аномалий сетевого трафика требуют анализа многомерных потоков телекоммуникационных данных, формирующих нестационарные информационные последовательности. Качество моделей зависит от методов преобразования значений исходного потока в обучающую выборку [1]. Разбиение на окна часто выбирают эвристически, в результате масштаб сегментации становится параметром, снижающим устойчивость классификации, особенно при дисбалансе классов.

Основная часть

Сегментация рассматривается как самостоятельный управляемый этап обработки последовательностей. Предлагается фреймворк, отличающийся адаптивным выбором моделей обработки в зависимости от характеристик сегментов трафика и внешних факторов. Поток преобразуется в сегменты фиксированной длины с регулируемым шагом [2]. Для каждого сегмента строится вектор признаков на основе агрегирования характеристик трафика, включающих в себя статистических значений числовых признаков. Выполняется исследование влияния длины сегмента и шага на качество классификации: перебор конфигураций, оценка по стандартным метрикам, проверка устойчивости на нескольких базовых моделях, а также сравнение схем агрегирования результатов обработки. Апробация на открытом датасете сетевого трафика демонстрирует устойчивое улучшение качества относительно несегментированного обучения и позволяет выделить оптимальный диапазон длин сегментов: слишком малые окна теряют информативные признаки, а чрезмерно крупные могут содержать шумовые фрагменты.

Выводы

Сегментация трафика позволяет повысить качество обнаружения аномалий. Практический эффект достигается без значительного усложнения модели - за счёт выбора параметров сегментации и устойчивых правил агрегирования. Дальнейшая работа направлена на повышение показателя качества обработки, использующей адаптивную сегментацию, учитывающую изменчивость трафика.

Литература

1. Lebedev I., Sukhoparov M., Semenov V., Khasanov D. Adaptive Segmentation of Information Sequences for Machine Learning Modular Regression Models. *Emerging Science Journal*, 2025, Vol. 9, No. 5, pp. 2420–2438. doi:10.28991/ESJ-2025-09-05-08.
2. Wang C., Li X., Zhou T., Cai Z. Unsupervised Time Series Segmentation: A Survey on Recent Advances. *Computers, Materials & Continua*, 2024, Vol. 80, No. 2, pp. 2657–2673. doi:10.32604/cmc.2024.054061.