

УКД: 004.896

## **ИМИТАЦИОННАЯ МОДЕЛЬ ГЕНЕРАТОРА ТРАФИКА, УЧИТЫВАЮЩАЯ ОСОБЕННОСТИ РЕАЛЬНОГО СЕТЕВОГО ТРАФИКА ДЛЯ МАШИННОГО ОБУЧЕНИЯ ЭЛЕМЕНТОВ СЕТИ СВЯЗИ**

**Ануфренко А.В., Снятков М.А. (Санкт-Петербург)**

В условиях взаимодействия большого количества разных по своей специфике сетей связи важное значение имеет возможность эффективного распределения потоков информации в сетевых узлах путем оперативного реагирования на возникающие деструктивные воздействия.

Мониторинг характеристик сетевого трафика позволяет анализировать состояние элементов сети связи, а также своевременно обнаруживать различного рода аномалии трафика. Возможность формирования на основе мониторинга трафика так называемых датасетов (массивов данных) тесно связана с возможностью использования механизмов машинного обучения в области телекоммуникаций, а именно с возможностью обучать телекоммуникационное оборудование реагировать на характеристические особенности телетрафика с целью дальнейшего формирования управляющих воздействий для поддержания оптимальных режимов работы сетей связи.

С одной стороны, формирование массивов данных, подходящих для машинного обучения часто связано с разного рода сложностями (недостаток выборки данных, зашумленность данных, аномалии данных и др.), влияющих на эффективность процесса самого обучения. С другой стороны, эту проблему возможно обойти путем использования искусственного (синтетического) трафика, созданного имитационной моделью. Другими словами, генератор трафика способен сформировать любое количество чистых, детальных данных, покрывающих любое нужное множество характерных особенностей сетевого трафика, чего не всегда можно добиться при анализе реальных датасетов.

Таким образом, задача разработки имитационной модели генератора трафика, учитывающей особенности реального сетевого трафика и используемой в дальнейшем для машинного обучения элементов сети связи является актуальной.

В статье предложен подход по созданию имитационной модели генератора трафика, который состоит из двух шагов (рис. 1):

1 шаг: создание модели генератора трафика;

2 шаг: разработка механизма обучения нейронной сети на основе искусственного трафика, синтезируемого имитационной моделью.

Первый шаг включает в себя четыре самостоятельных этапа, связанных с подготовкой исходных данных для моделирования и непосредственной разработкой имитационной модели генератора трафика.

Второй шаг состоит из двух этапов:

– выделение классификационных признаков сетевого трафика с учетом определения структуры сети, в которой циркулирует классифицируемый трафик;

– реализация интерфейсного взаимодействия между моделью генератора и нейронной сетью с последующим обучением нейронной сети.

В данной статье раскрываются вопросы, связанные только с созданием модели генератора трафика (первые четыре этапа). Вопросы, связанные с обучением нейронной сети на основе разработанного генератора трафика, будут изложены в последующих публикациях.

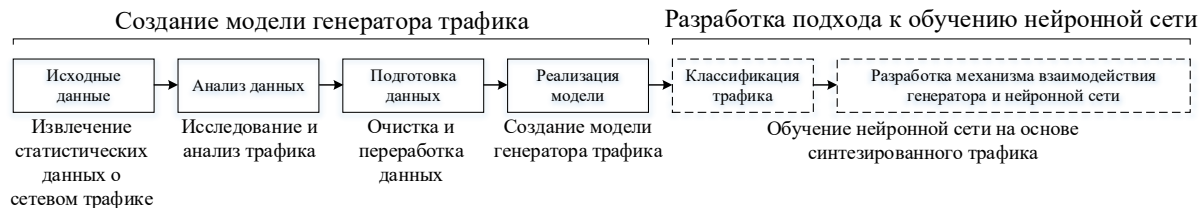


Рис. 1. Этапы разработки модели

Итак, рассмотрим первые четыре этапа предложенного подхода.

Этап «Исходные данные» представляет собой сбор статистики, который предлагается осуществлять с использованием утилиты NfSen на основе протокола NetFlow [1]. Сбор статистических данных включает в себя 2 шага:

1. Выделение из сетевого трафика отдельных потоков, характеризующихся значимыми параметрами: IP-адрес источника, IP-адрес назначения, порт источника для UDP и TCP, порт назначения для UDP и TCP, тип и код сообщения для ICMP, номер протокола IP (тип протокола транспортного уровня), сетевой интерфейс, значения поля ToS (уровень приоритета).

2. Формирование профиля трафика и базы статистических данных в формате .csv (рис.2).

Date first seen	Proto	Src IP Addr:Port	Dst IP Addr:Port	X-Src IP Addr:Port	X-Dst IP Addr:Port	In Byte
2022-03-22 14:33:17.010	UDP	10.56.156.47:26468	-> 10.88.1.38:50048	0.0.0.0:0	-> 0.0.0.0:0	63000
2022-03-22 14:33:17.010	UDP	10.88.1.38:50048	-> 10.56.156.47:26468	0.0.0.0:0	-> 0.0.0.0:0	66120
2022-03-22 14:33:17.006	UDP	10.23.66.35:123	-> 10.79.72.68:123	0.0.0.0:0	-> 0.0.0.0:0	76
2022-03-22 14:33:17.006	UDP	10.79.72.68:123	-> 10.23.66.35:123	0.0.0.0:0	-> 0.0.0.0:0	76
2022-03-22 14:33:17.012	UDP	10.56.156.47:26470	-> 10.88.1.38:50194	0.0.0.0:0	-> 0.0.0.0:0	124871

Рис. 2. Пример создаваемого утилитой NfSen файла со статистическими данными протокола NetFlow

Этап «Анализ данных» представляет собой анализ полученных статистических данных и извлечение необходимой информации для формирования шаблонов, необходимых для последующего обучения телекоммуникационного оборудования.

Анализ собранной статистики о телетрафике включает в себя следующую последовательность действий:

1. Изучение назначения и принципов функционирования рассматриваемой сети связи.
2. Определение IP-адресов элементов сети.
3. Фильтрация «мусорного» трафика (трафика не принадлежащего сети).
4. Определение распределения трафика по транспортным протоколам.
5. Определение распределения трафика по приложениям (службам).
6. Определение среднего размера пакета трафика фиксированной службы/приложения (порта) по направлениям передачи (клиент (запрос) / сервис (ответ)).
7. Определение распределения трафика конкретной службы/приложения (порта) по информационным направлениям.

8. Определение распределения трафика фиксированного сервиса / услуги / приложения (порта) по направлениям передачи (входящий/исходящий) по информационным направлениям.

9. Визуализирование трафика фиксированной службы/приложения по информационному направлению.

10. Анализ сеансов связи по фиксированной службе/приложению на фиксированном информационном направлении.

Этап «Подготовка данных» предназначен очистки и переработки статистических данных. На данном этапе осуществляется:

– формирование датасетов из собранной статистики, исключая содержание аномального трафика [2];

– подбор законов распределения характеристик трафика, описывающих интенсивность поступления трафика на порты телекоммуникационного оборудования;

– расчет значений коэффициента самоподобия.

Необходимо отметить, что для любой функционирующей сети объем данных, предоставляемый протоколом NetFlow, даже за относительно небольшое время исследования сети, позволяет произвести соответствующий анализ и сделать выводы о распределении характеристик трафика, при этом необходимо пользоваться правилами обработки ограниченного по объему статистического материала, изложенными в [3].

В рамках определения законов распределения характеристик трафика необходимо произвести расчет значений коэффициента самоподобия, который необходим для определения «пачечности» реального трафика с целью дальнейшего учета данных в разрабатываемой модели генератора трафика.

Расчет коэффициента самоподобия включает в себя следующие шаги [2]:

1. Разбиение полученного статистического ряда из  $N$  элементов на группы по  $k$  элементов ( $k > 10$ ) путем целочисленного деления  $N$ . В каждой группе вычисляется среднее арифметическое значение.

2. Расчет величины размаха накопленных отклонений по каждой группе

3. Расчет стандартного отклонения по каждой группе.

4. Расчет нормированного размаха для данного разбиения (берется отношение средних величин размаха и стандартного отклонения).

5. Поиск делителя  $N$  такого, что для  $(i + 1)$ -го разбиения  $k_{i+1} > k_i$ . Далее алгоритм повторяется для нового  $k$ .

6. Построение уравнения линейной регрессии для полученного набора данных.

Нахождение коэффициента самоподобия (одного из параметров линейной регрессии) осуществляется с помощью метода наименьших квадратов.

Весь перечисленный функционал этапа подготовки данных осуществляется с помощью разработанной на языке программирования C# библиотеки, методы которой подготавливают собранные статистические данные и «погружают» их в имитационную модель генератора трафика, разработанную в среде моделирования GPSS Studio.

Этап «Реализация модели» представляет собой разработку имитационной модели генератора трафика.

За основу модели предлагается взять хорошо известную On/Off модель [4], которая на языке системы массового обслуживания представляет собой неприоритетную, неэкспоненциальную нелинейную разомкнутую двухфазную систему массового обслуживания.

Разрабатываемая модель должна включать в себя четыре функциональных блока:

– блок очистки и переработки данных;

– блок расчета коэффициента самоподобия;

- блок генерации трафика;
- блок чередования On/Off периодов.

Первые два блока модели разрабатываются на языке C# и предназначены для подготовки части исходных данных (очистка данных, подбор закона распределения, расчет коэффициента самоподобия) для моделирования.

Остальные два блока отвечают за генерацию искусственного трафика, повторяющего характеристики телекоммуникационного трафика, статистика о котором была исследована на фрагменте сети.

Блок-схемы алгоритма функционирования модели представлены на рис. 3. В его основу положен, алгоритм описанный в [4]. Отличие заключается в процедуре подготовки исходных данных (описана выше), учитываемых в представленном алгоритме.

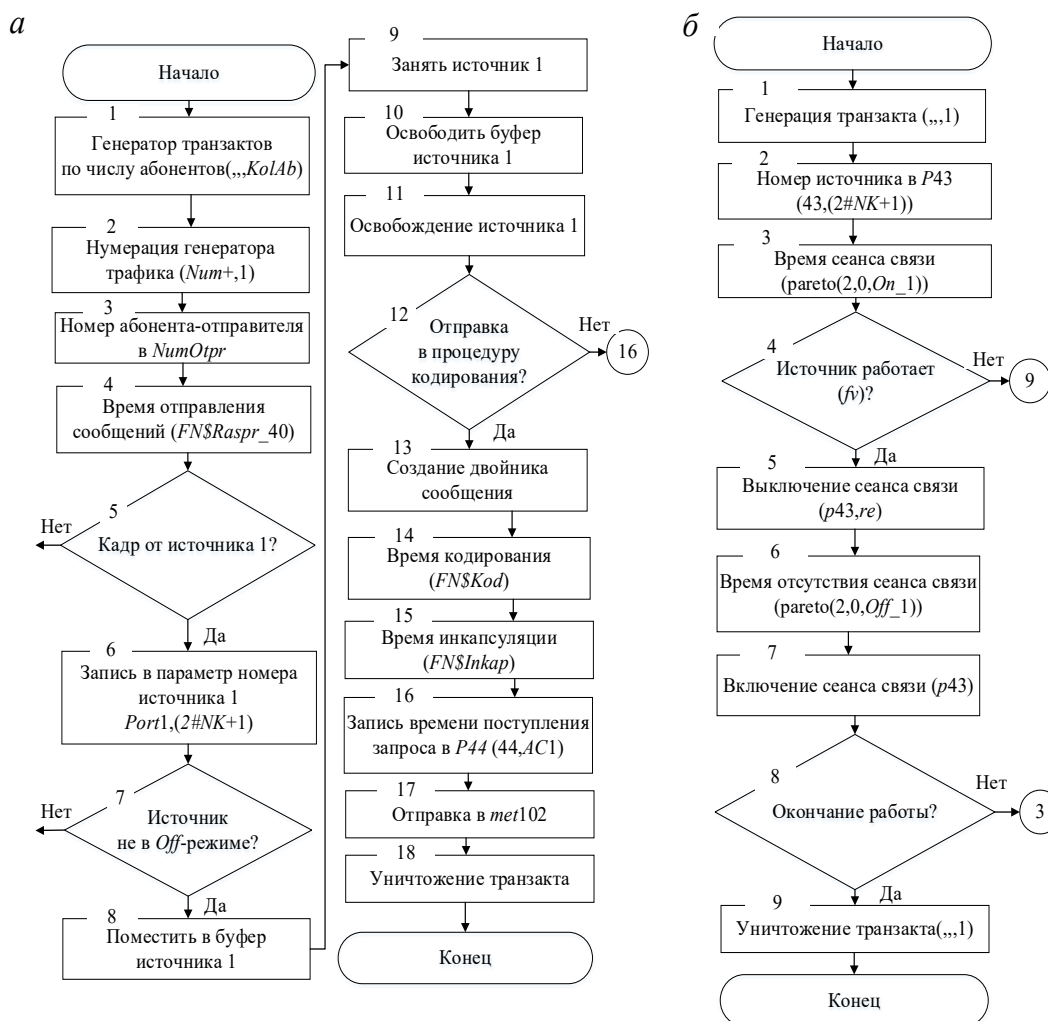


Рис. 3. Блок-схема алгоритма функционирования генератора трафика

Исходные данные, учитываемые в модели, представлены в таблице 1.

Таблица 1

Основные нагрузочные параметры, учитываемые в модели генератора трафика

Скорость кодирования $Vl$ , кбит/с	+
Алгоритмическая задержка кодирования $Tkod$ , мс	+
Задержка инкапсуляции $Tink$ , мс	+

Закон распределения длительности задержки инкапсуляции	+
Количество байт в <i>Ethernet</i> кадре, $W1$	+
Интенсивность передачи кадров $\lambda$ , кадр/с	+
Закон распределения интенсивности кадров	+
Длительность <i>ON</i> -периода $T1$ , с	+
Длительность <i>OFF</i> -периода $T2$ , с	+
Закон распределения <i>ON/OFF</i> периодов	+
Коэффициент Хёрста	+
Параметры утилиты $NfSen$	+

Имитационную модель генератора трафика предлагается реализовать в среде моделирования GPSS Studio, исходя из следующих ее преимуществ:

- высокое быстродействие при вычислении кода на языке GPSS;
- высокая апробированность языка GPSS;
- возможность объединения всего жизненного цикла модели в рамках общего проекта;
- обеспечение хранения модели и результатов экспериментов в единой базе данных;
- методическая и информационная связность всех этапов исследования в рамках единого комплекса программ;
- применение единого стандарта языка общения с исследователем.

### Вывод

Таким образом, представленный в статье подход позволяет разработать имитационную модель генератора трафика, учитывающую статистические особенности реального телекоммуникационного трафика на исследуемом фрагменте сети связи.

На следующем шаге предстоит решить вопросы, связанные с:

выделением классификационных признаков сетевого трафика с учетом определения структуры сети, в которой циркулирует классифицируемый трафик;

реализацией интерфейсного взаимодействия между моделью генератора и нейронной сетью с последующим обучением нейронной сети.

Дальнейшим этапом исследования является разработка методики применения имитационных моделей генератора трафика для предиктивного анализа состояния элементов сети связи (рис. 4).

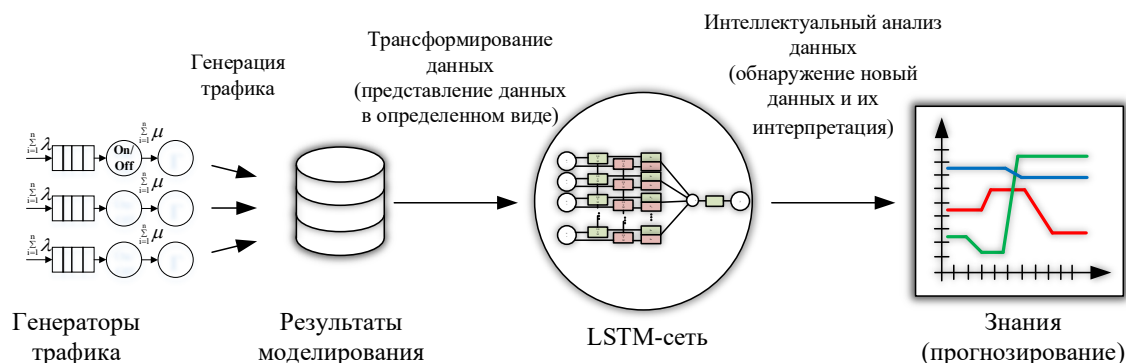


Рис. 4. Схема применения генераторов трафика и LSTM-сети для прогнозирования трафика

Реализация такого подхода находит свое применение в набирающем популярность в настоящее время предиктивном обслуживании (Predictive Maintenance), которое проводится с целью оценки состояния оборудования путем осуществления периодического (автономного) или непрерывного (онлайн) мониторинга.

Прогнозирование поведения элементов сети связи позволит заблаговременно реагировать на неполадки сети, способствуя тем самым повешению эффективности ее функционирования.

### Литература

1. <https://nfsen.sourceforge.net/>
2. Шелухин О.И., Ерохин С.Д., Ванюшина А.В. Классификация IP-трафика методами машинного обучения / под редакцией Шелухина О.И. Горячая линия-Телеком, 2018. 282 с.
3. Вентцель Е.С. Теория вероятностей: учеб. для вузов. 6-е изд. стер. М.: Высш. шк., 1999. 576 с.: ил.
4. Ануфренко, А.В. Имитационное моделирование узла агрегации сети связи / А.В. Ануфренко, И.С. Мешков, М.А. Снятков // Т-Сomm: Телекоммуникации и транспорт. 2018. Т. 12, № 4. С. 56–61. DOI 10.24411/2072-8735-2018-10066. EDN ХОХХКН.