

**ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ ИСТОЧНИКОВ НАГРУЗКИ
В СЕТЯХ ПЕРЕДАЧИ ДАННЫХ С КОММУТАЦИЕЙ ПАКЕТОВ**

**AN IMITATING MODELLING OF THE LOAD SOURCES
IN A PACKET DATA NETWORKS**

Аннотация. В статье рассматривается метод имитационного моделирования источников нагрузки в сетях передачи данных с пакетной коммутацией, основанный на «динамическом моделировании Маркова».

Summary. This article describes an imitating modelling method for the load sources in a packet networks, which is based on a technic called “dynamic markov modelling”.

Имитационное моделирование сетей передачи данных можно свести к моделированию следующих их компонент:

1. Источников нагрузки.
2. Узлового оборудования.
3. Каналов передачи данных.

Традиционно моделирование статистических характеристик источников сводится к применению аналитических моделей, т.е. выбору какого-либо распределения, характеризующего статистику источника с последующим определением параметров этого распределения. Одной из альтернатив этого подхода является имитационное моделирование источников нагрузки. В этом случае источник нагрузки моделируется при помощи некоторого алгоритма, который использует информацию о значениях моделируемой характеристики источника нагрузки в прошлом для прогнозирования этой характеристики в будущем. Такой подход обладает значительно большей универсальностью применения, чем аналитическое задание распределения и его параметров, и позволяет проводить обучение модели на реальных сетях с интересующим набором сетевых приложений. В конечном итоге можно создать модель, позволяющую прогнозировать значения величины интенсивности нагрузки моделируемого источника, а также синтезировать последовательности значений этой величины, подобные тем, что имеют место на практике [3].

Целью работы является изложение общих принципов и предпосылок моделирования источников нагрузки при помощи динамической имитационной модели, использующей автоматы с конечным числом состояний.

При реализации динамической имитационной модели будем полагать, что:

- моделируемой характеристикой источника является интенсивность нагрузки, создаваемой источником;
- интенсивность нагрузки рассматривается в рамках заранее выбранного временного интервала;
- существует зависимость между предыдущими и последующими значениями интенсивности нагрузки, т.е. рассматривается источник сообщений с памятью.

Предлагаемая модель базируется на технике «динамического моделирования Маркова» (Dynamic Markov Modelling, DMM), которая хорошо себя зарекомендовала в архиваторах как основополагающая часть метода «динамического сжатия Маркова» (Dynamic Markov Compression, DMC) [1, 2].

Рассмотрим поток символов, создаваемый источником, удовлетворяющим сформулированным выше требованиям. На каждом шаге работы модели обрабатывается очередной символ входного потока. Имитационная модель DMM представлена автоматом с конечным набором состояний. Такой автомат можно изобразить в виде ориентированного графа, называемого диаграммой состояний. Каждой дуге такого графа соответствует переход в другое состояние, определяемое информацией из входного потока, а также набор вероятностей появления символов входного потока на текущем шаге. Эти вероятности являются также вероятностями перехода в другие состояния.

Вычисление вероятностей переходов реализуется при помощи частотных счетчиков переходов [4]:

$$P_{пер\ A \rightarrow B} = \frac{Счетчик_{пер\ A \rightarrow B}}{\sum_{по\ всем\ i} Счетчик_i}$$

Работа имитационной модели начинается с некоторой начальной диаграммы состояний (например, с единственного состояния) при равных значениях счетчиков, причем $Счетчик_i \geq 1, \forall i$.

При получении некоторого символа входного потока происходит переход модели в следующее состояние, соответствующее пришедшему символу и модификация частотных счетчиков, соответствующих состоянию, из которого произошел переход:

$$Счетчик_{пер\ A \rightarrow B} = Счетчик_{пер\ A \rightarrow B} + 1.$$

На начальном шаге модель состоит из одного состояния, все переходы замкнуты. По мере поступления символов из входного потока, происходит корректировка диаграммы состояний модели. Корректировка состоит в клонировании состояния, в которое осуществляется переход, и изменении маршрутов перехода. Чем больше происходит делений состояний, тем больше увеличивается порядок контекстов, используемых моделью для прогноза. Поэтому деление необходимо производить только в том случае, когда есть уверенность в существовании соответствующих зависимостей от этих контекстов.

Для определения необходимости клонирования состояния существуют разные подходы, один из них заключается в сравнении количества переходов из текущего состояния в копируемое и количества переходов в копируемое из всех остальных состояний, с наперед заданными константами $c1$ и $c2$, являющимися параметрами модели. Если указанные количества переходов превышают эти пороговые константы, то производится клонирование состояния, в которое осуществляется переход. При этом счетчики частот символов делятся между двумя состояниями – копируемым и новым образованным в результате клонирования. Одним из вариантов распределения счетчиков является пропорциональное распределение между новым и копируемым состоянием пропорционально отношению количества переходов: а) из предыдущего состояния в копируемое и б) из всех остальных состояний в копируемое. Для случая, когда переход осуществляется из начального состояния в него же, в качестве критерия клонирования используется сравнение количества переходов по данной дуге (вызванных данным символом входного потока) и количества переходов по другим дугам.

Алгоритм динамического построения диаграммы состояний проиллюстрирован на рис. 1:

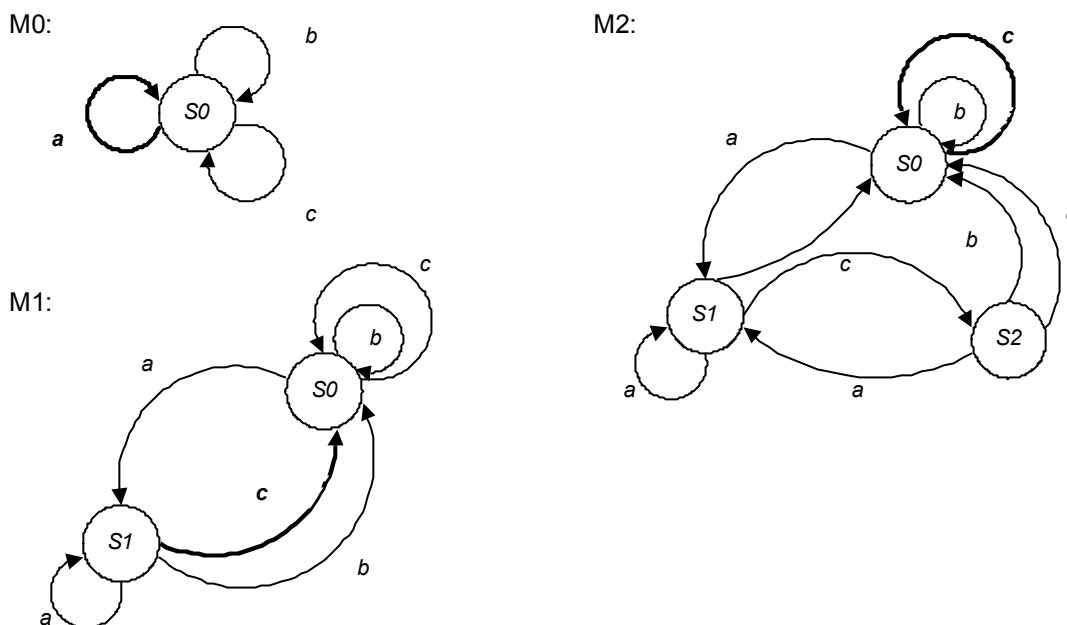


Рисунок 1 – Динамическое построение диаграммы состояний:

M_i – модель на i -м шаге, жирным выделены пришедшие на вход символы и произошедшие переходы

Входной алфавит полагается трехсимвольным – a, b, c . Трех возможным символам входного потока соответствуют переходы, представленные на рисунке дугами. Предполагается, что на нулевом шаге M_0 из входного потока пришел символ « a ». На следующем шаге M_1 (пришедший символ « c ») на рисунке представлена диаграмма состояний после первого клонирования состояния

S_0 (предполагается, что условия клонирования выполнены). На шаге **M2** (пришедший символ «с») представлена модель после второго клонирования состояния S_0 .

Выходом модели является набор вероятностей появления символов, вычисляемый на основании значений частотных счетчиков на каждом шаге по приведенной выше формуле.

При использовании данного метода для моделирования источников нагрузки, входным потоком является поток значений моделируемой случайной величины нагрузки, создаваемой исследуемым приложением. При этом может потребоваться модификация алгоритма для работы с очень большим алфавитом, в том случае, если множество возможных значений интенсивности нагрузки моделируемого источника велико.

Для апробации предложенного подхода было разработано специальное программное обеспечение, позволяющее проводить анализ, визуализацию и моделирование нагрузки, создаваемой реальными источниками сетевой нагрузки. В качестве входных данных ПО использует трассы сетевого обмена, получаемые при помощи программного анализатора протоколов Sniffer Pro, компании NAI. Данный анализатор позволяет проводить захват кадров канального уровня (например, Ethernet) и осуществлять декодирование большого числа протоколов канального, сетевого и транспортного уровней. Данные декодирования могут быть сохранены в журнальный файл для последующей обработки другими программами.

Frame	Src. Address	Dest. Address	Size	Rel. Time	Delta Time	Abs. Time	Summary
1 M	[195.5.27.19]	[195.5.27.18]	60	0:00:00.000	0.000.000	14.12.2001 15:16:10	TCP: D=3140 S=135 RST ACK=4167119396 WIN=0
2	[195.5.27.19]	[195.5.27.18]	60	0:00:00.468	0.468.642	14.12.2001 15:16:11	TCP: D=3140 S=135 RST ACK=4167119396 WIN=0
3	[195.5.27.19]	[195.5.27.18]	60	0:00:00.970	0.501.583	14.12.2001 15:16:11	TCP: D=3140 S=135 RST ACK=4167119396 WIN=0
4	[195.5.27.19]	[195.5.27.18]	271	0:00:02.692	1.721.971	14.12.2001 15:16:13	WINS: R ID=33295 OP=QUERY STAT=OK
5	[195.5.27.19]	[195.5.27.18]	60	0:00:03.832	1.139.905	14.12.2001 15:16:14	TCP: D=3142 S=135 RST ACK=4168131577 WIN=0
6	[195.5.27.19]	[195.5.27.18]	60	0:00:04.280	0.448.719	14.12.2001 15:16:14	TCP: D=3142 S=135 RST ACK=4168131577 WIN=0
7	[195.5.27.19]	[195.5.27.18]	60	0:00:04.782	0.501.575	14.12.2001 15:16:15	TCP: D=3142 S=135 RST ACK=4168131577 WIN=0
8	[195.5.27.19]	[195.5.27.18]	271	0:00:06.511	1.729.405	14.12.2001 15:16:17	WINS: R ID=33297 OP=QUERY STAT=OK
9	[195.5.27.19]	[195.5.27.18]	93	0:00:40.712	34.200.318	14.12.2001 15:16:51	CIFS/SMB: R Tree Disconnect Status= OK
10	[195.5.27.19]	[195.5.27.18]	93	0:00:40.712	0.000.792	14.12.2001 15:16:51	CIFS/SMB: R Tree Disconnect Status= OK
11	[195.5.27.19]	[195.5.27.18]	93	0:00:40.713	0.000.829	14.12.2001 15:16:51	CIFS/SMB: R Tree Disconnect Status= OK

Рисунок 2 – Пример журнального файла Sniffer Pro

На основании данных журнального файла, созданное ПО позволяет строить выборку интенсивностей нагрузки для заданного временного промежутка с заданной точностью, т.е. преобразовывать данные журнала в последовательность величин интенсивности нагрузок для выбранного разбиения (по умолчанию используется 100 мс).

ПО позволяет просматривать гистограмму полученных интенсивностей нагрузки, а так же вычислять ряд статистических параметров: среднее выборочное, дисперсию, коэффициенты асимметрии и эксцесса.

На основании данных загруженной сетевой трассы может быть проведено обучение динамической модели, а также ее последующее моделирование источника. ПО позволяет варьировать параметры создаваемой модели (константы деления, интервалы выборки, алгоритмы распределения счетчиков при клонировании). Полученные результаты моделирования визуализируются либо сохраняются в текстовом файле для дальнейшей обработки.

Для оценки качества моделирования используется следующий показатель:

$$K_{\text{прогноза}} = \frac{\sum_{i=0}^{N-1} P_i(i)}{N},$$

где i – шаг моделирования, $P_i(i)$ – предсказанная вероятность символа, появившегося на i -м шаге, N – количество шагов моделирования.

Результаты моделирования представлены на рис. 3.

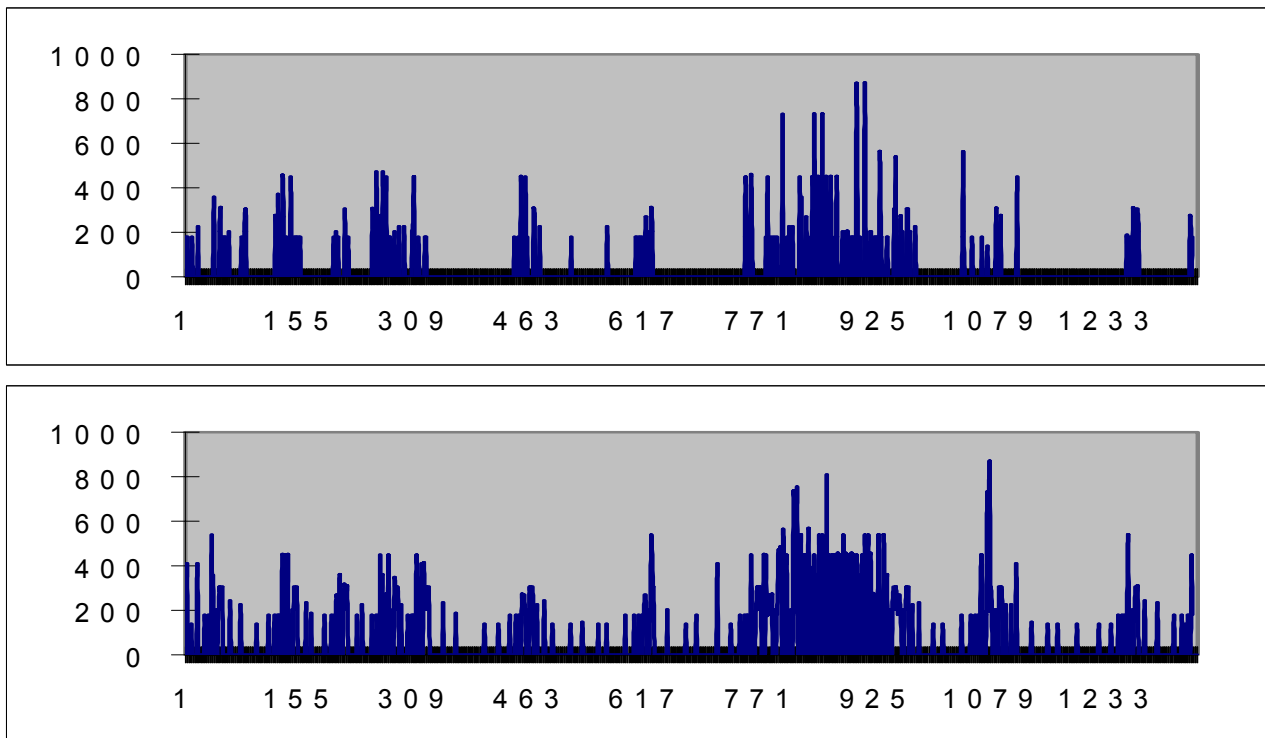


Рисунок 3 – Результаты моделирования для голосового трафика G.723.

Вверху – исходный трафик, внизу – данные, полученные при помощи построенной модели

Таблица 1 – Параметры моделирования

Время наблюдения (ч:мин:с.мс)	0:02:15.087
Интервал выборки, $t_{\text{изм}}$ (мс)	100
Количество временных интервалов разбиения	1345
Максимальная зарегистрированная нагрузка (байт/ $t_{\text{изм}}$)	872
Минимальная зарегистрированная нагрузка	0
C1	10
C2	20
Коэффициент прогноза	0,5496
Число состояний модели	60

Таблица 2 – Сравнение статистических характеристик исходного и генерируемого моделью трафика

Среднее выборочное/Дисперсия для исходного трафика	6.98052044609665E+0001	5.71013667634960E+0004
Среднее выборочное/Дисперсия для трафика, полученного при помощи модели	6.77249070631970E+0001	5.29603786299480E+0004

Исходя из приведенных в табл. 2 данных, можно сделать вывод о том, что полученная модель, в данном случае, позволила осуществить генерацию трафика, достаточно близкого к исходному по дисперсии и средневывборочному значению.

Использование построенной модели проходит по следующему сценарию, состоящему из двух фаз:

1. Обучение с использованием трасс работы исследуемых сетевых приложений и сервисов.
2. Использование модели для генерации трафика подобного трафику реальных приложений и сервисов, что может быть использовано в имитационных моделях сетей передачи данных.

Реализация модели была апробирована на различных трассах сетевого обмена, в их числе: трассы передачи голосовых данных с использованием Ethernet\IP\UDP\RTP\G723, передача файлов по протоколу Netbios (Ethernet\IP\UDP\RTP\SMB), вебсерфинг (Ethernet\IP\TCP\HTTP).

Литература

1. *Cormack G.V., Horspool R.N.* Data Compression Using Dynamic Markov Modelling. The Computer Journal. – Vol.30. – No.6, 1987. – PP. 541-550.
2. *Захаров М.* Описание метода Динамического сжатия Маркова, 1993.
3. *Бертсекас Д, Галлагер Р.* Сети передачи данных. – М.: Мир, 1989.
4. *Феллер В.* Введение в теорию вероятностей и её приложения. – М.: Мир, 1984.