

## Сравнительный анализ подходов к моделированию трафика в сетях передачи данных

А.И. Хобня

Представлены результаты сравнительного анализа подходов к моделированию трафика в сетях передачи данных, в частности в сетях нового поколения NGN. Рассматриваются преимущества и недостатки как ранних, так и современных методов. Предлагаются пути решения проблем и направления развития современных методов моделирования трафика в сетях NGN.

**Ключевые слова:** сетевой трафик, моделирование, сети, NGN.

The results of comparative analysis of networks traffic simulation approaches and traffic simulation in Next Generation Networks (NGN) in particular are presented. Advantages and disadvantages of both early and modern approaches are considered. The ways of solving problems and future work directions for modern approaches of traffic simulation are presented.

**Keywords:** network traffic, simulation, networks, NGN.

**Введение.** Целями моделирования сетей NGN являются определение оптимальной топологии, выбор соответствующего сетевого оборудования, определение характеристик работы сети и возможностей для ее будущего развития.

В процессе моделирования сети могут выясняться следующие параметры:

- предельная пропускная способность определенного фрагмента сети;
- зависимость потерь пакетов от нагрузки на отдельных узлах сети;
- время отклика серверов в различных режимах, в том числе в режимах экстремальной нагрузки;
- влияние подключения новых серверов на перераспределение загрузки различных участков сети;
- оптимальная топология сети;
- оптимальные протоколы маршрутизации и их параметры;
- максимальное число пользователей того или иного сервера сети;
- степень влияния мультимедийного трафика на работу сети.

Существуют аналитические и имитационные модели сетей. Аналитические модели сетей как правило строятся на основе аппарата теорий вероятности, массового обслуживания и марковских процессов. Также в данных моделях применяются методы диффузной аппроксимации, алгебраические и дифференциальные уравнения. Несмотря на значительные достижения в области математического моделирования, многие реальные системы невозможно представить с помощью соответствующих математических моделей. При наличии возможности формализации какой-либо системы посредством построения математической модели, полученная на ее основе задача оптимизации может являться слишком сложной для современных алгоритмов решения задач данного класса.

Альтернативным подходом моделирования сложных систем является *имитационное моделирование (ИМ)*, метод исследования, при котором изучаемая система заменяется моделью, с достаточной точностью описывающей реальную систему, с которой проводятся эксперименты с целью исследования данной системы.

**Подходы к моделированию трафика и их проблемы.** Одним из важнейших аспектов моделирования сетей NGN является моделирование трафика. Моделирование сетевого трафика используется как основа для разработки сетевых приложений и для планирования мощностей сетевых систем.

Существуют две основные характеристики, генерируемые моделями сетевого трафика: распределение длин пакетов и распределение длин временных интервалов между прибытиями пакетов. Другие параметры, такие как маршруты, распределения мест и т. д., имеют меньшее значение.

Проблема распределения размера пакетов довольно хорошо изучена на данное время. Существующие модели размеров пакета являются действенными и простыми. Большинство моделей размера пакетов не рассматривают проблему порядка размеров. Например, дегамма TCP в одном направлении, скорее всего, последует крошечной АСК в другом направлении спустя около половины одного времени обхода (RTT).

Проблема распределения интервалов времени между прибытиями пакетов гораздо сложнее. Понимание сетевого трафика значительно менялось со временем, что привело к серии эволюций в подходах к моделированию сетевого трафика.

Самые первые модели были основаны на основе простых распределений вероятности с предположением, что они справедливы для больших выборок. Например, распределение трафика Пуассона (с экспоненциальными интервалами между прибытиями пакетов) часто использовались в различных сетевых моделях [1]. Данные модели использовались в литературе с появлением компьютерных сетей, а до этого использовались в области телекоммуникаций [2]. Модели без использования памяти очень привлекательны с аналитической точки зрения [3], и при правильном подборе параметров модели Пуассона могут показывать достаточно точное соответствие для большинства последовательностей сетевого трафика на достаточно коротких промежутках времени. Чистые модели Пуассона имеют один параметр – средняя скорость прибытия пакета ( $\lambda$ ). Таким образом, интервалы между прибытиями пакетов имеют экспоненциальное распределение со средним равным  $1/\lambda$ . Прибытия пакетов на интервале  $[t_1, t_2]$ , если  $t=t_2-t_1$ , имеют распределение Пуассона со средними  $t\lambda$ .

Модели Пуассона популярны в теории массового обслуживания, потому что они имеют привлекательные аналитические свойства: они не имеют памяти (будущее поведение не имеет связи с прошлым или последним поведением), а также сумма нескольких потоков Пуассона генерирует новый Пуассона поток с  $\lambda'=\sum\lambda$ . Благодаря этим привлекательным свойствам, модели Пуассона были предложены в качестве основы для моделирования сетевых коммуникаций [3].

Составная модель Пуассона расширяет базовую возможность доставки партий пакетов. Интервалы времени прибытия пакетов внутри партии распределены экспоненциально, в то время как размеры партий распределены геометрически [4]. Данная модель позволяет генерировать неравномерный пульсирующий трафик. Составная модель Пуассона имеет некоторые из аналитических преимуществ чистой модели Пуассона: она по-прежнему не использует память, сумма потоков по-прежнему генерирует поток Пуассона, и стационарное уравнение по-прежнему достаточно просто рассчитать.

Однако, простые модели Пуассона демонстрировали слишком слабую пульсацию для реалистичного сетевого трафика. В частности в телекоммуникационных сетях наблюдаются всплески прибытия пакетов (когда люди говорят) и пауза между всплесками (когда люди молчат). В 1986 г. голос и данные начали делить телекоммуникационные каналы связи, таким образом исследования трафика начали рассматривать суперпозицию трафика из нескольких типов источников.

Реальный процесс, лежащий в основе голосовой сети, слишком сложен для эффективного моделирования. Исследователи стремились создать упрощенную модель. Необходима была модель трафика с изменяющейся скоростью прибытия пакетов. Тем не менее, в соответствии с голосовой моделью, имеет смысл изменять скорость прибытия квантовым образом. Очевидным решением стали цепи Маркова Пуассона (MMPPs) [5]. Непрерывная цепь Маркова определяет скорость прибытия модели Пуассона. Цепь Маркова является цепью с двумя состояниями, каждое с соответствующей скоростью  $\lambda$  и с определенным средним временем действия  $\tau$ . Таким образом MMPP определяется кортежем  $(\lambda_1, \lambda_2, R_1, R_2)$ . Для определения этих четырех параметров необходимы измерения реального сетевого трафика. Параметры MMPP выбираются таким образом, чтобы соответствовать реальному трафику для следующих характеристик:

- скорость прибытия пакетов;
- отношение дисперсии к среднему для числа прибывших пакетов в течение короткого периода времени;

– отношение дисперсии к среднему для числа прибывших пакетов в течение длительного периода времени;

– третий момент для числа прибывших пакетов в течение короткого периода времени.

К сожалению, модели трафика на основе MMPPs были валидированы путем симуляции с параметрами, которые были выбраны для голосового трафика, идущего по телекоммуникационным каналам связи. Как и ожидалось, результаты аналитических исследований модели соответствовали результатам имитационного моделирования.

Описанные выше модели, как правило, были недостаточно точны, если применялись для достаточно длительных промежутков времени. В начале 1990-х гг. были проведены всесторонние исследования большого количества высококачественных измерений сетевого трафика. Анализ данных показал четкие доказательства самоподобия [6]. Это объясняло расхождение между моделями и реальным поведением сетевого трафика. Модели трафика на основе самоподобия сложнее создавать и анализировать, также сложно вычислить параметры для самоподобных моделей, используя измерения реального сетевого трафика. Тем не менее, несколько моделей были улучшены путем использования генерации самоподобного трафика [7]. Впоследствии были созданы системы, упрощающие анализ существующих измерений трафика для вычисления реалистичных параметров самоподобных моделей [8].

Когда самоподобные модели трафика были впервые введены, не было каких-либо эффективных поддающихся анализу процессов для генерации данного типа моделей. Затем был разработан стохастический процесс для создания модели с самоподобным вводом и постоянной скоростью вывода. Данная модель основана на нормализованном броуновском движении [9]. Несмотря на то, что данная начальная модель была непрерывной, а не дискретной, она была эффективной и простой.

Все ранние самоподобные модели трафика имеют один существенный недостаток: оценки параметров самоподобия реального сетевого трафика требуют огромных объемов данных и тяжелых вычислений. Метод, используемый в первоначальном исследовании требует сходимости тысяч наблюдений [10]. Наиболее современный метод, вейвлет-анализ нескольких разрешений [8], является более эффективным, но все еще очень дорогостоящим. Это, безусловно, нежелательно для моделирования трафика. В более поздних исследованиях авторы пытаются построить систему, которая фиксирует пакетные взаимодействия, извлекает распределения из различных аспектов трафика, а затем моделирует реалистичный трафик с теми же характеристиками [11].

Метод SWING использует сравнительно простую модель для анализа сетевого трафика и дальнейшей симуляции. Модель анализирует характеристики пользователей, обмена запрос-ответ (RRE), соединений, отдельных пакетов и всей сети в целом. В рамках данной модели не предпринимается никаких попыток проанализировать характеристики самоподобия трафика. Самоподобие в созданном данной моделью трафике приходит естественно, благодаря агрегации множества периодических включений и выключений различных источников.

Метод SWING проходил валидацию на данных реального трафика путем извлечения коэффициентов Хёрста от генерируемого и реального трафика. Данная модель работает очень хорошо даже на близких к RTT временных масштабах. Можно по-прежнему ожидать, что математически сгенерированный трафик не будет проявлять самоподобие на чрезвычайно больших периодах времени. Тем не менее, с практической точки зрения, SWING работает исключительно хорошо для моделирования web-трафика.

**Моделирование трафика сетей NGN.** В сетях NGN различным видам трафика присваивается различный приоритет. Поэтому очень важно иметь возможность моделирования различных видов трафика отдельно. При моделировании существующей сети NGN можно воспользоваться одним из существующих алгоритмов самоподобия, использующих измерения характеристик трафика в реальной сети. Однако данные методы не могут помочь исследователю для моделирования будущей сети или поведения сети в других эксплуатационных условиях. Например, в процессе моделирования сети NGN могут выясняться следующие вопросы:

– как изменится нагрузка после подключения новых пользователей;

- как изменится нагрузка после введения новых сервисов;
- как изменится нагрузка после изменения среднего времени разговора (например, при предоставлении услуг телефонии офисам, колл-центрам и т. д.).

Для моделирования описанных выше ситуаций имеет смысл использовать алгоритм моделирования трафика, который в качестве параметров использует характеристики внешней для сети среды (средняя продолжительность разговора, средняя интенсивность разговора, средняя интенсивность использования других типов услуг NGN сети и т.д.), т.е. алгоритм на основе поведенческих моделей пользователей. Алгоритм такого типа позволит предсказывать характеристики трафика в зависимости от условий эксплуатации NGN сети пользователями:

- какие сервисы NGN сети используются пользователями;
- как часто данные сервисы используются;
- каким образом пользователи используют данные сервисы (длительность, интенсивность и т. д.).

В качестве основы для моделирования трафика в сети NGN может быть взят алгоритм SWING. Однако оригинальный алгоритм SWING направлен в большей степени на моделирование web-трафика. Для моделирования VoIP и других типов трафика необходимо осуществить определенные модификации данного алгоритма, учитывающие специфику протоколов данных типов трафика. Например, для VoIP трафика, как правило, одно соединение соответствует одному пользовательскому сеансу, который является VoIP-вызовом. Однако количество передаваемых данных неравномерно в пределах одного соединения, т.к. современные VoIP протоколы поддерживают VAD (Voice Activity Detection). VAD позволяет экономить пропускную способность каналов связи, т.к. перерывы в речи не оцифровываются и не передаются по сети. Для наиболее точного моделирования VAD следует учитывать наличие возможности осуществлять VoIP-вызовы с участием более двух пользователей. Количество пользователей в конкретном VoIP-вызове влияет на продолжительность пауз, в пределах которых не генерируется исходящий трафик, т.к. пользователь слушает другого участника вызова. Следует учитывать возможные различия параметров распределения продолжительности и частоты пауз в рамках различных VoIP-вызовов. Для более точного моделирования VoIP трафика необходимо использовать наиболее реалистичные распределения характеристик пользовательских сеансов. Современные работы [12] в области моделирования VoIP трафика используют экспоненциальные распределения для продолжительности вызова и интервалов между вызовами, при этом исследования данных характеристик для телефонных сетей обнаруживают иные зависимости. Для некоторых типов протоколов следует учитывать возможности балансировки нагрузки и распределенной обработки информации [13], [14].

Таким образом, моделируя необходимые реальные эксплуатационные условия, возможно осуществить правильный выбор сетевого оборудования и определить этапы будущего развития моделируемой сети NGN.

**Заключение.** Рассмотрены преимущества и недостатки различных методов моделирования сетевого трафика. Обозначена специфика моделирования трафика сетей нового поколения NGN. Предложено учитывать наличие возможности осуществлять VoIP-вызовы с участием более двух пользователей и различия параметров распределения продолжительности и частоты пауз в рамках различных VoIP-вызовов при моделировании работы VAD, а также использовать распределения, полученные в результате исследований телефонных сетей, при моделировании VoIP трафика в сетях NGN.

## Литература

1. Marathe, M. Predicted Capacity of Ethernet in a University Environment / M. Marathe, W. Hawe // Proceedings of Southcon. – 1982. – P. 1–10.
2. Mandelbrot, B. Self-similar error clusters in communication systems and the concept of conditional stationarity / B. Mandelbrot // IEEE Transactions on Communication Technology. – 1965. – Т. 13. – № 1. – P. 71–90.
3. Frost, V. Traffic Modeling for Telecommunications Networks / V. Frost, B. Melamed // IEEE Communications Magazine. – 1994. – Т. 32. – № 3. – P. 70–81.

4. Jain, R. Packet Trains - Measurements and a New Model for Computer Network Traffic / R. Jain, S.A. Routhier // *IEEE Journal on Selected Areas in Communications*. – 1986. – Т. 4. – № 6. – P. 986–995.
5. Heffes, H. A Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance / H. Heffes, D.M. Lucantoni // *IEEE Journal on Selected Areas in Communications*. – 1986. – Т. 4. – № 6. – P. 856–868.
6. Willinger, W. The Discovery of Self-Similar Traffic / W. Willinger // *In Performance Evaluation: Origins and Directions* / Под. ред. G. Haring, C. Lindemann, M. Reiser. – London : Springer-Verlag, 2000. – Vol. 2. – P. 513–527.
7. Erramilli, A. Self-Similar Traffic and Network Dynamics / A. Erramilli, M. Roughan, D. Veitch, W. Willinger // *In Proc. of the IEEE*. – 2002. – Т. 90. – № 5. – P. 800–819.
8. Abry, P. Wavelet analysis of long-range-dependent traffic / P. Abry, D. Veitch // *IEEE Transactions on Information Theory*. – 1998. – Т. 44. – № 1. – P. 2–15.
9. Norros, I. A storage model with self-similar input / I. Norros // *Queueing Systems*. – 1994. – Т. 16. – № 3–4. – P. 387–396.
10. Leland, W.E. On the self-similar nature of Ethernet traffic (extended version) / W.E. Leland, M.S. Taqqu, W. Willinger, D.V. Wilson // *Networking, IEEE / ACM Transactions*. – 1994. – Т. 2. – № 1. – P. 1–15.
11. Vishwanath, K.V. Swing : Realistic and Responsive Network Traffic Generation / K.V. Vishwanath, A. Vahdat // *Networking, IEEE / ACM Transactions*. – 2009. – Т. 17. – № 3. – P. 712–725.
12. Ji, L. Conversational Model Based VoIP Traffic Generation / L. Ji, X. Yin, X. Shi, Z. Wang // *Networking and Services, 2007. ICNS. Third International Conference on*. – 2007. – 14 p.
13. Демиденко, О.М. Имитационное моделирование распределенной обработки информации в ЛВС / О.М. Демиденко, И.В. Максимей, С.Ф. Маслович, В.И. Селицкий // *Известия Гомельского государственного университета им. Ф. Скорины*. – 2008. – № 5 (50). – С. 140–143.
14. Быченко, О.В. Имитационные модели режимов обработки информации в ЛВС / О.В. Быченко, О.М. Демиденко, В.Д. Левчук, С.Ф. Маслович, В.И. Селицкий // *Известия Гомельского государственного университета им. Ф. Скорины*. – 2008. – № 5 (50). – С. 3–6.

Гомельский государственный  
университет им. Ф. Скорины

Поступила в редакцию 14.10.2014