

В.С. Замятин

Использование статистического подхода при решении задач анализа и управления компьютерными сетями

В свете современных достижений в области телекоммуникационных технологий особую актуальность приобретают проблемы анализа и моделирования компьютерных сетей, в том числе анализа и оптимизации трафика.

Сложность и высокая стоимость современных и развивающихся телекоммуникационных систем (ТС) постепенно приводят к вытеснению эвристических подходов к формированию их архитектуры, выбору основных конструктивных параметров и оценке характеристик, заменяя их предварительным анализом и обоснованием с использованием формальных методов. При этом исследования, направленные на совершенствование методов управления трафиком, позволяют повысить эффективность функционирования ТС, увеличить пропускную способность каналов связи и скорость передачи данных.

В последние годы в США, Японии, западноевропейских странах ведутся активные математические исследования современных телекоммуникационных технологий. В исследования вовлечены десятки университетских групп и крупные исследовательские центры. Однако в свете имеющихся публикаций по математическому исследованию подобных систем следует сказать, что разрабатываемые ими аналитические модели в большинстве случаев носят графовый характер и возникают почти исключительно в задачах оптимизации структуры сетей. Математические методы, применяемые при этом, сводятся к известным методам теории массового обслуживания и теории потоков в сетях.

Существуют некоторые трудности использования уже наработанных моделей в рамках теории систем и сетей массового обслуживания, связанные с дефицитом и гетерогенностью информации о функционировании быстроразвивающихся сетей. Поэтому в некоторых случаях более предпочтительным является статистический подход к анализу загрузки телекоммуникационных каналов. Он основан на значительно меньшей информации, чем требует теория массового обслуживания, поскольку в качестве объекта исследования берется только величина загрузки самого канала, ко-

торая фиксируется через определенные равные интервалы времени.

Трафик (объем загрузки) каждого канала связи телекоммуникационной сети является важным фактическим показателем ее работы. Анализ трафика позволяет оценивать фактическую загрузку сети и необходимую емкость ее каналов, выяснять устойчивость работы сети и оперативность реакции на различные нештатные ситуации, судить о динамике развития сети и планировать сроки ее модернизации.

К базовым параметрам функционирования каналов передачи данных (параметрам мониторинга) относятся следующие показатели, сгруппированные как в целом по трафику, так и по основным информационным сервисам Интернет (HTTP-data, HTTP-request, SMTP, FTP, KNS-data и KNS-request):

- общее число соединений в течение пяти минут;
- общий объем переданной информации в течение пяти минут;
- общее число переданных пакетов в течение пяти минут;
- общее время соединений в течение пяти минут.

Из имеющихся базовых показателей легко строятся производные показатели, такие как скорость передачи данных, величина загрузки канала и др.

Исследования показали, что задача анализа трафика является весьма сложной со статистической точки зрения. Это связано с такими различными типами нарушения стационарности, присущими трафику, как колебания дисперсии наблюдений в течение суток, сезонными эффектами различной периодичности, нехарактерными наблюдениями и т.п. Для ее решения необходимы и уже разработанные методы и программы, и создание новых алгоритмов.

Ускоренное развитие сетевой активности в России в условиях ограниченных возможностей сетевой инфраструктуры вызывает повышенный интерес к поведению базовых показателей функционирования компьютерных сетей. Одним из важнейших показателей функцио-

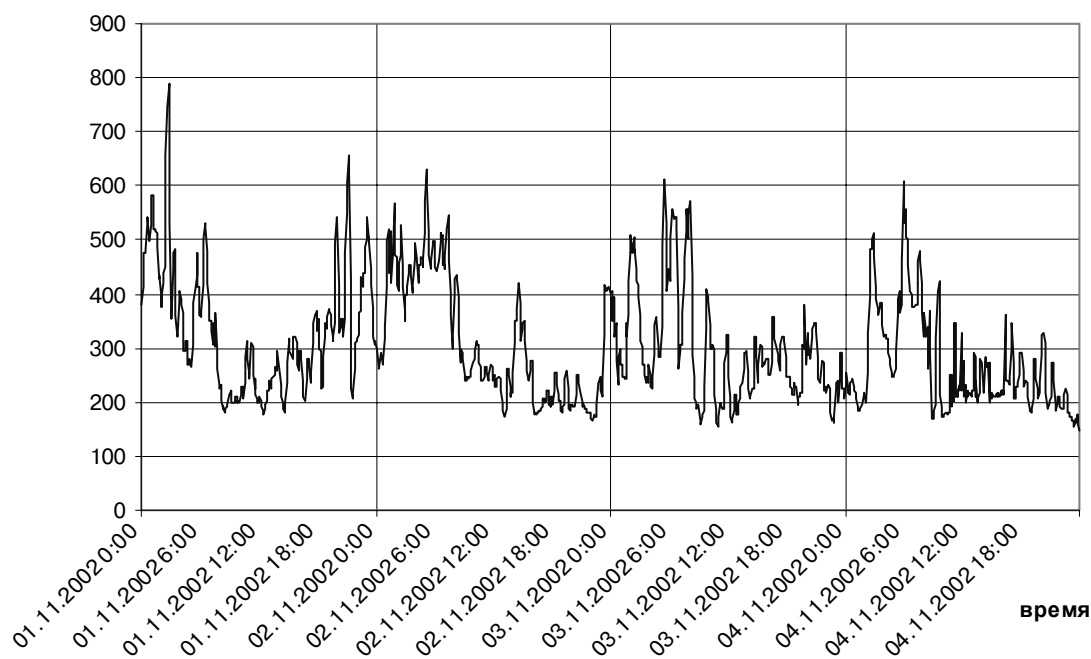


Рис. 1. Скорость передачи данных по протоколу HTTP-data в период с 1 по 4 ноября 2002 г.

нирования компьютерных сетей является скоростью передачи данных конечным пользователям. Под этой величиной понимается отношение количества переданной информации (в байтах) к суммарному времени передачи информации по всем соединениям за фиксированный промежуток времени. Таким образом, речь идет об усредненной скорости передачи данных конечным пользователям за выбранный промежуток времени. Причем усреднение ведется не только за промежуток времени, но и по всем пользователям, использовавшим сеть в этот промежуток времени. Учет характеристик данной кривой дает сетевому администратору и конечным пользователям возможность корректировать сетевую активность с учетом текущего состояния сети. Определение усредненной скорости передачи данных приобретает особую значимость, когда необходимо оценить эффективность применяемых к сети методов оптимизации трафика.

Для решения этих задач разработана статистическая модель внутрисуточных колебаний скорости передачи данных. Суть модели заключается в том, что данные о скорости передачи, для которой наблюдаются значительные колебания в соседние промежутки времени, усредняются по определенному алгоритму. Полученная в результате усреднения функция более информативна при исследовании результатов воздействия на сеть.

Учитывая, что скорость передачи данных зависит от активности пользователей, она является нестационарным временным рядом,

имеющим различные временные характеристики в зависимости от времени суток, дня недели и сезона. В нашем случае не имеет значения динамика ряда, существенными являются такие характеристики, как поведение процесса в течение суток и временной тренд (медленно меняющаяся составляющая), а также закон распределения, который учитывается при вычислении доверительных интервалов.

В наших исследованиях данные по скорости снимались автоматически для каждого ТСР-соединения, в течение всего анализируемого периода, агрегировались суммированием за пятиминутный интервал и накапливались в базе данных. Количество членов суммы в агрегате в зависимости от времени суток составляло от 2 до 7 тыс. Характер распределения агрегированных данных является нормальным.

Как показал анализ временного ряда скорости передачи данных, данный ряд имеет периодичность, и сходные особенности ряда повторяются каждые 24 часа. На рисунке 1 приведен график, иллюстрирующий данную закономерность.

Вычисление сезонной компоненты производилось по однофакторной статистической модели процесса, описывающего почасовое изменение скорости передачи данных в течение суток. В качестве 24 уровней фактора в этой модели рассматриваются различные часы суток.

В качестве оценки сезонной компоненты в каждый момент времени в аддитивной модели временного ряда можно рассматривать среднее арифметическое разности между средне-

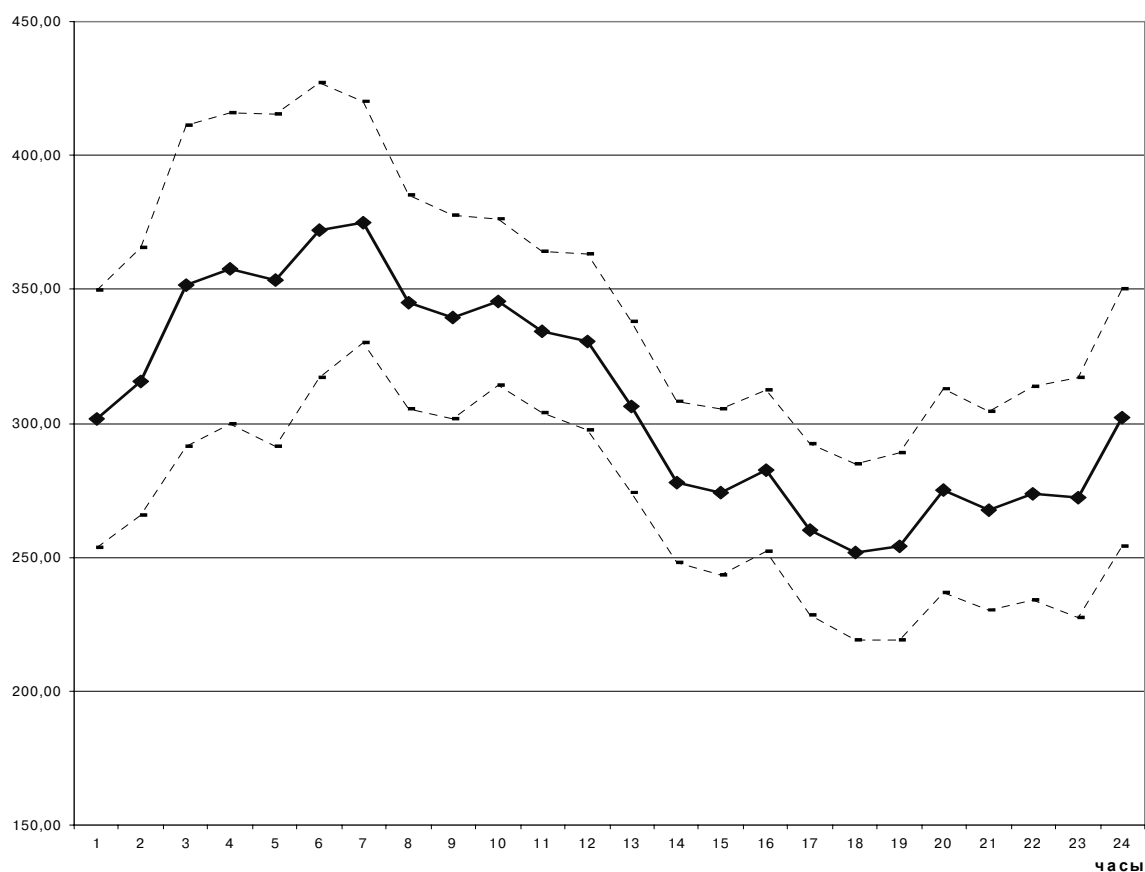


Рис. 2. Средняя скорость (профиль суток) передачи данных конечным пользователям на входе канала Бийск-Барнаул по протоколу HTTP-data в период с 30 октября по 5 ноября 2002 г.

часовым значением скорости передачи данных и соответствующим значением ряда скользящих средних.

Оценку сезонных эффектов можно также производить, вычисляя профиль суток, путем усреднения данных отдельно по каждому часу. В этом случае вместо отклонений от соответствующих значений ряда скользящих средних используются нормированные значения ряда. Как показали вычисления, ошибка оценки сезонной составляющей находится в допустимых пределах в случае отсутствия тренда и циклической компоненты временного ряда. Это имеет место в случае, если значения временного ряда скользящих средних близки к среднему значению исходного ряда.

В ходе проведенных исследований на бийском фрагменте RUNNet с помощью системы сбора статистической информации был осуществлен сбор данных о функционировании канала связи Бийск-Барнаул в период, предшествующий применению методов управления трафиком.

С помощью разработанной модели внутрисуточных колебаний скорости передачи дан-

ных конечным пользователям рассчитаны соответствующие значения факторов эффекта часа суток. Проведена оценка точности определения сезонной компоненты. Визуализация результатов вычислений осуществлена с помощью графиков, отражающих текущий профиль суток по скорости передачи данных. Построены также соответствующие доверительные интервалы значений средней скорости при уровне доверительной вероятности 0,95. Пример подобного графика приведен на рисунке 2.

Анализ статистической информации показал, что в общем объеме значительно преобладает трафик протокола HTTP-data (примерно 80–90%).

Выявлено, что скорость передачи данных по протоколу HTTP-data значительно превышает скорость передачи по всем остальным информационным протоколам. Также следует обратить внимание на существенное увеличение скорости передачи данных в ночные часы, по сравнению с дневными, связанное с тем, что ночью работает меньшее число пользователей.

В результате анализа расчетов сделан вывод о том, что для повышения эффективности

Использование статистического подхода при решении задач анализа...

функционирования сильно загруженного канала целесообразным является изменение параметров настройки, обеспечивающих повышение скорости передачи данных, передаваемых по протоколам SMTP и KNS.

Вследствие проведенных мероприятий по ограничению полосы пропускания для прото-

колов HTTP-data и FTP и созданию системы приоритетов пакетов достигнуто более чем двухкратное увеличение средней скорости передачи данных по протоколам SMTP и KNS. При этом наблюдалось незначительное уменьшение скорости передачи данных по протоколу HTTP-data.