

РАЗРАБОТКА АНАЛИТИЧЕСКОЙ МОДЕЛИ ПРОЦЕССА ПРОГНОЗИРОВАНИЯ ТРАФИКА

Зайниддинов Бобиржон Гофирович^{1.}

1. Декан факультета «Компьютерный инжиниринг» Нурафшанского филиала Ташкентского университета информационных технологий имени Мухаммада аль - Хоразмий, доктор философии по технических науках (PhD).

Куватова Нилуфар Шералиевна^{2.}

2. Стажер преподаватель кафедры «Информационных технологий» Нурафшанского филиала Ташкентского университета информационных технологий имени Мухаммада аль -Хоразмий.

Дусанова Гулшод Мирзаевна^{3.}

3. Ассистент кафедры «Технология пищевых продуктов» Шахрисабзского филиала Ташкентского химико-технологического института

АННОТАЦИЯ

В статье рассматривается процесс прогнозирования интернет-трафика с помощью методов математического моделирования. Представлены методы аналитического моделирования, учитывающие все факторы, и методы идентификации, использующие прогнозирование.

Ключевые слова: Аналитическое моделирование, трафик, пргноз, математическое ожидание.

В настоящее время наиболее перспективным способом создания и отладки при проектировании сети является её отработка на математической модели. В этом случае в системе реальный объект – прогнозирование трафика, объект регулирования замещается математической моделью. Это позволяет отрабатывать самые разнообразные и сложные режимы без проведения затратных и трудоёмких наладочных запусков. Конечно, без реальных запусков обойтись нельзя, однако сократить количество таких запусков, используя математическое моделирование, вполне реально.

Одним из важнейших направлений обработки экспериментальных данных, полученных при измерении интенсивности трафика, является построение математических моделей, адекватно описывающих её свойства.

Одной из основных характеристик трафика является его *объем* V , измеряемый в битах или байтах, переданных или прошедших через данный элемент сети за *фиксированное время* T . Тогда *интенсивность трафика* D определяется отношением:

$$D=V/T$$

Основной проблемой при проектировании вычислительной сети является прогнозирование трафика и распределение нагрузки сети. Это связано с тем, что трафик в сетях быстро превышает предел пропускной способности.

Разработчики сети прогнозируют трафик с учетом его полного планирования и моделирования бюджета. Учитывая специфическую топологию сети, планировщик сети использует прогнозирование трафика, для того чтобы оценивать ожидаемую загрузку соединения. Планировщик сети может также использовать прогнозирование трафика, для того чтобы оценить различные топологии и разработать структуру сети согласно этой оценке.

Предметом изучения теории трафика или телетрафика [1-3] стали процессы обслуживания системой связи потоков сообщений, поступающих на отдельные изучаемые объекты и сети, и их количественные характеристики. Этими объектами могут быть рабочие станции, коммутационные узлы, сети связи и их отдельные части. Математический аппарат, исследуемый теорией телетрафика, включает четыре основных элемента:

- поток поступающих сообщений;
- систему распределения информации, т.е. систему обслуживания;
- характеристики качества системы;
- дисциплина обслуживания.

Поток сообщений включает понятие о модели потока вызовов (требование на установление соединения), законах распределения длительности обслуживания сообщений, а также типах занимаемых для передачи сообщений каналов.

Система обслуживания характеризуется структурой построения и набором структурных параметров.

К характеристикам качества обслуживания поступающих сообщений относятся:

- вероятность явной потери сообщений;
- вероятность условной потери сообщений;
- среднее время задержки сообщений;
- вероятность потери поступающего вызова;
- интенсивность обслуженной нагрузки и т.д.

Под дисциплиной обслуживания понимают:

- способ обслуживания (с явными потерями, с ожиданием, с повторением или комбинированный);
- порядок обслуживания (в порядке очередности, в случайном порядке и с приоритетом);
- режим искания выходов коммутационной системы (свободный, групповой и индивидуальный).

Планировщику сети важны оценки роста нагрузки на сеть. Эти оценки объединяются со статистическими данными трафика, чтобы создать модель прогнозирования трафика.

При проектировании компьютерных сетей требуются подробные данные о многих параметрах отрасли, количестве пользователей и соединительных линий связи, ёмкости узлов связи, а также варианты схем распределения трафика сети. Эти сведения определяются на основе таких предпроектных материалов, как число пользователей, трафик сети, объём информации, сетевое оборудование и т.д. Материалы эти можно получить и с помощью прогнозирования, а отсутствие прогнозов на сетевой трафик считается главной причиной необоснованности планов развития сети [4].

Наиболее важным критерием является минимизация среднего времени задержки пакета даже при самой большой нагрузке сети. Рассмотрим, от каких параметров зависит это значение.

Первым параметром является средняя длина пакета. Задержка пакета - это время, которое пакет проводит в сети. Задержка пакета, хотя и зависит от его размера пакета, но определяется также и количеством пользователей сети, создаваемой ими нагрузкой, используемыми путями передачи пакетов и т.д., то есть текущим состоянием сети. Очевидно, что задержка пакета есть случайная величина, поэтому целесообразно рассматривать среднюю длину пакета, которая для любой конкретной сети с фиксированной емкостью является обобщенной характеристикой и в конечном итоге определяет пропускную способность системы. Таким образом, определение средней длины пакета, способов ее расчета и минимизация времени задержки пакета является важнейшей задачей для любой сети передачи данных.

Вторым параметром является число пользователей.

В [5] был рассмотрен алгоритм расчета прогнозирования трафика при проектировании распределенных систем обработки и хранения информации. При расчете использованы следующие данные: S - размер запрашиваемого пакета;

s_0 - минимальный размер пакета (64 байтов);

c - параметр распределения;

$P(S/S_0)$ - вероятность превышения размера запрашиваемого пакета над показателем S_0 в (S/S_0) раз;

В качестве распределения систем обработки используется распределение Парето. Особая важность такого анализа состоит в том, что распределения Парето с указанными моментными свойствами хорошо отражают поведение трафиков многих современных компьютерных сетей и одновременно обладают свойством долговременной зависимости (долгой памяти). Вероятность превышения размера пакета в $[s/s_0]$ раз аппроксимируется распределением Парето [6], имеющим в данной задаче следующий вид:

$$P(s/s_0) = \begin{cases} 0, & s < s_0 \\ c s_0^c s^{-(c+1)}, & s \geq s_0 > 0, \end{cases}$$

где, $s_0 = 64$ байтов, $c = 0,9$.

Выбор параметра (c) в работе [6,7] предлагается выполнять в зависимости от вида файлов. Из предложенного интервала [7] по протоколу FTP выбрана нижняя граница потому что, меньшее значение параметра соответствует большей выраженности свойств случайной величины. В результате расчета с шагом 100 байтов получены вероятности запроса пакета определенного размера (табл. 1).

Таблица 1. Вероятность запроса пакета

Размер пакета (байтов)	Вероятность запроса	Размер пакета (байтов)	Вероятность запроса
64	0,0140625	864	0,000100987
164	0,00235288	964	8ДЕ-05
264	0,000952263	1064	6,7Е-05
364	0,000517264	1164	5,6Е-05
464	0,000326152	1264	4.8Е-05
564	0,000225099	1364	4.2Е-05
664	0,000165076	1464	3.6Е-05

Математическое ожидание [8] (m_x) среднего размера пакета (в байтах) приближенно равно среднему арифметическому наблюдаемых значений случайной величины (тем точнее, чем больше число испытаний).

$$m_x \approx \frac{1}{j} \sum_{i=1}^j x_i$$

где, j - число испытаний; x_i - размер пакета с шагом 100 байтов.

В нашем случае $j=15$ (от 64 байтов до 1500 байтов с шагом 100 байтов), то есть $m_x = 764$. Ниже представлена формула и результаты расчета проектной пропускной способности.

Интенсивность запросов равна вероятности появления запросов, поступающих на вход узлов в единицу времени и имеет вид:

$$\lambda = \frac{P}{t}$$

где, λ - интенсивность запросов (1/с); P - вероятность появления запросов; t - период измерений (с).

В таблице 2. рассчитана интенсивность запросов, предложенных в качестве испытания, период измерений 1 час и вероятность появления запросов с шагом 0.05.

Таблица 2. Интенсивность запросов

Вероятность появления запросов, P	Интенсивность запросов, λ	Вероятность появления запросов, P	Интенсивность запросов, λ
0.05	1.38889E-05	0.55	0.000152778
0.1	2.77778E-05	0.6	0.000166667
0.15	4.16667E-05	0.65	0.000180556
0.2	5.55556E-05	0.7	0.000194444
0.25	6.94444E-05	0.75	0.000208333
0.3	8.33333E-05	0.8	0.000222222
0.35	9.72222E-05	0.85	0.000236111
0.4	0.000111111	0.9	0.00025
0.45	0.000125	0.95	0.000263889
0.5	0.000138889	0.999999	0.000277778

Пользуясь этими частотами запросов, можно определить требуемую пропускную способность $G_{п.с}$:

$$G_{п.с} = \frac{t \cdot m_x \cdot n \cdot \lambda}{t_e}$$

где, m_x - математическое ожидание среднего размера пакета (байт); n - количество узлов; t_e - время выполнения запроса (мс).

В [9] обосновано среднее время выполнения запроса 150 мс для протоколов HTTP, POP3, SMTP и FTP, и 60 мс для SIP. Процентные отношения [69] сетевого трафика составляют 26 % для email или данных и 62% для видео трафика. Для вышеуказанных двух типов трафиков получается время выполнения $t_g = 76$ мс. В результате расчета с шагом 100 узлов (от 100 до 1000) получена расчетная проектная пропускная способность (рис.1) при фиксированной вероятности появления запросов.

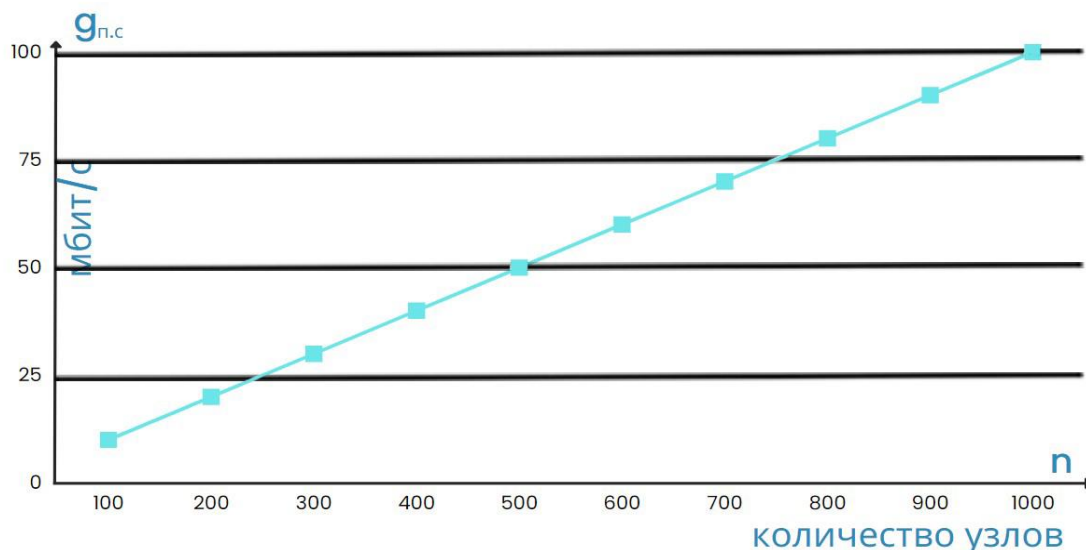


Рис.1. Зависимость $g_{n,s}$ от количества n узлов в сети

На рис.2 показана проектная пропускная способность $d_{g_{n,s}}$ для различных вероятностей P запросов при фиксированном количестве узлов. При этом, рассчитано среднее количество узлов, то есть 550.

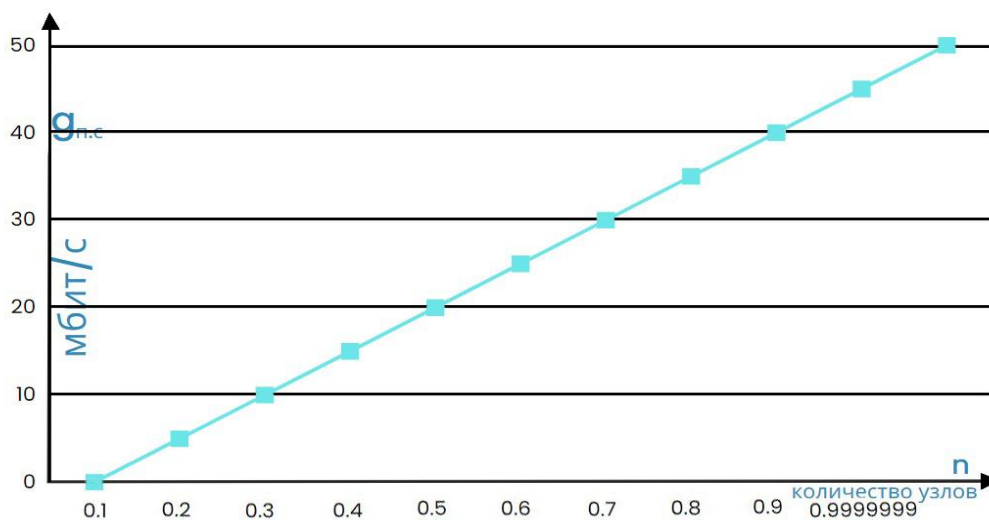


Рис.2. Зависимость $g_{n,s}$ от вероятности P появления запросов

На основании приведенных расчетов можно делать выводы о необходимой пропускной способности среды передачи данных.

При этом необходимо учитывать рост числа пользователей. Сегодня интернет уже достаточно глубоко проник в жизнь людей, позволяя работать, общаться и получать доступ к медиа развлечениям практически из любого места на планете. Рост числа пользователей сети делает распространение глобальной сети подлинно повсеместным.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ: (REFERENCES)

1. Абросимов Л.И., Мясников В.А., Мельников Ю.Н., Методы Автоматизированного проектирования систем телеобработки данных, Москва, 1992.
2. Абросимов Л.И., Анализ и проектирование вычислительных сетей. Учебное пособие, МЭИ, Москва, с. 3-4, 2000.
3. Антамошкин А.Н., Золотарев В.В., Алгоритм расчета прогнозируемого трафика при проектировании распределенных систем обработки и хранения информации. Вестник Сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнева, г. Красноярск, Выпуск 1 (8) 2006.
4. Анализ трафика, Cisco // http://www.cisco.com/support/RU/customer/content/9/97422/tech_tk652_tk701_tech_white_paper09186a00800d6b74.shtml
5. Аппаратура компьютерных сетей // <http://www.ixbt.com/comm/ev/nq04.html>
6. Боровиков В.П., Боровиков И.П. STATISTICA-Статистический анализ и обработка данных в среде Windows. Москва. 1998.
7. Брехов О.М., Звонарева Г.А., Корнеевкова А.В., Имитационное моделирование ЭВМ, Москва, 2008.
8. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. Выл.2. Москва. МИР, 1974.
9. Боев В.Д., Моделирование систем инструментальные средства GPSS WORLD, Санкт-Петербург, 2004.