

## ОПТИМИЗАЦИЯ КОМПЬЮТЕРНОЙ СЕТИ С ОБЕСПЕЧЕНИЕМ QoS НА ОСНОВЕ ЧИСЛЕННОГО КРИТЕРИЯ

Институт компьютерных технологий Национального авиационного университета

*Предложен подход к решению задачи оптимального распределения информационных файлов по узлам компьютерной сети при передаче больших объемов данных. В качестве численного критерия оптимальности используется комбинация параметров трафика по каналам связи сети, что позволяет обеспечить минимизацию общего трафика по сети за единицу времени, среднего времени ожидания запросов в очереди на обслуживание.*

### Введение

Оптимизация компьютерных сетей предусматривает определение:

- критерия эффективности работы сети;
- множества варьируемых параметров сети, прямо или косвенно влияющих на критерий эффективности;
- порога чувствительности для значений критерия эффективности.

На выбранный критерий эффективности влияют параметры различных типов. Например, на производительность сети в большей степени влияют:

- коммуникационные протоколы и их параметры (номинальная и эффективная пропускная способность протокола, размер и время жизни пакета, параметры квитирования);
- доля и характер широковещательного трафика, создаваемого различными протоколами;
- топология сети и используемое коммуникационное оборудование;
- интенсивность возникновения и характер ошибочных ситуаций;
- конфигурация программного и аппаратного обеспечения конечных узлов и др.

В современных компьютерных сетях передачи данных предусмотрено включение механизмов поддержки, учитывающих новые требования к приложениям сети, и предоставляющих необхо-

димое качество обслуживания (*Quality of Service, QoS*). Методы обеспечения качества обслуживания базируются на поддержке определенной полосы пропускания; сокращении вероятности потери кадров; исключении или управляемости сетевых перегрузок; возможности конфигурирования сетевого трафика; установке количественных характеристик трафика и др. [1-4].

Качество обслуживания компьютерной сети, как правило, является интегральным показателем, составляющие которого изменяются в зависимости от требований к сети. Одним из основных параметров услуг является среднее время выполнения запросов пользователей, определяющееся как интервал между завершением ввода запроса пользователем до получения ответа и завершения всех процессов, связанных с данным запросом. Интервал состоит из минимального времени полного выполнения запроса в случае, когда этот запрос мог использовать все ресурсы системы в монопольном режиме и суммы задержек, возникающих на разных этапах в процессе обработки запроса из-за совместного использования ресурсов сети.

Ранние реализации технологии QoS опирались главным образом на алгоритмы организации очередей, которые устанавливались и поддерживались сетевыми

маршрутизаторами и другими устройствами. При этом не обеспечивалось непосредственное управление непрерывными потоками трафика.

Для получения количественных характеристик создаваемых компьютерных сетей можно использовать средства моделирования, с помощью которых разрабатываются модели, воссоздающие информационные процессы, протекающие в сетях. При этом можно использовать методы натурального (физического) моделирования, создавая пилотные проекты сети, а также методы математического моделирования: имитационные модели и модели систем массового обслуживания (СМО) [5-7] и др.

В структуре компьютерной сети выделяются подсети:

- традиционного сетевого управления (конфигурацией, производительностью, безопасностью);
- моделирования функционирования сети, включая анализ нагрузок на отдельные участки и поддержку принятия решений по перепланированию.

В работах [8, 9] исследованы вопросы распределения информационных ресурсов по узлам компьютерных систем и сетей. При этом рассматривались алгоритмы размещения информационных файлов, ориентированные на оптимизацию. В качестве критериев оптимальности рассматривались средний объем данных, пересылаемых по каналам связи в единицу времени; общее время обработки запросов; общая стоимость трафика сети и др.

### Постановка задачи

Для описания процесса разделения ресурсов со случайным потоком запросов, когда могут возникать очереди, обычно используются модели СМО. Модель считается заданной, если полностью описаны ее компоненты: входящий поток запросов; количество и типы обслуживающих устройств; емкости накопителей, где запросы ожидают начала обслуживания;

время обслуживания запросов на устройствах.

Можно построить модель СМО, описывающую процессы появления задержек в результате возникновения очередей в узле сети. Однако построить модель, которая бы описывала среднее время выполнения запросов пользователей для всей сети, на основе аппарата теории СМО, достаточно сложно. Еще сложнее на основе полученной модели за приемлемое время определить оптимальное распределение файлов. Оценка качества модели должна приниматься на основе некоторого критерия оптимальности, который должен выражаться количественно и быть единственным.

### Выбор численного критерия

Возникает потребность в выборе численного критерия оптимальности, определяющего среднее время выполнения запросов пользователей и являющимся удобным для оптимального размещения файлов. Выбор такой характеристики СМО обусловлен тем, что пользователи, как правило, заинтересованы не в минимизации размера очереди или каких-то других характеристик СМО, а в том, чтобы их запросы обрабатывались как можно меньше время.

При определении среднего времени ожидания запросов  $W$  в очереди на обслуживание рекомендуется [10] использовать следующую формулу:

$$W = \frac{\rho^2}{\lambda(1-\rho)},$$

где  $\rho$  – коэффициент загрузки обслуживающего устройства ( $0 \leq \rho < 1$ );  $\lambda$  – интенсивность потока запросов (среднее число пакетов, которые претендуют на передачу за единицу времени).

При постановке задачи оптимизации размещения файлов среди узлов сети с целью получения высокого качества об-

служивания  
времени обсл  
времени ожи  
вить постоян  
мещения фай  
 $\lambda$  и пропус

вающего уст

отметить, что  
вую роль в об

Обычно  
время ожидан  
обслуживание  
максимально  
загрузки обслу  
ределяется из

При расп  
N обслуживан  
ходимо миним  
этом маршрут  
нее неизвестен  
ываться как о  
ройством, так  
сколькими об  
вами.

Построим  
терий качества)  
метров трафика

где  $C_i$  – весовы  
вающие среднее

Представим  
степенного ряда

$W(\rho) = b\rho$   
В отличие  
дуры минимиза

служивания можно величину  $b$  среднего времени обслуживания заявки (без учета времени ожидания обслуживания) оставить постоянной и не зависящей от размещения файлов. Величина  $\rho$  зависит от  $\lambda$  и пропускной способности обслуживающего устройства  $\mu$ :  $\rho = \frac{\lambda}{\mu}$ . Следует отметить, что параметр  $\rho$  играет ключевую роль в образовании очереди.

Обычно максимально допустимое время ожидания запросов в очереди на обслуживание  $M$  – постоянно, поэтому максимально допустимый коэффициент загрузки обслуживающего устройства определяется из выражения:

$$\rho_{\max} = \frac{M}{M + b}.$$

При распределении запросов между  $N$  обслуживающими устройствами необходимо минимизировать величину  $W$ , при этом маршрут прохождения запроса заранее неизвестен, т.е. запрос может обрабатываться как одним обслуживающим устройством, так и последовательно несколькими обслуживающими устройствами.

Построим целевую функцию (критерий качества) в виде комбинации параметров трафика по каналам связи в сети:

$$Q(\rho) = \sum_{i=1}^N C_i \rho_i, \quad (1)$$

где  $C_i$  – весовые коэффициенты, учитывающие среднее время обслуживания пакетов каналом связи:  $C_i = b_i$ ;  $b = \frac{\rho}{\lambda}$ .

Представим функцию  $W(\rho)$  в виде степенного ряда

$$W(\rho) = b\rho + b\rho^2 + b\rho^3 + \dots$$

В отличие от применяемой процедуры минимизации среднего времени

ожидания запросов в очереди на обслуживание в целом по сети предлагается минимизировать функцию определяющих параметров трафика по каналам связи (зависящую от  $\rho_i$ ). Зададим некоторую величину  $\varepsilon \geq O(\rho^2)$ , где  $O(\dots)$  – функция, отражающая временную сложность алгоритма. Величина  $\varepsilon$  должна быть такой, чтобы с практической точки зрения разница во времени ожидания запросов на обслуживание отдельным устройством, которая меньше  $\varepsilon$ , была величиной второго порядка малости. Для решения данной задачи необходимо определить ограничивающие условия. С этой целью представим функцию  $W(\rho)$  в виде

$$W(\rho) = b\rho + O(\rho^2).$$

Отсюда  $\frac{b\rho^2}{1-\rho} \leq \varepsilon$  и, следовательно, получим ограничивающее условие:

$$\rho \leq \frac{-\varepsilon + \sqrt{\varepsilon^2 + 4b\varepsilon}}{2b}.$$

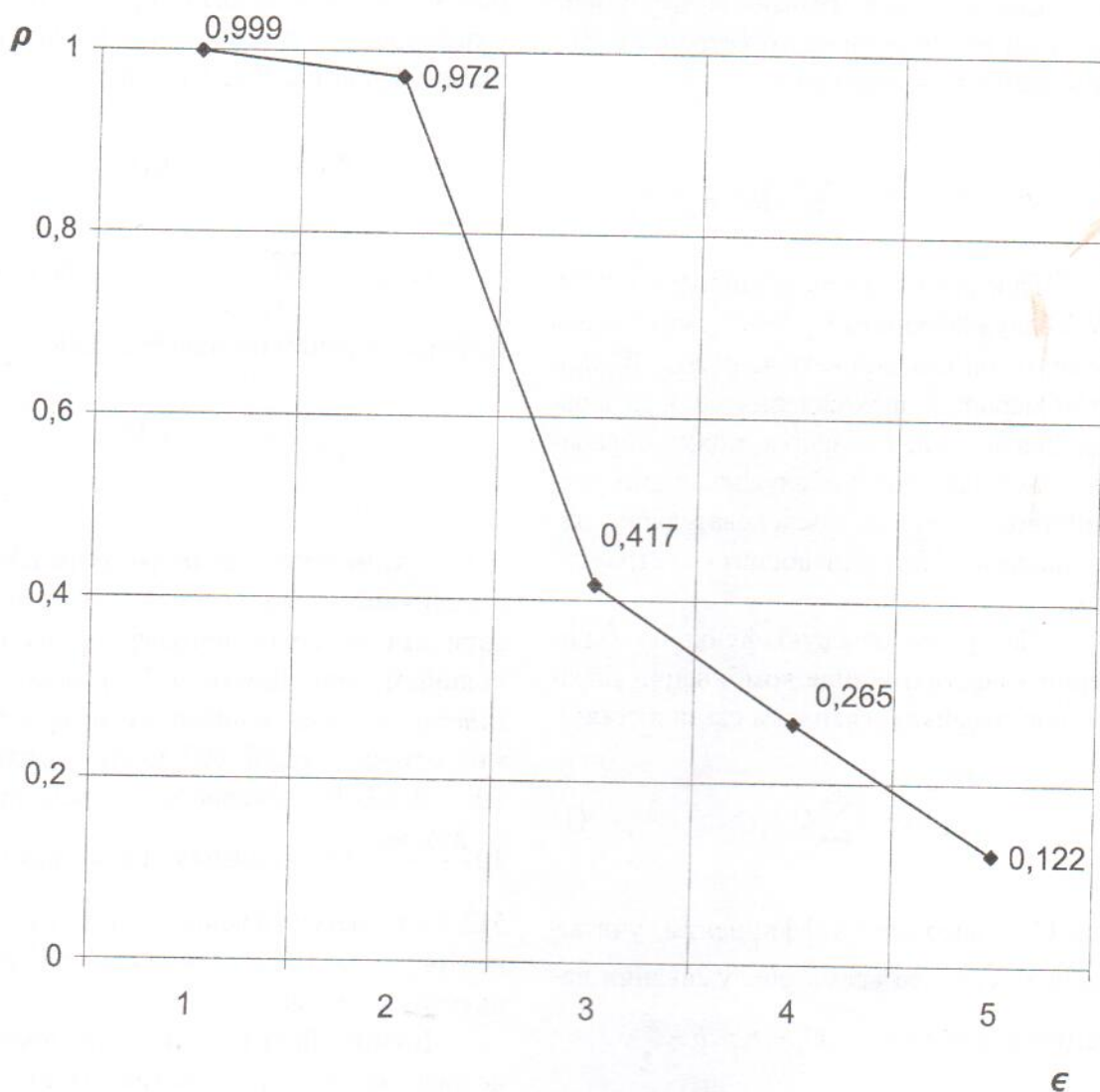
Рассмотрим пример определения ограничивающего условия для базового варианта наиболее популярной на сегодняшний день (более 90% рынка) сети Ethernet со следующими характеристиками: стандарт IEEE 802.3, спецификация 10 BASE-T; скорость передачи –  $10 \frac{\text{Мбит}}{\text{с}}$ ; минимальная длина пакета – 512 бит; максимальное количество абонентов – до 1024; начальный кадровый интервал – 9,6 мс.

Время обслуживания, как правило, не зависит от характеристик заявок (длины пакета). Поэтому можно считать, что пропускная способность в каждом канале связи одинакова и постоянна:  $\mu_i = \mu = \text{const}$ .

На рисунке представлена зависимость изменения  $\rho$  от  $\epsilon$ . Анализ показывает, что на загрузку каналов связи основное влияние оказывает величина  $\epsilon$ . Поэтому при решении задачи оптимизации среднего времени ожидания запросов в очереди на обслуживание каналами связи можно минимизировать функцию (1). При этом практически точность не теряется. В случае, если необходимо учитывать ограничения на среднюю загрузку каналов связи, можно увеличить коэффициенты целевой функции.

### Выводы

Предложен новый подход к построению математических моделей оптимального распределения файлов среди узлов компьютерной сети. Подход, в отличие от известных, базируется на использовании критерия оптимальности, на который метод обработки запросов в узлах сети влияет не существенно. При этом математические модели оптимального распределения файлов в большинстве случаев сводятся к решению обобщенной задачи.



Зависимость изменения коэффициента загрузки обслуживающего устройства  $\rho$  от величины  $\epsilon$

Проблема  
Выбор  
методов  
влияние на  
связи сети  
• ме  
сети за с  
• ме  
ожидание  
жизненные  
времени в  
телей;  
• не  
от схемы с  
Мате  
универсал  
дуре обра  
виде комб  
каналом с  
ных, котор  
связи за с  
шего траф  
ки запрос  
при, порог  
каналом с  
Испол  
воллет пол  
ния, посл  
туру и хар  
ти. С прак  
ченные ре  
вснем обр  
функцион  
Список  
1. Ку  
тивных сет  
тер, 2000.  
2. Жу  
Обеспечив  
обслужива

Выбор численного критерия оптимизации целевой функции в виде комбинации параметров трафика по каналам связи сети обеспечивает:

- минимизацию общего трафика по сети за единицу времени;
- минимизацию среднего времени ожидания запросов в очереди на обслуживание каналом связи, а также среднего времени выполнения запросов пользователей;
- независимость целевой функции от схемы обработки запросов в сети.

Математические модели являются универсальными по отношению к процедуре обработки запросов. К функциям в виде комбинации параметров трафика по каналам связи сети относятся объем данных, которые пересылаются по линиям связи за единицу времени, стоимость общего трафика сети, общее время обработки запросов в сети и пересылки информации, порожденной этими запросами по каналам связи.

Использование таких моделей позволяет получать квазиоптимальные решения, последовательно корректируя структуру и характеристики компьютерной сети. С практической точки зрения, полученные результаты, позволяют существенным образом повысить эффективность функционирования компьютерных сетей.

### Список литературы

1. Кульгин М. Технология корпоративных сетей. Энциклопедия. – С.Пб: Питер, 2000. – 704 с.
2. Жуков И. А., Клименко И. А. Обеспечение заданного уровня качества обслуживания в объединенных сетях //

Проблеми інформатизації та управління. – К.: НАУ, 2005. – Вип. 13. – С.5–14.

3. Жуков И. А., Виноградов М. А., Дровозов В. И., Халимон Н. Ф. Основы теории сетей передачи та розподілу даних: Навчальний посібник. – К.: Книжкове вид-во НАУ, 2006. – 272 с.

4. Кулаков Ю. А., Гайдукова Л., Халиль Х. А. Аль Шкерат. Способы повышения эффективности качества обслуживания (QoS) в многофункциональных сетях // Вісник НГУУ «КПІ». Інформатика, управління та обчислювальна техніка. – К.: ТОВ «ВЕК+», 2002. – № 39. – С. 132–141.

5. Вишневицкий В. М. Теоретические основы проектирования компьютерных сетей. – М.: Техносфера, 2003. – 512 с.

6. Ивницкий В. А. Теория сетей массового обслуживания. – М.: Изд-во Физико-математической литературы, 2004. – 772 с.

7. Кулаков Ю. О. Синтез компьютерных сетей керування виробництвом // Вимірювальна та обчислювальна техніка в технологічних процесах. – Хмельницький, 1998. – №2. – С. 29–32.

8. Жуков И. А., Зайченко Ю. П., Печурин Н. К. Модель распределения информационных ресурсов в компьютерных сетях // Проблеми інформатизації та управління. – К.: НАУ, 2005. – Вип. 3 (14). – С. 9–14.

9. Цегелик Г. Г. Системы распределенных баз данных. – Львов: Свит, 1990. – 168 с.

10. В. Столлингс. Современные компьютерные сети. 2-е изд. – С.Пб.: Питер, 2003. – 783 с.