

ИССЛЕДОВАНИЕ МЕТОДА ЦЕЛЕВОЙ КООРДИНАЦИИ В ЗАДАЧАХ УПРАВЛЕНИЯ ОЧЕРЕДЯМИ НА УЗЛАХ ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ

Семеняка М.В.

Харьковский национальный университет
радиоэлектроники, Украина.

E-mail: maxisemen@gmail.com

Abstract

Recent research in the area of providing quality of service in telecommunication networks have proved the necessity of new approaches, researches and development for solving problems related to increasing quality of service for traffic flows of different customers. The paper presents the hierarchical concept of congestion management on the telecommunications network nodes in which the method of goal coordination has been determined. Proposed model is suitable for network nodes with multiprocessor or multi-core architecture. Each processor was assigned to a macro-queue for which the problem of packet distribution and bandwidth allocation was solved (first level of hierarchy) while on the node coordinator (second level of hierarchy) those results were verified and coordinated using iterative information transfer with first level sequentially improving values of Lagrange multipliers, applying the gradient algorithm. Proposed flow-based hierarchical goal coordination queuing model presented in decomposing form of differential equations.

Важной задачей развития современных телекоммуникационных сетей является повышение качества обслуживания (Quality of Service, QoS) все возрастающего количества запросов пользователей. Среди множества методов повышения качества обслуживания важное место занимают задачи управления очередями, которые позволяют эффективно повысить такие показатели как средняя задержка доставки пакетов, джиттер, количество отброшенных пакетов без существенных затрат на модернизацию существующей инфраструктуры сети [1]. В свою очередь особенности и эффективность решения задач управления очередями зависит от тех математических моделей и методов, которые в них заложены. В этой связи актуальным видится исследование новых подходов управления очередями на узлах телекоммуникационной сети.

Современная телекоммуникационная сеть имеет большую распределенную структуру, происходит переход к распределенным, облачным технологиям, которые требуют соблюдения в своей работе принципов согласования, координации элементов. Математическим аппаратом способным решить поставленную задачу согласования и координации распределенной сети для повышения качества обслуживания трафика пользователей являются решения многоуровневых, иерархических систем [2].

Иерархические структуры (очереди) лучше всего подходят для масштабирования сети, в случае увеличения количества потоков, которые необходимо обслужить. Этот математический аппарат позволяет более эффективно использовать многоядерные (многопроцессорные) системы, когда общий ресурс распределяется параллельно и каждый процессор занимается решением задачи обслуживания для своей группы очередей.

В области управления трафиком уже существуют решения иерархических систем в области маршрутизации. Математические модели представлены дифференциальными и алгебраическими уравнениями [3,4]. В данной работе предложена потоковая модель иерархических очередей в декомпозиционной форме в виде дифференциальных уравнений.

Условимся, что с точки зрения повышения масштабируемости решений задачи управления очередями определим двухуровневую функциональную иерархию расчетов:

- на нижнем уровне – искомое распределение потоков подлежит расчету независимо для каждой макроочереди для трафиков, которые поступают от сети доступа;
- на верхнем уровне – полученные на нижнем уровне решения задач распределения потоков и выделения пропускной способности для каждой очереди координируются с целью предотвращения вероятной перегрузки трактов передачи сети.

Тогда для каждой r -й макроочереди в качестве искомым выступают переменные x_{ij}^r , которые характеризуют интенсивность i -го трафика, направленного в j -ю очередь, а также переменные b_j^r , под которыми понимается пропускная способность выделенная для j -й очереди r -й макроочереди. С целью недопущения потерь пакетов в узле сети в ходе расчета переменных необходимо обеспечить выполнение условий сохранения потока:

$$\sum_{j=1}^N x_{ij}^r = a_i \quad (i = \overline{1, M}), \quad (1)$$

где $a_i \quad (i = \overline{1, M})$ – интенсивность трафика i -го класса, поступающего на обслуживание.

Условие (1) должно выполняться для каждого i -го потока, поступающего в произвольную макроочередь для обслуживания. Кроме того с целью предотвращения возможной перегрузки тракта передачи сети важно выполнить условие:

$$\sum_{r \in L} \sum_{i \in M} \sum_{j \in N} x_{ij}^r \leq b, \quad (2)$$

где b – пропускная способность исходящего КС.

Стоит учесть, что при децентрализованном расчете искомым переменных на каждой отдельно взятой макроочереди условие (2) в явном виде учесть не представляется возможным. Причина тому состоит в том, что распределение потоков для каждой макроочереди рассчитывается только на основе имеющейся у нее информации о поступивших в нее трафиках и пропускной способности канала связи, без данных о результатах распределения трафика и выделения пропускной способности на других макроочередях. В этой связи условие (2) целесообразно записать в следующем виде:

$$\sum_{i \in M} \sum_{j \in N} x_{ij}^r \leq b - \sum_{\substack{s \in L \\ s \neq r}} \sum_{i \in M} \sum_{j \in N} x_{ij}^s. \quad (3)$$

Смысл выражения (3) заключается в том, что интенсивность трафика, поступившего на обслуживание в r -ю макроочередь, не должна превышать пропускную способность канала связи, которая осталась после обслуживания трафиков в других макроочередях.

Согласно физическому содержанию искомая переменная x_{ij}^r может принимать только два значения – или ноль, или λ_{ij}^r , т.е.

$$x_{ij}^r \geq 0 \quad (i = \overline{1, M}, \quad j = \overline{1, N}), \quad (4)$$

$$x_{ij}^r \in \{0, \lambda_{ij}^r\}, \quad (5)$$

где λ_{ij}^r – интенсивность трафика i -го класса, поступившего на обслуживание в j -ю очередь r -й макроочереди.

В дальнейшем условие (3) удобно будет использовать в векторно-матричной форме:

$$B_r \bar{x}_r \leq b - \sum_{\substack{s \in K \\ s \neq r}} B_s \bar{x}_s, \quad (6)$$

где $B_r, B_s \quad (r, s \in K; r \neq s)$ – согласующие матрицы, координаты $\{0;1\}$ которых выбираются в соответствии с выражением (3).

Для улучшения показателей качества обслуживания критерием оптимальности выступала аддитивная функция минимизации средней длины очереди в узле:

$$\min_x F \text{ при } F = \sum_{r \in K} \sum_{j \in N} \left[\frac{\rho_j^r}{1 - \rho_j^r} - \rho_j^r \right], \quad (7)$$

где $\rho_j^r = \frac{\sum_{i=1}^M x_{ij}^r}{b_j^r}$ – коэффициент использования пропускной способности очереди.

Решение оптимизационной задачи, связанной с минимизацией функции (7) при наличии ограничений (1)–(6), будет положено в основу предлагаемого метода иерархически-координационного распределения очередей. Тогда, в соответствии с теоремой о строгой дуальности задач на условный экстремум, переходя к задаче на безусловный экстремум, имеет место равенство:

$$\min_{x,b} F = \max_{\mu} \Phi,$$

определим в соответствии с формулой (7) двойственную функцию $\Phi(\mu)$ таким образом:

$$\Phi(\mu) = \{\min_{x,b} L(x,b,\mu)\},$$

где

$$L = \sum_{r \in K} \sum_{j \in N} \left[\frac{\rho_j^r}{1 - \rho_j^r} - \rho_j^r \right] + \sum_{r \in K} \bar{\mu}_r^t (B_r \bar{x}_r + \sum_{\substack{s \in K \\ s \neq r}} B_s \bar{x}_s - b). \quad (8)$$

Применение иерархическо-координационного подхода к решению поставленной задачи предполагает обоснованный выбор координирующих векторов (вектора). При решении оптимизационной задачи по управлению многоуровневой системой в основу иерархическо-координационных методов могут быть положены принципы координации такие, как принцип целевой координации, принцип оценки взаимодействий и принцип прогнозирования взаимодействия [2, 5].

В рамках предложенной потоковой модели наиболее обоснованным являлся принцип целевой координации, которых определял порядок учета ограничений на взаимодействие макроочереди и распределения расчетных задач между уровнями иерархии. При этом в ходе декомпозиции решения задачи (8) векторы искомым переменных x_{ij}^r и b_j^r будут рассчитываться непосредственно на процессах каждой макроочереди, а множители Лагранжа $\bar{\mu}_r$ будут рассчитываться на верхнем уровне.

Таким образом, для решения сформулированной оптимизационной задачи используем принцип целевой координации [2, 5], в рамках которого лагранжиан (8) представим в виде:

$$L = \sum_{r \in K} \sum_{j \in N} \left[\frac{\rho_j^r}{1 - \rho_j^r} - \rho_j^r \right] + \sum_{r \in K} \bar{\mu}_r^t B_r \bar{x}_r + \sum_{r \in K} \bar{\mu}_r^t \sum_{\substack{s \in K \\ s \neq r}} B_s \bar{x}_s - \sum_{r \in K} \bar{\mu}_r^t b. \quad (9)$$

Предположив, что величины $\bar{\mu}_r$ являются фиксированными, можно предпоследнее слагаемое в выражении (9) привести к виду [5]:

$$\sum_{r \in K} \bar{\mu}_r^t \sum_{\substack{s \in K \\ s \neq r}} B_s \bar{x}_s = \sum_{r \in K} \sum_{\substack{s \in K \\ s \neq r}} \bar{\mu}_r^t B_s \bar{x}_s,$$

тогда выражение (9) можно записать следующим образом:

$$L = \sum_{r \in K} L_r,$$

где

$$L_r = \left[\frac{\rho_j^r}{1 - \rho_j^r} - \rho_j^r \right] + \bar{\mu}_r^t B_r \bar{x}_r + \sum_{\substack{s \in K \\ s \neq r}} \bar{\mu}_r^t B_s \bar{x}_s - \bar{\mu}_r^t b. \quad (10)$$

Таким образом, функция (9) приобретает сепарабельную форму, а общая проблема распределения нагрузки между очередями, выделения пропускной способности оказалась декомпозицированной на ряд подзадач (по числу макроочереди), где каждая задача распределения на r -й макроочереди, состоящая в расчете векторов x_{ij}^r и b_j^r , свелась к оптимизации лагранжиана L_r (10). При этом лагранжиан (10) является функцией переменных только одного индекса, определяющего номер макроочереди.

Решение задачи по минимизации выражения (10) определяет нижний уровень расчетов, результатом которого является формирование распределения трафиков и выделения пропускной способности на каждой конкретной макроочереди согласно содержанию искомым векторов x_{ij}^r и b_j^r . На верхнем уровне, основной задачей которого является координация решений, полученных на нижнем уровне с

целью недопущения перегрузки тракта передачи (6), осуществляется модификация вектора множителей Лагранжа в ходе выполнения следующей градиентной процедуры:

$$\mu_r(\alpha + 1) = \mu_r(\alpha) + \nabla \mu_r,$$

где $\nabla \mu_r$ – градиент функции, рассчитывается, исходя из получаемых на нижнем уровне результатов решения задачи распределения потоков трафика и выделения пропускной способности на каждой конкретной r -й макроочереди ($r \in K$), т.е.

$$\nabla \mu_r(x_{ij}^r, b_j^r) = \sum_{r \in K} B_r \bar{x}_r - b.$$

Результаты решения задач на каждой отдельной макроочереди собираются координатором (верхний уровень иерархии), где осуществляется их анализ и координация путем расчета (корректировки) векторов множителей Лагранжа $\mu_r (r \in K)$ в ходе оптимизации лагранжиана (9). Полученные новые решения снова "спускаются" на первый уровень управления для итерационной оптимизации процесса распределения трафика и выделения пропускной способности (рис.1).

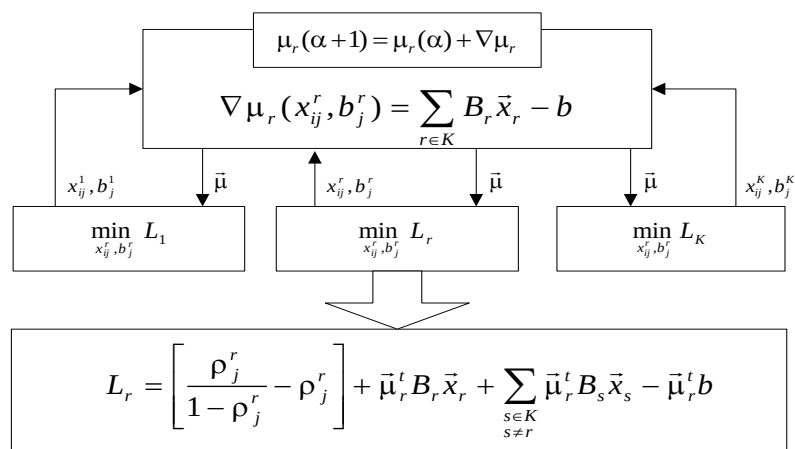


Рис. 1. Схема двухуровневого управления очередями

Применение данной структуры позволило обеспечить более высокий уровень адекватности (системности) процесса управления [6], более высокое качество решения задач управления перегрузками на узлах сети, адаптировав их под современную многопроцессорную архитектуру сетевых узлов. Использование параллельно работающих процессоров для вычислений на первом уровне позволило увеличить скорость нахождения оптимального решения. На технологическом уровне это позволило обеспечить более эффективное использование буферного ресурса, улучшить межконцевые значения средней задержки, джиттера, количества отброшенных пакетов.

Литература:

1. Вегенша III. Качество обслуживания в сетях IP. : Пер. с англ. – М.: Издательский дом «Вильямс», 2003. – 386 с.
2. Месарович М. Теория иерархических многоуровневых систем. : Пер. с англ. под ред. И. Ф. Шахнова. Предисл. чл.-кор. АН СССР Г.С.Поспелова. – М.: «Мир», 1973. – 344 с.
3. Лемешко А.В., Ахмад М. Хайлан Многоуровневое управление трафиком в сети MPLS-TE DiffServ на основе координационного принципа прогнозирования взаимодействий [Электронный ресурс] // Проблемы телекоммуникаций. - 2010. - №1(1). - С. 35-44. - Режим доступа: http://pt.journal.kh.ua/2010/1/1/101_lemeshko_traffic.pdf.
4. Беленков А.Г., Евсеева О.Ю., Лемешко А.В. Метод распределения нагрузки в иерархических телекоммуникационных сетях на базе декомпозиционных принципов предсказания взаимодействий и целевой координации // Праці УНДІРТ. – 2005. – №2(42). – С. 11-16.
5. Сингх М., Титли А. Системы: декомпозиция, оптимизация и управление. Пер. с англ. проф. Запорожца А.В. – М.: Машиностроение, 1986. – 494 с.
6. Семеняка М.В., Симоненко А.В., Али С. Али Разработка и исследование метода динамической балансировки очередей на маршрутизаторах мультисервисной телекоммуникационной сети [Электронный ресурс] // Проблемы телекоммуникаций. – 2012. – № 3 (8). – С. 64 – 78. – Режим доступа: http://pt.journal.kh.ua/2012/3/1/123_semenyaka_balancing.pdf.