
М.В. Семеняка

**ИССЛЕДОВАНИЕ МЕТОДА ДВУХУРОВНЕВОГО ОБСЛУЖИВАНИЯ ОЧЕРЕДЕЙ
НА УЗЛАХ ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ**

Харьковский национальный университет радиоэлектроники, г. Харьков, Украина

Ключевые слова: обслуживание очередей, качество обслуживания, двухуровневая структура, многоядерные системы.

Предложен метод двухуровневого обслуживания очередей. В рамках предложенного метода задача обслуживания очередей была сформулирована в виде оптимизационной задачи, представленной двухуровневой иерархической структурой. На первом уровне решалась задача распределения потоков между очередями, на втором уровне решения, полученные на первом уровне, координировались с целью предотвращения вероятной перегрузки канала связи. При управлении многоуровневой структурой в основу был положен принцип целевой координации, которых определял порядок учета ограничений на взаимодействие очередей и распределения расчетных задач между уровнями иерархии. Координация решений осуществлялась с использованием метода множителей Лагранжа. В результате имитационного моделирования проведено исследование результатов сходимости предложенного метода.

M.V. Semenyaka

**RESEARCH OF TWO-LEVEL QUEUE MANAGEMENT METHOD ON
TELECOMMUNICATION NETWORK NODES**

Kharkiv National University of Radioelectronics, Kharkiv, Ukraine

Keywords: queue management, Quality of Service, two-level structure, multi-core systems.

The paper presents two-level queue management method. Within the proposed method queuing problem has been reduced to solving of optimization problem that was presented as two-level hierarchical structure. At the first level required distribution of flows and bandwidth allocation was calculated independently for each macro-queue, at the second level solutions obtained on lower level for each queue were coordinated in order to prevent probable network link overload. For managing multilevel structure the method of goal coordination has been determined, this allowed to define the order for consideration of queue cooperation restrictions and calculation tasks distribution between levels of hierarchy. Decisions coordination was performed using a the method of Lagrange multipliers. Using simulation modeling the study of method convergence has been undertaken.

В настоящее время важной задачей развития современных телекоммуникационных сетей является повышение качества обслуживания (Quality of Service, QoS) все возрастающего количества запросов пользователей. Среди множества средств повышения качества обслуживания значимое место занимают методы управления очередями, позволяющие эффективно улучшить такие показатели как средняя задержка пакетов, джиттер, количество отброшенных пакетов без существенных затрат на модернизацию существующей инфраструктуры сети [1,4]. Современная телекоммуникационная сеть имеет большую многоуровневую структуру, происходит переход к распределенным, облачным технологиям, которые требуют соблюдения в своей работе принципов

согласования, координации элементов [2]. Математическим аппаратом способным решить поставленную задачу согласования и координации распределенной сети для повышения качества обслуживания потоков пользователей является теория многоуровневых, иерархических систем [5].

Иерархические структуры (очереди) 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 своей группы очередей. Таким образом актуальной является задача исследования новых методов обслуживания очередей представленных иерархическими структурами. В данной работе предложена потоковая модель иерархических очередей в декомпозиционной форме в виде алгебраических уравнений.

В сетевом узле в ходе решения задачи обслуживания очередей необходимо определить распределение множества потоков между очередями, а также порядок обслуживания потоков пакетов заданной интенсивности [3,6]. Для повышения масштабируемости решений задачи обслуживания очередей определим двухуровневую функциональную иерархию расчетов:

- на нижнем уровне – искомое распределение потоков подлежит расчету независимо для каждой макроочереди для потоков, которые поступают от сетей доступа или других сетевых узлов;
- на верхнем уровне – полученные на нижнем уровне решения задач распределения потоков и выделения пропускной способности для каждой очереди координируются с целью предотвращения вероятной перегрузки канала связи ввиду децентрализации принимаемых решений на каждой конкретной макроочереди.

Общая схема предлагаемого метода представлена на рис. 1.

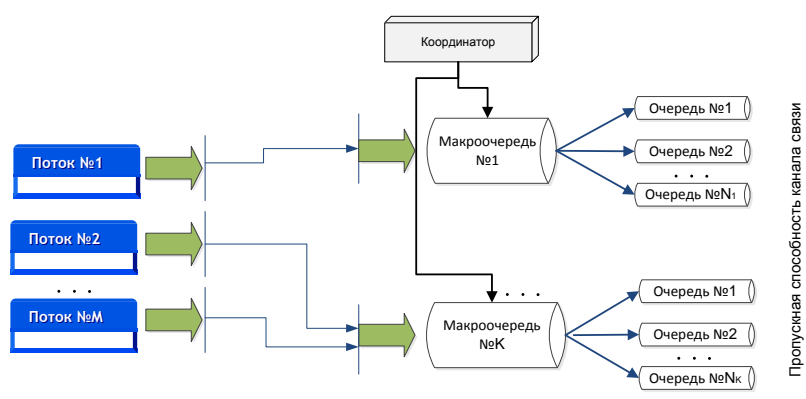


Рисунок 1. Общая схема двухуровневого метода иерархическо-координационного обслуживания очередей

Предположим, что для каждой r -й ($r = \overline{1, K}$) макроочереди в качестве искомым выступают переменные x_{ij}^r , которые характеризуют долю интенсивности i -го ($i = \overline{1, M}$) потока, направленного в j -ю ($j = \overline{1, N_r}$) очередь r -й макроочереди, а также переменные b_j^r , под которыми понимается пропускная способность, выделенная для j -й очереди r -й макроочереди.

Согласно физическому содержанию искомой переменной x_{ij}^r , а также с целью недопущения потерь пакетов в узле сети необходимо обеспечить выполнение следующих условий:

$$x_{ij}^r \in \{0,1\}, \quad \sum_{r=1}^K \sum_{j=1}^{N_r} x_{ij}^r = 1 \quad (i = \overline{1, M}). \quad (1)$$

Отсутствие потерь на сетевом узле должно гарантировать отсутствие потерь в сети в целом. Условие (1) должно выполняться для каждого i -го потока, поступающего в произвольную

макроочередь для обслуживания. Интенсивность суммарного потока направленного на обслуживание сетевого узла не должна превышать ПС канала связи:

$$\sum_{r=1}^K \sum_{j=1}^{N_r} \sum_{i=1}^M a_i x_{ij}^r \leq b \quad (2)$$

где a_i – интенсивность потока i -го класса, поступающего на обслуживание сетевого узла, b – пропускная способность исходящего канала связи.

Для решения задачи распределения трафика между очередями к использованию предлагается нелинейная функция минимизации средней длины очереди в узле:

$$\min_{x,b} F \quad \text{при } F = \sum_{r=1}^K \sum_{j=1}^{N_r} \gamma_j \left[\frac{\rho_j^r}{1-\rho_j^r} - \rho_j^r \right], \quad (3)$$

где γ_j – приоритет обслуживания пакетов j -й очереди, $\rho_j^r = \frac{\sum_{i=1}^M a_i x_{ij}^r}{b_{ij}^r}$ – коэффициент использования пропускной способности, выделенной для j -й очереди. Предложенная целевая функция (3) позволяет наиболее полно сохранить физический смысл искомым переменных.

Для решения поставленной задачи перехода к более гибкой, масштабируемой иерархической структуре обслуживания очередей описанную выше централизованную постановку задачи необходимо представить в декомпозиционной форме. В соответствии с теоремой о строгой дуальности задач на условный экстремум перехода к задаче на безусловный экстремум, имеет место равенство:

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

определив двойственную функцию $\Phi(\mu)$ следующим образом:

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

$$L = \sum_{r=1}^K L_r; \quad L_r = \sum_{j=1}^{N_r} \gamma_j \left[\frac{\rho_j^r}{1-\rho_j^r} - \rho_j^r \right] + \mu B_r \bar{x}_r - \bar{\mu}_r^t b \quad (4)$$

где $B_r, B_s (r, s \in K; r \neq s)$ – согласующие матрицы, координаты которых выбираются в соответствии с интенсивностями потоков, поступающих в сетевой узел; μ_r – множитель Лагранжа направленный в r -ю макроочередь для координации решений.

Таким образом общая проблема распределения нагрузки между очередями, выделения ПС оказалась декомпозицированной на ряд подзадач (по числу макроочередей), 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 (4). При этом лагранжиан (4) является функцией переменных только одного индекса, определяющего номер макроочереди.

Решение задачи по минимизации выражения (4) определяет нижний уровень расчетов, результатом которого является формирование распределения трафика и выделения ПС каждой конкретной макроочереди согласно содержимому 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 канала связи (2), осуществляется модификация вектора множителей Лагранжа в ходе выполнения следующей градиентной процедуры:

$$\mu_r(\alpha + 1) = \mu_r(\alpha) + \nabla \mu_r; \quad \nabla \mu_r(x_{ij}^r, b_j^r) = \sum_{r=1}^K B_r \bar{x}_r - b \quad (5)$$

где $\nabla\mu_r$ – градиент функции, рассчитывается, исходя из получаемых на нижнем уровне результатов решения задачи распределения потоков и выделения ПС каждой конкретной r -й макроочереди ($r \in K$). 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 (4). Полученные новые решения снова "спускаются" на первый уровень управления для координации следующей итерации расчетов.

В статье предложен двухуровневый метод иерархическо-координационного обслуживания очередей на узлах телекоммуникационной сети. В основу метода положен принцип целевой координации, а задача обслуживания очередей сведена к решению оптимизационной задачи, представленной в декомпозиционной форме. При решении задачи обслуживания очередей это позволило обеспечить более высокий уровень адекватности (системности) процесса управления, позволило адаптировать задачи управления перегрузками под современную многопроцессорную архитектуру сетевых узлов. Предложенный математический аппарат подходит для решения задач распределения ресурсов в распределенных, многоядерных, многопроцессорных, многомодульных системах. Использование параллельно работающих процессоров для вычислений на первом уровне иерархии позволит увеличить скорость нахождения оптимального решения.

Литература:

1. Вегенша Ш. Качество обслуживания в сетях IP: Пер. с англ. – М.: Издательский дом «Вильямс», 2003, 386 с.
2. Badger M., Grance T., Patt-Corner R., Voas J. Cloud Computing Synopsis and Recommendations: Recommendations of the National Institute of Standards and Technology // CreateSpace Independent Publishing Platform, 2012, 82 p.
3. Лемешко А.В., Семеняка М.В., Али С. Али Особенности проведения лабораторного эксперимента по исследованию процессов управления очередями в телекоммуникационных сетях // Цифрові технології, 2011, №10, с.5-17.
4. М.В. Семеняка, А.В. Симоненко, Али С. Али Разработка и исследование метода динамической балансировки очередей на маршрутизаторах мультисервисной телекоммуникационной сети // Проблеми телекомунікацій, 2013, №3(8), с. 66-78, [Электронный ресурс], режим доступа: http://pt.journal.kh.ua/2012/3/1/123_semenyaka_balancing.pdf
5. М. Сингх , А. Титли Системы: декомпозиция, оптимизация и управление, – М.: Машиностроение, 1986. – 494 с.
6. Tsai T.-Y., Chung Y.-L., Tsai Z. Introduction to Packet Scheduling Algorithms for Communication Networks // Communications and Networking, Jun Peng, 2010.