

УДК 621.391

МОДЕЛЬ И МЕТОД ПРЕДОТВРАЩЕНИЯ ПЕРЕГРУЗКИ С АКТИВНЫМ УПРАВЛЕНИЕМ ОЧЕРЕДЬЮ НА УЗЛАХ ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ



[А.В. ЛЕМЕШКО](#), [М.В. СЕМЕНЯКА](#)

Харьковский национальный
университет радиоэлектроники

Abstract – One of the most important challenges for telecommunication networks nowadays are toughening of requirements to quality of service (QoS) for delivering flows of different types to growing number of users. This results in the need of finding new methods for traffic management in the areas of routing, queues management, packet marking etc. This work is devoted to improving QoS using queue management (QM) mechanisms and presents congestion management (CM) model and method that includes active queue management (AQM) technique. Combining these two important methods (CM and AQM) of queue management within a single approach allows to get coordinated decision of solving flow distribution task, bandwidth allocation for each queue and preventive limitation of queue length to increase QoS. At the same time, formulated task was reduced to solving of a single optimization problem with usage of nonlinear programming. With help of analytical modeling methods performed research of the given model and suggested values of the control variables for investigated structure to get better QoS for user flows. It was taken into account the quality of obtained solutions for different input data (number of flows, their intensity and arrival time distribution, number of allocated queues, link capacity, penalty coefficients). Within a method the centralized queue management task, that is described in the model, was presented in a hierarchical form in order to take into account multicore, multiprocessor architecture of modern network equipment, which also increased the scalability of the final solutions. The main feature of the proposed model and method is the capability to control packet distribution to provide differentiated serving for flows with different class with simultaneous control of queue length to prevent overloading of each queue. This allows to improve firsts of all such QoS indicators as packet loss, delay, jitter without substantial upgrading of the existing network.

Анотація – Запропоновані модель та метод запобігання перевантаження з активним управлінням чергою на вузлах телекомунікаційної мережі. Об'єднання цих двох важливих способів управління чергами в рамках однієї моделі дозволило погоджено вирішити завдання щодо розподілу потоків між чергами, виділення пропускної здатності інтерфейсу для кожної черги, а також превентивного обмеження інтенсивності потоків на інтерфейсі маршрутизатора. З метою адаптації запропонованих рішень до багатоядерної та/або багатопроцесорної архітектури сучасних маршрутизаторів здійснено декомпозиційне представлення розробленої моделі та представлено ієрархічно-координаційний метод запобігання перевантаження з активним управлінням чергою. В результаті це дозволило підвищити масштабованість кінцевих рішень та сприяти покращенню якості обслуговування в цілому.

Аннотация – Предложены модель и метод управления предотвращения перегрузки с активным управлением очередью на узлах телекоммуникационной сети. Объединение этих двух важных способов управления очередями в рамках одной модели позволило согласованно решить задачи по распределению потоков между очередями, выделению пропускной способности интерфейса для каждой очереди, а также превентивного ограничения интенсивности потоков на интерфейсе маршрутизатора. С целью адаптации предлагаемых решений к многоядерной и/или многопроцессорной архитектуре современных маршрутизаторов осуществлено декомпозиционное представление разработанной модели и представлен иерархическо-координационный метод предотвращения перегрузки с активным управлением очередью. В результате это позволило повысить масштабируемость конечных решений и способствовать улучшению качества обслуживания в целом.

Введение

Важным направлением развития телекоммуникационных сетей (ТКС) сегодня является дальнейшее совершенствование средств обеспечения качества обслуживания (Quality of Service, QoS) потоков пользователей. Среди комплекса мер по обеспечению заданного уровня QoS важное место занимают решения по управлению очередями (Queue Management), т.к. именно очереди пакетов на маршрутизаторах

являются основным источником возникновения задержек, джиттера и возможных потерь пакетов [1]. В свою очередь особенности и эффективность технологических решений в области управления очередями зависят от тех математических моделей и методов, которые в них заложены. Среди множества механизмов управления очередями можно выделить две группы решений, которые получили наибольшее распространение:

- механизмы управления перегрузками (Congestion Management);
- механизмы предотвращения перегрузок (Congestion Avoidance), также называемые механизмами активного управления очередями (Active Queue Management).

Технологические средства управления перегрузками, представленные механизмами First In – First Out (FIFO), Priority Queues (PQ), Custom Queues (CQ), Weighted Fair Queue (WFQ), Class-Based Weighted Fair Queues (CBWFQ), Low Latency Queues (LLQ), имеют ряд недостатков, среди которых преобладающей является статичность настроек, проявляющаяся в необходимости вмешательства администратора при конфигурировании механизма; неспособность к адаптации под изменения в состоянии интерфейса и сети в целом; несогласованность получаемых решений с другими средствами управления очередями, например, механизмами предотвращения перегрузок Random Early Detection (RED), Weighted Random Early Detection (WRED), Random Exponential Marking (REM), Adaptive Virtual Queue (AVQ), Blue and Stochastic Fair Blue (SFB) и др., что способно нивелировать достоинства как первой, так и второй группы решений [2]. Поэтому актуальным видится направление исследований, связанное с разработкой новых моделей и методов управления очередями, которые могли быть положены в основу соответствующих управляющих механизмов на узлах как проводной, так и беспроводной телекоммуникационной сети [3, 4].

В этой связи в данной работе предлагается потоковая модель управления очередями, в рамках которой обеспечивается согласованное решение следующих важных задач для повышения качества обслуживания потоков пользователей:

- распределение поступающих потоков между отдельными очередями сетевого узла (Congestion Management);
- выделение пропускной способности (ПС) интерфейса для каждой из очередей (Resource Allocation);
- превентивное ограничение интенсивности поступающих на интерфейс потоков пакетов (Congestion Avoidance).

I. Модель предотвращения перегрузки с активным управлением очередью

Пусть в рамках предлагаемой модели на обслуживание сетевого узла поступает M потоков со следующими характеристиками: a_i – интенсивность i -го потока, pr_i – приоритет пакетов i -го потока ($i = \overline{1, M}$). Отличительной особенностью модели является возможность превентивного ограничения интенсивности потоков, поступающих на интерфейс маршрутизатора. Тогда в предлагаемую модель вводится множе-

ство управляющих переменных первого типа α_i , ($i = \overline{1, M}$), каждая из которых характеризует долю i -го потока, получившего отказ в обслуживании на интерфейсе маршрутизатора при реализации функций Congestion Avoidance. По своему физическому смыслу переменные α_i численно определяют вероятность отбрасывания пакетов i -го потока на рассматриваемом интерфейсе маршрутизатора.

Пакеты, поступившие на обслуживание, должны распределяться между N очередями в ходе решения задач Congestion Management путем расчета множества переменных второго типа $x_{i,j}$ ($i = \overline{1, M}$, $j = \overline{1, N}$), каждая из которых характеризует долю i -го потока, направленного на обслуживание в j -ю очередь. Для решения задач класса Resource Allocation в рамках предлагаемой модели необходимо обеспечить расчет множества переменных третьего типа – b_j , каждая из которых определяет величину пропускной способности интерфейса, выделенную для обслуживания j -й ($j = \overline{1, N}$) очереди. Пусть b – общая пропускная способность канала связи. Общая схема предлагаемого управления очередями представлена на рис. 1.

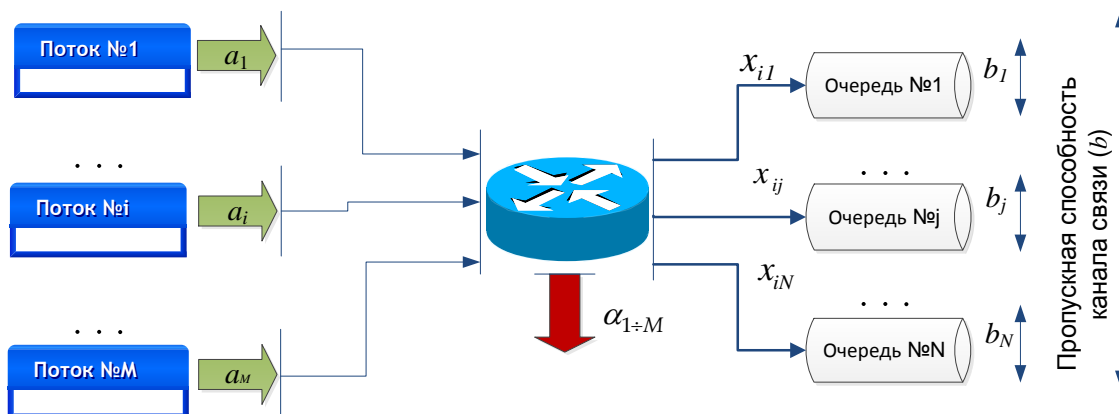


Рис. 1. Общая схема предлагаемого управления очередями

Искомые переменные всех трех типов удобно представить в виде соответствующих управляющих векторов:

$$\vec{x} = \begin{bmatrix} x_{1,1} \\ x_{1,2} \\ \dots \\ x_{1,N} \\ \dots \\ x_{M,N} \end{bmatrix}, \quad \vec{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_M \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_N \end{bmatrix}. \quad (1)$$

Тогда обобщенный вектор управляющих переменных можно представить в следующем виде:

$$\vec{Y} = \begin{bmatrix} \vec{x} \\ \vec{a} \\ \vec{b} \end{bmatrix}. \quad (2)$$

Размер вектора \vec{Y} определяется структурой векторов (1), т.е. зависит от числа потоков, поступивших на обслуживание, и количества формируемых очередей:

$$N_{\vec{Y}} = M \cdot N + M + N. \quad (3)$$

Для обеспечения приблизительно равного качества обслуживания пакетов одного и того же потока каждый их пакетов целесообразно обрабатывать в рамках одной из очередей. Таким образом, в соответствии с физикой решаемой задачи переменная $x_{i,j}$ является булевой, т.е.

$$x_{i,j} \in \{0,1\}. \quad (4)$$

Дифференциация качества обслуживания обеспечивается за счет того, что потоки с разными приоритетами (требованиями к качеству обслуживания) обрабатываются в различных очередях. Согласно физическому смыслу переменных α_i и b_j на них накладываются следующие ограничения:

$$0 \leq \alpha_i \leq 1, (i = \overline{1, M}), \quad (5)$$

$$\sum_{j=1}^R b_j \leq b, \quad 0 \leq b_j, (j = \overline{1, N}). \quad (6)$$

Взаимосвязь между управляющими переменными первого и второго типа, развивая подход, предложенный в работах [5-7], можно отразить в рамках следующих условий сохранения потока на интерфейсе маршрутизатора:

$$\sum_{j=1}^N x_{ij} + \alpha_i = 1, (i = \overline{1, M}). \quad (7)$$

Согласно условию (7) пакеты i -го потока могут быть либо направлены на обслуживание одной из очередей интерфейса, либо отброшены. Предложенный в данной работе подход также использован при решении задач маршрутизации с профилированием трафика [8].

С целью обеспечения управляемости процессом предотвращения перегрузки важно удовлетворить условие:

$$\sum_{i=1}^M a_i x_{i,j} < b_j, (j = \overline{1, N}), \quad (8)$$

выполнение которого гарантирует, что суммарная интенсивность потоков, направленных на обслуживание в j -ю очередь, не превысит пропускную способность интерфейса, выделенную для данной очереди. Условия (8) отражают функциональную взаимосвязь между переменными второго и третьего типа.

Для предотвращения перегрузки очереди по ее длине (в пакетах) условия (2)-(8) могут также быть дополнены нелинейными ограничениями:

$$\bar{n}_j < n_j^{\max}, \quad (9)$$

где n_j^{\max} – максимальный размер j -й очереди; \bar{n}_j – средняя длина j -й очереди, которая в зависимости от характеристик, формирующих данную очередь потоков, поддерживаемой дисциплины обслуживания пакетов, может быть рассчитана с использованием выражений, описанных в [9].

Расчет управляющего вектора \vec{Y} (2) целесообразно осуществить в ходе решения оптимизационной задачи, связанной с минимизацией как использования пропускной способности интерфейса, так и возможных отказов в обслуживании, вызванных превентивным ограничением интенсивности потоков

$$\min_{x,b,\alpha} F.$$

К использованию предлагались два типа целевых функций. Одна из них была представлена линейной формой вида:

$$F = \sum_{i=1}^M \sum_{j=1}^N h_{i,j}^x x_{i,j} + \sum_{i=1}^M h_i^\alpha \alpha_i + \sum_{j=1}^N h_j^b b_j, \quad (10)$$

где $h_{i,j}^x$ – условная стоимость (метрика) обслуживания i -го потока j -й очередью; h_i^α – условная стоимость отбрасывания пакетов i -го потока; h_j^b – условная стоимость выделения пропускной способности интерфейса для j -й очереди. Метрики $h_{i,j}^x$, h_i^α и h_j^b должны прямо пропорционально зависеть от приоритета пакетов обслуживаемых потоков pr_i , ($i = \overline{1, M}$).

К достоинствам использования линейной целевой функции (10) стоит отнести ее хорошую наглядность и относительно невысокую вычислительную сложность последующих расчетов, что существенно упрощает технологическую реализацию модели. Однако основным ее недостатком является то, что в первую очередь будет загружаться наиболее «дешевый» ресурс, вплоть до его полной перегрузки. Кроме того, в случае возможной перегрузки интерфейса первоначально ограничение будет испытывать низкоприоритетный поток вплоть до полного отказа в обслуживании. Поток с более высоким приоритетом не будет ограничиваться в обслуживании до тех пор, пока можно отказать низкоприоритетному.

Также в качестве целевой функции рекомендуется использовать квадратичную форму вида:

$$F = \vec{x}^t H_x \vec{x} + \vec{\alpha}^t H_\alpha \vec{\alpha} + \vec{b}^t H_b \vec{b}, \quad (11)$$

где $(\cdot)^t$ – операция транспонирования, H_x – диагональная матрица, содержащая весовые коэффициенты $h_{i,j}^x$:

$$H_x = \begin{vmatrix} h_{1,1}^x & 0 & \dots & 0 \\ 0 & h_{1,2}^x & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & h_{M,N}^x \end{vmatrix}, \quad (12)$$

H_α – диагональная матрица коэффициентов h_i^α :

$$H_\alpha = \begin{vmatrix} h_1^\alpha & 0 & \dots & 0 \\ 0 & h_2^\alpha & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & h_M^\alpha \end{vmatrix}, \quad (13)$$

H_b – диагональная матрица коэффициентов h_j^b :

$$H_b = \begin{vmatrix} h_1^b & 0 & \dots & 0 \\ 0 & h_2^b & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & h_N^b \end{vmatrix}. \quad (14)$$

Использование функции (11), несмотря на некоторое усложнение порядка расчета, позволяет обеспечить более справедливый (по сравнению с линейной формой (10)) характер использования канального ресурса интерфейса и возможных отказов в обслуживании [5]. Это проявлялось, например, в том, что в случае возможной перегрузки интерфейса отказы в обслуживании касаются всех потоков, но в меньшей степени высокоприоритетных, в большей – низкоприоритетных.

Расчетные решения, полученные в ходе использования модели (4)-(9), (14), должны обеспечивать адаптивный характер ограничения интенсивности потока при перегрузке интерфейса. Кроме необходимости учета приоритета пакетов обслуживаемых потоков в ходе решения важно обеспечить превентивный характер отказов в обслуживании по аналогии с функционалом механизма RED/WRED. Это сопряжено с обоснованием параметров самой модели (4)-(9), (14), что, прежде всего, касается выбора численных значений и соотношения весовых коэффициентов (12)-(14) в целевой функции (11), отвечающих за величины удельной стоимости при согласованном решении задач Congestion Management, Congestion Avoidance и Active Queue Management.

II. Исследование предложенной модели предотвращения перегрузки с активным управлением очередью

Целью исследования модели (4)-(9), (14) является определение таких соотношений между весовыми коэффициентами – координатами матриц (12)-(14), чтобы обеспечивался превентивный характер отказов в обслуживании пакетов по аналогии с возможностями механизма RED/WRED. При этом в случае низкой загруженности

интерфейса (до 50%) отбрасывание пакетов производить нецелесообразно. В случае средней (51-70%) и высокой (71-100%) загрузки интерфейса необходимо организовать превентивное отбрасывание пакетов (Congestion Avoidance) для предотвращения эффекта глобальной синхронизации, присущего механизму Tail Drop.

Исследование предложенной модели основывалось на результатах аналитического моделирования, в ходе которого минимизация функции (11) с учётом ограничений (4)-(9) производилась с использованием пакета Optimization Toolbox среды Matlab 7. Особенности решения задач по управлению очередями будут продемонстрированы на численном примере для исходных данных, представленных на рис. 2.

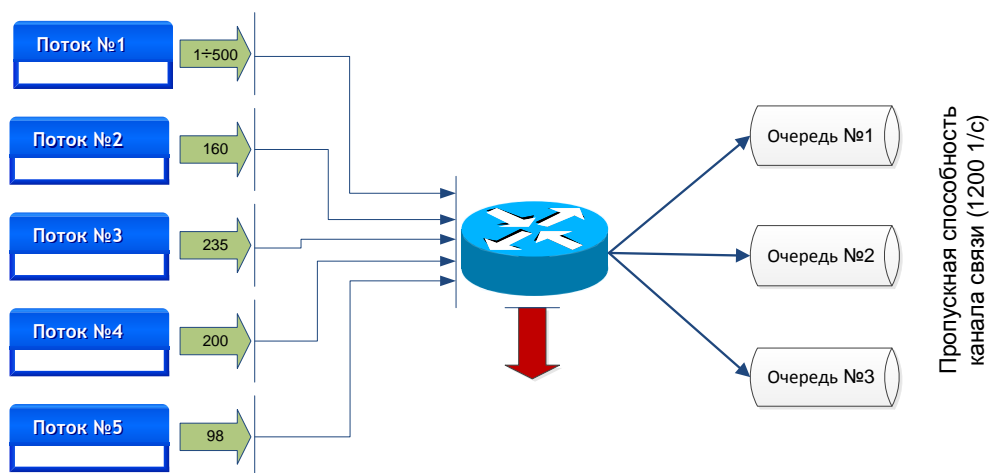


Рис. 2. Исходные данные для расчета

Пусть на вход маршрутизатора поступал трафик, состоящий из пяти потоков. Интенсивность первого потока изменялась от 1 до 500 1/с ($a_1 = 1 \div 500$ 1/с), тогда как интенсивность остальных потоков оставалась постоянной ($a_2 = 160$ 1/с, $a_3 = 235$ 1/с, $a_4 = 200$ 1/с, $a_5 = 98$ 1/с). Пропускная способность канала связи (интерфейса) составляла 1200 1/с. Для примера, обслуживание потоков предложено организовать с использованием трёх очередей.

Результаты исследования показали, что в случае, если $h_{i,j}^x \ll h_i^\alpha$, т.е. соотношение $h = h_{i,j}^x / h_i^\alpha$ составляло 1:600 и выше, ограничение интенсивности потока наблюдалось лишь в условиях перегрузки, когда интенсивность поступающего на интерфейс трафика превосходит пропускную способность канала связи, что аналогично поведению механизма Tail Drop. В случае, когда условная стоимость за отбрасывание пакетов меньше, чем метрика распределения пакетов между очередями ($h_i^\alpha < h_{i,j}^x$), пакеты отбрасывались даже в случае наличия на интерфейсе достаточного объема свободных ресурсов по пропускной способности. На рис.3 представлена зависимость вероятности отбрасывания пакетов 1-го потока от его интенсивности для различных соотношений метрик $h = h_{i,j}^x / h_i^\alpha$: 1:200, 1:300, 1:400, 1:500, 1:550. Вероятность отбрасывания пакетов i -го потока численно оценивалась через параметр α_i .

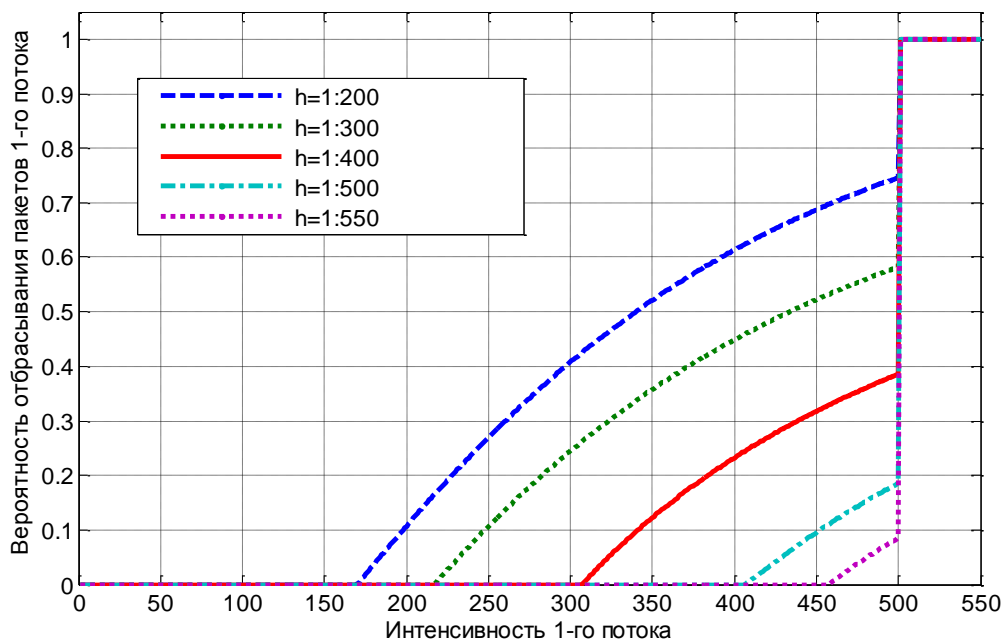


Рис. 3. Зависимость вероятности потери пакетов 1-го потока от его интенсивности

Как следует из рис. 3, изменением соотношения весовых коэффициентов можно регулировать степень превентивности при отбрасывании пакетов в зависимости от загруженности интерфейса. В рамках приведенных исходных данных можно представить следующие рекомендации по выбору координат матриц (12)-(14):

- в области низких нагрузок на интерфейс (при интенсивности потока 1÷150 1/с) отбрасывание пакетов нецелесообразно ввиду наличия достаточного объема пропускной способности, весовые коэффициенты должны выбираться из соотношения $h_{ij}^x / h_i^\alpha = 1:200$ и выше;

- в областях средних (151÷300 1/с) и высоких нагрузок (301÷500 1/с) можно организовать ограничение интенсивности потоков пакетов по аналогии с механизмом WRED, когда вероятность отбрасывания зависит от приоритета пакета.

Например, для потоков низкого приоритета, весовые коэффициенты стоит выбирать в пределах $h_{ij}^x / h_i^\alpha \approx 1:200 \div 1:400$, для потоков с более высоким приоритетом стоит выбирать весовые коэффициенты в пределах $h_{i,j}^x / h_i^\alpha \approx 1:400 \div 1:600$. Важно отметить, что для данного примера при соотношении $h_{i,j}^x / h_i^\alpha \approx 1:550$ вероятность отбрасывания пакетов будет соответствовать результатам работы механизма RED с параметрами, установленными «по умолчанию», т.е. для знаменателя граничной вероятности, равного десяти. Согласно результатам, представленным на рис. 3, в табл. 1 представлены рекомендованные значения для соотношения весовых коэффициентов h для потоков пакетов с различными IP-приоритетами ($pr_i, i = \overline{1, M}$).

Таблица 1. Соотношения весовых коэффициентов для потоков различного приоритета

| IP-приоритет | Рекомендуемое отношение весовых коэффициентов, $h = h_{i,j}^x / h_i^\alpha$ |
|--------------|---|
| 0 | 1:200 |
| 1 | 1:250 |
| 2 | 1:300 |
| 3 | 1:400 |
| 4 | 1:450 |
| 5 | 1:500 |
| 6 | 1:550 |
| 7 | 1:600 |

Предложенная модель (4)-(9), (14) ориентирована на централизованное управление очередями, когда согласованный расчет управляющих переменных, представленных вектором (2), осуществляется в рамках однопроцессорной архитектуры маршрутизатора.

III. Декомпозиционное представление модели предотвращения перегрузки с активным управлением очередью

Отличительной особенностью аппаратурных решений в области управления трафиком является использование многоядерных, многопроцессорных, многомодульных архитектур сетевых узлов, которые применяются с целью повышения производительности маршрутизаторов в ядре ТКС. Например, некоторые модели современных маршрутизаторов (Cisco XR 12000 series, Juniper MX2020) могут использовать по несколько десятков модулей передачи данных (Data Plane), каждый из которых использует многоядерную архитектуру [10]. Поэтому представленную выше модель (1)-(14) необходимо модифицировать под распределенный характер расчетов, дабы учесть, что каждый процессор, ядро или модуль сетевого узла отвечали за определенную группу очередей (макроочередь). С этой целью модель (1)-(14) будет приведена к декомпозиционной форме по количеству макроочередей. Это, посредством разбиения одной сложной задачи на ряд более простых, также позволит снизить вычислительную сложность расчетов и повысить масштабируемость предлагаемых решений в целом.

Первым шагом на пути решения поставленной задачи является сбор информации о состоянии интерфейсов маршрутизатора, которая включает данные об интенсивности поступающих потоков, доступной пропускной способности, классах (приоритетах) поступивших потоков. В дальнейшем будем считать, что задача по определению числа макроочередей и очередей в них предварительно решены. Тогда вектор искомым переменных каждой для r -й ($r = \overline{1, R}$) макроочередей будет иметь вид:

$$\vec{Y}_r = \begin{bmatrix} \vec{x}_r \\ \vec{\alpha}_r \\ \vec{b}_r \end{bmatrix}, \quad (15)$$

где \vec{x}_r – вектор переменных распределения потоков между очередями r -й макроочереди; $\vec{\alpha}_r$ – вектор переменных превентивного ограничения интенсивности потоков, поступающих в r -ю макроочередь; \vec{b}_r – вектор переменных выделения ПС интерфейса для обслуживания очередей r -й макроочереди.

Таким образом, условия предотвращения перегрузки очередей по пропускной способности и длине очереди будут иметь вид:

$$\sum_{i=1}^M a_i^r x_{i,j}^r < b_j^r, \quad (r = \overline{1, R}; j = \overline{1, N_r}) \quad (16)$$

$$\sum_{r=1}^R \sum_{j=1}^{N_r} b_j^r \leq b, \quad (17)$$

$$\bar{n}_j^r < n_{j \max}^r, \quad (18)$$

где a_i^r – интенсивность i -го потока, поступающего в r -ю макроочередь; b_j^r – пропускная способность интерфейса, выделенная для обслуживания пакетов j -й очереди r -й макроочереди; $x_{i,j}^r$ – доля интенсивности i -го потока, направленного в j -ю очередь r -й макроочереди; N_r – число очередей в r -й макроочереди; $n_{j \max}^r$ – максимальный размер j -й очереди r -й макроочереди; \bar{n}_j^r – средняя длина j -й очереди r -й макроочереди, которая вычисляется по аналогии со слагаемыми выражения (9) [9].

На переменные $x_{i,j}^r$, b_j^r и α_i^r накладываются ограничения, подобные (4)-(7). Переменные $x_{i,j}^r$ и α_i^r будут рассчитываться на процессорах (ядрах, модулях) для каждой макроочереди отдельно, т.е. независимо друг от друга, тогда условие (17) целесообразно представить в декомпозиционной форме:

$$b_j^r \leq b - \sum_{s=1}^R \sum_{i=1}^{N_s} b_i^s, \quad (r = \overline{1, R}, j = \overline{1, N_r}). \quad (19)$$

В правой части неравенства (19) суммирование производится для случая, когда $i \neq j$ при $s = r$. Таким образом, в условии (19) учитывается децентрализация при расчете порядка распределения пропускной способности интерфейса между очередями макроочереди, так как каждая макроочередь управляется без информации о результатах расчета по другим макроочередям. Смысл выражения (19) заключается в том, что ПС, выделенная для обслуживания потоков j -й очереди r -й макроочереди, не должна превышать ПС, которая осталась после обслуживания остальных очередей/макроочереди. В векторно-матричной форме условие (20) будет иметь вид:

$$\bar{b}_r \leq I_r \cdot b - \sum_{s=1}^R C_{r,s} \bar{b}_s, \quad (r = \overline{1, R}), \quad (20)$$

где $C_{r,s}$ – согласующая матрица, координаты которой выбираются в соответствии с выражением (21); I_r – вектор, состоящий из единиц и имеющий размер такой же, как и вектор \bar{b}_r .

IV. Иерархическо-координационный метод предотвращения перегрузки с активным управлением очередью

В основу предлагаемого метода будет положено решение оптимизационной задачи, связанной с минимизацией целевой функции (11) при наличии условий-ограничений (4)-(7), (18), (20), (22). В ходе решения задачи условной оптимизации (11) с целью учета условий на взаимодействие макроочереди, переходя к задаче на безусловный экстремум, имеет место равенство:

$$\min_{x, \alpha, b} F = \max_{\mu} \Phi, \quad (23)$$

где двойственная функция $\Phi(\mu)$ определяется следующим образом:

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

при

$$L = \sum_{r=1}^R \left[(\bar{x}_r)^t H_x^r \bar{x}_r + (\bar{\alpha}_r)^t H_\alpha^x \bar{\alpha}_r + (\bar{b}_r)^t H_b^r \bar{b}_r + \bar{\mu}_r^t (\bar{b}_r + \sum_{s=1}^R C_{r,s} \bar{b}_s - I_r b) \right], \quad (24)$$

$\bar{\mu}_r$ – вектор множителей Лагранжа.

Лагранжиан (24) при фиксированных значениях $\bar{\mu}_r$, учитывая тождество

$$\sum_{r=1}^R \bar{\mu}_r^t \sum_{s=1}^R C_{r,s} \bar{b}_s = \sum_{r=1}^R \sum_{s=1}^R \bar{\mu}_r^t C_{r,s} \bar{b}_s,$$

можно представить в виде:

$$L = \sum_{r=1}^R L_r,$$

где

$$L_r = (\bar{x}_r)^t H_x^r \bar{x}_r + (\bar{\alpha}_r)^t H_\alpha^x \bar{\alpha}_r + (\bar{b}_r)^t H_b^r \bar{b}_r + \bar{\mu}_r^t \bar{b}_r + \left(\sum_{\substack{s=1 \\ s \neq r}}^R \bar{\mu}_s^t C_{s,r} \right) \bar{b}_r - \bar{\mu}_r^t I_r b. \quad (25)$$

Для решения оптимизационной задачи (23)-(25) использован принцип прогнозирования (предсказания) взаимодействия [11], в рамках которого вводится следующая иерархия задач:

- задача нижнего уровня иерархии: определение порядка распределения потоков ($x_{i,j}^r$) и отказов в обслуживании (α_i^r) на процессорах или ядрах каждой из макроочереди путем минимизации выражения (25);

- задача верхнего уровня иерархии (координатора интерфейса): координация решений нижнего уровня, заключающаяся в вычислении переменных, отвечающих за выделение ресурсов ПС каждой очереди (b_j^r), и вектора координирующих коэффициентов ($\vec{\mu}_r$) с учетом поступивших с нижнего уровня значений $x_{i,j}^r$ и α_i^r для обеспечения выполнения условий (19) в ходе минимизации (24).

Важным отличием используемого принципа прогнозирования взаимодействия является возможность получения в любой момент времени реализуемых (допустимых) значений управляющих переменных, т.е. на любой итерации канал связи не будет перегружен, но оптимальное решение будет найдено, когда значения координат векторов $\vec{\mu}_r$ и \vec{b}_r на предыдущем шаге будут незначительно отличаться от значений на текущей итерации:

$$|\mu(k+1) - \mu(k)| \leq \xi_\mu \quad (\xi_\mu \rightarrow 0), \quad (26)$$

$$|b_j^r(k+1) - b_j^r(k)| \leq \xi_b \quad (\xi_b \rightarrow 0), \quad (27)$$

где ξ – погрешность, с которой находится оптимальное решение, $|\cdot|$ – модуль числа. Этим принцип прогнозирования взаимодействий выгодно отличается от принципа целевой координации [6, 7], при использовании которого получаемое решение становится допустимым, т.е. удовлетворяет условиям на взаимодействие (19), лишь по окончании координирующей процедуры.

Полученные на верхнем уровне результаты расчета вновь "спускаются" на нижний уровень для очередной итерации. При этом, если условие (20) удовлетворяется на одной из промежуточных итераций, целесообразно реализовать так называемую секвенционную координацию, т.е. использовать промежуточные (субоптимальные) результаты, что на практике позволит уменьшить время расчета искомых переменных, представленных вектором (15).

Вычислительная структура предложенного иерархическо-координационного метода предотвращения перегрузки с активным управлением очередью представлена на рис. 4.

Выводы

В работе предложена модель предотвращения перегрузки с активным управлением очередью на маршрутизаторе телекоммуникационной сети. Новизна модели состоит в обеспечении согласованного решения задач по распределению поступающих потоков между отдельными очередями сетевого узла (Congestion Management); выделению пропускной способности (ПС) интерфейса для каждой из очередей (Re-

source Allocation) и превентивному ограничению интенсивности поступающих на интерфейс потоков пакетов (Congestion Avoidance).

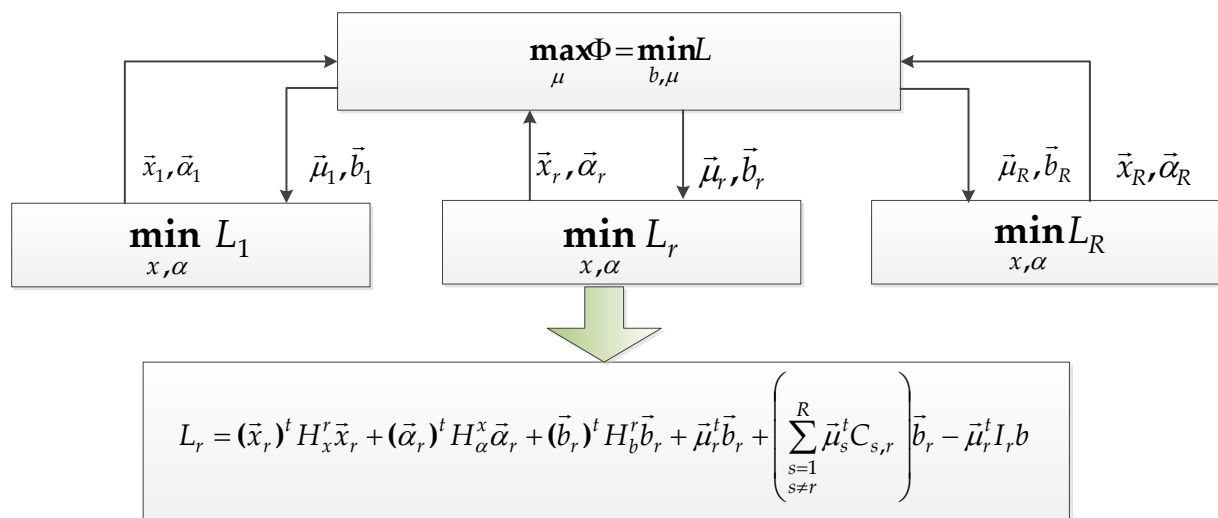


Рис. 4. Вычислительная структура иерархическо-координационного метода предотвращения перегрузки с активным управлением очередью

Согласованный расчет управляющих переменных $(x_{i,j}, b_j, \alpha_i)$, отвечающих за перечисленные задачи, осуществлялся в ходе решения оптимизационной задачи смешанного целочисленного нелинейного программирования, связанной с минимизацией целевой функции (11) при наличии ограничений (4)-(9).

В ходе исследования предложенной модели сформулированы рекомендации по выбору весовых коэффициентов в целевой функции (11) для обеспечения превентивного ограничения интенсивности потоков, поступающих на интерфейс маршрутизатора для борьбы с его перегрузкой. Путем настройки соотношения между этими весовыми коэффициентами удалось организовать ограничение интенсивности потоков по аналогии с механизмом WRED, когда вероятность отбрасывания пакетов зависела от его приоритета.

С целью адаптации к многоядерной, многопроцессорной архитектуре построения современного сетевого оборудования ориентированная на централизованное управление очередями модель (4)-(11) получила свое развитие и была представлена в декомпозиционной форме (15)-(20). Преобразование коснулось условий предотвращения перегрузки (6) и (8). В рамках данной декомпозиционной модели задача управления очередями была сформулирована как задача иерархического управления многоуровневой системой, для решения которой был использован принцип прогнозирования взаимодействий, положенный в основу предлагаемого метода предотвращения перегрузки с активным управлением очередью. На нижнем уровне (процессоре или ядре маршрутизатора) решались задачи распределения потоков между очередями и превентивного ограничения их интенсивности. На верхнем уровне (координаторе) решалась задача выделения ПС очередям интерфейса.

Использование предложенных решений позволило обеспечить согласованность при решении основных интерфейсных задач, связанных с управлением очередями, а также повысить масштабируемость конечных решений за счет реализации иерархическо-координационной стратегии управления.

Список литературы:

1. *Vegesna S.* IP Quality of Service // Cisco press, 2001. – 368 p.
2. *Tsai T.-Y., Chung Y.-L., Tsai Z.* Introduction to Packet Scheduling Algorithms for Communication Networks // Communications and Networking, Jun Peng. – 2010. – 434 p.
3. *Silva A.P., Burleigh S., Hirata C.M., Obraczka K.* Congestion control in disruption-tolerant networks: a comparative study for interplanetary networking applications // CHANTS'14 Proceedings of the 9th ACM MobiCom workshop on Challenged networks. – 2014. – P. 65-68.
4. *Michopoulos V., Guan L., Oikonomou G., Phillips I.* A Comparative Study of Congestion Control Algorithms in IPv6 Wireless Sensor Networks // International Conference on Distributed Computing in Sensor Systems and Workshops (DCOSS). – 2011. – P. 1-6.
5. *Lemeshko A.V., Ali S. Ali, Semenyaka M.V.* Results of the dynamic flow-based queue balancing model research // Modern Problems of Radio Engineering Telecommunications and Computer Science (TCSET'2012). – 2012. – P. 318-319.
6. *Lemeshko A.V., Semenyaka M.V.* Researching of mathematical models based on optimal control approaches for congestion control in telecommunication network // East-West Design & Test Symposium (EWDTS'2012). – 2012. – P. 341-344.
7. *Lemeshko A.V., Semenyaka M.V., Simonenko A.V.* Researching and designing of the dynamic adaptive queue balancing method on telecommunication network routers // The experience of designing and application of CAD Systems in Microelectronics (CADSM'2013). – 2013. – P. 204-207.
8. *Лемешко А.В., Добрышкин Ю.Н., Щербинин С.А.* Исследование модели управления трафиком с анализом областей превентивного ограничения его интенсивности на границе сети // Моделювання та інформаційні технології: Збірник наукових праць.– К.: Інститут проблем моделювання в енергетиці Національної АН України. – 2008. – Вип. 49. – С. 65-71.
9. *Семеняка М.В.* Двухуровневый метод иерархическо-координационного обслуживания очередей на узлах телекоммуникационной сети // Научно-технический вестник информационных технологий, механики и оптики. – 2014. – Вып. № 4 (92). – С. 98-105.
10. *Wood R.* Next-Generation Network Services // Cisco Press, 2005. – 624 p.
11. *Сингх М, Титли А.* Системы: декомпозиция, оптимизация и управление. – М.: Машиностроение. – 1986. – 494 с.