

МЕТОД ДИНАМИЧЕСКОГО УПРАВЛЕНИЯ ОЧЕРЕДЯМИ В СЕТИ MPLS-TE

Лемешко А.В., Али С. Али

Харьковский национальный университет радиоэлектроники
61166, Харьков, пр. Ленина, каф. телекоммуникационных систем,

E-mail: avlem@ukr.net, конт. тлф. 702-13-20

We propose a method and flow-based model balancing queues at the nodes of MPLS-network. The novelty of the model is that it is in contrast to the previously known models take into account the features of the technology Traffic Engineering Queues, aimed at achieving a balanced load buffer resource node.

Важным технологическим средством поддержки и повышения качества обслуживания (Quality of Service, QoS) в современных телекоммуникационных системах и сетях являются механизмы обслуживания очередей, возникающих на узлах (маршрутизаторах, маршрутизирующих коммутаторах) транспортной сети при перегрузках ее каналов связи [1, 2]. Именно задержки пакетов в очередях и их потеря при переполнении очереди определяющим образом влияют на межконцевые показатели QoS – среднюю задержку, джиттер, вероятность доставки пакетов и др.

Несмотря на обилие известных механизмов обслуживания очередей (PQ, CQ, CBQ, FQ, WFQ, LLQ и др.), поддерживаемых в современном сетевом оборудовании различных фирм производителей, проблема обеспечения эффективного решения данной задачи стоит все еще достаточно остро. При этом основное внимание при совершенствовании перечисленных механизмов уделяется следующим аспектам (требованиям):

- поддержка дифференциации пакетов, принадлежащих различным типам трафика;
- повышение уровня автоматизации при решении данной задачи с постепенным снижением роли административного вмешательства;
- переход от по пакетной стратегии обслуживания очередей к потоковой, что должно снизить время обработки пакетов в очереди;
- обеспечение согласованности решений с другими задачами управления трафиком (маршрутизации, распределения пропускной способности (ПС) каналов связи (КС), ограничения длины очереди, резервирования ресурсов и др.).

К сожалению, сдерживающим фактором на пути реализации отмеченных требований является несовершенство, а то и полное отсутствие четкого формализованного описания процесса обслуживания очередями, т.к. большинство механизмов основывается на преимущественно эвристических решениях, под которые можно и нужно подвести необходимую теоретическую базу в виде математических моделей и методов. В этой связи, в настоящем докладе основное внимание будет уделено вопросам выбора математической модели и разработки метода управления очередями на узлах сети MPLS-TE, т.к. именно эта технология де-факто является основной транспортной платформой при построении и функционировании сетей следующего поколения (Next Generation Network, NGN).

При описании модели условимся, что число агрегированных потоков трафика известно и равно M , что соответствует принятым на практике решениям в рамках известных методов маркировки пакетов. Наряду с этим примем, что максимальное число очередей на сетевом узле также фиксировано (N). Кроме того, обозначим через a_i ($i = \overline{1, M}$) – интенсивность трафика i -го класса, поступающего на обслуживание сетевым узлом. Кроме того, пусть b_j ($j = \overline{1, N}$) – часть пропускной способности исходящего КС, которая выделена j -й очереди ($j = \overline{1, N}$). Переменные b_j ($j = \overline{1, N}$) будут рассчитывать динамично, адаптируясь к изменению состояния сетевого узла, а не административно, как, например, в большинстве известных решений.

В ходе управления очередями необходимо выполнить условие отсутствия перегрузки канала связи:

$$\sum_{j=1}^N b_j \leq b, \quad (1)$$

и условие предотвращения перегрузки сетевого узла:

$$\sum_{i=1}^M a_i \leq b. \quad (2)$$

где b – пропускная способность исходящего КС.

Придать динамический характер процессу обслуживания очередей в рамках предлагаемой модели можно путем введения управляющей переменной x_{ij} , под которой подразумевалась доля i -го трафика, поступающего для обслуживания в j -ю очередь. Согласно смыслу x_{ij} имеют место следующие условия:

$$x_{ij} \in \{0,1\} \quad (i = \overline{1,M}, j = \overline{1,N}), \quad (3)$$

$$\sum_{j=1}^N x_{ij} = 1 \quad (i = \overline{1,M}), \quad (4) \quad \sum_{i=1}^M a_i x_{ij} \leq b_j \quad (j = \overline{1,N}). \quad (5)$$

Выполнение условия (4) гарантирует отсутствие потерь пакетов на рассматриваемом сетевом узле. Условия (5) вводятся для предотвращения перегрузки пропускной способности КС, выделяемой для передачи пакетов той или иной очереди сетевого узла в процессе управления. В качестве искомого вектора выберем вектор

$$\bar{x} = \begin{bmatrix} x_{ij} \\ \dots \\ b_j \end{bmatrix} \quad (i = \overline{1,M}; j = \overline{1,N}), \quad (6)$$

в ходе расчета которого удастся обеспечить согласованность в решении задач обслуживания очередей и динамического распределения за ними пропускной способности исходящего канала связи.

Для каждой очереди определим ее текущую загруженность и максимальную емкость, обозначив их соответственно через \bar{n}_j и n_j^{\max} ($j = \overline{1,N}$). Кроме того, дополним условия предотвращения перегрузки отдельных очередей по их пропускной способности (5) условиями предотвращения перегрузки очередей по их длине. В общем виде искомые условия будут иметь вид:

$$\bar{n}_j \leq n_j^{\max} \quad (j = \overline{1,N}), \quad (7)$$

и задача теперь сводится лишь к выбору (обоснованию) аналитического выражения для расчета средней длины очереди в процессе обслуживания.

Средняя длина очереди, опуская индекс очереди, может выражаться различными аналитическими зависимостями. Например для СМО М/М/1/ n^{\max} :

$$\bar{n} = \frac{\rho^2 \left[1 - (n^{\max} + 1)\rho^{n^{\max}} + n^{\max}\rho^{n^{\max}+1} \right]}{(1 - \rho^{n^{\max}+2})(1 - \rho)} - \rho, \quad \text{где } \rho = \frac{\sum_{i=1}^M a_i x_{ij}}{b_j}.$$

Целесообразно задачу, связанную с расчетом вектора (6), сформулировать в виде оптимизационной. Основным требованием к целевой функции является учет физики протекающих на узле процессов обслуживания пакетов (1)–(7), а также соответствие получаемых решений принципам концепции Traffic Engineering Queues, касающихся обеспечения сбалансированной загрузки буферных ресурсов. С этой целью выше изложенную модель важно дополнить следующими условиями

$$f(p_j, d_j) \cdot n_j \leq \alpha, \quad (j = \overline{1,N}), \quad (8)$$

где α – верхний динамически управляемый предел загруженности очередей на узле ТКС, $f(p_j, d_j) = \frac{p_j}{v \cdot d_j}$ – функция от характеристик j -го потока: его приоритета p_j и длины пакета d_j , а v – некоторый нормировочный коэффициент, который должен сглаживать различие в порядке значений приоритета ($0 \div 7$) и длины пакета в байтах.

Тогда задача обслуживания очередей может быть сведена к задаче балансировки их длины в ходе минимизации следующей целевой функции:

$$\min_{x, b, \alpha} \alpha, \quad (9)$$

что соответствует минимизации верхнего порога загруженности очередей на узле ТКС, взвешенного относительно таких характеристик потока, как длина пакета и его приоритет, что способствует сбалансированной загруженности всех очередей в соответствии с требованиями технологии Traffic Engineering Queues.

В рамках модели данной модели (1)-(9) предложен метод динамического управления очередями на узлах сети, который основывается на решении следующих важных задач:

- сбор и обработка информации о состоянии (числе и загруженности) очередей, а также характеристиках обслуживаемого трафика (интенсивность, длина и приоритет пакетов). При этом, процесс сбора информации может носить как периодический характер (по аналогии с дистанционно-векторной маршрутизацией), так и аperiodический, т.е. «по требованию»;

- расчет управляющих переменных, представленных вектором (6), в ходе решения оптимизационной задачи (9), что позволит придать процессу обслуживания очередей динамический характер. В результате расчетов формируется таблица обслуживания очередей пакетов в соответствии с их длиной и приоритетом;

- анализ полученных решений с последующей оценкой состояния очередей и характеристик трафика (т.е. возврат к п.1). Если решение обеспечить в рамках принятых ограничений невозможно, что обычно связано с переполнением очереди, то задействуются стандартные средства ограничения длины очереди.

Таким образом, в докладе предложены потоковая модель и метод динамического управления очередями на узлах сети MPLS-TE. Новизна модели состоит в том, что она в отличие от ранее известных моделей учитывает особенности технологии Traffic Engineering Queues, нацеленной на обеспечение сбалансированной загруженности буферного ресурса – очередей сетевого узла. Важной особенностью предлагаемого решения является то, что балансировку в рамках предлагаемой модели планируется осуществлять с учетом приоритета и длины образующих ту или иную очередь пакетов.

Сама технологическая по своей сути задача обслуживания очередей в общем случае была сведена к оптимизационной задаче смешанного математического программирования, связанной с минимизацией линейной функции (9) при наличии в т.ч. нелинейных ограничений (7) и (8), а также булевой природы некоторых рассчитываемых переменных (3). Решение данной задачи предполагает использование хорошо апробированных методов решения – округления (Rounding-off), ветвей и границ (Branch-and-bound), последовательной линеаризации (SLP), штрафных функций (Penalty function), множителей Лагранжа (Lagrangian relaxation) и различных смешанных методов.