

МАТЕМАТИЧЕСКИЕ МОДЕЛИ И УПРАВЛЕНИЕ В ТЕЛЕКОММУНИКАЦИОННЫХ СИСТЕМАХ

УДК 621.391

А.В. ЛЕМЕШКО, д-р техн. наук, А.В. СИМОНЕНКО, С.И. СИВАЩЕНКО, канд. техн. наук

ПОТОКОВАЯ МОДЕЛЬ ДИНАМИЧЕСКОЙ БАЛАНСИРОВКИ ОЧЕРЕДЕЙ НА УЗЛАХ ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ

Введение

Как показал проведенный анализ [1, 2], ключевая роль в обеспечении качества обслуживания (Quality of Service, QoS) в современных мультисервисных телекоммуникационных сетях (ТКС) отводится механизмам управления очередями на узлах (маршрутизаторах, маршрутизирующих коммутаторах) ТКС. К основным требованиям, которые выдвигаются к подобного рода механизмам, следует отнести динамический характер решения задачи (по аналогии с протоколами динамической маршрутизации), обеспечение справедливого обслуживания пакетов различных очередей, недопущение перегрузки каждой отдельной очереди и сетевого узла в целом. Динамическое управление очередями определяет необходимость разработки соответствующих моделей и методов (алгоритмов), использование которых позволило бы осуществлять перерасчет порядка обслуживания очередей зависимости от их загруженности, а также характеристик поступающего на узел трафика. Учет интенсивности трафика предполагает необходимость использования именно потоковых моделей (flow-based model), в которых основной учитываемой характеристикой трафика является его интенсивность наряду с другими важными параметрами – длины пакета, приоритета и т.д. Это особенно актуально ввиду того, что многие средства управления сетевыми ресурсами в настоящее время стали потоковоориентированными, например Flow-Based WFQ, Flow-Based Distributed WFQ, Flow-Based Routing или Flow-Based Weighted Random Early Detection.

В этой связи заслуживает внимания подход, основанный на балансировке очередей на принципах технологии инжиниринга трафика (Traffic Engineering Queues) [3, 4], предложенного для управления очередями в сетях MPLS (MultiProtocol Label Switching). Реализация принципов Traffic Engineering (TE) при управлении различными сетевыми ресурсами (трафиком, очередями, пропускной способностью каналов связи и т.д.), по мнению разработчиков данной технологии, должна обеспечить рост качества обслуживания в сети в целом.

Модель балансировки очередей на принципах Traffic Engineering Queues

В процессе моделирования условимся, что число отдельных трафиков или агрегированных по классам или приоритетам потоков 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 или назначается административно. Например, в алгоритме приоритетного обслуживания (Priority Queuing, PQ) может быть выделено до четырех очередей ($N = 4$), а в алгоритме взвешенного справедливого обслуживания (Weighted Fair Queuing, WFQ) по умолчанию число очередей равно 256, а в алгоритме CBWFQ (Class-based WFQ) число очередей соответствует количеству классов обслуживаемых трафиков и равно 64 [1, 2].

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 канала связи, которая выделена j -й очереди ($j = \overline{1, N}$), что типично, например, для алгоритма CBWFQ. Одно из ключевых отличий предлагаемого решения будет состоять в том, что переменные b_j ($j = \overline{1, N}$) будут рассчитываться адаптивно к изменению состоянию сетевого узла, а не административно, как, например, в CBWFQ.

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

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

где b – пропускная способность исходящего КС.

Кроме того, с целью предотвращения перегрузки сетевого узла необходимо обеспечить выполнение следующего условия:

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

Выполнение условия (2) определяет необходимость превентивного ограничения интенсивности суммарного (агрегированного) потока пакетов, поступающих на сетевой узел, чтобы она не превышала пропускную способность исходящего канала связи. Эта функция, как правило, алгоритмов произвольного раннего обнаружения перегрузки и ограничения длины очереди (RED – Random Early Detection, WRED – Weighted RED).

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

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

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

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

Выполнение условия (4) гарантирует отсутствие потерь пакетов на рассматриваемом сетевом узле. Условия (5) вводятся для предотвращения перегрузки пропускной способности отдельных очередей сетевого узла в процессе управления. По аналогии с моделью, рассмотренной в работах [5, 6], 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.

Формализация условий предотвращения перегрузки очередей на узле ТКС

В ходе выполнения условий (5) ввиду случайного и нестационарного характера сетевого трафика на узле возникают очереди и связанные с ними задержки пакетов. С целью введения верхней границы подобных задержек на узлах ТКС общую буферную емкость, как правило, ограничивают. Таким образом, для каждой очереди определим ее текущую загруженность и максимальную вместимость, обозначив их соответственно через \bar{n}_j и n_j^{\max} ($j = \overline{1, N}$), и

дополним условия предотвращения перегрузки отдельных очередей по их пропускной способности (5) условиями предотвращения перегрузки очередей по их длине. В самом общем виде искомые условия будут иметь следующий 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.

Для формулировки искомых условий необходимо задаться моделью трафика (его характеристиками – интенсивность, длина пакета и т.д.) и моделью обслуживания пакетов в рамках отдельно взятой очереди, в качестве которой на практике, как правило, реализуется модель FIFO (First In, First Out). Кроме того, выбирая гипотезу относительно характеристик обслуживаемого трафика, можно описать отдельную очередь известной системой массового обслуживания (СМО), в рамках которой уже можно корректно рассчитать ее вероятностно-временные характеристики, в том числе время обслуживания и среднюю (текущую) длину очереди. Так, например, при моделировании процесса обслуживания в отдельной очереди различными типами систем массового обслуживания средняя длина очереди, опуская индекс очереди, будет выражаться следующими аналитическими зависимостями [7]:

СМО M/M/1:
$$\bar{n} = \frac{\rho}{1-\rho};$$

СМО M/M/1/ n^{\max} :
$$\bar{n} = \frac{\rho^2 \left[1 - (n^{\max} + 1)\rho^{n^{\max}} + n^{\max}\rho^{n^{\max}+1} \right]}{(1 - \rho^{n^{\max}+2})(1 - \rho)};$$

СМО M/D/1:
$$\bar{n} = \frac{\rho}{1-\rho} - \frac{\rho^2}{2(1-\rho)};$$

СМО M/D/1/ n^{\max} :
$$\bar{n} = \frac{\sum_{n=0}^{n^{\max}} n \left[\sum_{j=0}^n \frac{e^{j\rho} (-j\rho)^{n-j}}{(n-j)!} - \sum_{j=0}^{n-1} \frac{e^{j\rho} (-j\rho)^{n-j-1}}{(n-j-1)!} \right]}{\sum_{j=0}^{n^{\max}+1} \frac{e^{j\rho} (-j\rho)^{n^{\max}-j+1}}{(n^{\max}-j+1)!}};$$

СМО самоподобного трафика:
$$\bar{n} = \frac{\rho^{1/2(1-H)}}{(1-\rho)^{H/(1-H)}},$$

где $\rho = \frac{\sum_{i=1}^M a_i x_{ij}}{b_j}$, $0.5 \leq H \leq 1$ – параметр самоподобия (Херста).

При этом каждому типу трафика, а значит и каждой очереди, может соответствовать своя модель обслуживания, не обязательно соответствуя перечню вариантов СМО, приведенных выше. Кроме того, нетрудно заметить, что условие (7) является более строгим, чем требование (5).

Формулировка оптимизационной задачи по управлению очередями на сетевом узле

В связи с тем, что, в общем случае, выбор x_{ij} и b_j в рамках ограничений (1), (3), (4) и (7) можно произвести множеством случаев, то целесообразно задачу, связанную с расчетом искомых переменных (6), сформулировать в виде оптимизационной задачи. 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, касающихся обеспечения сбалансированной загрузки буферных ресурсов. С этой целью изложенную выше модель важно дополнить следующими условиями

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

где α – верхний динамически управляемый предел загруженности очередей на узле ТКС.

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

$$\min \alpha, \quad (9)$$

что соответствует минимизации верхнего порога загруженности очередей на узле ТКС, т.е. способствует сбалансированной загруженности всех очередей в соответствии с требованиями технологии Traffic Engineering Queues.

Выводы

В статье предложена потоковая модель балансировки очередей на узлах телекоммуникационной сети. 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 сетевого узла. Сама технологическая задача управления очередями была сведена к оптимизационной задаче смешанного математического программирования, связанной с минимизацией линейной функции (9) при наличии в т.ч. нелинейных ограничений (7) – (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 линейризации (SLP), штрафных функций (Penalty function), множителей Лагранжа (Lagrangian relaxation), имитации отжига (Simulated annealing), а также генетический алгоритм (Genetic algorithm) и различные смешанные (гибридные) методы [8–10].

Список литературы: 1. *Вегенша Ш.* Качество обслуживания в сетях IP: Пер. с англ. М.: Изд. дом «Вильямс», 2003. 386 с. 2. *Справочник по телекоммуникационным технологиям:* Пер. с англ. М.: Изд. дом «Вильямс», 2004. 640 с. 3. *Li, Y. Panwar, S. Liu, C.J.* On the Performance of MPLS TE Queues for QoS Routing // Simulation series. 2004. Vol 36; part 3. PP. 170-174. 4. *Huerta, M., Padilla, J. J., Hesselbach, X., Fabregat, Ramon, Ravelo O.* Buffer Capacity Allocation: A method to QoS support on MPLS networks // Proc. EATIS2006-Euro American Conference on Telematics and Information Systems, 2006. PP. 14-28. 5. *Лемешко А.В., Симоненко А.В., Ватти Махмуд.* Потоковая модель управления очередями с динамическим распределением пропускной способности исходящего канала связи // Наукові записки УНДІЗ. 2008. №3(5). С. 34-39. 6. *Симоненко А.В., Ахмад Хайлан, Али Али* Модель динамического управления очередями и пропускной способностью канала связи на маршрутизаторах мультисервисной сети // Радиотехника. 2008. Вып. 155. С. 164-168. 7. *Петров М.Н.* Вероятностно-временные характеристики в сетях и системах передачи интегральной информации. Красноярск, 1997. 220 с. 8. *Tawarmalani M., Sahinidis N.V.* Global optimization of mixed-integer nonlinear programs: A theoretical and computational study // Mathematical Programming. 2004. Vol. 99(3). PP. 563-591. 9. *Tawarmalani M., Sahinidis N.V.* Convexification and global optimization in continuous and mixed-integer nonlinear programming: theory, algorithms, software, and applications, nonconvex optimization and its applications series, 65. Boston MA: Kluwer Academic Publishers, 2002. 499 p. 10. *Kocis G.R., Grossmann I.E.* Computation Experience with DICOPT: solving MILNP problems in process systems engineering // Computers Chem. Eng. 1989. Vol. 13. PP. 307-315.

Харьковский национальный
университет радиотехники
Харьковский университет Воздушных Сил
имени Ивана Кожедуба

Поступила в редколлегию 15.09.2009