

## ДИНАМИЧЕСКАЯ МАРШРУТИЗАЦИЯ В ПАКЕТНЫХ СЕТЯХ С ГАРАНТИРОВАННЫМ КАЧЕСТВОМ ОБСЛУЖИВАНИЯ

Широкая и повсеместная реализация достижений научно-технического прогресса в области новых информационных технологий значительно повышает уровень требований к современным телекоммуникационным системам за счет увеличения числа пользователей, расширения спектра предоставляемых им услуг связи, а также ужесточения требований к качеству обслуживания. В этой связи все большее число пользователей, наряду с традиционными услугами (передача речи, обмен данными, факсимильная связь и др.), заинтересованы в новых видах услуг, требующих для воплощения широкополосные цифровые каналы. К их числу стоит отнести, прежде всего, видеоконференцсвязь, ТВ, видеотелефон, цветное факсимиле, видеопочту, поиск видеоинформации, большинство из которых являются услугами с комплексным предоставлением информации и в соответствии с рекомендациями ITU-T определяются как мультимедиа.

Основное место в качестве единой транспортной подсистемы, способной адаптивно к изменению уровня требований обеспечить интегральное обслуживание пользователей в рамках вышперечисленных услуг, отводится сетям ATM, реализующих принципы ячеечной коммутации и поддерживающих соединения виртуальных каналов (VCC) и путей (VPC). Неоспоримым достоинством таких сетей является встроенный механизм обеспечения гарантированного качества обслуживания (МОГКО) с требуемыми показателями QoS (например, вероятность обслуживания с заданным качеством, среднее или максимальное время обслуживания и др.), базирующийся на использовании протокола PNNI Phase 1 [1]. Однако относительная дороговизна, а также сложность настройки и сопровождения оборудования сетевой среды ATM привели к необходимости ее совершенствования на аппаратном и программном уровнях и, одновременно, поиску приемлемых решений в области смежных телекоммуникационных технологий.

Бурное продвижение IP-технологий, связанное, прежде всего, с развитием Internet, совершенствованием инфраструктуры глобальных сетей и принятием ряда новых мультимедийных сервисов, например, аудио- и видео- по запросу, IP-телефонии, де-факто определило альтернативное направление развития концепции цифровых сетей интегрального обслуживания [2]. Этот шаг подтверждается некоторым расширением традиционной для IP-сетей дейстаграмной службы с принятием в качестве стандартов протокола передачи видео- и аудиоинформации в реальном масштабе времени RTP и протокола резервирования ресурсов RSVP, направленных на обеспечение заданного уровня QoS для обслуживаемых приложений. Заметное усложнение всегда отличавшихся своей простотой и прозрачностью протоколов IP-сетей создало условия для разработки и принятия эффективных концепций в области конвергенции IP и ATM технологий, базирующихся на реализации схемы IP-over-ATM [3], в рамках которой IP-сеть развертывается поверх высокоскоростной транспортной инфраструктуры канального уровня ATM. Такое решение должно позволить отображать классы IP-трафика на различные сервисные категории сети ATM и осуществлять передачу информации о резервировании ресурсов с уровня IP на уровень ATM. Ярким примером подобного подхода в настоящее время является технология мультипротокольной коммутации меток MPLS, принятая в качестве отраслевого стандарта под эгидой IETF [4] и основанная на принципах разделения функций пересылки пакетов в транспортной подсистеме (уровень ATM) и управления этим процессом (уровень IP).

Таким образом, претворение в жизнь новых телекоммуникационных технологий неразрывно связано с реализацией действенных МОГКО запросов пользователей, которые не должны представлять собой лишь компиляцию ранее известных разработок (RSVP, PNNI). Задачи адекватной оценки доступных ресурсов все еще требуют детальной системной проработки, являясь предметом многочисленных научных и прикладных исследований. Следует отметить, что применение МОГКО не является автономной прикладной задачей, а составляет обеспечивающий фундамент для решения не менее важных взаимосвязанных между собой задач маршрутизации, коммутации и управления в сети. В этой связи в настоящей работе будет предложен подход к решению задач динамической маршрутизации в пакетных сетях, ориентированных на виртуальные соединения, с реализацией МОГКО пользователей. Важность поставленных задач диктует необходимость применения для их решения целостного подхода, базирующегося на использовании системы математических моделей, адекватно описывающих структуру транспортной подсистемы (ТПС) и протекающие в ней процессы информа-

ционного обмена. В случае необходимости, системные решения могут дополняться эвристическими наработками, а не наоборот, что часто имеет место на современном этапе развития систем телекоммуникаций.

Основным требованием к математической модели решения поставленной задачи является необходимость учета особенностей механизмов резервирования сетевых ресурсов (пропускной способности цифровых трактов, емкости буферных устройств коммутаторов транспортной подсистемы, параметров виртуальных соединений и параметров сетевого контроля) на нужды отдельных пользовательских потоков и их взаимозависимости друг от друга.

Структурную модель транспортной подсистемы можно представить в виде ориентированного взвешенного графа  $\Gamma(R, L)$ , множество вершин которого  $R = \{R_i\}$ ,  $(i = \overline{1, N})$  составляют коммутаторы, где  $N$  - общее количество коммутаторов в ТПС, а множество дуг  $L = \{L_{i,j}\}$ ,  $(i, j = \overline{1, N}; i \neq j)$  - цифровые тракты между коммутаторами ТПС.

Учитывая высокую динамичность процессов информационного обмена, адекватно описать функциональную сторону ТПС возможно только на основе соответствующих динамических моделей. Анализ основных подходов к разработке функциональной модели ТПС определил приоритетное использование аппарата дифференциальных или разностных управляемых уравнений состояния [5,6]. Тогда динамику информационного обмена можно представить в виде следующей системы неавтономных уравнений загрузки буферов очередей на коммутаторах ТПС:

$$x_{i,j}(k+1) = x_{i,j}(k) - \sum_{\substack{l=1, \\ l \neq i}}^N b_{i,l}(k) \cdot u_{i,l}^j(k) + \sum_{\substack{m=1, \\ m \neq i,j}}^N b_{m,i}(k) \cdot u_{m,i}^j(k) + y_{i,j}(k), \quad (1)$$

где  $b_{m,i}(k) = c_{m,i}(k) \cdot \Delta t$ ,  $y_{i,j}(k) = \zeta_{i,j}(k) \cdot \Delta t$ ,  $(k=0,1,2,\dots; \Delta t = t_{k+1} - t_k)$ ;  $x_{i,j}(k)$  - объем данных, находящихся на коммутаторе  $R_i$  и предназначенных для передачи коммутатору  $R_j$  в момент времени  $t_k$ , трактуемый в дальнейшем как переменная состояния;  $c_{i,j}(k)$  - скорость передачи данных от коммутатора  $R_i$  к коммутатору  $R_j$  в момент времени  $t_k$  в тракте  $(i,j)$ ;  $u_{i,l}^j(k)$  - доля пропускной способности тракта  $(i,l)$ , выделенная пользовательскому трафику с адресом  $R_j$  в момент времени  $t_k$  и трактуемая в дальнейшем как маршрутная переменная;  $\zeta_{i,j}(k)$  - интенсивность поступления данных на коммутатор  $R_i$  в момент времени  $t_k$  с адресатом  $R_j$  от пользователей сети;  $\Delta t$  - период перерасчета маршрутных переменных.

С целью предотвращения перегрузки элементов ТПС, ввиду ограниченности буферов очередей на коммутаторах и пропускных способностей цифровых трактов передачи, на переменные состояния и маршрутные переменные накладываются ограничения вида

$$0 \leq x_{i,j}(k) \leq x_{i,j}^{\max}; \quad 0 \leq u_{i,l}^j(k); \quad \sum_{n=1}^N u_{i,l}^n(k) \leq 1, \quad (2)$$

где  $x_{i,j}^{\max}$  - емкость буфера очереди для трафиков с адресатом  $R_j$  на коммутаторе  $R_i$ .

С целью описания в рамках предлагаемой функциональной модели ТПС особенностей режима виртуальных соединений, используемого при реализации МОГКО, введем ряд дополнительных обозначений. Пусть также  $R_i^S = \{R_i^S(q)\}$  - множество всех виртуальных путей (маршрутов) доведения с гарантированным качеством пользовательских трафиков от коммутатора  $R_i$  до коммутатора  $R_S$  в рассматриваемой транспортной подсистеме, где  $R_i^S(q) = [R_i, R_j, \dots, R_k, R_S]$  - произвольный  $q$ -й виртуальный путь ВРС длины  $v$ , т.е.  $|R_i^S(q)| = v$ . В соответствии с тем, что одновременно на комму-

таторе  $R_i$  ТПС может обслуживаться целый ряд пользовательских запросов  $M$ , то под  $y_{i,s}^{(n)}$  условимся понимать объем произвольного ( $n$ )-го пользовательского трафика от коммутатора  $R_i$  с адресом  $R_s$ , претендующего на обслуживание с определенным показателем качества QoS, причем  $y_{i,s} = \sum_{n=1}^M y_{i,s}^{(n)}$ ;  $R_i^s(q, n)$  - искомый виртуальный канал VCC передачи ( $n$ )-го пользовательского трафика по  $q$ -му VPC  $R_i^s(q)$ ;  $U_i^s(q, n) = [u_{i,j}^s, u_{j,l}^s, \dots, u_{k,s}^s]$  - вектор загрузки виртуального пути  $R_i^s(q)$  виртуальным каналом  $R_i^s(q, n)$  размерности  $v$ ;

Для наглядности последующего изложения в рамках рассматриваемого трафика ( $n$ ) и рассчитанного виртуального канала нумерацию вершин будем производить относительно исходящей (первой) вершины виртуального пути, ограничиваясь его размерностью, т.е. вышевведенные обозначения примут вид:

$$R_1^v = [R_1, R_2, \dots, R_{v-1}, R_v], U_1^v = [u_{1,2}^v, u_{2,3}^v, \dots, u_{v-1,v}^v], y_{i,s}^{(n)} = y^{(n)}.$$

В качестве параметра QoS выберем максимальное время обслуживания (доставки)  $T_{обсл}^{\max}(n)$  пользовательского трафика ( $n$ ) в транспортной подсистеме, представляющего собой граничную сумму задержек в коммутаторах и трактах ТПС вдоль маршрута доведения. Этот показатель используется применительно к заявкам категорий CBR (постоянная скорость передачи) и rt-VBR (переменная скорость передачи информации в реальном масштабе времени). При решении задачи динамической маршрутизации с гарантированным качеством обслуживания в пакетных сетях, ориентированных на виртуальные соединения, в зависимости от значения  $T_{обсл}^{\max}(n)$  ниже рассмотрим следующие два случая.

В первом случае, пусть максимальное время обслуживания пользовательского трафика  $T_{обсл}^{\max}(n)$  не превышает период перерасчета маршрутных таблиц  $\Delta t$ , т.е. имеет место неравенство

$$T_{обсл}^{\max}(n) \leq \Delta t, \quad (3)$$

т.е. на каждом участке маршрута (тракте  $(i, i+1)$ ) будет выполняться неравенство

$$\frac{y^{(n)}}{u_{i,i+1}^v \cdot c_{i,i+1} \cdot \Delta t} < 1. \quad (4)$$

Тогда в дополнение к ограничениям (2) на маршрутные переменные накладываются комплексные ограничения вида

$$y^{(n)} \sum_{i=1}^{v-1} \frac{1}{u_{i,i+1}^v \cdot c_{i,i+1}} \leq T_{обсл}^{\max}(n). \quad (5)$$

Во втором случае, пусть максимальное время обслуживания пользовательского трафика  $T_{обсл}^{\max}(n)$  превышает период перерасчета содержимого маршрутных таблиц  $\Delta t$ , т.е. имеет место неравенство

$$T_{обсл}^{\max}(n) > \Delta t. \quad (6)$$

Условие (6) гарантирует, что на одном или более трактах рассчитанного виртуального канала условие (4) выполняться не будет, т.е. на этих участках произойдет перерасчет (и, возможно, неодно-

кратно) маршрутных переменных. Количество подобных перерасчетов  $m_{i,i+1}$ , где  $(i, i+1)$ - тракт в рассчитанном виртуальном пути, можно определить исходя из следующего выражения:

$$\frac{1}{m_{i,i+1}} \sum_{k=1}^{m_{i,i+1}} \frac{y_{i,i+1}^{(n)}(k)}{u_{i,i+1}^v(k) \cdot c_{i,i+1} \cdot \Delta t} \leq 1, \quad (7)$$

где  $y_{i,i+1}^{(n)}(k)$  - доля трафика  $y^{(n)}$ , передаваемого на  $k$ -м временном интервале в  $(i, i+1)$ - тракте виртуального канала, для которого равенство  $\sum_{k=1}^{m_{i,i+1}} y_{i,i+1}^{(n)}(k) = y^{(n)}$  гарантирует передачу всего пользовательского трафика  $(n)$  на протяжении всех  $m_{i,i+1} + 1$  интервалов существования маршрутных переменных (таблиц) через цифровой тракт  $(i, i+1)$ .

Результирующее ограничение, которое должно выполняться на всем протяжении рассчитываемого виртуального пути для второго случая (6), принимает вид

$$\sum_{i=1}^{v-1} m_{i-1,i} + m_{i,i+1} \sum_{k=m_{i-1,i}} \frac{y_{i,i+1}^{(n)}(k)}{u_{i,i+1}^v(k) \cdot c_{i,i+1}} \leq T_{обсл}^{\max}(n), \text{ где } m_0 = 1. \quad (8)$$

Нетрудно заметить, что условие (5) представляет собой частный случай условия (8) для  $m_{i,i+1} < 1$ ,  $(i = \overline{1, v-1})$ .

Для нахождения искомых виртуальных путей  $R_i^S(q)$  и каналов  $R_i^S(q, n)$  доведения  $(n)$ -го пользовательского трафика в виде вектора загрузки  $U_i^S(q, n)$  в соответствии с ограничениями (5) или (8) представим систему разностных уравнений состояния (1) в векторно-матричной форме:

$$X(k+1) = X(k) + B(k) \cdot U(k) + Y(k), \quad (9)$$

где  $X(k) = [x_{1,2}(k), \dots, x_{i,j}(k), \dots, x_{N,N-1}(k)]^T$  - вектор загрузки буферных устройств на коммутаторах транспортной подсистемы в момент времени  $t_k$  размерности  $N \cdot (N-1)$ ;  $U(k) = [u_{1,2}^2(k), \dots, u_{i,l}^j(k), \dots, u_{N,N-1}^{N-1}]^T$  - вектор маршрутных переменных размерности  $N \cdot (N-1)^2$ , в соответствии с которыми в ТПС реализуется процесс маршрутизации в момент времени  $t_k$ ;  $B(k)$  - матрица пропускных способностей трактов между коммутаторами ТПС в момент времени  $t_k$  размерности  $N \cdot (N-1) \times N \cdot (N-1)^2$ , а ее элементы формируются в виде линейных комбинаций  $b_{i,j}(k)$ , исходя из выражения (1);  $Y(k) = [y_{1,2}(k), \dots, y_{i,j}(k), \dots, y_{N,N-1}(k)]^T$  - вектор пользовательской нагрузки на коммутаторы ТПС в момент времени  $t_k$  размерности  $N \cdot (N-1)$ .

В качестве критерия оптимальности, проводимой в ТПС маршрутизации информационных потоков, выберем минимум квадратичного функционала вида

$$J = \sum_{k=0}^{a-1} [X^T(k) \cdot Q_X \cdot X(k) + U^T(k) \cdot Q_U \cdot U(k)], \quad (10)$$

где  $Q_X$  - диагональная неотрицательно определенная весовая матрица, определяемая приоритетностью очередей на коммутаторах ТПС;  $Q_U$  - диагональная неотрицательно определенная весовая матрица, определяемая важностью цифровых трактов передачи данных в ТПС;  $a$  - количество интервалов  $\Delta t$ , для которых осуществляется расчет маршрутных переменных.

Физически квадратичный функционал (10) характеризует суммарные затраты по загрузке буферных устройств коммутаторов, пропускных способностей трактов ТПС на протяжении цикла оптимизации  $T = a \cdot \Delta t$  и функционально связан с объемом своевременно переданных пользовательских данных. Формулировка оптимизируемого функционала (10) позволяет реализовать свойство прогнозирования проводимой в ТПС маршрутизации информационных потоков. Причем период прогнозирования совпадает по своему смыслу с величиной интервала оптимизации  $T$ .

На основании проведенного выше структурного и функционального описаний ТПС с помощью соответствующих математических моделей (1-9) выбранного критерия оптимальности процесса маршрутизации (10) решение исходной прикладной задачи по расчету маршрутных переменных можно свести к решению многошаговой оптимизационной задачи по минимизации векторного функционала (10) с учетом динамических ограничений (9), а также ограничений на переменные состояния и маршрутные переменные (2), (8). И хотя наличие перечисленных ограничений существенно затрудняет получение искомого решения в аналитическом виде, для получения необходимых и достаточных условий существования оптимума (10) воспользуемся иерархическо-координационным методом целевой координации [6,7]. Его реализация нацеливает на значительное снижение вычислительных затрат при расчете маршрутных переменных за счет возможности распараллеливания вычислений.

В соответствии с выбранным методом решение исходной задачи минимизации функционала (10) может быть заменено решением двойственной задачи максимизации по  $\lambda$  функционала  $\Phi(\lambda)$ , где

$$\Phi(\lambda) = \min_{x,u} L(X,U,\lambda),$$

$$L(X,U,\lambda) = \sum_{k=0}^{a-1} \left\{ X^T(k) \cdot Q_X \cdot X(k) + U^T(k) \cdot Q_U \cdot U(k) + \lambda^T(k) \cdot [-X(k+1) + X(k) + B(k) \cdot U(k) + Y(k)] \right\}. \quad (11)$$

Многоуровневый характер решения задачи позволяет обеспечить вычисление лагранжиана  $L(X,U,\lambda)$  для заданных на вышестоящем уровне  $\lambda = \lambda^*$  независимо для каждого момента времени, определяемого индексом  $k$ . С этой целью определим гамильтониан

$$H(k) = X^T(k) \cdot Q_X \cdot X(k) + U^T(k) \cdot Q_U \cdot U(k) + \lambda^T [X(k) + B(k) \cdot U(k) + Y(k)]. \quad (12)$$

Тогда с учетом выражения (12) уравнение (11) принимает вид

$$L(X,U,\lambda) = \min_{x,u} \sum_{k=0}^{a-1} \left[ H(k) - \lambda^T(k-1) \cdot X(k) \right] = \sum_{k=0}^{a-1} L_k(\lambda),$$

$$L_k(\lambda) = \min_{x,u} \left[ H(k) - \lambda^T(k-1) \cdot X(k) \right]. \quad (13)$$

Для получения численного решения задачи необходимо рассчитать независимо друг от друга значения функций  $L_k$  для заданных  $\lambda = \lambda^*$  и затем максимизировать  $L$ , используя градиентные процедуры, где градиент функции  $L$  имеет вид

$$\nabla L(\lambda) \Big|_{\lambda = \lambda^*} = -X^*(k+1) + X^*(k) + B(k) \cdot U^*(k) + Y(k). \quad (14)$$

Здесь  $X^*, U^*$  - решения, минимизирующие  $L_k$  при  $\lambda = \lambda^*$ .

Задача минимизации  $L_k$  в выражении (13) при наличии ограничений (2) и (8) в отсутствии динамических ограничений (9) является задачей параметрической оптимизации и решается методами математического программирования.

Таким образом, задача минимизации целевого функционала (10) может быть решена с помощью следующей двухуровневой процедуры. На нижнем иерархическом уровне при заданных  $\lambda = \lambda^*$  ми-

минимизируются вне зависимости друг от друга функции  $L_k$ ,  $k = \overline{0, a-1}$  в выражении (13). Полученные оптимальные решения  $X^*$ ,  $U^*$  используются на верхнем уровне для вычисления градиента (14), который, в свою очередь, применяется для нахождения вектора  $\lambda$ , максимизирующего  $L$ . Общий оптимум функционала  $J$  достигается при близости градиента (14) к нулю.

В результате расчета маршрутных переменных, составляющих содержание искомым маршрутных таблиц, будет получен ответ на следующих два основных вопроса. Во-первых, существует ли решение задачи по гарантированному обслуживанию запросов пользователей. Во-вторых, по какому виртуальному пути будет происходить доведение принятого к обслуживанию трафика и каким образом будет производиться резервирование буферов очередей на коммутаторах и пропускных способностей тактов передачи вдоль рассчитанного VCC.

Достоинства предложенного подхода к решению задач динамической маршрутизации обусловлены преимуществами применяемых моделей и методов решения. Введение понятия вектора загрузки виртуального пути, ограничений (5) или (8) с выбранным в качестве показателя QoS максимальным временем обслуживания  $T_{обсл}^{max}(n)$  пользовательского трафика позволяет сориентировать применяемую модель на учет особенностей маршрутизации в пакетных сетях с виртуальными соединениями и реализованными механизмами обеспечения гарантированного качества обслуживания. Наличие интегральных ограничений на маршрутные переменные (2) позволяет в явном виде обеспечить учет влияния на выбор маршрута доведения уже существующих виртуальных путей и каналов.

Реализация предложенных моделей и процедур решения маршрутных задач может быть проведена как для случая централизованного управления в ТПС, так и на случай распределенного и требует информированности элементов управляющей подсистемы о существующих и создаваемых виртуальных путях и каналах, а также их основных характеристиках: времени существования и задействованных (зарезервированных) ресурсах. Подобная информированность позволяет максимально эффективно использовать свойство прогнозирования принятых моделей, критериев и процедур расчета при нахождении вектора загрузки виртуального пути. Это достигается путем варьирования величин использования пропускных способностей трактов передачи данных за счет рационального использования вновь освободившихся ресурсов, которые на момент расчета искомого VCC были заняты смежными приложениями пользователей.

Как известно [1], для работы протокола PNNI Phase 1 требуется в 10 раз больше процессорного времени, чем для известного протокола OSPF, применяемого в маршрутизаторах IP-сетей. В этой связи неоспоримым достоинством предложенной двухуровневой процедуры расчета маршрутных переменных является возможность распараллеливания вычислений. Это в сумме с использованием многопроцессорных вычислительных комплексов на коммутаторах ТПС позволит значительно сократить расход процессорного времени при доведении особо чувствительных к задержкам пользовательских трафиков, например, с категорией обслуживания rt-VBR.

Предложенный вариант организации процесса динамической маршрутизации с МОГКО не претендует на законченность и может, при необходимости, быть дополнен другими важными показателями QoS, придавая механизму гарантированного обслуживания комплексный характер путем введения дополнительных ограничений на переменные состояния и маршрутные переменные в рамках принятых моделей за счет изменения вида или формы критериев оптимальности маршрутизации в ТПС или применения более эффективных процедур расчета маршрутных переменных.

**Список литературы:** 1. ATM Forum. Private Network-network Interface Specification. Version 1.0, AF-PNNI-0055.000. Math 1996. 2. *Алексеев И.* Интегрированные услуги нового поколения Internet // Сети. Network World: глобальные сети и телекоммуникации. 1999. №10. С.102-108. 3. *Сатовский Б.Л.* MPLS – технология маршрутизации: нового поколения сетей общего пользования // Сети и системы связи. 2001. №3. С.56-65. 4. *Вудс Д.* MPLS: новый регулятор движения на сетевых магистралях // Сети и системы связи. 2000. №12. С.80-83. 5. *Segall A.* The modeling of adaptive routing in data-communications networks. // IEEE Trans. on communications, 1975. Vol. 25. №1. P.85-95. 6. *Лемешко А.В.* Многопроцессорные вычисления при решении задач динамической маршрутизации информационных потоков в сетях передачи данных // Вестник Харьк. гос. политехн. ун-та. 1998. №15. С.95-98. 7. *Сингх М., Титли А.* Системы: декомпозиция, оптимизация и управление. М.: Машиностроение, 1986. 494 с.