УДК 621.396

А.А. Илюхин

СПОСОБ ЭФФЕКТИВНОГО РАСПРЕДЕЛЕНИЯ РЕСУРСА ПРОПУСКНОЙ СПОСОБНОСТИ СПУТНИКОВЫХ СЕТЕЙ ИНТЕРАКТИВНОГО ДОСТУПА

Рассмотрены механизмы статистического резервирования ресурса обратных каналов спутниковыми терминалами в сетях интерактивного доступа. Усовершенствован способ распределения доступной пропускной способности в виде постановки и решения задачи стохастического нелинейного целочисленного программирования. Предложены непараметрические и эвристические процедуры, позволяющие повысить вычислительную эффективность и качество решения оптимизационной задачи.

Ключевые слова: *спутниковая сеть, интерактивный доступ, пропускная способность, динамическое резервирование, стохастическое программирование.*

Введение. Интерактивные мультимедийные спутниковые сети, соответствующие спецификации стандартов DVB-RCS (ETSI EN 301 790 v1.3.1, 2003 г.), IPoS (TIA-1008, 2003 г.), становятся весьма привлекательным и высокоэффективным решением для ведомств, крупных предприятий и компаний с территориально-распределенной инфраструктурой, крайне заинтересованных в снижении затрат на оплату услуг связи и предпочитающих создавать собственные более экономичные технологические сети связи.

Реализованные в таких сетях механизмы интерактивного доступа (рисунок 1) основаны на периодической передаче спутниковыми терминалами (СТ) в моменты времени t_1, t_5,.... запросов на резервирование пропускной способности обратных (СТ – HUB) каналов, учитывающих текущее состояние очередей буферов различных классов сервиса (голосовые приложения, передача данных, видеоконференцсвязь и т. д.). HUB-станция, анализируя доступный ресурс пропускной способности, формирует частотно-временной план (ЧВП), регламентирующий объем и порядок передачи терминалами пачек пакетных данных на длительности очередного фрейма. Спустя некоторое время планирования, ЧВП циркулярно рассылается получившим регистрацию активным спутниковым терминалам. Длительность процедуры распределения (планирования) ресурсов $[t_2, t_3]$ в реализованных современных VSAT-технологиях достигает 100…200 мс и с учетом времени распространения сигнала 500 мс существенно увеличивает реакцию сети на обработку запросов динамического резервирования.

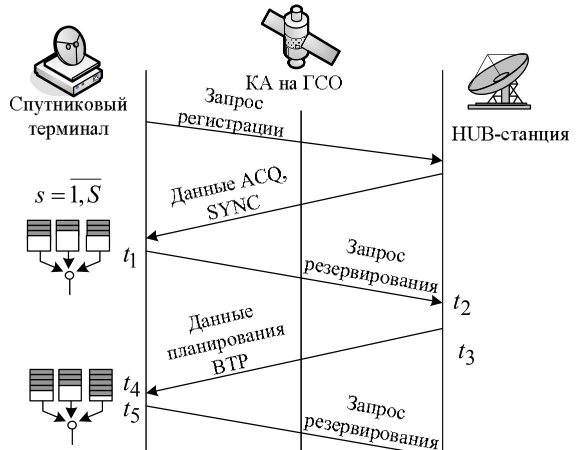

Рисунок 1 – Временная диаграмма процесса резервирования пропускной способности

Большинство предлагаемых способов планирования заключается в нахождении вида интегральной функции качества сетевого обслуживания по всем классам сервиса от распределяемых ресурсов и непосредственном решении известными методами задачи нелинейного целочисленного программирования [1–4]. Ряд работ [5, 6] предполагает включение в интегральную функцию качества прогнозных оценок состояния очередей либо оцениваемых параметров известных статистических моделей потоков. Уровень вычислительной сложности предлагаемых алгоритмов, особенно при решении задач большой размерности, не соответствует условиям реализации в реальном времени, что не позволило довести их до практического применения в современных VSAT-технологиях.

Отсутствие математически строгого решения подобных задач, обладающего приемлемой для реализации в реальном времени вычислительной сложностью, исключающего использование известных статистических моделей потоков данных и допускающего применение механизмов обеспечения дифференцированных требований к качеству сервиса QoS (Quality of Service) для различных приложений, обусловливает необходимость совершенствования методов их решения.

Постановка задачи. Общедоступный частотно-временной ресурс обратных каналов согласно спецификации стандарта DVB-RCS разделяется на фреймы (рисунок 2), каждый фрейм, в свою очередь, – на частотно-временные слоты.

Передача пакетного трафика спутниковыми терминалами осуществляется по одному или последовательно по нескольким частотным каналам на временных позициях трафиковых тайм-слотов. Последовательность тайм-слотов в пределах фрейма на каждой несущей имеет следующий порядок: тайм-слот CSC общеканальной сигнализации, тайм-слот ACQ частотной коррекции, тайм-слот SYNC синхронизации и n_{trf} трафиковых тайм-слотов для передачи пользовательских данных.

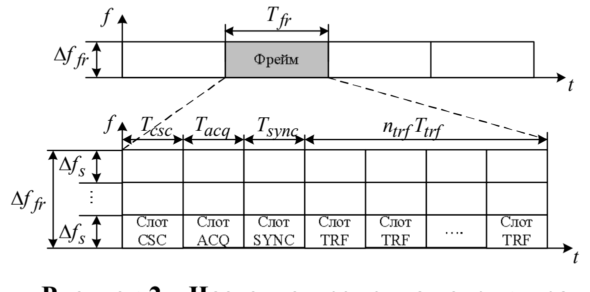

Рисунок 2 – Частотно-временная структура фрейма

Распределение доступных трафиковых тайм-слотов в очередном k-ом фрейме

$$M(k) = \Delta f_{fr} / \Delta f_s \cdot n_{trf} - a(k),$$

где $\Delta f_{fr} / \Delta f_s$ – количество частотных несущих на фрейм; $a(k)$ – число слотов, резервированных запросами предыдущих фреймов, осуществляется с учетом известного либо прогнозируемого статистического распределения количества протокольных блоков данных (PDU – Protocols Data Unit) в очередях буферов MAC-уровня активных спутниковых терминалов $f_s^I : Z^+ \cup \{0\}$. В общем случае $f_s^I(y) > 0$, $\forall y \in [0, \bar{y}_s]$, $s = \overline{1,S}$, где $\bar{y}_s$ – максимально возможное число PDU соответствующего класса сервиса.

Вектор переменных решения $\vec{x} = (x_1,....,x_S)^T$, заданных на множестве допустимых целочисленных значений $X^I = \{x_s \mid x_s \geq 0\}$, $\forall s = \overline{1,S}$, определяет количество обслуживаемых PDU по всем очередям классов сервиса. Для обслуживания одного PDU соответствующего класса сервиса HUB-станцией выделяется необходимый объем трафиковых тайм-слотов $\vec{d} = (d_1,....,d_S)^T$.

Уровень требований к обеспечению дифференцированного качества QoS (Quality of Service) по классам сервиса устанавливается вектором удельных весовых коэффициентов (штрафов) $\vec{\varpi} = (\varpi_1,....,\varpi_S)^T$ за отказ в обслуживании одного PDU соответствующего класса сервиса s.

Формирование частотно-временной структуры фрейма с резервированными слотами и ее циркулярная рассылка в виде сервисных таблиц состава фрейма (FCT) и структуры слотов (TCT) осуществляются HUB-станцией сети для всех активных спутниковых терминалов.

Решение оптимизационной задачи заключается в достижении наиболее высокого уровня обеспеченности обслуживанием потоков данных активных спутниковых терминалов с учетом дифференцированных требований к качеству их сервиса путем минимизации интегральной штрафной функцией $Q(\vec{x})$.

Например, для класса сервиса s при наличии в буфере y PDU и выделении тайм-слотов под обслуживание всего лишь x_s PDU весовая функция штрафа выражается как

$$q_s(y, x_s) = \varpi_s \max\{0, y - x_s\}. \qquad (1)$$

Тогда ожидаемый штраф для класса сервиса s в случае известной дискретной (гистограммной) плотности распределения f_s^I, когда обеспечены тайм-слотами x_s PDU, выражается функцией

$$Q_s^I(x_s) = \sum_{y=x_s}^{\infty} q_s(y, x_s) f_s^I(y) = \varpi_s \sum_{y=x_s}^{\infty} (y - x_s) f_s^I(y), \qquad (2)$$

а общий ожидаемый штраф $Q(\vec{x}) = \sum_s Q_s(x_s)$.

Общее количество назначаемых трафиковых тайм-слотов всем классам сервиса должно быть не более числа доступных тайм-слотов во фрейм

$$H(\vec{x}) \leq 0, \qquad (3)$$

где $H(\vec{x}) = \vec{x}^T \vec{d} - \grave{I}$.

С учетом ограничения (3) задача распределения тайм-слотов между потоками всех классов сервиса активных спутниковых терминалов математически формализуется в следующем виде:

$$min Q(\vec{x}, f_s^I(y), s),\ H(\vec{x}) \le 0,\ \vec{x} = (x_1, ..., x_S),\ \vec{x} \in \mathbf{X}^I. \quad (4)$$

Задача относится к классу задач стохастического нелинейного целочисленного программирования. Известные методы, применяемые к задачам такого класса, не позволяют достичь приемлемой для решения в реальном времени вычислительной сложности и приводят к длительным процедурам поиска оптимальных решений [7]. Для упрощения решения предлагается привести задачу целочисленного программирования к задаче нелинейной безусловной оптимизации с ослаблением ограничения целочисленности переменных до непрерывных значений и заключительного этапа поиска близких к оптимальным целочисленных решений на основе эвристических процедур.

Однако ввиду недифференцируемости функции $f_s^I(y)$ целочисленных переменных для применения градиентных процедур методов нелинейной оптимизации проведем ее замену некоторой аппроксимирующей функцией $f_s : R^+ \cup \{0\}$ плотности вероятности числа PDU, допускающей условие двойной дифференцируемости $f_s(y)$ в интервале $(0, \infty)$.

В условиях непараметрической априорной неопределенности относительно вида распределения количества PDU на основе наблюдаемой в дискретные моменты времени однородной выборки $Y_s = \{y(t_1), ..., y(t_N)\}$ по каждому классу сервиса $\forall s = \overline{1, S}$ могут быть сформированы регрессионные оценки плотности вероятности [8]

$$\hat{f}_s(y) = \frac{\sum_{n=1}^{N} K\left(\frac{y - y(t_n)}{\alpha}\right)}{N\alpha}, \quad (5)$$

где $K(\alpha)$ – ядро, удовлетворяющее ряду известных свойств [8]; α – параметр сглаживания, значение которого может быть определено из условия максимизации функции правдоподобия, определяемой по методу перекрестной проверки

$$\Lambda(\alpha) = \prod_{n=1}^{N} \hat{f}(y(t_n)), \quad (6)$$

где $\hat{f}(y(t_n)) = \dfrac{\sum_{\substack{i=1 \\ i \ne n}}^{N} K\left(\frac{y - y(t_i)}{\alpha}\right)}{(N-1)\alpha}$.

Значения элементов выборки $Y_s = \{y(t_1), ..., y(t_N)\}, \forall s = \overline{1, S}$ обновляются в ходе реализации процедуры скользящего окна.

Представим составляющие (2) исходной целевой функции с целочисленными переменными в форме записи с действительными переменными

$$Q_s(x_s) = \varpi_s \int_{x_s}^{\infty} (y - x_s) \hat{f}_s(y) dy \quad (7)$$

и проведем анализ свойств задачи нелинейной оптимизации, использующей непараметрические регрессионные оценки плотности.

Анализ свойств задачи нелинейной оптимизации

Утверждение 1. Пусть на выпуклом множестве $\mathbf{X} \subset R^n$ существует некоторая функция f : $\mathbf{X} \mapsto R$, дважды непрерывно дифференцируемая по $\mathbf{X}$. Если матрица Гессе $\nabla^2 f(\mathbf{x})$ является положительно полуопределенной, то $f(\mathbf{x})$ выпукла.

Утверждение 2. Интегральная функция $Q(\mathbf{x})$ s переменных с непараметрической регрессионной оценкой плотности распределения вероятности $\hat{f}_s(y)$, $s = \overline{1, S}$ является выпуклой по множеству X.

Доказательство. Определим первые и вторые производные функции $Q_s(x_s)$ по x_s в интервале $(0, \infty)$ для всех классов сервиса $s = \overline{1, S}$

$$\frac{dQ_s(x_s)}{dx_s} = \varpi_s \left(\int_0^{x_s} \hat{f}_s(y) dy - 1 \right), \quad (8)$$
$$\frac{d^2 Q_s(x_s)}{dx_s^2} = \varpi_s \hat{f}_s(x_s).$$

Таким образом, матрица первых производных и матрица Гессе для $\mathbf{Q}(\mathbf{x})$ примут вид:

$$\nabla_{\mathbf{x}} Q(\mathbf{x}) = \begin{bmatrix} \varpi_1 \left(\int_0^{x_1} \hat{f}_1(y) dy - 1 \right) \\ \cdots\cdots\cdots\cdots \\ \varpi_S \left(\int_0^{x_S} \hat{f}_S(y) dy - 1 \right) \end{bmatrix}, \quad (9)$$
$$\nabla^2_{\mathbf{xx}} Q(\mathbf{x}) = \begin{bmatrix} \varpi_1 \hat{f}_1(x_1) & & \\ & \cdots\cdots & \\ & & \varpi_s \hat{f}_s(x_s) \end{bmatrix},$$

где недиагональные элементы $\nabla^2_{\mathbf{xx}} Q(\mathbf{x})$ равны нулю.

Поскольку $\hat{f}_s(x_s) \ge 0$, $\forall s = \overline{1, S}$, то матрица

Гессе $\nabla^2_{\mathbf{xx}}Q(\mathbf{x})$ положительно полуопределена. Утверждение доказано.

Следствие 1. Поскольку функция $\hat{f}_s(x_s)$ положительна ($\hat{f}_s(x_s)>0$) для всех $x_s>0$, $\forall s=\overline{1,S}$, то интегральная функция $Q(\mathbf{x})$ строго выпукла.

Доказательство. Если функция $\hat{f}_s(x_s)>0$ для всех $x_s>0$, $s=\overline{1,S}$, то получаем выполнение условия положительной определенности матрицы Гессе $\nabla^2_{\mathbf{xx}}Q(\mathbf{x})$ для функции

$$\mathbf{y}^T\nabla^2_{\mathbf{xx}}Q(\mathbf{x})\mathbf{y}=\sum_s y_s^2\varpi_s f_s(x_s)>0,\ \forall\mathbf{y}\neq 0,\quad (10)$$

где $\mathbf{y}=(y_1,y_2,...,y_s)$. Следствие доказано.

Следствие 2. Интегральная функция $Q(\mathbf{x})$ устойчиво выпукла по $\overline{\mathbf{X}}$.

Следствие 3. Ограничивающее неравенство $H(\mathbf{x})\leq 0$ является связывающим при оптимальном решении $\mathbf{x}^*\in\mathbf{X}$.

Доказательство. Из полученных матриц (9) очевидно, что $Q_s(x)$ является убывающей функцией от x для всех $s=\overline{1,S}$. Исходя из этого свойства целевой функции при $\grave{I}\rightarrow\infty$, задача имеет неограниченное множество решений $\mathbf{x}^*\rightarrow(\infty,...,\infty)$. Таким образом, введенное ограничение $H(\mathbf{x})\leq 0$ порождает ограничение и множества допустимых решений. Отметим, что ограничивающее неравенство $H(\tilde{\mathbf{o}})\leq 0$ может не быть связывающим при оптимальном целочисленном решении $\mathbf{x}^*\in\mathbf{X}^I$.

Способ решения задачи целочисленного программирования. Для решения задачи предлагается использовать метод множителей Лагранжа, позволяющий привести задачу условной оптимизации к безусловной в виде функции Лагранжа

$$L(\mathbf{x},\lambda)=Q(\mathbf{x})+\lambda H(\mathbf{x}),\ \lambda\geq 0.\quad (11)$$

Утверждение 3. Допустим $\tilde{o}^*$ и λ^* удовлетворяют следующим условиям:

$$\nabla_{\mathbf{x}}L(\mathbf{x}^*,\lambda^*)=0,\quad (12)$$

$$\nabla_{\lambda}L(\mathbf{x}^*,\lambda^*)=0,\quad (13)$$

$$\mathbf{y}^T\nabla^2_{\mathbf{xx}}L(\mathbf{x}^*,\lambda^*)\mathbf{y}>0\quad (14)$$

для всех $\mathbf{y}\neq\mathbf{0}$ с $\mathbf{y}\in\left\{\mathbf{y}\,|\,\nabla H(\mathbf{x}^*)^T\mathbf{y}=\mathbf{0}\right\}$, тогда $\mathbf{x}^*$ является устойчивым локальным минимумом.

Доказательство. Функция Лагранжа, представленная выражением (11), является выпуклой. В этом случае с учетом полученных первых и вторых производных (8) матрица Гессе $\nabla^2_{\mathbf{xx}}L(\mathbf{x}^*,\lambda^*)$ для Лагранжевой функции также является положительно определенной. Таким образом, найденные значения $\mathbf{x}^*$ и λ^*, удовлетворяющие условиям (12) и (13), обеспечивают устойчивый локальный минимум функции $L(\mathbf{x},\lambda)$, который означает также, что $\mathbf{x}^*$ обеспечивает устойчивость локального минимума $Q(\mathbf{x})$ с ограничением $H(\mathbf{x})\leq 0$.

Оптимальные решения $(\mathbf{x}^*\in\mathbf{X}^I,\lambda^*)$ для функционала (11) вычисляются в два этапа. На первом этапе определяем действительные решения $\mathbf{x}^*$ и λ^*, удовлетворяющие условиям (12–14)

$$\nabla_{\mathbf{x}}L(\mathbf{x},\lambda)=\begin{bmatrix}\varpi_1\left(\int_0^{x_1}\hat{f}_1(y)dy-1\right)+\lambda d_1\\ \dots\dots\dots\\ \varpi_s\left(\int_0^{x_s}\hat{f}_s(y)dy-1\right)+\lambda d_s\end{bmatrix},\quad (15)$$

$$\nabla_{\lambda}L(\mathbf{x},\lambda)=H(\mathbf{x}),\ \nabla^2_{\mathbf{xx}}L(\mathbf{x},\lambda)=\nabla^2_{\mathbf{xx}}Q(\mathbf{x}).\quad (16)$$

Из условия (12) и выражения (15) получаем

$$\mathbf{x}^*=\begin{bmatrix}F_1^{-1}\left(1-\dfrac{\lambda^* d_1}{\varpi_1}\right)\\ \dots\dots\dots\\ F_s^{-1}\left(1-\dfrac{\lambda^* d_s}{\varpi_s}\right)\end{bmatrix},\quad (17)$$

где $F_s^{-1}(y)=\inf\{x\,|\,F_s(x)=y\}$, $F_s(x)=\int_0^x\hat{f}_s(y)dy$, а λ^* может быть получено подстановкой (17) в (16) для $\nabla_{\lambda}L(\boldsymbol{x},\lambda)$

$$\Psi(\lambda)=\sum_{s=1}^{S}d_sF_s^{-1}\left(1-\frac{\lambda d_s}{\varpi_s}\right).\quad (18)$$

Из условия $\Psi(\lambda)=M$ получаем

$$\lambda^*=\Psi^{-1}(M),\quad (19)$$

где $\Psi^{-1}(y)=\inf\{x\,|\,\Psi_s(x)=y\}$.

Из (17, 19) окончательно получаем

$$\mathbf{x}^*=\begin{bmatrix}F_1^{-1}\left(1-\dfrac{d_1}{\varpi_1}\Psi^{-1}(M)\right)\\ \dots\dots\dots\\ F_s^{-1}\left(1-\dfrac{d_s}{\varpi_s}\Psi^{-1}(M)\right)\end{bmatrix}.\quad (20)$$

На втором этапе с использованием в качестве начальных оптимальных решений первого этапа определяем ближайшие целочисленные решения. Для эффективного поиска близких к оптимальным решений предлагается использовать эвристические процедуры, имеющие низкую вычислительную сложность.

1. Определяем множество номеров классов сервиса с нецелочисленными решениями $A = \left\{s \mid x_s^* > [x_s^*], s = \overline{1,S}\right\}$. Если множество A пустое, т. е. $x_s^* = [x_s^*]$, $s = \overline{1,S}$, то решение найдено, в противном случае вычисляем освободившийся ресурс тайм-слотов $a = \sum_{s \in A} \left(x_s^* - [x_s^*]\right) d_s$ вследствие ограничений на целочисленность.

2. Для всех номеров классов сервиса $s \in A$ вычисляем коэффициент чувствительности частных штрафных функций относительно добавляемой разности к ближайшей целочисленной переменной

$$b_s = \frac{Q_s\left(x_s^*\right) - Q_s\left([x_s^*] + 1\right)}{d_s\left([x_s^*] + 1 - x_s^*\right)}.$$

3. Определяем класс сервиса с наибольшим коэффициентом чувствительности $j = \arg\max\left\{b_s, s = \overline{1,S}\right\}$. Контролируя условие $a \geq d_j$, обновляем значения $x_j^* = [x_j^*] + 1$, $a = a - d_j$, $A = A - \{j\}$ до тех пор, пока $A = 0$.

Анализ эффективности и результаты моделирования. Эффективность предлагаемого способа распределения ресурса пропускной способности оценивалась по критерию относительной близости оптимального Q^* и полученного эвристической процедурой значений интегральной целевой функции Q

$$\eta = \frac{Q - Q^*}{Q^*} (\%) \tag{21}$$

методом статистических испытаний.

Исходные данные для моделирования. В составе спутниковой сети интерактивного доступа спутниковые терминалы, представленные потоками данных четырех классов сервиса $\vec{s} = (1, 2, ..., 4)$, осуществляют работу через HUB-станцию. Распределение количества PDU в очередях по каждому классу сервиса в пределах фрейма моделируется экспоненциальным законом с вектором параметров $\vec{\mu}_s$, изменяемых от фрейма к фрейму случайным образом, равномерно в диапазоне (4, 10) [PDU/фрейм]. Для аппроксимации плотности распределения используется непараметрическая регрессионная оценка с ядром Епанечникова

$$K(z) = \begin{cases} \frac{3}{4}(1 - z)^2 \text{ при } |z| \leq 1; \\ 0, \text{ при } |z| > 1, \end{cases} \tag{22}$$

и $\alpha = 3$. Затраты ресурсов и уровень требований к качеству обслуживания представлены векторами $\vec{d}$ = (8, 6, 4, 2) [тайм-слотов/PDU] и $\vec{\varpi}$ = (4, 3, 2, 1) соответственно. Изменение доступного ресурса пропускной способности M [тайм-слотов] в каждом фрейме моделировалось равномерным законом в диапазоне (100, 200). Результаты моделирования на длительности 200 фреймов представлены на рисунках 3, 4.

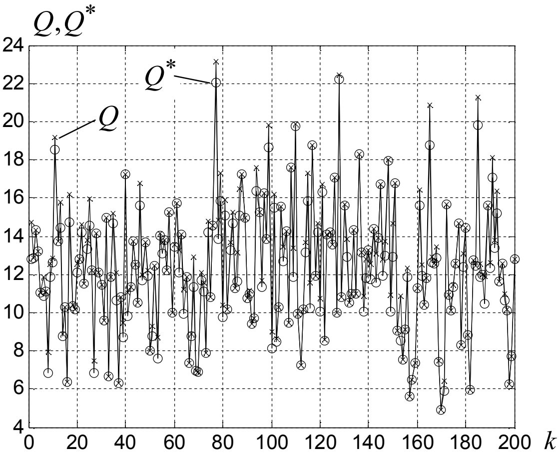

Рисунок 3 – Соотношение оптимальных и полученных эвристическим решением значений интегральной целевой функции

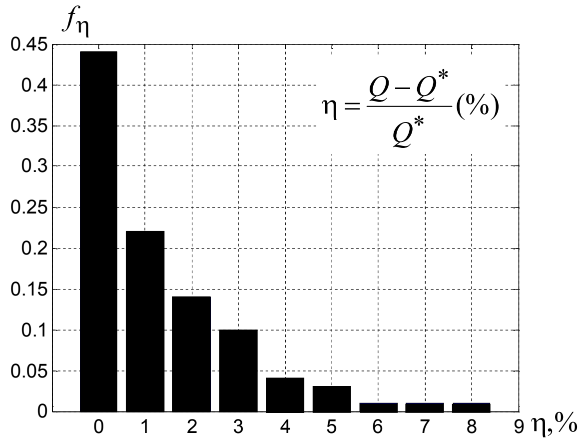

Рисунок 4 – Плотность распределения показателя эффективности

Анализ результатов показывает, что в 44 % от общего количества фреймов полученные предлагаемым алгоритмом решения соответствуют оптимальным значениям интегральной целевой функции Q^*. Достаточно высокая степень близости по показателю $\eta \leq 5$ % обеспечивается в 90 % тестовых задач. Использование непараметрических регрессионных оценок позволяет с приемлемой точностью на малых объемах выборки аппроксимировать плотность распределения количества PDU в буфере s-класса сервиса с учетом ее нестационарного изменения.

Заключение. Предлагаемый способ основан на процедуре взвешенного распределения пропускной способности частотно-временного плана между спутниковыми терминалами, учитывающего уровень дифференцированных требований к качеству обслуживания источников мультимедийных данных. Динамический и стохастический характер поступления пакетных данных в буферы различных классов сервиса СТ и инерционность процесса формирования ЧВП обусловили необходимость применения непараметрических регрессионных оценок плотности распределения количества PDU и упрощения вычислительных процедур решения задачи целочисленного нелинейного программирования. Результаты моделирования алгоритма распределения пропускной способности на примере 200 тестовых задач подтверждают возможность получения близких к оптимальным решений на уровне показателя эффективности $\eta \leq 5$ % в 90 % случаях при среднем времени вычисления не более 20 мс.

Библиографический список

1. *Neale J., Green R., Landovskis J.* Interactive channel for multimedia satellite networks, IEEE Commun. Mag., PP. 192-198, Mar. 2001.

2. *Lee K.-D., Cho Y.-H., Lee S. J., Lee H.-J.* Optimal design of su-perframe pattern for DVB-RCS return link, ETRIJ., vol. 24, no. 3, PP. 251-254, 2002.

3. *Lee K.-D., Cho Y.-H., Lee H.-J., Oh D. G.* Improving efficiency of timeslot assignment for nonrealtime data in a DVB-RCS return link: Modeling and algorithm, ETRIJ., vol. 25, no. 4, 2003.

4. *Lee K.-D., Lee H.-J., Cho Y.-H., Oh D. G.* Throughput-maximizing timeslot scheduling for interactive satellite multi-class services, IEEE Commun. Lett., vol. 7, PP. 263-265, June 2003.

5. *Zhang T., Berg E., Chennikara J., Agrawal P., Chen J.-C., Ko-dama T.* Local predictive resource reservation for handoff in multimedia wireless IP networks, IEEE J. Select. Areas Commun., vol. 19, PP. 1931-1941, Oct. 2001.

6. *Leung K. K., Massey W. A., Whitt W.* Traffic models for wireless communication networks, IEEE J. Select. Areas Commun., vol. 12, Oct. 1994.

7. *Химмельблау Д.* Прикладное нелинейное программирование / под ред. М. Л. Быховского. - М.: Мир, 1975. - 534 с., ил.

8. *Хардле В.* Прикладная непараметрическая регрессия. - М.: Мир, 1993. - 349 с.