

Моделирование процессов обслуживания абонентов в сети передачи данных на базе космических аппаратов на низкой круговой орбите. I

А.А. Маслов*, Г.В. Себекин*, М.С. Степанов**, С.Н. Степанов**,
А.О. Щурков*,

* *Московский физико-технический институт (национальный исследовательский университет),
Москва*

** *Московский технический университет связи и информатики, Москва*

Поступила в редколлегию 06.11.2024 г. Принята 07.12.2024 г.

Аннотация—Управление процессами обслуживания мультисервисного трафика в сетях передачи данных на базе космических аппаратов на низкой круговой орбите является сложной задачей, требующей моделирования с учетом реальных условий функционирования сетей. В статье разработаны аналитическая и имитационная модели такой сети для трудного с точки зрения реализации требований по качеству предоставления сервисов случая концентрации большого количества абонентов на относительно небольшом участке. Аналитическая модель строится на основе многомерных ступенчатых марковских процессов. Имитационная модель позволяет проверить приемлемость допущений, принятых при аналитическом моделировании, и оценить показатели качества при большом количестве состояний сети. Модели могут быть использованы для оценки возможностей сетей по обслуживанию районов высокой концентрации абонентов.

КЛЮЧЕВЫЕ СЛОВА: космические аппараты на низкой круговой орбите, каналный ресурс, мультисервисный трафик, трафик реального времени, эластичный трафик, абонентские терминалы.

DOI: 10.53921/18195822_2024_24_4_335

1. ВВЕДЕНИЕ

Рассматривается сеть спутниковой связи (ССС) на базе низколетящих космических аппаратов (КА), в которой обеспечивается обмен данными между абонентскими терминалами (АТ) и региональной станцией (РС), сопряженной с наземными телекоммуникационными сетями, путем ретрансляции сигналов АТ линейными КА либо непосредственно на РС (примерами могут служить системы Globalstar [1], «Марафон-IoT» [2]), либо на РС через геостационарные КА (система «Гонец» при обеспечении ретрансляции через КА «Луч» [3]). Обобщенная схема сети представлена на рисунке 1.

Хотя число космических аппаратов (КА) в орбитальной группировке (ОГ) может составлять сотни и даже тысячи, абоненту для получения услуг по широкополосному доступу в Интернет и обмену данными в конкретной точке в текущий момент времени доступны лишь те КА, которые находятся в его зоне радиовидимости. Это количество изменяется во времени от единиц до десятков и зависит от координат абонента, параметров ОГ, характеристик КА и АТ. Естественно, на ресурс пропускной способности КА претендуют все активные АТ в окрестности конкретного абонента, и потому в районах с высокой плотностью АТ возможно снижение качества обслуживания трафика. Более того, гарантированные в СССР показатели QoS определяются значениями в регионах с высокой концентрацией абонентов. Влияние фактора высокой

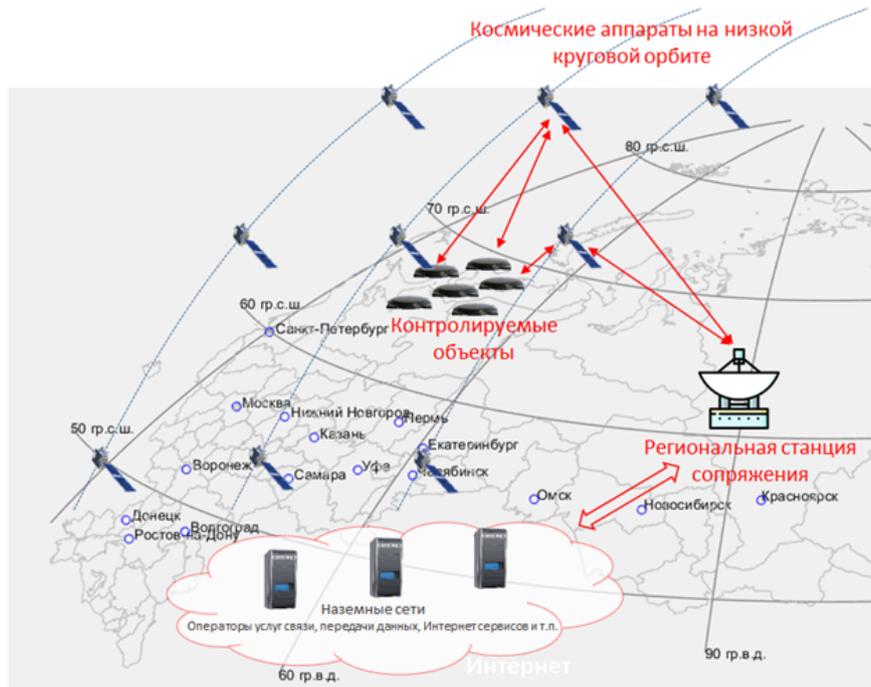


Рис. 1. Обобщенная схема ССС

концентрации на показатели QoS необходимо учитывать, как при проектировании ССС, так и в процессе ее эксплуатации. В частности, при проектировании сети исходными данными являются группы потенциальных потребителей, требуемые услуги, особенности размещения потребителей по регионам. В ходе эксплуатации оператор должен отслеживать возможности предоставления услуг с учетом выполнения нормативов по показателям QoS. Для этого он накапливает данные по статистике работы абонентов, прогнозирует численность сети и трафик по каждой услуге. Таким образом, и на этапе проектирования сети, и на этапе эксплуатации необходим математический аппарат, обеспечивающий оценку показателей QoS в зависимости от ожидаемых характеристик мультисервисного трафика, его пространственной неравномерности, характеристик ОГ, КА и АТ.

С учетом перечисленных обстоятельств целью настоящей статьи является моделирование процессов функционирования ССС в регионе с высокой концентрацией абонентов, позволяющее оценить выполнение требований по показателям QoS и выбрать предпочтительные параметры обслуживания трафика каждого типа. При этом под показателями QoS понимается доля необслуженных заявок каждого типа и среднее время доставки файлов [4]. Ресурс, как правило, измеряется целым числом единиц скорости передачи информации, выраженной в битах в секунду [4, 5].

Из соображений качественного характера можно предположить, что самым неудобным с точки зрения реализации нормативов по качеству обслуживания трафика является вариант концентрации всех АТ в маленькой области (практически в одной точке). Если бы, к примеру, все абоненты ССС находились в одном месте, то для реализации требований по QoS потребовалось бы количество КА, многократно превышающее их необходимое количество в случае относительно равномерного распределения. Подобный подход предлагается применить и для отдельного региона с высокой концентрацией АТ. Как показывает анализ, в регионе есть некоторое место наиболее высокой плотности размещения АТ (например, какие-то месторождения, большой, по меркам региона, населенный пункт), и при удалении от него плотность

размещения АТ снижается. Ведь две трети территории РФ имеют плотность населения менее 3 чел/км² [6]. Соответственно будем считать, что реальное размещение АТ можно для пессимистической оценки показателей QoS заменить на размещение всех АТ в ограниченной области высокой концентрации АТ, и полагать, что остальная часть площади региона не содержит АТ, или поток оттуда очень мал. Это позволяет допустить, что все АТ региона обслуживаются в каждый момент времени одним и тем же множеством КА. С учетом длительности сессий связи АТ, измеряемой минутами, и скорости движения КА по орбите, измеряемой километрами в секунду, можно считать, что диаметр ограниченной области высокой концентрации может составлять до 50-100 км.

Моделированию процессов обслуживания мультисервисного трафика, включающего несколько типов трафика реального времени и эластичный трафик данных, посвящен ряд работ, например, [4, 7–16], среди которых есть и работы, относящиеся к спутниковым сетям [14–16]. Общий подход к построению аналитических моделей предусматривает принятие предположения о пуассоновском характере входных потоков заявок, экспоненциальном распределении времен обслуживания и объемов передаваемых файлов, что позволяет строить модели на основе многомерных ступенчатых марковских процессов, которые учитывают основную специфику анализируемых систем. Оценки показателей качества основываются на численном решении систем уравнений равновесия (СУР) с числом переменных, составляющем тысячи и даже миллионы. При этом сходимость результатов численных решений исследуется косвенными методами на основе анализа близости последовательных приближений и выполнения известных теоретических соотношений [4]. Однако, в этих работах не учитывается специфика ССС на базе низколетящих КА, состоящая в стохастическом изменении доступного для обмена информацией канального ресурса. Учет фактора изменения доступного ресурса осуществляется в ряде работ либо в виде обслуживания ненадежными приборами или систем массового обслуживания, функционирующих в случайной среде [17–21]. Здесь, однако, оцениваются показатели, не относящиеся к анализируемой сети, не учитываются перечисленные выше факторы, например, характеристики ОГ, либо рассматриваются моносервисные потоки заявок. Чтобы остаться в рамках общего подхода к разработке аналитической модели в настоящей статье делается предположение о допустимости замены потока пролетов КА над абонентами марковским процессом с эквивалентными параметрами в части вероятностей состояний и интенсивностей переходов между состояниями. Проверка допустимости такой замены требует, в свою очередь, разработки имитационной модели, которая, кроме того, необходима и для оценки показателей качества при большом количестве переменных в СУР, с которым «не справляются» аналитические модели. Для достижения поставленной цели в разделе 2 дается краткое описание сети с учетом принятых предположений. В разделе 3 проводится оценка зависимостей количества доступных абонентам КА от времени. В разделе 4 определено пространство состояний сети, а в разделе 5 формируется и решается система уравнений равновесия для сети, а также приводятся соотношения для расчета показателей качества. Раздел 6 посвящен имитационному моделированию процессов функционирования сети.

2. ОПИСАНИЕ СЕТИ

Полагается, что орбитальное построение ССС предусматривает использование некоторого количества равномерно разнесенных колец (орбитальных плоскостей) с одинаковым наклоном и одним и тем же числом КА в каждом кольце, которые распределены по кольцу через равные промежутки (рисунок 2). Пусть общее количество КА составляет $N_{\text{ка}}$ и каждому КА присвоен номер $n_{\text{ка}} = 1, 2, \dots, N_{\text{ка}}$. АТ размещены в ограниченной области диаметром до 50-100 км и обслуживаются в каждый момент времени одним и тем же множеством КА. Ресурс одного КА примем равным $v_{\text{ка}}$ единиц канальной скорости. Соответственно, ресурс, который

может использоваться для предоставления сервисов абонентам в рассматриваемой области, равен $v_{ка}n(t)$, где $n(t)$ - количество доступных КА в момент времени t , изменяющееся в диапазоне от n_{min} до n_{max} . Будем считать, что используемые в сети протоколы обмена данными и управления сетью позволяют оперативно перераспределять ресурсы между активными АТ.

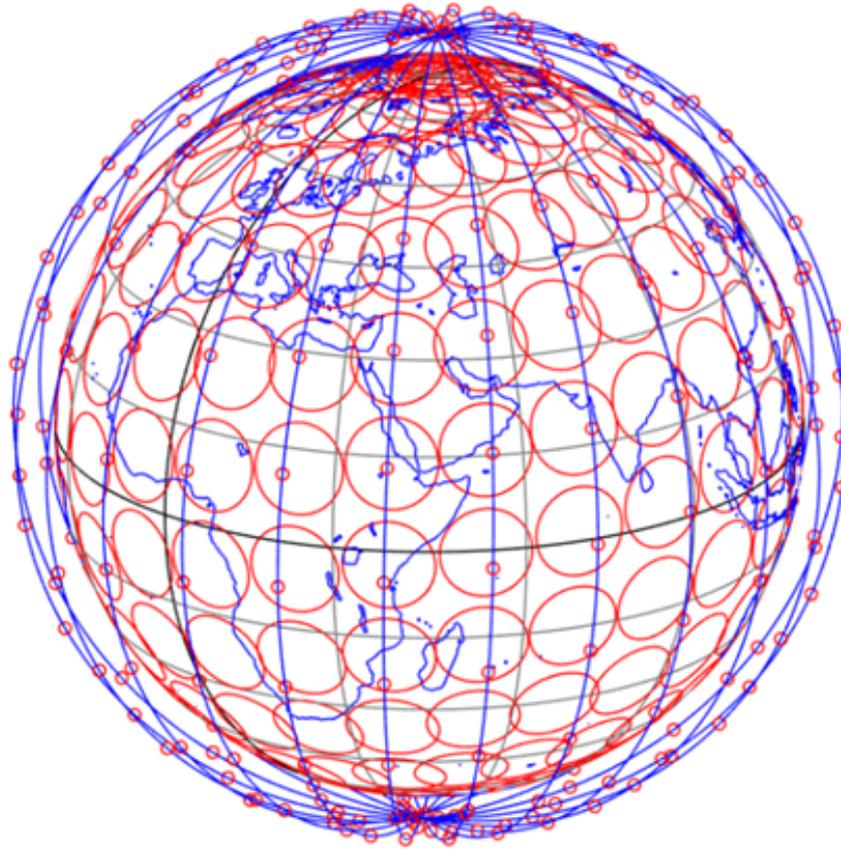


Рис. 2. Примерный вид моментального положения КА относительно Земли

При получении абонентами сервисов осуществляется передача трафика реального времени K типов и эластичного трафика данных, для которого скорость передачи может быть переменной в зависимости от доступного ресурса. Поток заявок на передачу трафика каждого типа является пуассоновскими и имеют интенсивности λ_k , $k = 1, 2, \dots, K$, и λ_e . Обслуживание одной заявки на передачу трафика реального времени k -го типа требует выделения b_k единиц ресурса на случайное время, имеющее экспоненциальное распределение с параметром μ_k . Обозначим через $i_k(t)$ число обслуживаемых заявок на предоставление сервиса k в текущий момент времени t . Для эластичного трафика данных на обслуживание каждой заявки выделяется равный максимально возможный ресурс в диапазоне от b_{min} до b_{max} . Время обслуживания одной заявки с минимальной скоростью (т.е. при выделении ресурса b_{min}) является экспоненциально распределенным с параметром μ_e . Обозначим через $d(t)$ число обслуживаемых заявок на передачу данных в момент времени t . Динамика изменения состояния сети описывается случайным процессом $r(t) = (i_1(t), \dots, i_K(t), d(t), n(t))$, определенном на конечном пространстве состояний S , куда входят состояния $s = (i_1, \dots, i_K, d, n)$ с компонентами

$$i_1 = 0, 1, \dots, \left\lfloor \frac{v_{ка}n}{b_1} \right\rfloor; \dots; i_K = 0, 1, \dots, \left\lfloor \frac{v_{ка}n - i_1 b_1 - \dots - i_{K-1} b_{K-1}}{b_K} \right\rfloor; \quad (1)$$

$$d = 0, 1, \dots, \left\lfloor \frac{v_{\text{ка}}n - i_1b_1 - \dots - i_Kb_K}{b_{\text{min}}} \right\rfloor; n = n_{\text{min}}, n_{\text{min}} + 1, \dots, n_{\text{max}}.$$

Для удобства обозначений примем, что $\mathbf{i} = (i_1, i_2, \dots, i_K, d)$ и $l = \sum_{k=1}^K i_k b_k$. Тогда состояния сети записываются в сокращенном виде $s = (\mathbf{i}, n)$. Сеть может изменять состояние при наступлении событий поступления заявок на обслуживание, завершения обслуживания, изменения количества доступных КА. Возможны случаи отказа в обслуживании и недообслуживания заявок. В частности, поступившая заявка на передачу трафика реального времени k -го типа, $k = 1, 2, \dots, K$, получает отказ, если нет достаточного ресурса для ее обслуживания, т.е. $v_{\text{ка}}n - l - db_{\text{min}} < b_k$. Аналогично, для заявки на передачу данных отказ имеет место при $v_{\text{ка}}n - l - db_{\text{min}} < b_{\text{min}}$. Случаи недообслуживания заявок могут возникать при уменьшении количества доступных КА с n до $n - 1$ при условии

$$v_{\text{ка}}(n - 1) < l + db_{\text{min}} \leq v_{\text{ка}}n. \tag{2}$$

При этом система управления связью РС должна прервать некоторые из сеансов обслуживания заявок с суммарным ресурсом не менее $l + db_{\text{min}} - v_{\text{ка}}(n - 1)$. Правило выбора прерываемых сеансов $F(s)$ зависит от существующей системы приоритетов. Пусть для s выполняется условие (2) и количество заявок k -го типа, для которых прерывается обслуживание, равно $F_k(s) \geq 0$, а количество соответствующих заявок на передачу данных равно $F_e(s)$. Тогда сеть переходит из состояния $s = (i_1, i_2, \dots, i_K, d, n)$ в состояние $s' = (i_1 - F_1(s), i_2 - F_2(s), \dots, i_K - F_K(s), d - F_e(s), n - 1)$.

Полагаем, что показателями качества обслуживания трафика в сети являются:

- доли необслуженных заявок на передачу трафика реального времени k -го типа π_k , $k = 1, 2, \dots, K$, и заявок на передачу данных π_e ;
- среднее время доставки файла W .

3. ОЦЕНКА ЗАВИСИМОСТЕЙ КОЛИЧЕСТВА ДОСТУПНЫХ АБОНЕНТАМ КА ОТ ВРЕМЕНИ

Определим поток пролетов КА над абонентом $n(t)$, как изменяющееся во времени количество КА, через которые абонент может при наличии на них свободного канального ресурса осуществлять передачу трафика. Доступность КА для абонента зависит от высоты орбиты h , минимально допустимого угла места для АТ α и ширины диаграммы направленности бортовой антенны КА 2β . Пусть радиус зоны обслуживания КА – это угол γ между направлениями из центра Земли на КА и любую из точек на краю зоны. При этом в зависимости от соотношения между 2β и α возможны два случая, когда γ определяется параметром α (рисунок 3) или параметром 2β (рисунок 4).

Из несложных геометрических соотношений получаем радиус зоны, вычисленный на основе α :

$$\gamma_1 = \arccos \frac{r \cdot \cos \alpha}{r + h} - \alpha. \tag{3}$$

Здесь r – радиус Земли. Радиус, вычисленный на основе β :

$$\gamma_2 = 90^\circ - \beta - \arccos \frac{(r + h) \sin \beta}{r}. \tag{4}$$

С учетом соотношений (3) и (4) имеем

$$\gamma = \min(\gamma_1, \gamma_2). \tag{5}$$

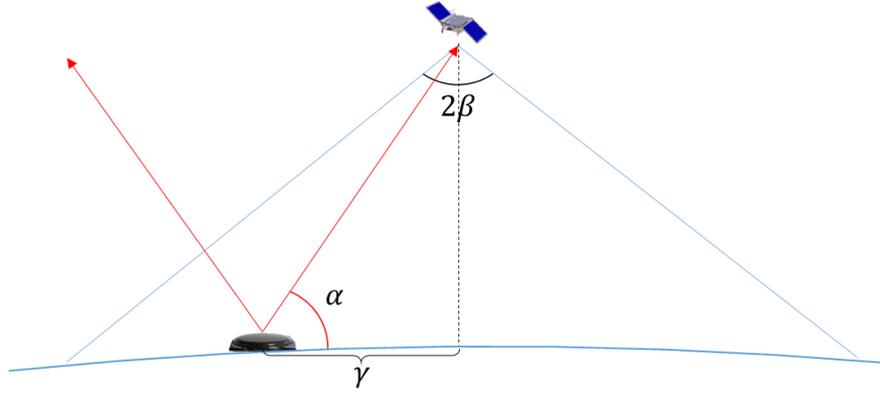


Рис. 3. Радиус зоны обслуживания КА определяется минимально допустимым углом места для абонента

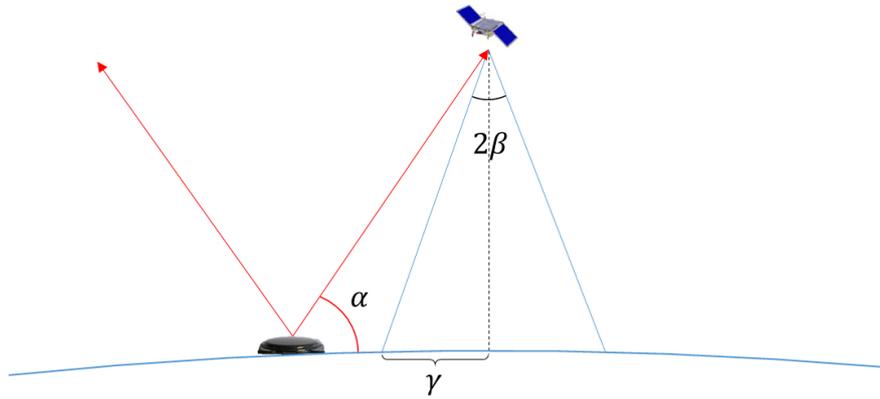


Рис. 4. Радиус зоны обслуживания КА определяется шириной диаграммы направленности антенны КА

Итак, КА доступен для АТ, если угол с вершиной в центре Земли между направлениями на КА и на АТ γ_T не превышает γ . Если КА находится над точкой земной поверхности с широтой $\phi_{\text{ш}}^{\text{ка}}$ и долготой $\phi_{\text{д}}^{\text{ка}}$, а АТ имеет координаты $\phi_{\text{ш}}^{\text{ат}}$ и $\phi_{\text{д}}^{\text{ат}}$, то с учетом известных соотношений сферической геометрии имеем

$$\gamma_T = \arccos(\sin \phi_{\text{ш}}^{\text{ат}} \sin \phi_{\text{ш}}^{\text{ка}} + \cos \phi_{\text{ш}}^{\text{ат}} \cos \phi_{\text{ш}}^{\text{ка}} \cos (\phi_{\text{д}}^{\text{ка}} - \phi_{\text{д}}^{\text{ат}})). \quad (6)$$

При известных функциях изменения во времени широты и долготы точек проекции на Землю каждого КА ($\phi_{\text{ш}}^{\text{ка}}(n_{\text{ка}}, t)$ и $\phi_{\text{д}}^{\text{ка}}(n_{\text{ка}}, t)$, $n_{\text{ка}} = 1, 2, \dots, N_{\text{ка}}$) множество доступных абоненту в точке ($\phi_{\text{ш}}^{\text{ат}}, \phi_{\text{д}}^{\text{ат}}$) КА $S_{\text{д}}(t)$ с учетом соотношений (5) и (6) можно записать в виде:

$$S_{\text{д}}(t) = \{n_{\text{ка}} : \gamma_T(\phi_{\text{ш}}^{\text{ка}}(n_{\text{ка}}, t), \phi_{\text{д}}^{\text{ка}}(n_{\text{ка}}, t), \phi_{\text{ш}}^{\text{ат}}, \phi_{\text{д}}^{\text{ат}}) \leq \gamma, n_{\text{ка}} = 1, 2, \dots, N_{\text{ка}}\}. \quad (7)$$

Искомая реализация процесса $n(t)$ представляет собой количество элементов множества $S_{\text{д}}(t)$, определяемого формулой (7), или $n(t) = |S_{\text{д}}(t)|$.

С целью моделирования процесса $n(t)$ разработан модуль в среде Matlab, учитывающий упомянутые выше характеристики орбитальной группировки и антенн АТ и КА. Модуль обеспечивает:

- оценку периода обращения КА вокруг Земли по высоте орбиты h ;
- построение исходного кольца орбиты на плоскости с равномерным размещением КА в точках орбиты;

- в) пересчет координат из двухмерных в трехмерные с учетом наклона орбиты и нулевой долготы восходящего узла;
- г) построение колец путем поворота по долготе плоскости орбиты, полученной на предыдущем шаге, с обеспечением равного угла между соседними кольцами;
- д) получение размещения КА в произвольный момент времени с учетом того, что КА делает полный оборот вокруг Земли за рассчитанный период обращения, а Земля вращается вокруг своей оси с угловой скоростью 360° в сутки;
- е) построение зон обслуживания КА для заданного размещения КА на орбитах;
- ж) собственно построение $n(t)$ для заданного (например, суточного) интервала с учетом динамики изменения положения КА на орбитах, моментальных зон обслуживания и заданных координат АТ.

Примерный вид моментального положения КА относительно Земли, полученный с использованием модуля, иллюстрируется рисунком 2. Расчет проведен для высоты орбиты 750 км, 12 колец по 22 КА, наклона орбиты 89° . Здесь также показаны зоны обслуживания КА при $\alpha = 30^\circ$ и $2\beta = 180^\circ$. На рисунке 5 приведен фрагмент искомой зависимости $n(t)$.



Рис. 5. Примерный вид зависимости количества доступных КА от времени $n(t)$ на интервале длиной 1500 секунд для АТ в точке 20° с. ш. 57° в. д.

Полученные зависимости $n(t)$ для требуемых точек дислокации АТ могут использоваться при имитационном моделировании процессов функционирования ССС. Чтобы использовать существующую методологию исследований на основе многомерных ступенчатых марковских процессов, целесообразно и $n(t)$ аппроксимировать ступенчатым марковским процессом.

Пусть реализация $n(t)$ определена на интервале от 0 до T с шагом Δt , достаточно малым, чтобы за время Δt количество доступных КА изменялось не более чем на единицу. Реализации соответствует диапазон изменения $n(t)$ от n_{min} до n_{max} . Создадим квадратную матрицу \mathbf{Q} размерности $(n_{max} - n_{min} + 1) \times (n_{max} - n_{min} + 1)$, каждый элемент которой $q_{i,j}$ равен количеству переходов процесса $n(t)$ из состояния $i + n_{min} - 1$ в состояние $j + n_{min} - 1$. Для заполнения матрицы достаточно программно просканировать $n(t)$ в моменты завершения квантов времени с первого по последний (с номером $\lfloor T/\Delta t \rfloor$). Очевидно, при $|i - j| > 1$ $q_{i,j} = 0$. Каждый диагональный элемент матрицы $q_{i,i}$ равен суммарному времени пребывания сети в состоянии $i + n_{min} - 1$, выраженному в Δt . При этом общее число интервалов времени пребывания $n(t)$ в состоянии $i + n_{min} - 1$ равно $q_{i-1,i}I(i > 1) + q_{i+1,i}I(i < n_{max} - n_{min} + 1)$, где индикаторная функция $I(\text{условие } A)$ равна 1, если условие выполнено, и 0 - в противном случае. Следовательно, среднее время пребывания $n(t)$ в состоянии $i + n_{min} - 1$ равно $t_{cpi} = q_{i,i}\Delta t / (q_{i-1,i}I(i > 1) + q_{i+1,i}I(i < n_{max} - n_{min} + 1))$, а при аппроксимации $n(t)$ ступенчатым марковским процессом интенсивность выхода процесса из данного состояния

$$\lambda_{i+n_{min}-1}^{ka} = 1/t_{cpi}. \tag{8}$$

Соответственно, при $i > 1$ интенсивность перехода из состояния $i + n_{min} - 1$ в состояние $i + n_{min} - 2$ равна

$$\lambda_{i+n_{min}-1, i+n_{min}-2}^{ка} = \frac{\lambda_{i+n_{min}-1}^{ка} q_{i, i-1}}{(q_{i, i-1} I(i > 1) + q_{i, i+1} I(i < n_{max} - n_{min} + 1))} \quad (9)$$

и при $i < n_{max} - n_{min} + 1$ интенсивность перехода из состояния $i + n_{min} - 1$ в состояние $i + n_{min}$ равна

$$\lambda_{i+n_{min}-1, i+n_{min}}^{ка} = \frac{\lambda_{i+n_{min}-1}^{ка} q_{i, i+1}}{(q_{i, i-1} I(i > 1) + q_{i, i+1} I(i < n_{max} - n_{min} + 1))}. \quad (10)$$

Здесь можно отметить, что с уменьшением Δt в матрице Q диагональные элементы $q_{i, i}$ будут увеличиваться (примерно обратно пропорционально Δt), а элементы $q_{i, i-1}$ и $q_{i, i+1}$ будут оставаться неизменными (количество переходов $n(t)$ между состояниями не изменяется).

Использование соотношений (8)-(10) для $i = 1, 2, \dots, n_{max} - n_{min} + 1$ позволяет определить все элементы матрицы $\Lambda^{ка}$ интенсивностей перехода между состояниями $n(t)$, которая имеет вид:

$$\left\| \begin{array}{ccccccc} -\lambda_{n_{min}}^{ка} & \lambda_{n_{min}, n_{min}+1}^{ка} & 0 & \dots & 0 & 0 & 0 \\ \lambda_{n_{min}+1, n_{min}}^{ка} & -\lambda_{n_{min}+1}^{ка} & \lambda_{n_{min}+1, n_{min}+2}^{ка} & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -\lambda_{n_{max}-1}^{ка} & \lambda_{n_{max}-1, n_{max}}^{ка} & 0 \\ 0 & 0 & 0 & \dots & \lambda_{n_{max}, n_{max}-1}^{ка} & -\lambda_{n_{max}}^{ка} & 0 \end{array} \right\|. \quad (11)$$

Сначала в соответствии с (8) вычисляются диагональные элементы, а затем с использованием (9) и (10) элементы, находящиеся слева и справа от каждого диагонального элемента матрицы (11).

4. ПРОСТРАНСТВО СОСТОЯНИЙ СЕТИ И ПОДМНОЖЕСТВА СОСТОЯНИЙ ЭТОГО ПРОСТРАНСТВА

С учетом условий (1) и определенного в разделе 3 диапазона изменения количества доступных КА пространство состояний сети S можно записать в виде:

$$S = \{(\mathbf{i}, n) : i_k \geq 0, k = 1, 2, \dots, K; d \geq 0; l + db_{min} \leq \nu_{ка} n, n = n_{min}, n_{min} + 1, \dots, n_{max}\}. \quad (12)$$

Это пространство можно представить как объединение непересекающихся подмножеств состояний S_n с фиксированным количеством доступных КА $n = n_{min}, n_{min} + 1, \dots, n_{max}$, т. е. $S = S_{n_{min}} \cup S_{n_{min}+1} \cup \dots \cup S_{n_{max}}$, где

$$S_n = \{(\mathbf{i}, n) : i_k \geq 0, k = 1, 2, \dots, K; d \geq 0; l + db_{min} \leq \nu_{ка} n\}. \quad (13)$$

На интервалах, где количество доступных КА неизменно и равно n , текущее состояние сети s при наступлении событий поступления или завершения обслуживания заявок изменяется в рамках S_n . Если же количество доступных КА уменьшается или увеличивается, то s переходит в другое подмножество, $s \in S_{n-1}$ или $s \in S_{n+1}$ соответственно.

Для оценки показателей качества необходимо определить подмножества состояний, в которых возможен отказ в обслуживании или прерывание начатой сессии.

Пусть $U_{k, n}$ (где $k = 1, 2, \dots, K$ и $n = n_{min}, n_{min}+1, \dots, n_{max}$) – подмножество состояний сети, для которых поступающие заявки сервиса реального времени k получают отказ при условии доступности n КА,

$$U_{k,n} = \{(\mathbf{i},n) : (\mathbf{i},n) \in S_n; l + db_{min} > \nu_{ка} \cdot n - b_k\}. \quad (14)$$

Соответственно, $U_{e,n}$ – подмножество состояний сети, для которых поступающие заявки на передачу данных получают отказ,

$$U_{e,n} = \{(\mathbf{i},n) : (\mathbf{i},n) \in S_n; l + db_{min} > \nu_{ка} \cdot n - b_{min}\}. \quad (15)$$

Подмножество состояний сети U_n^{int} , где $n = n_{min} + 1, \dots, n_{max}$, при выходе из которых по событию уменьшения количества доступных КА происходит прерывание обслуживания части заявок, записывается в виде

$$U_n^{int} = \{(\mathbf{i},n) : (\mathbf{i},n) \in S_n; l + db_{min} > \nu_{ка}(n - 1)\}. \quad (16)$$

При этом остаются необслуженными $F_k(s)$ заявок на предоставление сервисов реального времени, $k = 1, 2, \dots, K$, и $F_e(s)$ заявок на передачу данных, а состояние сети изменяется на $s' = (i_1 - F_1(s), i_2 - F_2(s), \dots, i_K - F_K(s), d - F_e(s), n - 1)$. Очевидно, что в состояние $s' \in S_{n-1}$ возможен переход из нескольких состояний $s \in S_n$, количество которых зависит как от правила выбора прерываемых сеансов $F(s)$, так и от s . При составлении уравнений равновесия осуществляется приравнивание интенсивностей входа в каждое состояние и выхода из него. Поэтому принципиально важным для формирования СУР является определение подмножеств состояний $U''(s)$, из которых можно попасть в s по событию уменьшения количества доступных КА.

Определим $U''(s)$. Пусть $s = (i_1, i_2, \dots, i_K, d, n) \in S_n, n = n_{min}, n_{min} + 1, \dots, n_{max} - 1$. Попасть в s по событию уменьшения количества доступных КА с прерываниями хотя бы одного сеанса можно из состояний вида $s' = (i'_1, i'_2, \dots, i'_K, d', n+1) \in U_{n+1}^{int}$ при выполнении условий $i'_1 - F_1(s') = i_1, i'_2 - F_2(s') = i_2, \dots, i'_K - F_K(s') = i_K, d' - F_e(s') = d$. Следовательно, можно записать

$$U''(s) = \{s' : s' \in U_{n+1}^{int}; i'_1 - F_1(s') = i_1, i'_2 - F_2(s') = i_2, \dots, i'_K - F_K(s') = i_K, d' - F_e(s') = d\}. \quad (17)$$

5. СИСТЕМА УРАВНЕНИЙ РАВНОВЕСИЯ ДЛЯ СЕТИ НА БАЗЕ НИЗКОЛЕТЯЩИХ КА И ПОКАЗАТЕЛИ КАЧЕСТВА ОБСЛУЖИВАНИЯ ТРАФИКА

Обозначим через $p(s)$ вероятность нахождения системы в состоянии s , а через $P(s)$ ненормированную вероятность указанного состояния. Ненормированные вероятности используются в итерационных методах решения СУР. Связь $p(s)$ и $P(s)$ с учетом условия нормировки следующая:

$$p(s) = \frac{P(s)}{\sum_{s \in S} P(s)}. \quad (18)$$

СУР с учетом соотношений (12)-(17) имеет следующий вид:

$$\begin{aligned} & P(i_1, \dots, i_K, d, n) \left[\sum_{k=1}^K \left(\lambda_k I((i_1, \dots, i_K, d, n) \in S_n \setminus U_{k,n}) + i_k \mu_k \right) + \right. \\ & \left. + \lambda_e I((i_1, \dots, i_K, d, n) \in S_n \setminus U_{e,n}) + d \mu(l, d, n) + \lambda_{n,n-1}^{ка} I(n > n_{min}) + \lambda_{n,n+1}^{ка} I(n < n_{max}) \right] = \\ & = \sum_{k=1}^K [P(i_1, \dots, i_k - 1, \dots, i_K, d, n) \lambda_k I(i_k > 0) + \end{aligned} \quad (19)$$

$$\begin{aligned}
& +P(i_1, \dots, i_k + 1, \dots, i_K, d, n)(i_k + 1)\mu_k I\left((i_1, \dots, i_K, d, n) \in S_n \setminus U_{k,n}\right) + \\
& +I(d > 0)P(i_1, \dots, i_K, d - 1, n)\lambda_e + P(i_1, \dots, i_K, d + 1, n)(d + 1)\mu(l, d + 1, n) \times \\
& \times I\left((i_1, \dots, i_K, d, n) \in S_n \setminus U_{e,n}\right) + P(i_1, \dots, i_K, d, n - 1)\lambda_{n-1,n}^{\text{ка}} I(l + db_{\min} \leq \nu_{\text{ка}}(n - 1)) + \\
& +P(i_1, \dots, i_K, d, n + 1)\lambda_{n+1,n}^{\text{ка}} I(n < n_{\max}) I(l + db_{\min} \leq \nu_{\text{ка}}n) + \\
& +I(n < n_{\max}) I(U''(s) \neq \emptyset) \sum_{s' \in U''(s)} P(s') \lambda_{n+1,n}^{\text{ка}}, \\
& (i_1, \dots, i_K, d, n) \in S.
\end{aligned}$$

Здесь интенсивность обслуживания заявок данных

$$\mu(l, d, n) = \begin{cases} 0, & \text{если } d = 0, \\ \min\left[\max\left(\left\lfloor \frac{\nu n - l}{d} \right\rfloor, b_{\min}\right), b_{\max}\right] \mu_e / b_{\min}, & d > 0. \end{cases} \quad (20)$$

В отличие от СУР для сетей с фиксированным объемом ресурса пропускной способности в настоящей СУР присутствуют составляющие, связанные с изменением количества доступных КА, в т. ч.:

- из состояния (i_1, \dots, i_K, d, n) можно выйти с интенсивностью $\lambda_{n,n-1}^{\text{ка}}$ при уменьшении количества КА (при $n > n_{\min}$) и с интенсивностью $\lambda_{n,n+1}^{\text{ка}}$ при увеличении количества КА (при $n < n_{\max}$);
- в состояние (i_1, \dots, i_K, d, n) можно попасть:
 - а) из состояния $(i_1, \dots, i_K, d, n - 1)$ при условии, что такое состояние возможно, т. е. $l + db_{\min} \leq \nu_{\text{ка}}(n - 1)$, с интенсивностью $\lambda_{n-1,n}^{\text{ка}}$ при увеличении количества КА;
 - б) из состояния $(i_1, \dots, i_K, d, n + 1)$ с интенсивностью $\lambda_{n+1,n}^{\text{ка}}$ при уменьшении количества КА (при $n < n_{\max}$) и при $l + db_{\min} \leq \nu_{\text{ка}}n$ (т. е. без прерывания обслуживания заявок);
 - в) из состояний, относящихся к множеству $U''(s)$, если это множество не пусто, с интенсивностями $\lambda_{n+1,n}^{\text{ка}}$ при уменьшении количества КА (с прерыванием обслуживания части заявок).

Система (19) может быть решена численно с использованием, например, метода Гаусса-Зейделя, который хорошо зарекомендовал себя при решении подобных СУР [4]. Затем вероятности состояний сети вычисляются с учетом условия нормировки (18).

С целью оценки показателей качества рассмотрим достаточно большой интервал времени функционирования сети T . Время пребывания сети в конкретном состоянии s оценивается величиной Tr_s (здесь для простоты состояние сети обозначено нижним индексом), а количество интервалов, на протяжении которых сеть находится в указанном состоянии, оценивается величиной $Tr_s \lambda_s$, где λ_s – суммарная интенсивность выхода сети из состояния s (по событиям поступления заявок, завершения обслуживания или изменения количества доступных КА) и $1/\lambda_s$ – средняя продолжительность интервала пребывания сети в состоянии s .

Доли необслуженных заявок каждого типа можно оценить двумя способами, приводящими к одному результату.

Первый способ

Очевидно, потери заявок k -го типа возможны в двух случаях:

- 1) Состояние s принадлежит множеству $U_{k,n}, n = n_{\min}, \dots, n_{\max}$, т.е. при поступлении заявки отказ в обслуживании является следствием отсутствия достаточного свободного ресурса

$(l + db_{min} > \nu_{ka} \cdot n - b_k)$. Общее количество заявок, получивших отказ, можно оценить величиной

$$\sum_{s \in U_{k,n}, n=n_{min}, \dots, n_{max}} T p_s \lambda_k. \quad (21)$$

2) Для состояния s выполняется соотношение $l + db_{min} > \nu_{ka}(n - 1)$ и интервал пребывания в состоянии завершается уменьшением количества доступных КА (с вероятностью $\lambda_{n,n-1}^{ka} / \lambda_s$). При этом обслуживание части заявок прекращается вследствие уменьшения объема доступного ресурса. Обозначим количество недообслуженных заявок k -го типа $F_k(s)$. Общее количество недообслуженных заявок оценивается, как

$$\sum_{s \in U_n^{int}, n=n_{min}+1, \dots, n_{max}} T p_s F_k(s) \lambda_{n,n-1}^{ka}. \quad (22)$$

Долю необслуженных заявок k -го типа определим как отношение количества потерянных заявок (вследствие отказа и прерывания обслуживания) к общему числу поступивших заявок рассматриваемого типа за время T . С учетом выражений (21) и (22) получаем

$$\pi_k = \sum_{s \in U_{k,n}, n=n_{min}, \dots, n_{max}} p_s + \sum_{s \in U_n^{int}, n=n_{min}+1, \dots, n_{max}} p_s F(k,s) \lambda_{n,n-1}^{ka} / \lambda_k. \quad (23)$$

Аналогично, для заявок на передачу эластичного трафика данных имеем

$$\pi_e = \sum_{s \in U_{e,n}, n=n_{min}, \dots, n_{max}} p_s + \sum_{s \in U_n^{int}, n=n_{min}+1, \dots, n_{max}} p_s F_e(s) \lambda_{n,n-1}^{ka} / \lambda_e. \quad (24)$$

В соотношениях (23) и (24) первое слагаемое соответствует вероятности отказа в обслуживании при поступлении заявок π_k^{ref}, π_e^{ref} , второе слагаемое – вероятности прерывания обслуживания при уменьшении количества КА π_k^{int}, π_e^{int} .

Второй способ

Успешному обслуживанию заявки соответствует окончание интервала пребывания сети в одном из возможных состояний по событию завершения обслуживания. Интервал пребывания сети в состоянии s оканчивается успешным завершением обслуживания заявки k -го типа с вероятностью $i_k \mu_k / \lambda_s$. Поэтому количество обслуженных заявок k -го типа за время T составит

$$\sum_{s \in S} (T p_s \lambda_s) (i_k \mu_k / \lambda_s) = \sum_{s \in S} T p_s i_k \mu_k. \quad (25)$$

Поскольку количество поступивших заявок k -го типа равно $\lambda_k T$, получаем с учетом соотношения (25)

$$\pi_k = 1 - \sum_{s \in S} p_s i_k \mu_k / \lambda_k. \quad (26)$$

Аналогично для заявок эластичного трафика данных имеем

$$\pi_e = 1 - \sum_{s \in S} p_s \mu_e(s) / \lambda_e. \quad (27)$$

Здесь $\mu_e(s) = \mu(l, d, n) d$ – суммарная интенсивность обслуживания заявок эластичного трафика в состоянии s , вычисляемая с учетом выражения (20).

Среднее время доставки файла W в соответствии с формулой Литтла равно

$$W = \frac{y_e}{\lambda_e (1 - \pi_e)}. \quad (28)$$

В формуле (28) y_e – среднее количество заявок на передачу трафика данных в системе, определяемое соотношением

$$y_e = \sum_{s \in S} p_s d.$$

6. ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ

Модель реализована на языке Matlab и является дискретно-событийной. Как и в аналитической модели, состояние сети в произвольный момент времени t определяется в виде $s = (\mathbf{i}, n)$, где $\mathbf{i} = (i_1, i_2, \dots, i_K, d)$. События изменяют состояние сети. В частности:

- по событию поступления заявок в случае выполнения условий приема на обслуживание одна из компонент вектора \mathbf{i} увеличивается на единицу;
- по событию завершения обслуживания соответствующая компонента вектора \mathbf{i} уменьшается на единицу;
- при изменении количества доступных КА n увеличивается или уменьшается ресурс сети. При этом возможна ситуация недостаточности дальнейшего обслуживания части заявок, требующая отработки процедуры выбора заявок, обслуживание которых подлежит прерыванию.

В соответствии с [4] генерация экспоненциально распределенных интервалов между событиями поступления заявок и продолжительностей обслуживания (объемов файлов) осуществляется с помощью встроенного в Matlab датчика равномерно распределенных случайных чисел $r \in (0,1)$ и преобразования $x = -\ln r/\lambda$, где λ - параметр экспоненциального распределения.

Для генерации динамики изменения количества доступных КА используется подход, описанный в разделе 3 и иллюстрируемый рисунком 5.

Используемые для работы модели входные данные совпадают с данными для аналитической модели с той разницей, что зависимость количества доступных КА от времени является расчётной. В результате моделирования оцениваются значения вероятностей отказа в обслуживании при поступлении заявок π_k^{ref}, π_e^{ref} и прерывания обслуживания при уменьшении количества КА π_k^{int}, π_e^{int} . Для построения доверительных интервалов для этих оценок применен подход, описанный в [4] и основанный на использовании совокупности выборочных интервальных оценок средних значений и дисперсий указанных вероятностей, а также таблиц распределения Стьюдента.

Имитационная модель выполняет следующие функции:

- выбор очередного события для обработки в соответствии с текущим модельным временем;
- генерацию интервалов до поступления очередных заявок каждого типа в начальный момент времени и после выбора, как ближайших по времени, моментов поступления заявок;
- проверку выполнения условий принятия поступающих заявок на обслуживание с учетом наличия достаточного объема доступного ресурса (а при необходимости с учетом группового поступления заявок на передачу данных, дополнительных ограничений по предельному количеству обслуживаемых заявок каждого типа);
- включение принятых заявок в список обслуживаемых заявок соответствующего типа с генерацией времени обслуживания (для заявок на передачу данных времени обслуживания на минимальной скорости), а также накопление интервальной статистики по количеству отклоненных заявок;
- после наступления события изменения количества доступных КА в случае возникновения дефицита ресурса отработку процедуры выбора заявок, обслуживание которых подлежит прерыванию, и набор статистики по количеству прерванных сеансов обслуживания заявок каждого типа;

- после наступления каждого события корректировку состояния сети, расчет ресурса на обслуживание эластичного трафика, а в случае потребности оценки среднего времени пребывания на обслуживании заявок на передачу данных накопление статистики по суммарному времени пребывания указанных заявок в сети;
- оценку вероятностей отказа в обслуживании и прерываний сеансов, а также среднего времени доставки файла путем обработки набранной статистики и построение доверительных интервалов для заданной доверительной вероятности.

7. ЗАКЛЮЧЕНИЕ

Для сетей на базе космических аппаратов на низкой круговой орбите для случая высокой концентрации абонентов на относительно небольшом участке построены модели совместного обслуживания мультисервисного трафика реального времени и эластичного трафика данных, включая аналитическую и имитационную модели. Получены соотношения для расчета зависимости количества доступных абонентам космических аппаратов от времени с учетом характеристик орбитальной группировки, параметров антенн космического аппарата и абонентского терминала. Предположение о том, что при оценке показателей качества обслуживания трафика расчетный поток пролетов космических аппаратов может быть заменен на марковский процесс с эквивалентными параметрами переходов между состояниями, позволило построить аналитическую модель на основе многомерных ступенчатых марковских процессов. Для оценки показателей качества сформирована система уравнений равновесия, которая может быть решена с использованием метода Гаусса-Зейделя. Разработан сценарий оценки показателей в среде Matlab. Для проверки допустимости сделанного предположения потребовалась разработка имитационной модели процессов обслуживания трафика в сети, которая также реализована в среде Matlab. Построенная модель и методы расчёта её характеристик могут быть использованы для численного анализа процессов обслуживания мультисервисного трафика, а также для оценки эффективности совместной передачи трафика сервисов реального времени и данных в сетях спутниковой связи. Соответствующее исследование будет выполнено в [22]

СПИСОК ЛИТЕРАТУРЫ

1. Сайт компании Globalstar [Электронный ресурс] – URL: <https://www.globalstar.com/en-us> (дата обращения 26.09.2024).
2. Кузовников А., Семкин П. Перспективные ССС: «Марафон-IoT», «СКИФ», «Гонец-М1», «Экспресс-РВ». Доклад на XVI Международном навигационном форуме КОНГРЕСС СФЕРА, 13.04.2023, ЦВК «Экспоцентр».
3. Сайт АО «Спутниковая система «Гонец» [Электронный ресурс] – URL: <https://gonets.ru/rus/> (дата обращения 26.09.2024).
4. Степанов С.Н. Теория телетрафика: концепции, модели, приложения. – М.: Горячая линия – Телеком, 2015. – 868 с.
5. Башарин Г.П. Лекции по математической теории телетрафика: Учеб. пособ. – М.: РУДН, 2009. – 342 с.
6. Сайт Федеральной службы государственной статистики [Электронный ресурс] – URL: <https://rosstat.gov.ru/> (дата обращения 26.09.2024).
7. Stepanov M.S., Kanishcheva M.G., Malikova E.E., Ndayikunda J., Kaigorodov N.G. The Development and Analysis of a Service Model for the Traffic of a Surveillance System Operator by a Dedicated Resource of an LTE Cell, 2022 Systems of Signals Generating and Processing in the Field of on Board Communications, 2022, pp. 1-6.

8. Stepanov M.S., Stepanov S.N., Kroshin F.S. Effective Algorithm of Estimation the Performance Measures of Group of Servers with Dependence of Call Repetition on the Type of Call Blocking DCCN 2022. Lecture Notes Computer Science, Springer, Cham., DCCN 2022. Lecture Notes Computer Science, Springer, Cham., 2022.
9. Gibadullina Elena E., Viskova Elena V., Stepanov Sergey N. Automated Service Configuration Management in IP/MPLS Networks // 4th International Science and Technology Conference «Modern Network Technologies – 2022», MoNeTec – 2022.
10. Andrabi U.M., Kanishcheva M., Stepanov S.N. Observation system resource planning in presence of access control based on volume of resource occupied by traffic flows// T-Comm, 2022, vol. 16, no.8, pp.54-62.
11. Dawood T., Stepanov M.S., Naoussi C., Joao B., Yahia K. The Mathematical Model of the Internet of Things Traffic Servicing in Case of its Impulse Nature, 2023 Systems of Signals Generating and Processing in the Field of on Board Communications, Moscow, Russian Federation, 2023, pp. 1-8.
12. Ndimumahoro F., Stepanov M.S., Muzata A.R., Tammam D., Popov V.G. Using the Principles of Mobile Systems Modeling for LoRaWAN Characteristics Estimation, 2022 Systems of Signals Generating and Processing in the Field of on Board Communications, 2022, pp. 1-8.
13. Степанов С.Н., Степанов М.С. Приближенный метод оценки характеристик совместного обслуживания трафика реального времени и эластичного трафика данных в мультисервисных узлах доступа // Автомат. и телемех., – 2023, – № 11, – С. 93–114.
14. Maslov A.A., Sebekin G.V., Stepanov S.N., Shchurkov A.O., Vasilyev A.P. Model of processes for joint maintenance of real-time multiservice traffic and elastic data traffic in a network of low-power mobile subscriber terminals based on high-throughput satellites // T-COMM. – 2024 vol. 18, – № 5, – P. 41–49. – ISSN: 2072-8735: Модели процессов совместного обслуживания
15. Маслов А.А., Себекин Г.В., Степанов М.С., Степанов С.Н., Щурков А. О. Модель резервирования канального ресурса при обслуживании разнородного трафика в сети маломощных мобильных абонентских терминалов на базе КА с высокой пропускной способностью. Информационные процессы – 2024. – Т. 24 – № 1. – С. 1–15.
16. Себекин Г.В., Маслов А.А., Щурков А.О. Моделирование совместного обслуживания мультисервисного трафика реального времени и эластичного трафика данных в сетях на базе космических аппаратов с высокой пропускной способностью // Информационно-измерительные и управляющие системы. – 2024. – Т. 22. – № 2. – С. 11-22.
17. Afanaseva L., Bashtova E. Queueing Systems with Unreliable Servers in a Random Environment в сборнике Topics in Statistical Simulation, серия Springer Proceedings in Mathematics & Statistics, издательство Springer New York (New York, N.Y., United States), 2014, том 116, с. 1-9.
18. Afanaseva L.G., Grishunina S.A. Stability conditions for a multiserver queueing system with a regenerative input flow and simultaneous service of a customer by a random number of servers в журнале Queueing Systems, издательство Baltzer Science Publishers B.V. (Netherlands), 2020, том 94, с. 213-241.
19. Руденко И.В. Системы массового обслуживания с ненадежными и восстанавливающимися приборами. Диссертация на соискание ученой степени кандидата технических наук — Москва, – 2012. — 106 с.
20. Adou Y., Markova E., Gudkova I. Performance Measures Analysis of Admission Control Scheme Model for Wireless Network, Described by a Queueing System Operating in Random Environment // Emerging Technologies for Connected Society: 2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT): Proceedings (Moscow, 5–9 November 2018). – IEEE, 2018. P. 262–267. DOI: 10.1109/ICUMT.2018.8631263.
21. Adou Y., Markova E., Gudkova I. Approximate Product Form Solution for Performance Analysis of Wireless Network with Dynamic Power Control Policy. Lecture Notes in Computer Science. 2019. Т. 11965 LNCS. С. 379-390.

22. Маслов А.А., Себекин Г.В., Степанов М.С., Степанов С.Н., Щурков А. О. Моделирование процессов обслуживания абонентов в сети передачи данных на базе космических аппаратов на низкой круговой орбите. II. Информационные процессы – 2025. – Т. [25] – № [1].

**Model for Reserving Channel Resource for the Joint Maintenance
of Real-time Traffic and Elastic Data Traffic
in a Network of Low-power Mobile Subscriber Terminals
Based on High-Throughput Satellites
in Geostationary or Highly Elliptical Orbits. I**

A.A. Maslov, G.V. Sebekin, M.S. Stepanov, S.N. Stepanov, A.O. Shchurkov

Managing the processes of servicing multiservice traffic in data transmission networks based on spacecraft in a low circular orbit is a complex task that requires modeling taking into account the real conditions of network operation. The article develops analytical and simulation models of such a network for the case of concentration of a large number of subscribers in a relatively small area, which is difficult from the point of view of implementing requirements for the quality of service provision. The analytical model is based on multidimensional stepwise Markov processes. The simulation model allows you to check the acceptability of the assumptions made during analytical modeling and evaluate quality indicators for a large number of network states. The models can be used to assess the capabilities of networks to serve areas of high concentration of subscribers.

KEYWORDS: spacecraft in low circular orbit, channel resource, multiservice traffic, real-time traffic, elastic traffic, subscriber terminals.